American Petroleum Institute

STEP Strategies for Today's Environmental Partnership

A Monte Carlo Approach to Generating Equivalent Ventilation Rates in Population Exposure Assessments

Health and Environmental Sciences Department
Publication Number 4617
March 1995

One of the most significant long-term trends affecting the future vitality of the petroleum industry is the public's concerns about the environment. Recognizing this trend, API member companies have developed a positive, forward-looking strategy called STEP: Strategies for Today's Environmental Partnership. This program aims to address public concerns by improving our industry's environmental, health and safety performance; documenting performance improvements; and communicating them to the public. The foundation of STEP is the API Environmental Mission and Guiding Environmental Principles.

API ENVIRONMENTAL MISSION AND GUIDING ENVIRONMENTAL PRINCIPLES

The members of the American Petroleum Institute are dedicated to continuous efforts to improve the compatibility of our operations with the environment while economically developing energy resources and supplying high quality products and services to consumers. The members recognize the importance of efficiently meeting society's needs and our responsibility to work with the public, the government, and others to develop and to use natural resources in an environmentally sound manner while protecting the health and safety of our employees and the public. To meet these responsibilities, API members pledge to manage our businesses according to these principles:

- To recognize and to respond to community concerns about our raw materials, products and operations.
- To operate our plants and facilities, and to handle our raw materials and products in a manner that protects the environment, and the safety and health of our employees and the public.
- To make safety, health and environmental considerations a priority in our planning, and our development of new products and processes.
- To advise promptly, appropriate officials, employees, customers and the public of information on significant industry-related safety, health and environmental hazards, and to recommend protective measures.
- To counsel customers, transporters and others in the safe use, transportation and disposal of our raw materials, products and waste materials.
- To economically develop and produce natural resources and to conserve those resources by using energy efficiently.
- To extend knowledge by conducting or supporting research on the safety, health and environmental effects of our raw materials, products, processes and waste materials.
- To commit to reduce overall emission and waste generation.
- To work with others to resolve problems created by handling and disposal of hazardous substances from our operations.
- To participate with government and others in creating responsible laws, regulations and standards to safeguard the community, workplace and environment.
- To promote these principles and practices by sharing experiences and offering assistance to others who produce, handle, use, transport or dispose of similar raw materials, petroleum products and wastes.

A Monte Carlo Approach to Generating Equivalent Ventilation Rates in Population Exposure Assessments

Health and Environmental Sciences Department

API PUBLICATION NUMBER 4617

PREPARED UNDER CONTRACT BY:

TED JOHNSON AND MICHAEL MCCOY, JR.
IT CORPORATION
3710 UNIVERSITY DRIVE, SUITE 201
DURHAM, NC 27707

JANUARY 1995

American Petroleum Institute

FOREWORD

API PUBLICATIONS NECESSARILY ADDRESS PROBLEMS OF A GENERAL NATURE. WITH RESPECT TO PARTICULAR CIRCUMSTANCES, LOCAL, STATE, AND FEDERAL LAWS AND REGULATIONS SHOULD BE REVIEWED.

API IS NOT UNDERTAKING TO MEET THE DUTIES OF EMPLOYERS, MANUFACTURERS, OR SUPPLIERS TO WARN AND PROPERLY TRAIN AND EQUIP THEIR EMPLOYEES, AND OTHERS EXPOSED, CONCERNING HEALTH AND SAFETY RISKS AND PRECAUTIONS, NOR UNDERTAKING THEIR OBLIGATIONS UNDER LOCAL, STATE, OR FEDERAL LAWS.

NOTHING CONTAINED IN ANY API PUBLICATION IS TO BE CONSTRUED AS GRANTING ANY RIGHT, BY IMPLICATION OR OTHERWISE, FOR THE MANUFACTURE, SALE, OR USE OF ANY METHOD, APPARATUS, OR PRODUCT COVERED BY LETTERS PATENT. NEITHER SHOULD ANYTHING CONTAINED IN THE PUBLICATION BE CONSTRUED AS INSURING ANYONE AGAINST LIABILITY FOR INFRINGEMENT OF LETTERS PATENT.

ACKNOWLEDGMENTS

THE FOLLOWING PEOPLE ARE RECOGNIZED FOR THEIR CONTRIBUTIONS OF TIME AND EXPERTISE DURING THIS STUDY AND IN THE PREPARATION OF THIS REPORT:

API STAFF CONTACT

Will Ollison, Health and Environmental Sciences Department

MEMBERS OF THE EXPOSURE ASSESSMENT MULTI-YEAR TASK FORCE

Jack Hinton, Texaco

Lewis Cook, Chevron
Lee Gilmer, Texaco
Charles Lapin, ARCO
Donald Molenaar, Unocal
Joseph Yang, Mobil

RANCHO LOS AMIGOS MEDICAL CENTER

William Linn

CONTRACTOR ACKNOWLEDGMENTS

This report describes a research project conducted by IT Air Quality Services (ITAQS) for the American Petroleum Institute (API). The project consisted of eight work elements:

1. acquisition of four time/activity databases in which each documented diary event is associated with a measured pulse rate,
2. acquisition of clinical data relating subject pulse rate to ventilation rate,
3. statistical analysis of clinical data to determine appropriate procedures for converting pulse rate to equivalent ventilation rate (EVR),
4. conversion of each pulse-rate database into a corresponding database listing EVR by diary event,
5. statistical analysis of each EVR database to identify factors that affect EVR,
6. development of algorithms for predicting EVR according to population group,
7. testing of each algorithm by comparing model predictions with measured EVR values, and
8. preparation of this report.

Mike McCoy was the ITAQS project manager for the overall project and was primarily responsible for work elements 1, 2, 4, and 7. Ted Johnson was ITAQS technical director for the project and had primary responsibility for work elements 3, 5, 6, and 8. Doug Brinson assisted with work element 3. Joan Abernethy typed the report.

ITAQS work on this project was funded by API under ITAQS Project No. 465063-8. Dr. Will Ollison served as the API task assignment manager and provided technical guidance throughout the task.

The authors would like to express their appreciation to William Linn of the Rancho Los Amigos Medical Center for providing the databases acquired in work elements 2 and 3.

ABSTRACT

A number of researchers have developed computer-based models for simulating the exposure of human populations to air pollution. The probabilistic version of the National Ambient Air Quality Standards Exposure Model (pNEM) is typical of these models in that it characterizes each exposure by time period and pollutant concentration. Unlike most other exposure models, pNEM also characterizes each exposure by a measure of respiration, the equivalent ventilation rate (EVR). EVR is defined as ventilation rate divided by body surface area. In the current version of pNEM, EVR is determined by an algorithm that randomly selects values from lognormal distributions that are specific to age and breathing rate category. A research team directed by Jack Hackney and William Linn conducted four studies in Los Angeles which used time/activity diaries and heart rate monitors to obtain ventilation rate data representative of typical daily activities. IT Air Quality Services acquired the four Hackney/Linn databases and converted each into a file of EVR values, one EVR value for each diary event. Researchers analyzed these files and developed a series of algorithms for generating EVR values that are superior to those used in the current pNEM methodology. Each algorithm uses Monte Carlo (probabilistic) techniques to produce EVR values that vary according to age, gender, activity, breathing rate category (slow, medium, or fast), microenvironment, time of day, activity duration, and other variables present in the input time/activity data files. The algorithms were tested by applying them to representative time/activity databases that contained a measured EVR value for each diary record. In each test, analysts compared the distribution of generated EVR values with the corresponding distribution of measured EVR values. Results of these tests suggest that the algorithms produce realistic EVR distributions.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

EXECUTIVE SUMMARY

The U.S. Environmental Protection Agency (EPA) has developed the pNEM methodology as a means of evaluating current and proposed National Ambient Air Quality Standards (NAAQS). The pNEM approach provides exposure estimates for defined population groups based on activity data specific to each group. Similar to other exposure models, pNEM characterizes each exposure event by time period and pollutant concentration. Unlike most other exposure models, pNEM also characterizes each exposure event by equivalent ventilation rate (EVR), defined as ventilation rate divided by body surface area (BSA).

This report presents a series of algorithms that can be used by pNEM and similar models to estimate EVR values by exposure event. Each algorithm is optimized for use with one of four specific data types. The algorithms were developed through an analysis of data reported by a research team directed by Jack Hackney and William Linn. The Hackney/Linn team conducted four studies in Los Angeles to obtain ventilation rate data representative of typical daily activities (elementary school students, high school students, outdoor workers, and construction workers). The heart rate of each study subject was continuously monitored as the subject documented his or her activities in a special diary. Separate clinical trials were conducted in which the heart rate and ventilation rate of each subject were measured simultaneously. These measurements were used to develop a "calibration curve" for each subject relating heart rate to ventilation rate.

Existing calibration curves for three studies (elementary school, high school, and outdoor workers) were used to convert one-minute heart rate measurements into one-minute ventilation rates. The ventilation rate values were in turn divided by the subject's estimated body surface area to produce one-minute EVR values.

Hackney and Linn did not provide calibration curves for the construction worker study. An analysis of the applicable calibration data collected during the construction worker study produced two alternative sets of calibration curves, one based on a linear relationship (Model A) and the other on a log-log relationship (Model C). Each set of calibration curves was used to convert one-minute heart rate values into corresponding one-minute ventilation rates. The results were averaged to produce a final set of one-minute ventilation rates. The values were then divided by body surface area to produce a database listing one-minute EVR values for the construction workers.

Algorithms for predicting EVR were developed by applying a four-step procedure to each of the one-minute EVR databases. In Step 1, ITAQS processed each one-minute EVR database to produce a special "event EVR file." Each file provided a sequence of exposure events keyed to the activities documented by each subject. The listing for each event included the average EVR for the event and the values of 20 variables which were considered likely to influence EVR values.

In Step 2, ITAQS prepared tables of descriptive statistics for event EVR values which had been categorized by breathing rate, activity, microenvironment, time of day, and event duration. These statistics provided an initial means for identifying factors to be considered in developing the EVR prediction algorithms. These factors were compiled into sets of candidate variables, each set specific to a particular database type.

In Step 3, ITAQS developed one or more Monte Carlo models for each database type. Each model was specific to one of the following four demographic groups: elementary students, high school students, outdoor workers, and construction workers. Each model consisted of an algorithm capable of generating an EVR for each event in a time/activity data set of the specified database type and demographic group. Each algorithm predicted EVR as a function of six or more predictor variables which constituted a predictor set.

Each predictor set was developed by first defining a candidate variable set for the database type and then performing stepwise linear regression analyses to determine which of the candidate variables were significant predictors of EVR for a particular demographic group. The regression analyses were performed on the Hackney/Linn databases, as these were the only databases available which provided a "measured" EVR value for each exposure event. The results of the regression analyses determined the variables to be included in the predictor set and the coefficients of various terms in the associated Monte Carlo model.

The best overall predictor variable was found to be LGM, the natural logarithm of the geometric mean of all event EVR values associated with a subject. Statistical analysis of the LGM values indicated that the distribution of LGM values was approximately normal.

In addition to LGM, the regression analyses suggested that variables associated with microenvironment, daytime activities, the exertion level of activities, day of week, breathing rate, and duration of activity were generally useful in predicting event EVR. It should be noted that the duration variables were among the least significant of these predictors.

Each regression analysis produced a set of residual values, one for each EVR value. Statistical analysis of the residuals indicated that (1) the standard deviation of the residuals varied significantly from subject to subject and (2) the distribution of the subject-specific standard deviations was approximately lognormal.

In Step 4, ITAQS performed an initial validation of the Monte Carlo approach by applying the EVR-generator algorithm to each of the four Hackney/Linn databases.

Each application produced a distribution of event EVR values that could be compared with the distribution of "measured" values. For three of the databases (elementary school, high school, and construction workers), the modeled and observed distributions compared favorably with respect to mean, standard deviation, and percentiles up to the 99th or 99.5th percentiles. At higher percentiles, the algorithm tended to underestimate EVR for the elementary and high school databases and over estimate EVR for the construction worker database.

The algorithm did not perform as well for the outdoor worker database. In this case, the model significantly underestimated the standard deviation and the percentiles above the 90th percentile. Analysts found that the differences between modeled and observed distributions were significantly reduced when a particular atypical subject was removed from the analysis. Removal of atypical subjects did not account for all of the differences between the modeled and observed results, however.

Section 1
INTRODUCTION

During the past 15 years, a number of researchers have developed computer-based models for simulating the exposure of human populations to air pollution. These models are frequently used to estimate exposures under various regulatory control scenarios. For example, the U.S. Environmental Protection Agency (EPA) has developed an exposure model specifically to evaluate current and proposed National Ambient Air Quality Standards (NAAQS). A recent version of this model incorporating stochastic features (the probabilistic NAAQS Exposure Model or pNEM) has been used to estimate the exposures of urban populations to ozone[1,2] and carbon monoxide[3].

The pNEM approach provides exposure estimates for defined population groups based on activity data specific to each group. Similar to other exposure models, pNEM characterizes each exposure by time period and pollutant concentration. Unlike most other exposure models, pNEM also characterizes each exposure by a measure of respiration, the equivalent ventilation rate (EVR). EVR is defined as ventilation rate (liters per minute) divided by body surface area (square meters). Clinical research by EPA suggests that EVR exhibits less interpersonal variability than ventilation rate for a given level of exertion[4].

In pNEM, EVR values vary according to specified lognormal distributions with the constraint that EVR must not exceed a specific upper limit. The limit is assumed to vary with age, gender, and activity duration. A report by Johnson et al.[1] describes the most recent version of the algorithm that is used to determine the limiting values for EVR. A preliminary version of this algorithm was described by Johnson et al. in a report for the American Petroleum Institute[5].

In a typical application of pNEM, the population of a designated study area is divided into homogenous population groups called cohorts. Using a special file containing empirical activity diary data, the model generates a distinct activity pattern for each cohort. The pattern consists of a sequence of "exposure events". Each event assigns to the cohort a specific environmental setting and a specific breathing rate category. The environmental setting is characterized by geographic location (e.g., within 5 km of a fixed-site monitor) and a microenvironment (e.g., outdoors within 10 m of a road). Using ambient fixed-site monitoring data appropriate to the geographic location and a mass balance model to adjust monitored concentrations to the microenvironment, pNEM provides an estimate of the pollutant concentration associated with each exposure event.

In addition to the environmental setting, each exposure event assigns to the cohort one of four breathing rate categories: sleeping, slow, medium, and fast. These categories were selected to permit use of activity diary data collected during the Cincinnati Activity Diary Study[6]. In this study, over 900 subjects completed three-day time/activity diaries which used these breathing categories to characterize the exertion level associated with each activity.

During a three-year period (1989 to 1991), a research team directed by Jack Hackney and William Linn conducted four studies in Los Angeles to obtain ventilation rate data representative of typical daily activities.[7,8,9] The heart rate of each study subject was continuously monitored as the subject documented his or her activities in an activity dairy similar to that used in the Cincinnati study. Separate clinical trials were conducted in which the heart rate and ventilation rate of each subject were measured simultaneously. These measurements were used to develop a subject-specific "calibration curve" relating heart rate to ventilation rate.

Using this calibration curve, researchers can transform each heart rate measurement into an estimated ventilation rate. The ventilation rates for each subject can then be converted to corresponding EVR estimates by dividing by the subject's body surface area. Each of the resulting EVR values can be indexed according to the diary entries made by the subject for the associated time period. The EVR values can also be indexed by the subject's demographic characteristics (age, gender, etc.).

The four Hackney/Linn studies produced databases ideally suited for analyzing the relationships between activity diary entries and EVR. Under a contract with the American Petroleum Institute, IT Air Quality Services (ITAQS) acquired the four Hackney/Linn databases and converted each database into a file of EVR values, one EVR value per diary "event." ITAQS analyzed these EVR files and developed a series of algorithms that could be used to generate EVR values in pNEM and similar exposure models. Each algorithm uses Monte Carlo techniques to produce EVR values that vary according to diary entries and subject characteristics.

This report summarizes these research efforts by ITAQS. Section 2 describes the construction of the event EVR files. Section 3 provides descriptive statistics for EVR values classified according to breathing rate category (sleeping, slow, medium, or fast), activity, microenvironment, time of day and duration. Section 4 presents EVR-generating algorithms appropriate for various types of databases. Initial efforts to validate these algorithms are described in Section 5. Section 6 presents a summary of the report and recommendations for further research.

Section 2
CONSTRUCTION OF EVENT EVR FILES

ACQUISITION OF HACKNEY/LINN DATA SETS

ITAQS acquired data from four studies conducted by a research team under the guidance of Jack Hackney and William Linn. These studies, hereafter referred to as the "Hackney/Linn" studies, are identified according to the demographic characteristics of the subjects:

1. Elementary school students[7]
2. High school students[7]
3. Outdoor workers[8]
4. Construction workers[9]

Table 2-1 provides descriptive information concerning each study. Additional information concerning these studies can be found in the paper by Johnson, McCoy, Capel, Wijnberg, and Ollison included in Appendix A.

Each of the Hackney/Linn data sets contained a series of one-minute heart rate (MINHR) values measured during a period documented by a time/activity diary. The diary provided information concerning the subject's location, breathing rate category, and activity during each one-minute interval. Associated with each MINHR value is a one-minute ventilation rate value (MINVR) estimated through the use of a subject-specific calibration curve. ITAQS also developed a demographic profile (age, occupation, etc.) for each subject based on data from background questionnaires administered during each study.

Table 2-1. Characteristics of four time/activity studies conducted by the Hackney/Linn research team.

Study	Characteristics of subjects	Number of subject-days	Study calendar periods	Diary type	Nominal diary time period	Breathing rates reported?	Source of calibration curves
Los Angeles - elem. school	Elementary school students, 10 to 12 years	58	Oct. 1989	Real-time[a]	Midnight to midnight	Yes	Linn
Los Angeles - high school	High school students, 13 to 17 years	66	Sept. and Oct. 1990	Real-time[a]	Midnight to midnight	Yes	Linn
Los Angeles - outdoor worker	Adult outdoor workers, 19 to 50 years	60	Summer 1989	Real-time[a]	Midnight to midnight	Yes	Linn
Los Angeles - construction	Construction workers, 23 to 42 years	19	July - Nov. 1991	Real-time[a]	Subject wake up to subject returns home from work	Yes	ITAQS[b]

[a]Study employed the Cincinnati diary format.
[b]See Section 2 of this report.

Before beginning an analysis of the four data sets, ITAQS conferred with Linn to determine whether they had the most recent version of each data set. Linn confirmed that ITAQS had the most recent versions of the data sets associated with the elementary school, high school, and outdoor worker studies. He recommended that ITAQS independently develop a calibration model for the construction workers based on the reported calibration data. The results of ITAQS efforts to develop this model are discussed in the next section of this report.

DEVELOPMENT OF CALIBRATION CURVES FOR CONSTRUCTION WORKERS

Linn provided ITAQS with his best estimates of the one-minute EVR values for three of the four Hackney/Linn studies: elementary school, high school, and outdoor workers. It is important to note that the EVR values in these data sets were not measured directly. As indicated above, measured MINHR values were converted to MINVR through the use of subject-specific calibration curves. Each MINVR was divided by the subject's estimated BSA to produce a corresponding one-minute EVR value.

The calibration curves used in the three studies were all in the form:

$$\ln(\text{MINVR}) = a_0 + (a_1)(\text{MINHR}). \tag{1}$$

The parameters of each curve were determined by fitting Equation 1 to subject-specific data sets obtained from clinical tests. In these tests, MINVR and MINHR were measured simultaneously while the subject exercised at varying levels of exertion.

Equation 1 is a "log-linear" relationship. Raizenne and Spengler[10] have also used log-linear relationships as the basis for calibration curves to convert heart rate to ventilation rate.

Linn did not provide ITAQS with "best estimate" EVR values for the construction worker data set. He indicated to ITAQS that he was dissatisfied with the results of his initial attempts to develop calibration curves for these data. Linn recommended that ITAQS evaluate a variety of relationships and select one which appeared to produce reasonable ventilation estimates throughout the range of measured MINHR values. Table 2-2 presents descriptive statistics, including minimum and maximum values, for the MINHR values measured during the activity diary phase of the construction worker study.

ITAQS acquired the calibration data for the construction workers and attempted to fit each of the following four models to the data for each subject.

A. Linear: $MINVR = a_0 + (a_1)(MINHR)$ (2)

B. Quadratic: $MINVR = a_0 + (a_1)(MINHR) + (a_2)(MINHR)^2$ (3)

C. Log-log: $\ln(MINVR) = a_0 + (a_1)[\ln(MINHR)]$ (4)

D. Log-linear: $\ln(MINVR) = a_0 + (a_1)(MINHR)$ (5)

In these models, MINVR is minute ventilation rate and MINHR is minute heart rate.

Linear regression analysis was used to fit each of the four models to the calibration data for each subject. In each analysis, the parameter on the left side of the equals sign was considered the dependent variable. The parameter on the right side was considered the independent variable. For example, ln(MINVR) and ln(MINHR) were the dependent and independent variables, respectively, of the regression equation whenever the log-log model (C) was fit to data.

Table 2-2. Descriptive statistics for minute heart rate values measured during the activity diary phase of construction worker study.

PID	Number of heart rate measurements	Measured minute heart rate, beats/min				
		Arithmetic				
		Mean	Std. dev.	Minimum	Median	Maximum
1761	548	88	12.6	53	86	155
1763	270	98	12.3	69	99	130
1764	584	69	9.9	42	68	133
1765	446	91	22.6	40	91	168
1766	729	83	9.4	63	82	139
1767	558	78	20.8	40	79	183
1768	735	74	14.2	45	73	154
1769	638	100	20.6	66	96	197
1770	644	88	17.3	50	87	144
1771	595	109	17.4	62	112	142
1772	616	109	15.5	60	110	160
1773	687	100	17.3	69	98	154
1774	214	95	18.1	67	91	181
1775	651	99	16.1	57	100	141
1776	624	95	15.8	61	96	142
1778	641	100	16.0	63	100	159
1779	146	111	12.7	87	110	164
1780	529	88	10.5	61	89	121
1781	615	85	12.3	54	86	120

Regression analyses based on Models A, C, and D produced estimates of two regression coefficients: a_0 and a_1. Three regression coefficients were obtained from analyses using Model B (a_0, a_1, and a_2). Table 2-3 lists the values of these coefficients for each combination of subject and model. The table also provides a goodness-of-fit statistic for each fit.

The goodness-of-fit statistic is the R^2 value obtained from a regression analysis of measured MINVR values versus predicted MINVR values. The predicted MINVR values were determined by substituting each measured MINHR value into the specified model and using the model to calculate a corresponding MINVR value.

For the majority of subjects, the quadratic model (Model B) yielded the best goodness-of-fit value of the four tested models. This result was expected, as the quadratic is the only model tested that has three regression coefficients. The additional coefficient in the quadratic equation improves the model's ability to fit data.

The two-coefficient models (A,C, or D) are preferable as the basis for constructing calibration curves, as these models are more likely to produce relationships between MINVR and MINHR that increase monotonically throughout the range of applied MINHR values. If the three-coefficient quadratic (Model B) is excluded from consideration, the subject-specific breakdown of the best and worst-fitting models is as follows:

Model	Best fit	Worst Fit
Model A (linear)	5	10
Model C (log-log)	9	0
Model D (log-linear)	6	10

Note that Model C (the log-log model) provides the greatest number of "best" fits (9) among the 20 subjects and the smallest number of "worst" fits (0) of the three models.

Table 2-3. Results of fitting four general models to the construction worker calibration data.

PID	Model	a_0 Value	a_0 p	a_1 Value	a_1 p	a_2 Value	a_2 p	Goodness of fit statistic
1761	A	-52.01	0.0000	0.87	0.0000	-	-	0.8998
	B	14.34	0.7979	-0.64	0.6144	0.0084	0.2428	0.9082
	C	-11.21	0.0000	3.20	0.0000	-	-	0.9066[a]
	D	-0.091	0.7571	0.0361	0.0000	-	-	0.8984[b]
1763	A	-58.99	0.0000	0.86	0.0000	-	-	0.9356[b]
	B	26.76	0.5661	-0.94	0.3402	0.0093	0.0764	0.9468
	C	-13.34	0.0000	3.60	0.0000	-	-	0.9451[a]
	D	-0.536	0.0141	0.0376	0.0000	-	-	0.9379
1764	A	-40.28	0.0000	0.95	0.0000	-	-	0.9036[b]
	B	85.45	0.0135	-2.61	0.0083	0.0247	0.0008	0.9514
	C	-6.76	0.0000	2.36	0.0000	-	-	0.9320
	D	0.877	0.0000	0.0334	0.0000	-	-	0.9450[a]
1765	A	-36.88	0.0000	0.72	0.0000	-	-	0.9502[b]
	B	21.61	0.2050	-0.53	0.1437	0.0064	0.0014	0.9661
	C	-6.44	0.0000	2.16	0.0000	-	-	0.9642[a]
	D	1.183	0.0000	0.0228	0.0000	-	-	0.9633
1766	A	-56.30	0.0000	0.91	0.0000	-	-	0.8861[a]
	B	-39.78	0.6325	0.55	0.7556	0.0019	0.8415	0.8864
	C	-10.81	0.0000	3.10	0.0000	-	-	0.8812
	D	0.178	0.4658	0.0328	0.0000	-	-	0.8740[b]
1767	A	-9.81	0.0002	0.44	0.0000	-	-	0.9135[b]
	B	27.98	0.0062	-0.49	0.0419	0.0054	0.0004	0.9465
	C	-2.43	0.0000	1.29	0.0000	-	-	0.9222
	D	1.930	0.0000	0.0158	0.0000	-	-	0.9405[a]
1768	A	-23.93	0.0000	0.57	0.0000	-	-	0.9477[a]
	B	-28.93	0.0144	0.67	0.0043	-0.0005	0.6452	0.9481
	C	-4.74	0.0000	1.78	0.0000	-	-	0.9359
	D	1.684	0.0000	0.0171	0.0000	-	-	0.8929[b]
1769	A	-66.58	0.0000	1.07	0.0000	-	-	0.9713
	B	-7.62	0.6403	-0.17	0.6097	0.0062	0.0009	0.9810
	C	-11.16	0.0000	3.20	0.0000	-	-	0.9761[a]
	D	0.237	0.1667	0.0326	0.0000	-	-	0.9620[b]
1770	A	-17.80	0.0000	0.49	0.0000	-	-	0.8530[b]
	B	44.93	0.0287	-1.01	0.0374	0.0088	0.0030	0.8946
	C	-4.19	0.0000	1.66	0.0000	-	-	0.8692
	D	1.447	0.0000	0.0200	0.0000	-	-	0.8852[a]

(continued)

Table 2-3 (Continued)

PID	Model	Regression results						Goodness of fit statistic
		a_0		a_1		a_2		
		Value	p	Value	p	Value	p	
1771	A	-76.52	0.0000	0.93	0.0000	-	-	0.8414[a]
	B	-262.20	0.0033	4.34	0.0070	-0.0155	0.0297	0.8673
	C	-20.94	0.0000	5.12	0.0000	-	-	0.7581
	D	-1.994	0.0003	0.0461	0.0000	-	-	0.7327[b]
1772	A	-64.17	0.0000	0.93	0.0000	-	-	0.9587[b]
	B	42.17	0.2190	-1.06	0.1000	0.0091	0.0036	0.9700
	C	-9.71	0.0000	2.83	0.0000	-	-	0.9692[a]
	D	0.660	0.0000	0.0263	0.0000	-	-	0.9668
1773	A	-41.93	0.0000	0.66	0.0000	-	-	0.9341[a]
	B	-83.84	0.0021	1.48	0.0045	-0.0039	0.0962	0.9409
	C	-10.17	0.0000	2.89	0.0000	-	-	0.8988
	D	0.327	0.1008	0.0277	0.0000	-	-	0.8699[b]
1774	A	-15.63	0.0000	0.44	0.0000	-	-	0.9483[b]
	B	20.09	0.0017	-0.29	0.0205	0.0036	0.0000	0.9792
	C	-3.39	0.0000	1.46	0.0000	-	-	0.9636
	D	1.806	0.0000	0.0150	0.0000	-	-	0.9784[a]
1775	A	-76.98	0.0000	1.05	0.0000	-	-	0.8143[a]
	B	-7.87	0.9038	-0.40	0.7684	0.0075	0.2880	0.8221
	C	-21.19	0.0000	5.28	0.0000	-	-	0.8010
	D	-2.385	0.0000	0.0544	0.0000	-	-	0.7876[b]
1776	A	-45.23	0.0000	0.75	0.0000	-	-	0.8076[b]
	B	42.54	0.1752	-0.95	0.1145	0.0077	0.0068	0.8556
	C	-6.74	0.0006	2.16	0.0000	-	-	0.8442
	D	1.057	0.0066	0.0209	0.0000	-	-	0.8548[a]
1777[c]	A	-68.62	0.0000	0.80	0.0000	-	-	0.9629[b]
	B	24.39	0.1509	-0.73	0.0114	0.0061	0.0000	0.9832
	C	-14.21	0.0000	3.63	0.0000	-	-	0.9814[a]
	D	-0.421	0.0196	0.0292	0.0000	-	-	0.9701
1778	A	-49.25	0.0000	0.86	0.0000	-	-	0.8739
	B	-1.71	0.9578	-0.15	0.8253	0.0052	0.1433	0.8837
	C	-8.08	0.0000	2.53	0.0000	-	-	0.8815[a]
	D	0.826	0.0000	0.0269	0.0000	-	-	0.8565[b]
1779	A	-41.42	0.0000	0.77	0.0000	-	-	0.9828
	B	-11.08	0.2628	0.17	0.3902	0.0029	0.0037	0.9875
	C	-6.62	0.0000	2.20	0.0000	-	-	0.9862[a]
	D	1.323	0.0000	0.0215	0.0000	-	-	0.9730[b]

(continued)

Table 2-3 (Continued)

PID	Model	Regression results						Goodness of fit statistic
		a_0		a_1		a_2		
		Value	p	Value	p	Value	p	
1780	A	-40.68	0.0000	0.79	0.0000	-	-	0.8980[b]
	B	100.76	0.0000	-2.29	0.0000	0.0163	0.0000	0.9832
	C	-5.98	0.0000	2.08	0.0000	-	-	0.9388
	D	1.277	0.0000	0.0229	0.0000	-	-	0.9649[a]
1781	A	-25.39	0.0000	0.56	0.0000	-	-	0.8831
	B	-10.22	0.6868	0.23	0.6805	0.0018	0.5453	0.8847
	C	-6.05	0.0000	2.05	0.0000	-	-	0.8838[a]
	D	1.158	0.0000	0.0220	0.0000	-	-	0.8809[b]

[a]Best fit (excluding Model B).
[b]Worst fit.
[c]No activity diary data available.

To assist in selecting a reasonable general model to be applied to all subjects, ITAQS developed calibration curves for subjects chararacterized by good, average, and poor fits. Figure 2-1 presents the calibration curves for Subject No. 1779 produced by Models A, C, and D when MINHR is varied from 75 to 175 beats/minute. This subject was associated with high goodness-of-fit statistics (0.9730 to 0.9875) for all four models. Figure 2-2 presents calibration curves for Subject No. 1766 when MINHR is varied from 50 to 150 beats/minute. Subject No. 1766 was associated with average goodness-of fit statistics (0.8812 to 0.8864) for the four models. The lowest goodness-of-fit statistics were associated with Subject No. 1771 (0.7327 to 0.8673). Figure 2-3 provides calibration curves for this subject when MINHR is varied from 50 to 150 beats/minute. Note that the range of MINHR values appearing in each figure approximates the range of MINHR values measured during the activity diary phase of the study (Table 2-2).

Ideally, the calibration curve selected for a subject should produce an estimate of ventilation rate (V_e) at maximum aerobic power (MAP) that is typical of persons of the same age and gender. ITAQS estimated a <u>theoretical</u> value for V_e at MAP for each subject using the following procedure:

1. Determine the subject's age and gender. (Note: the three subjects in this exercise are males aged 31 to 40.)

2. Use the relationship presented in Figure 9-14 of Astrand and Rodahl[11] to estimate MINHR at MAP for indicated age and gender.

3. Use the relationship presented in Figure 10-11 of Astrand and Rodahl to estimate the oxygen uptake rate (VO_2) at MAP for the indicated heart rate.

4. Use Table 3 (males) or Table 4 (females) in an unpublished paper by Johnson and Adams (Appendix B) to estimate the V_e-to-VO_2 ratio at MAP.

5. Multiply the VO_2 value determined in Step 3 by the V_e-to-VO_2 ratio determined in Step 4 to estimate MINVR at MAP.

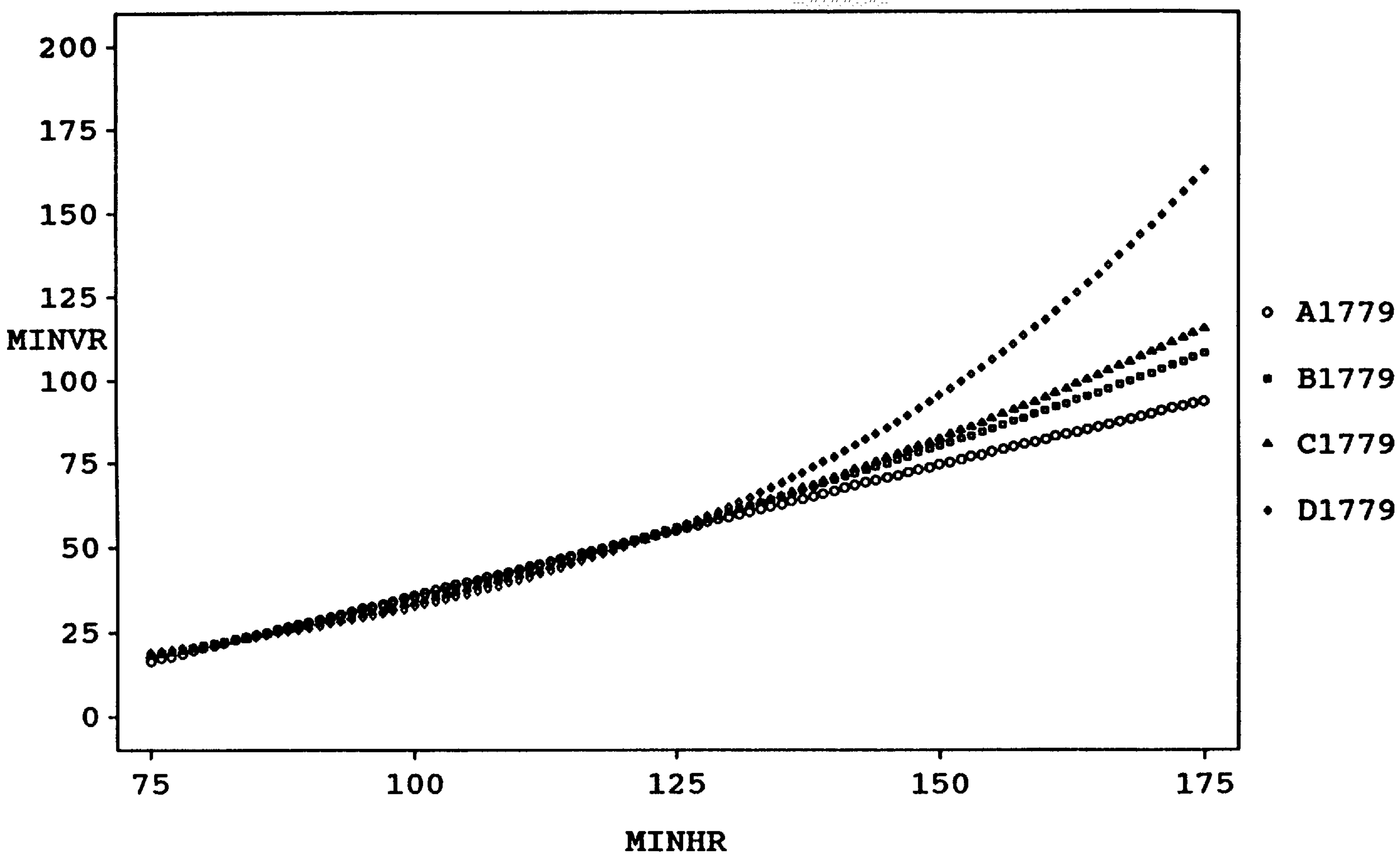

Figure 2-1. Calibration curves relating minute ventilation rate (MINVR) to minute heart rate (MINHR) developed for Subject No. 1779 through the use of Models A, B, C, and D.

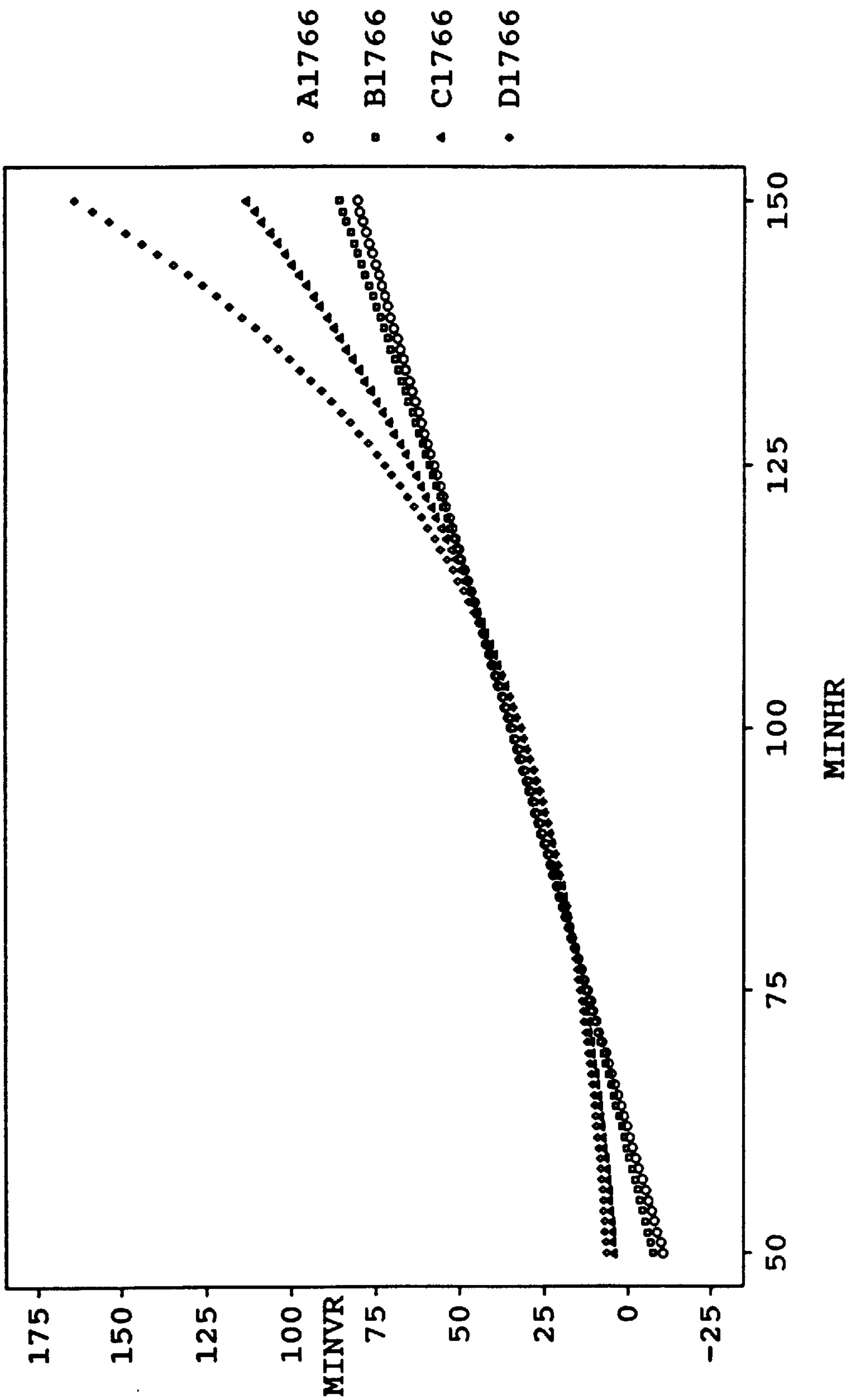

Figure 2-2. Calibration curves relating minute ventilation rate (MINVR) to minute heart rate (MINHR) developed for Subject No. 1766 through the use of Models A, B, C, and D.

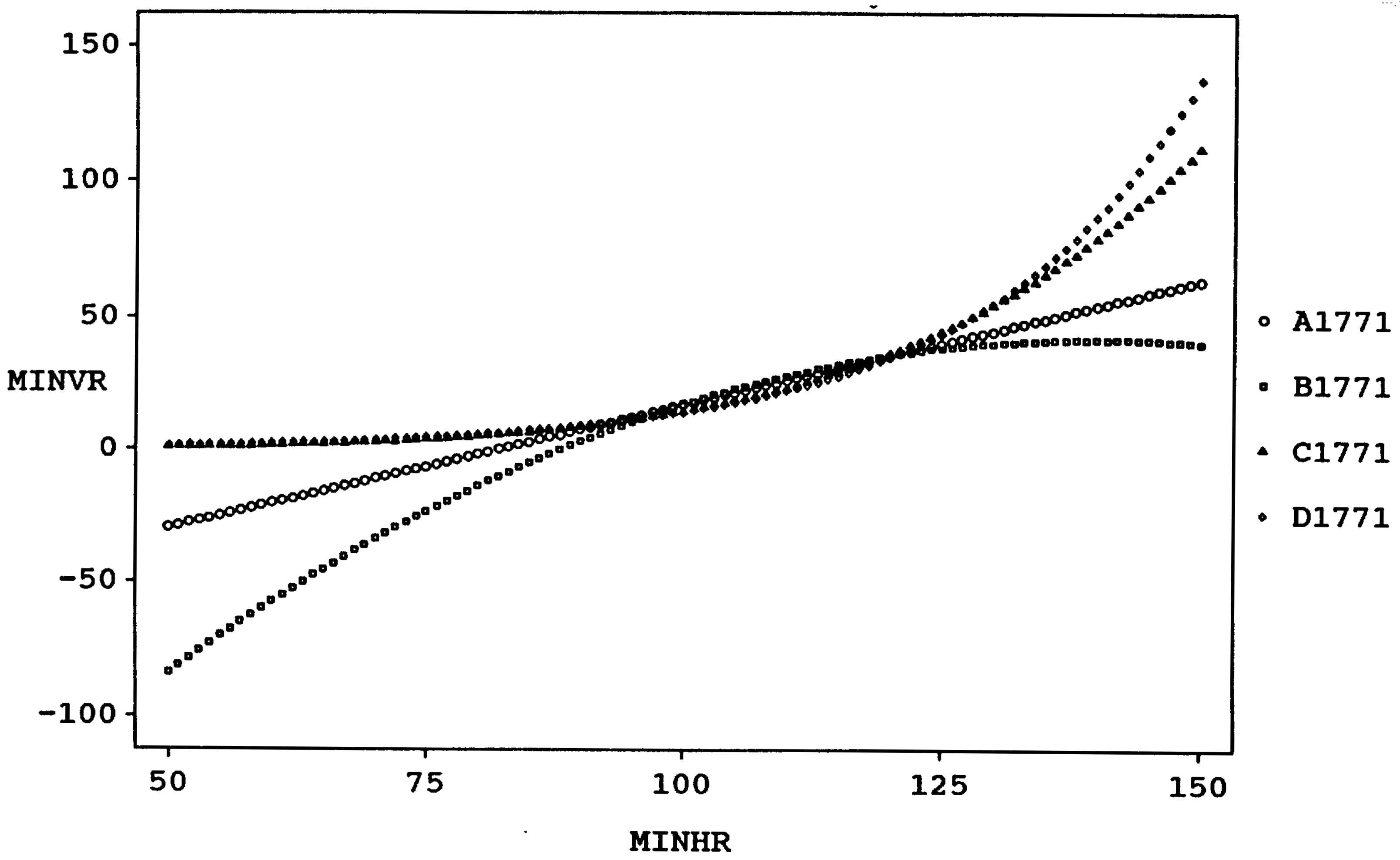

Figure 2-3. Calibration curves relating minute ventilation rate (MINVR) to minute heart rate (MINHR) developed for Subject No. 1771 through the use of Models A, B, C, and D.

Table 2-4 in this report presents the results of these calculations.

To evaluate the reasonableness of each model, researchers used the following procedure to estimate MINVR at MAP for each subject.

1. Note the value of MINHR at MAP determined by Step 2 above.

2. Insert the MINHR value in the calibration equation (A, B, C, or D) and calculate MINVR at MAP.

The resulting estimates are listed in Table 2-4 under the column heading "model-derived MINVR."

When the theoretical estimates of MINVR at MAP are compared to the model-derived estimates of MINVR at MAP, Model B produces the best matchups for Subject 1779 (the good fit example) and for Subject 1766 (the average fit example). Model A produces the best matchup for Subject 1781 (the poor fit example). Model C yields impossibly high values for MINVR at MAP for Subject 1766 (234.1 liters/minute) and for Subject 1771 (291.9 liters/minute). Model D yields even higher estimates for these two subjects (608 liters/minute for Subject 1766 and 575.5 liters/minute for Subject 1771). The Model B estimate for Subject 1771 is too low (15.5 liters/minute).

Analysts also used the four models to estimate the MINVR corresponding to the highest MINHR value reported for each of the three subjects during the diary period (164 beats/minute for Subject 1779, 139 beats/minute for Subject 1766, and 142 beats/minute for Subject 1771). Table 2-4 lists these estimates by subject and model.

The three calibration data sets share a common deficiency in that the largest MINHR value measured during the calibration phase of the study is less than one or more of the MINHR values measured during the activity diary phase. Consequently,

Table 2-4. Theoretical and model-derived estimates of minute ventilation rate (MINVR) associated with three subjects of the construction worker study.

Subject ID	Age	Fit category	Theoretical values at MAP: HR	VO_2	V_e/VO_2	V_e	Model	Theoretical maximum MINHR	Model-derived MINVR	Maximum measured MINHR	Model-derived MINVR
1779	40	Good	180	3.50	33.2	116	A	180	97.2	164	84.9
							B	180	113.5	164	94.8
							C	180	122.1	164	99.5
							D	180	137.0	164	99.5
1766	31	Average	190	4.00	33.2	133	A	190	116.6	139	70.2
							B	190	133.3	139	73.4
							C	190	234.1	139	88.8
							D	190	607.9	139	114.1
1771	39	Poor	181	3.60	33.2	120	A	181	91.8	142	55.5
							B	181	15.5	142	210.8
							C	181	291.9	142	84.3
							D	181	575.5	142	94.8

the calibration curve is being applied to MINHR values beyond the range of the calibration data. In this region, exercise physiologists would expect the slope of the relationship between MINVR and MINHR to be positive and to increase as MINHR increases. Models C and D are consistent with this expectation, as the slope of each curve increases as MINHR increases. However, these two models produce impossibly high MINVR estimates at MAP conditions for the three subjects listed in Table 2-4. Model B (the quadratic model) produced a negative slope between MINVR and MINHR for Subject 1771 in the region above the calibration range. Model A (the linear model) has a constant slope at all values of MINHR.

Models B and D perform too poorly to merit further consideration. Of the two remaining models, Model A underestimates MINVR at MAP and Model C overestimates MINVR at MAP. The unknown "correct" relationship is likely to be a monotonically increasing function that lies between these two models. Rather than attempt to guess this function, analysts applied Models A and C independently to the construction worker data to produce two separate EVR data sets. The EVR values derived using Model A are considered to represent lower bound estimates. The EVR values derived using Model C represent upper bound estimates.

THE EVENT EVR FILES

ITAQS combined the MINHR, MINVR, time/activity, and demographic data associated with each of the Hackney/Linn studies to produce a special "event EVR file" for the study. Each event EVR file lists average EVR values by event rather than by minute. In constructing each file, an event was assumed to begin whenever a subject changed activity, microenvironment, or breathing rate. An event was also assumed to begin whenever there was more than a one minute gap in the diary entries.

The EVR values listed for each subject were calculated by dividing the subject's MINVR values by the subject's estimated body surface area (BSA). BSA was estimated by the relationship

$$BSA = (0.007182)(H)^{0.725}\,(W)^{0.425} \tag{6}$$

where "H" is height in centimeters and "W" is weight in kilograms. This relationship was initially proposed by Dubois and Dubois[12] in 1916, confirmed by Anderson et al.,[13] and has been used in previous analyses by ITAQS[5] and EPA researchers.[4]

Table 2-5 lists the BSA values calculated for each of the 16 subjects of the elementary school study. It also provides subject-specific data on age, gender, race, height, and weight. In addition, Table 2-5 presents statistics by subject for the number of days monitored, the total number of events defined for the EVR file, the average number of events per day, and the average event duration. The average event duration provides a rough indicator of the level of detail available in the subject's diary, with smaller values indicating a greater level of detail. Tables 2-6, 2-7, and 2-8 present similar tabulations for the high school, outdoor worker, and construction worker studies, respectively.

Table 2-9 lists 20 variables that have been included in each of the special event EVR data sets. Data Items 6 through 12 were not available in the original data sets obtained from Dr. Linn. These items were added because (1) they were already available through a previous API task and (2) they provided additional information which assisted researchers in identifying patterns in the EVR data. All data items required special processing by ITAQS to produce a consistent data format across all four studies. This processing was performed on the NCC mainframe IBM computer using Fortran 77 programs developed specifically for this project. Appendix C provides additional information concerning each data item.

DESCRIPTIVE STATISTICS FOR EVENT EVR VALUES BY SUBJECT

Tables 2-10, 2-11, and 2-12 present descriptive statistics by subject for the event EVR values associated with the elementary school students, the high school students, and the outdoor workers, respectively. Table 2-13 presents descriptive statistics for the construction workers based on application of Model A. Model C results for the construction workers are presented in Table 2-14. Table 2-15 presents results where the Model A and Model C estimate for each event have been averaged.

The file containing elementary school data contained four event EVR values above 105 liters$\cdot$min$^{-1}\cdot$m^{-2}, all associated with Subject No. 1606. As these values appear to be outliers, Table 2-10 presents descriptive statistics for this subject with and without these four values.

Subject No. 1719 of the high school student study is associated with only 15 event EVR values. Each of the other subjects of this study has at least 60 event EVR values. Table 2-11 presents group descriptive statistics for the high school students with and without Subject No. 1719.

A comparison of Tables 2-13 and 2-14 indicates that Models A and C produce similar results for the construction workers with respect to measures of central tendency (means and medians). The models, however, differ significantly with respect to estimates of minimum and maximum values. The maximums produced by the two models differ by more than 10 percent for 11 of the 19 subjects. The difference exceeds 20 percent for six subjects and 50 percent for two subjects (No. 1769 and No. 1775). Note that Model A yields unreasonably low values for some of the minimums. For example, the minimum for Subject No. 1765 is 0.30 liters$\cdot$min$^{-1}\cdot$m^{-2}.

Table 2-5. Characteristics of the subjects of the elementary school study and associated data statistics.

	Subject characteristics						Data statistics			
PID	Age, years	Gender	Race[a]	Height, cm	Weight, kg	BSA[b]	Number of days monitored	Number of events	Average number of events per day	Average event duration, min
1601	11	F	3	142	28.6	1.0863	3	55	18.3	29.62
1602	11	F	3	142	30.8	1.1222	3	117	39.0	20.66
1603	11	F	3	152	36.7	1.2708	2	40	20.0	42.20
1604	11	F	3	160	71.7	1.7488	3	91	30.3	26.56
1605	11	F	3	155	39.0	1.3193	3	103	34.3	24.26
1606	11	M	3	142	29.0	1.0936	2	66	33.0	25.80
1607	11	F	4	147	44.9	1.3503	3	89	29.7	28.46
1609	10	F	4	155	26.3	1.3946	2	39	19.5	41.03
1610	10	F	1	163	54.4	1.5737	2	42	21.0	27.93
1611	11	F	2	160	48.1	1.4760	2	32	16.0	33.84
1612	12	M	2	163	51.3	1.5341	2	26	13.0	25.15
1614	12	F	3	150	53.5	1.4731	1	14	14.0	113.79
1615	12	F	3	152	42.6	1.3538	1	21	21.0	19.05
1616	12	M	3	155	39.0	1.3193	1	42	42.0	29.07
1617	12	M	2	165	44.9	1.4666	2	66	33.0	28.30
1618	12	F	3	160	49.0	1.4877	1	31	31.0	31.77

[a]1: white, 2: black, 3: Asian, 4: other, 5: other.
[b]Body surface area, m^2.

Table 2-6. Characteristics of the subjects of the high school study and associated data statistics.

	Subject characteristics						Data statistics			
PID	Age, years	Gender	Race[a]	Height, cm	Weight, kg	BSA[b]	Number of days monitored	Number of events	Average number of events per day	Average event duration, min
939	16	F	1	168	51.3	1.5678	3	123	41.0	23.21
1701	17	F	1	171	61.2	1.7196	3	179	59.7	15.20
1702	14	F	4	157	50.8	1.4935	2	65	32.5	14.60
1703	15	F	2	169	68.5	1.7841	3	105	35.0	20.39
1704	17	F	1	157	56.7	1.5648	3	112	37.3	17.71
1705	17	F	3	165	57.2	1.6249	3	190	63.3	13.36
1706	14	F	4	163	58.1	1.6175	3	85	28.3	16.74
1707	16	F	2	174	73.5	1.8781	3	121	40.3	20.60
1708	16	M	1	183	75.3	1.9675	3	60	20.0	28.15
1709	15	M	1	173	100.2	2.1318	3	133	44.3	15.71
1711	17	M	3	163	52.6	1.5512	3	99	33.0	26.41
1712	14	F	1	163	70.8	1.7594	3	64	21.3	34.23
1713	14	F	4	160	44.5	1.4276	3	122	40.7	23.27
1714	15	M	1	165	64.0	1.7044	3	78	26.0	19.18
1715	15	M	1	185	66.7	1.8872	3	139	46.3	17.34
1716	15	F	3	155	49.4	1.4591	3	149	49.7	14.60
1717	14	F	3	155	47.6	1.4361	3	104	34.7	17.52
1718	16	M	1	182	62.6	1.8098	3	112	37.3	19.02
1719[c]	13	M	3	165	49.4	1.5278	1	15	15.0	16.47

[a]1: white, 2: black, 3: Asian, 4: other, 5: other.
[b]Body surface area, m^2.
[c]Subject reported questionable diary data.

Table 2-7. Characteristics of the subjects of the outdoor worker study and associated data statistics.

	Subject characteristics						Data statistics			
PID	Age, years	Gender	Race[a]	Height, cm	Weight, kg	BSA[b]	Number of days monitored	Number of events	Average number of events per day	Average event duration, min
124	32	M	9	180	94.3	2.1436	3	113	37.7	22.01
986	38	M	1	168	72.1	1.8137	3	110	36.7	25.66
1073	28	M	4	157	63.5	1.6421	3	92	30.7	29.45
1244	21	M	1	175	57.6	1.7025	2	69	34.5	25.74
1246	38	M	4	178	83.0	2.0093	3	173	57.7	21.16
1290	32	M	1	175	68.9	1.8376	3	128	42.7	30.67
1291	19	F	4	160	42.2	1.3961	3	100	33.3	29.73
1299	28	M	1	185	92.5	2.1692	3	178	59.3	19.66
1300	23	F	1	155	52.2	1.4927	3	122	40.7	23.04
1304	35	M	1	173	68.0	1.8080	2	46	23.0	52.17
1306	21	F	4	157	52.6	1.5159	3	70	23.3	38.57
1308	33	M	1	173	77.6	1.9116	5	121	24.2	24.23
1319	44	M	2	183	88.0	2.1022	3	52	17.3	45.89
1350	26	M	2	183	96.6	2.1874	3	120	40.0	26.93
1361	29	M	4	178	97.1	1.1475	4	128	32.0	36.10
1372	26	M	1	180	54.0	1.6907	2	72	36.0	33.99
1404	50	M	1	173	74.8	1.8828	2	94	47.0	25.03
1490	24	F	1	170	66.2	1.7683	3	141	47.0	21.45
1494	32	M	1	185	91.2	2.1556	3	114	38.0	29.35
1496	24	F	1	165	65.8	1.7248	4	147	36.8	25.10

[a]1: white, 2: black, 3: Asian, 4: other, 5: other.
[b]Body surface area, m^2.

Table 2-8. Characteristics of the subjects of the construction worker study and associated data statistics.

	Subject characteristics						Data statistics			
PID	Age, years	Gender	Race[a]	Height, cm	Weight, kg	BSA[b]	Number of days monitored	Number of events	Average number of events per day	Average event duration, min
1761	26	F	1	180	81.6	2.0158	1	81	81.0	6.46
1763	29	M	3	160	61.2	1.6357	1	51	51.0	4.98
1764	32	F	2	180	74.8	1.9426	1	127	127.0	4.35
1765	30	M	1	185	65.8	1.8763	1	114	114.0	3.75
1766	31	F	4	170	77.1	1.8865	1	121	121.0	6.07
1767	34	F	1	188	99.8	2.2622	1	115	115.0	4.86
1768	32	M	2	175	70.3	1.8529	1	62	62.0	11.63
1769	32	M	1	196	104.3	2.3727	1	160	160.0	3.65
1770	26	M	1	175	81.6	1.9952	1	109	109.0	5.89
1771	39	M	1	168	68.0	1.7693	1	111	111.0	4.50
1772	32	M	1	180	117.9	2.3568	1	129	129.0	4.71
1773	39	M	1	175	77.1	1.9271	1	110	110.0	6.24
1774	23	M	4	173	68.0	1.8080	1	13	13.0	10.00
1775	42	M	1	170	68.0	1.7887	1	132	132.0	4.57
1776	29	F	4	178	81.6	1.9952	1	72	72.0	8.69
1778	35	M	4	193	99.8	2.3063	1	111	111.0	5.77
1779	40	M	1	178	79.4	1.9715	1	48	48.0	2.98
1780	37	M	4	191	109.8	2.3787	1	113	113.0	4.31
1781	38	F	4	165	74.8	1.8222	1	82	82.0	7.50

[a]1: white, 2: black, 3: Asian, 4: other, 5: other.
[b]Body surface area, m^2.

Table 2-9. Data items included in event EVR files prepared by ITAQS.

Item No.	Description	Item No.	Description
1	Personal ID	11	Air Conditioner
2	Month	12	Gas Stove
3	Day	13	Season
4	Year	14	Daytype
5	Starting Time	15	Maximum Temperature
6	Demographic Group	16	Activity Classification Code
7	Gender	17	Microenvironment
8	Race	18	Breathing Rate
9	Income	19	Duration
10	Attached Garage	20	Average EVR for Event

Table 2-10. Descriptive statistics for equivalent ventilation rates averaged by event obtained from elementary school student study.

Subject identification code	Number of events	Event equivalent ventilation rate, liters·min^{-1}·m^{-2}						
		Arithmetic		Geometric				
		Mean	Std. dev.	Mean	Std. dev.[a]	Minimum	Median	Maximum
1601	55	11.34	5.28	10.65	1.376	7.28	9.94	38.73
1602	117	12.82	2.48	12.61	1.194	8.79	12.20	21.22
1603	40	21.43	3.76	21.15	1.170	17.21	20.19	34.08
1604	91	9.54	3.75	9.07	1.344	5.46	8.58	30.83
1605	103	10.74	4.08	10.22	1.345	6.29	9.52	33.49
1606	66	22.76	25.63	16.54	1.972	8.39	12.78	107.76
1606[b]	62	17.36	14.50	14.67	1.648	8.39	12.70	86.04
1607	89	18.42	3.07	18.20	1.163	14.17	17.42	28.88
1609	39	15.86	4.64	15.19	1.350	8.64	14.90	24.38
1610	42	15.04	3.49	14.65	1.260	9.34	14.23	23.55
1611	32	11.71	2.02	11.56	1.168	8.64	11.31	18.77
1612	26	6.08	1.31	5.96	1.214	4.82	5.75	9.35
1614	14	8.06	1.31	7.95	1.189	5.32	8.13	9.78
1615	21	11.67	1.03	11.62	1.093	9.89	11.69	13.56
1616	42	10.42	2.37	10.17	1.246	7.05	9.85	16.91
1617	66	6.89	2.21	6.57	1.360	3.57	6.55	13.26
1618	31	4.06	1.16	3.94	1.267	2.80	3.94	8.74
All	874	12.88	9.10	11.32	1.618	2.80	11.26	107.76
All[b]	870	12.45	6.53	11.20	1.580	2.80	11.22	86.04
All[c]	808	12.07	5.29	10.97	1.565	2.80	11.13	38.73

[a]Dimensionless.
[b]Omits four largest values (105.31, 106.12, 106.92, and 107.77).
[c]Omits Subject No. 1606.

Table 2-11. Descriptive statistics for equivalent ventilation rates averaged by event obtained from high school student study.

Subject identification code	Number of events	Event equivalent ventilation rate, liters·min^{-1}·m^{-2}						
		Arithmetic		Geometric				
		Mean	Std. dev.	Mean	Std. dev.[a]	Minimum	Median	Maximum
939	123	7.70	2.48	7.38	1.327	4.46	7.01	15.94
1701	179	8.47	2.81	8.11	1.325	4.69	7.79	25.01
1702	65	10.64	4.87	9.79	1.487	4.02	8.93	29.68
1703	105	9.40	4.11	8.86	1.373	5.42	8.11	38.54
1704	112	11.09	1.80	10.96	1.166	8.31	10.69	16.62
1705	190	9.67	2.68	9.39	1.261	6.15	8.92	23.39
1706	85	7.76	1.19	7.67	1.170	4.95	7.84	10.92
1707	121	8.21	3.22	7.67	1.441	3.73	7.62	18.64
1708	60	14.40	10.79	12.03	1.721	6.98	9.67	48.67
1709	133	13.18	2.59	12.94	1.211	8.44	12.67	21.16
1711	99	7.66	2.71	7.26	1.384	3.74	7.00	17.57
1712	64	8.30	2.57	7.99	1.299	5.49	7.44	17.73
1713	122	8.06	2.61	7.72	1.332	4.36	7.36	17.65
1714	78	8.92	2.29	8.67	1.264	5.75	8.51	19.60
1715	139	8.59	2.86	8.26	1.304	4.75	7.95	24.96
1716	149	6.39	1.97	6.13	1.325	3.77	6.05	14.74
1717	104	10.59	2.42	10.35	1.238	6.84	10.28	19.96
1718	112	9.82	2.96	9.44	1.316	6.02	8.84	20.87
1719	15	7.44	0.91	7.39	1.129	6.04	7.32	9.32
All	2055	9.21	3.75	8.66	1.400	3.73	8.41	48.67
All[b]	2040	9.22	3.76	8.67	1.401	3.73	8.42	48.67

[a]Dimensionless.
[b]Omits Subject No. 1719.

Table 2-12. Descriptive statistics for equivalent ventilation rates averaged by event obtained from outdoor worker study.

Subject identification code	Number of events	Event equivalent ventilation rate, liters·min^{-1}·m^{-2}						
		Arithmetic		Geometric				
		Mean	Std. dev.	Mean	Std. dev.[a]	Minimum	Median	Maximum
124	113	10.30	12.29	8.02	1.745	2.25	6.99	74.67
986	110	8.73	13.42	6.85	1.649	3.75	6.26	118.51
1073	92	11.34	7.96	9.97	1.550	5.83	8.95	49.08
1244	69	13.77	6.43	12.82	1.420	7.65	11.99	44.88
1246	173	8.82	4.71	8.38	1.303	6.02	8.07	58.31
1290	128	13.04	16.23	9.12	2.055	3.96	6.55	109.65
1291	100	9.33	5.09	8.67	1.398	5.28	7.84	41.53
1299	178	18.97	21.45	13.13	2.127	5.65	9.39	119.47
1300	122	6.59	2.86	6.24	1.351	3.92	5.84	29.99
1304	46	6.10	1.58	5.92	1.273	3.60	5.85	11.72
1306	70	5.82	1.08	5.72	1.201	3.68	5.70	9.15
1308	121	12.14	6.90	11.10	1.454	7.31	9.91	45.83
1319	52	8.52	2.61	8.23	1.277	6.20	7.71	20.53
1350	120	8.90	4.09	8.33	1.395	4.57	7.67	31.91
1361	128	9.08	2.68	8.83	1.247	5.45	8.94	31.62
1372	72	5.34	1.85	5.11	1.322	3.25	4.90	16.08
1404	94	11.67	13.02	9.74	1.568	6.62	8.73	88.42
1490	141	5.09	1.02	4.99	1.214	3.14	4.87	9.12
1494	114	6.68	10.51	5.20	1.649	3.40	4.61	84.86
1496	147	9.96	5.51	9.22	1.403	5.51	8.72	40.26
All	2190	9.90	10.42	8.15	1.676	2.25	7.71	119.47

[a]Dimensionless.

Table 2-13. Descriptive statistics for equivalent ventilation rates averaged by event obtained from construction worker study (Model A).

Subject identification code	Number of events	Event equivalent ventilation rate, liters·min^{-1}·m^{-2}						
		Arithmetic		Geometric				
		Mean	Std. dev.	Mean	Std. dev.[a]	Minimum	Median	Maximum
1761	81	12.70	3.86	12.07	1.403	2.68	12.83	27.72
1763	51	15.87	5.11	15.01	1.419	6.00	15.20	30.33
1764	127	13.43	3.25	13.02	1.293	4.69	13.33	20.83
1765	114	15.92	6.64	14.42	1.696	0.30	15.32	44.04
1766	121	11.34	3.38	10.86	1.343	5.24	10.92	22.01
1767	115	10.33	3.12	9.84	1.382	3.83	10.22	21.34
1768	62	10.67	3.44	10.17	1.375	3.31	10.28	26.15
1769	160	17.35	7.14	16.02	1.493	6.17	16.13	46.35
1770	109	13.62	3.64	13.14	1.313	6.80	13.67	25.71
1771	111	15.87	7.14	12.97	2.297	0.38	16.50	28.24
1772	129	18.42	4.65	17.77	1.325	6.13	18.55	30.25
1773	110	13.92	5.52	12.84	1.514	4.27	13.09	27.22
1774	13	14.47	3.44	14.04	1.303	7.66	15.70	19.29
1775	132	18.50	8.44	15.16	2.223	0.40	19.96	36.50
1776	72	15.31	4.19	14.73	1.329	6.38	14.92	26.95
1778	111	18.04	5.35	17.29	1.344	7.40	17.29	37.94
1779	48	22.74	3.61	22.49	1.160	15.94	22.76	38.55
1780	113	12.67	2.93	12.28	1.303	5.55	12.79	20.10
1781	82	13.87	2.93	13.54	1.257	5.77	14.03	22.64
All	1861	15.00	5.89	13.74	1.587	0.30	14.17	46.35

[a]Dimensionless.

Table 2-14. Descriptive statistics for equivalent ventilation rates averaged by event obtained from construction worker study (Model C).

Subject identification code	Number of events	Event equivalent ventilation rate, liters·min^{-1}·m^{-2}						
		Arithmetic		Geometric				
		Mean	Std. dev.	Mean	Std. dev.[a]	Minimum	Median	Maximum
1761	81	12.56	4.48	11.88	1.395	4.46	12.52	34.51
1763	51	15.92	5.72	14.99	1.420	7.00	14.78	36.17
1764	127	13.82	3.09	13.48	1.253	6.69	13.50	21.34
1765	114	16.17	7.19	14.97	1.466	4.33	14.85	53.12
1766	121	10.80	2.91	10.46	1.285	6.40	10.18	22.32
1767	115	10.43	2.86	10.04	1.323	4.83	10.21	21.17
1768	62	10.96	3.09	10.61	1.280	5.49	10.45	26.20
1769	160	17.29	10.43	15.19	1.625	6.26	14.45	78.41
1770	109	14.01	3.76	13.53	1.302	7.59	13.81	28.06
1771	111	16.91	8.93	14.21	1.896	3.05	15.79	38.37
1772	129	18.36	5.17	17.63	1.340	7.52	17.91	34.78
1773	110	14.52	6.74	13.18	1.546	5.64	12.74	33.67
1774	13	14.56	3.12	14.23	1.259	8.64	15.73	19.19
1775	132	20.09	12.35	15.87	2.132	2.60	19.06	63.31
1776	72	12.97	3.13	12.61	1.267	7.34	12.45	22.64
1778	111	18.46	6.77	17.43	1.394	8.10	17.08	49.83
1779	48	22.10	4.34	21.76	1.189	15.17	22.11	43.31
1780	113	12.40	2.47	12.14	1.235	6.96	12.34	19.45
1781	82	13.47	2.88	13.17	1.244	6.64	13.43	23.27
All	1861	15.12	7.13	13.78	1.534	2.60	13.57	78.41

[a]Dimensionless.

Table 2-15. Descriptive statistics for equivalent ventilation rates averaged by event obtained from construction worker study (average of Models A and C).

Subject identification code	Number of events	Event equivalent ventilation rate, liters · min^{-1} · m^{-2}						
		Arithmetic		Geometric				
		Mean	Std. dev.	Mean	Std. dev.[a]	Minimum	Median	Maximum
1761	81	12.63	4.14	11.99	1.39	3.57	12.67	31.11
1763	51	15.90	5.40	15.01	1.42	6.50	14.86	33.25
1764	127	13.62	3.17	13.25	1.27	5.69	13.34	21.09
1765	114	16.04	6.89	14.79	1.51	2.31	15.22	48.58
1766	121	11.07	3.14	10.67	1.31	5.88	10.50	22.17
1767	115	10.38	2.99	9.95	1.35	4.33	10.23	21.25
1768	62	10.81	3.26	10.40	1.32	4.40	10.36	26.18
1769	160	17.32	8.71	15.63	1.55	6.24	15.14	62.38
1770	109	13.82	3.70	13.34	1.31	7.20	13.74	26.88
1771	111	16.39	7.98	13.80	1.96	1.71	16.30	33.30
1772	129	18.39	4.90	17.70	1.33	6.82	18.23	32.51
1773	110	14.22	6.11	13.02	1.53	4.96	12.92	30.44
1774	13	14.51	3.28	14.13	1.28	8.15	15.76	19.24
1775	132	19.30	10.28	15.68	2.11	1.50	19.62	49.91
1776	72	14.14	3.66	13.68	1.30	6.86	13.69	24.79
1778	111	18.25	6.04	17.37	1.37	7.75	17.19	43.88
1779	48	22.42	3.96	22.13	1.17	15.55	22.48	40.93
1780	113	12.54	2.70	12.22	1.27	6.26	12.57	19.77
1781	82	13.67	2.90	13.36	1.25	6.20	13.73	22.95
All	1861	15.06	6.45	13.80	1.54	1.50	13.85	62.38

[a]Dimensionless.

Section 3
DESCRIPTIVE STATISTICS FOR CATEGORIZED EVR VALUES

This section presents descriptive statistics for event EVR values that have been categorized according to breathing rate, activity, microenvironment, time of day, and event duration. These statistics provide an initial means for identifying factors that may be effective in predicting EVR.

BREATHING RATE CATEGORY

Four breathing rate categories were used to characterize the diary entries for each of the four studies: sleeping, slow, medium, and fast. Table 3-1 provides geometric means and standard deviations by breathing rate category for the elementary school, high school, and outdoor worker studies, respectively. Table 3-2 provides three sets of statistics for the construction worker study based on the method used to convert heart rate data to EVR data: linear (Model A), log-log (Model C), and average of linear and log-log. In each case the geometric means increase as the perceived breathing rate increases. Note that there were no EVR values characterized as "sleeping" in the construction worker study.

With respect to the construction worker geometric means, the linear and log-log values are comparable for the slow, medium, and fast categories. For example, the geometric means for the fast category are 16.3 liters $\cdot$ min^{-1} $\cdot$ m^{-2} for the linear data and 16.1 liters $\cdot$ min^{-1} $\cdot$ m^{-2} for the log-log data. The corresponding geometric mean for the averaged data is 16.2 liters $\cdot$ min^{-1} $\cdot$ m^{-2}.

ACTIVITY CATEGORY

A set of 45 activity codes (numbered 1 through 45) were used to characterize the diary entries in the elementary school, high school, and outdoor worker studies.

Table 3-1. Geometric means and standard deviations of event EVR values by breathing rate category (elementary school, high school, and outdoor worker studies).

Breathing rate category	Elementary school			High school			Outdoor workers		
	n[a]	GM[b]	GSD[c]	n	GM	GSD	n	GM	GSD
Sleeping	20	8.05	1.80	42	7.18	1.42	78	5.62	1.43
Slow	474	10.34	1.58	1,442	8.15	1.35	1,903	8.08	1.67
Medium	319	12.09	1.48	508	10.03	1.40	194	9.41	1.52
Fast	57	15.89	1.63	55	12.12	1.70	15	22.84	1.82

[a]Number of event EVR values.
[b]Geometric mean of EVR (liters · min^{-1} · m^{-2}).
[c]Geometric standard deviation of EVR (dimensionless).

Table 3-2. Geometric means and standard deviations of event EVR values by breathing rate category (construction worker study).

Breathing rate category	Construction (linear)			Construction (log-log)			Construction (average)		
	n[a]	GM[b]	GSD[c]	n	GM	GSD	n	GM	GSD
Sleeping	0	-	-	0	-	-	0	-	-
Slow	733	12.66	1.69	733	12.83	1.58	733	12.80	1.60
Medium	1,014	14.32	1.51	1,014	14.28	1.49	1,014	14.33	1.48
Fast	112	16.29	1.42	112	16.05	1.49	112	16.18	1.45

[a]Number of event EVR values.
[b]Geometric mean of EVR (liters · min^{-1} · m^{-2}).
[c]Geometric standard deviation of EVR (dimensionless).

Four additional codes (80, 81, 82, and 83) were defined for the construction worker study. Appendix C provides an activity description for each numerical code.

Table 3-3 provides geometric means and standard deviations for EVR values by activity code for the elementary school, high school, and outdoor worker studies, respectively. Table 3-4 provides similar statistics for the construction worker study according to calibration method. The activities associated with relatively large geometric means (given that $n \geq 10$) are listed below by study.

<u>High EVR Activities</u>

Elementary school:

meal preparation and cleanup (Code 8)
recess and physical education (27)
taking a walk (31)
other active leisure (33)

High school:

recess and physical education (27)
active sports and games outside school (28)
taking a walk (31)
other active leisure (33)

Outdoor worker:

yard work and outdoor chores (11)
active sports and games outside school (28)
jogging or bicycling (30)

Construction worker:

other personal needs (17),
uncertain of applicable code (41),
special code -- sitting or standing (80),
special code -- walking (81),
special code -- hand-carrying building supplies or equipment (82), and
special code -- working at trade (83)

Table 3-3. Geometric means and standard deviations of event EVR values by activity category (elementary school, high school, and outdoor worker studies).

Abbreviated activity category	Elementary school			High school			Outdoor worker		
	n[a]	GM[b]	GSD[c]	n	GM	GSD	n	GM	GSD
1 All destination oriented travel	258	11.71	1.57	942	8.95	1.36	933	8.63	1.80
2 Income-related work	0	-	-	14	6.75	1.32	224	9.25	1.70
3 Day-care	0	-	-	2	7.61	1.21	0	-	-
4 Kindergarten-12th grade	34	11.04	1.62	100	8.08	1.36	0	-	-
5 College or trade school	0	-	-	0	-	-	0	-	-
6 Adult education and special training	0	-	-	0	-	-	0	-	-
7 Homework	21	11.99	1.60	52	7.05	1.35	0	-	-
8 Meal preparation and cleanup	10	12.59	1.51	50	7.90	1.28	27	7.35	1.32
9 Laundry	3	19.95	1.09	7	9.22	1.24	10	5.61	1.11
10 Other indoor chores	14	7.94	1.38	34	8.29	1.40	25	8.12	1.32
11 Yard work and outdoor chores	1	11.69	-	3	10.22	1.18	19	11.01	1.71
12 Child care and child-centered activities	1	10.12	-	0	-	-	55	6.43	1.39
13 Errands and shopping	6	7.82	1.58	1	15.52	-	63	7.60	1.38
14 Personal care outside home (doctor, hair dresser, etc.)	0	-	-	1	12.23	-	1	8.18	-
15 Eating	65	11.26	1.55	114	7.87	1.26	112	8.22	1.53
16 Sleeping	20	8.05	1.80	42	7.18	1.42	78	5.62	1.43
17 Other	92	10.80	1.49	160	8.68	1.38	132	8.02	1.40
18 Religious activities	20	11.17	2.25	26	7.86	1.30	12	6.50	1.38
19 Meetings of clubs, organizations, etc.	0	-	-	0	-	-	4	12.85	1.70
20 Other collective participation	7	8.84	1.50	29	8.22	1.21	1	9.01	-
21 Spectator sports events	0	-	-	0	-	-	3	17.36	2.47
22 Movies, concerts, and other events outside home	5	14.32	1.26	22	9.22	1.28	6	5.85	1.30
23 Cafe, bar, tea room	0	-	-	0	-	-	16	6.09	1.34
24 Museums and exhibitions	0	-	-	0	-	-	0	-	-
25 Parties and receptions	0	-	-	3	9.55	1.04	1	9.97	-
26 Visiting with friends	7	9.14	1.43	26	8.33	1.27	29	7.91	1.36
27 Recess and physical education	25	15.89	1.59	22	15.22	1.31	0	-	-
28 Active sports and games outside school	4	11.18	1.65	18	20.27	2.11	27	10.31	1.72
29 Hunting, fishing, hiking	5	15.50	1.51	1	12.75	-	0	-	-

(continued)

Table 3-3 (Continued)

Abbreviated activity category	Elementary school			High school			Outdoor worker		
	n[a]	GM[b]	GSD[c]	n	GM	GSD	n	GM	GSD
30 Jogging or bicycling	3	13.56	1.22	6	14.95	1.77	29	15.09	2.23
31 Taking a walk	19	13.76	1.39	10	11.20	1.35	8	10.16	1.66
32 Artistic creations, music, and hobbies	30	10.94	1.41	15	7.64	1.29	1	5.82	-
33 Other active leisure	81	12.03	1.52	140	9.59	1.39	9	10.88	1.32
34 Reading	14	9.32	1.50	24	6.05	1.37	25	6.71	1.33
35 Television or radio	49	10.35	1.53	81	7.42	1.36	125	6.67	1.42
36 Conversation and correspondence	22	9.90	1.37	21	8.13	1.35	46	7.10	1.27
37 Relaxing, reflecting, thinking (no visible activity)	20	11.06	1.59	61	7.64	1.27	106	7.53	1.49
38 Other passive leisure	13	8.66	1.76	3	9.13	1.06	3	7.91	1.20
39 Asthma attack	0	-	-	0	-	-	0	-	-
40 Other sudden illness or injury	0	-	-	0	-	-	3	5.63	1.60
41 Uncertain of applicable code	11	8.33	1.48	7	9.79	1.15	17	7.24	1.42
42 No entry in diary	0	-	-	1	6.70	-	0	-	-
43 Interview	0	-	-	0	-	-	0	-	-
44 Wakeup	10	10.28	1.69	16	7.10	1.32	40	6.46	1.34
45 Baby crying	0	-	-	0	-	-	0	-	-
80 Sitting or standing (including driving on job site)	0	-	-	0	-	-	0	-	-
81 Walking	0	-	-	0	-	-	0	-	-
82 Hand-carrying building materials or equipment	0	-	-	0	-	-	0	-	-
83 Working at trade (hammering, sawing, framing, etc.)	0	-	-	0	-	-	0	-	-

[a]Number of event EVR values.
[b]Geometric mean of EVR (liters $\cdot$ m^{-1} $\cdot$ m^{-2}).
[c]Geometric standard deviation of EVR (dimensionless).

Table 3-4. Geometric means and standard deviations of event EVR values by activity category (construction worker study).

Abbreviated activity category	Construction (linear)			Construction (log-log)			Construction (average)		
	n[a]	GM[b]	GSD[c]	n	GM	GSD	n	GM	GSD
1 All destination oriented travel	74	8.08	2.32	74	9.33	1.87	74	8.83	1.98
2 Income-related work	3	5.51	2.42	3	6.75	1.82	3	6.18	2.06
3 Day-care	0	-	-	0	-	-	0	-	-
4 Kindergarten-12th grade	0	-	-	0	-	-	0	-	-
5 College or trade school	0	-	-	0	-	-	0	-	-
6 Adult education and special training	0	-	-	0	-	-	0	-	-
7 Homework	0	-	-	0	-	-	0	-	-
8 Meal preparation and cleanup	7	9.02	1.18	7	9.25	1.16	7	9.14	1.16
9 Laundry	0	-	-	0	-	-	0	-	-
10 Other indoor chores	0	-	-	0	-	-	0	-	-
11 Yard work and outdoor chores	0	-	-	0	-	-	0	-	-
12 Child care and child-centered activities	0	-	-	0	-	-	0	-	-
13 Errands and shopping	5	10.80	1.42	5	11.08	1.38	5	10.94	1.40
14 Personal care outside home (doctor, hair dresser, etc.)	0	-	-	0	-	-	0	-	-
15 Eating	28	9.88	1.38	28	9.99	1.31	28	9.95	1.34
16 Sleeping	0	-	-	0	-	-	0	-	-
17 Other	55	13.21	1.56	55	13.41	1.54	55	13.33	1.54
18 Religious activities	0	-	-	0	-	-	0	-	-
19 Meetings of clubs, organizations, etc.	2	7.17	1.40	2	7.58	1.30	2	7.37	1.35
20 Other collective participation	0	-	-	0	-	-	0	-	-
21 Spectator sports events	0	-	-	0	-	-	0	-	-
22 Movies, concerts, and other events outside home	0	-	-	0	-	-	0	-	-
23 Cafe, bar, tea room	0	-	-	0	-	-	0	-	-
24 Museums and exhibitions	0	-	-	0	-	-	0	-	-
25 Parties and receptions	0	-	-	0	-	-	0	-	-
26 Visiting with friends	0	-	-	0	-	-	0	-	-
27 Recess and physical education	0	-	-	0	-	-	0	-	-
28 Active sports and games outside school	0	-	-	0	-	-	0	-	-
29 Hunting, fishing, hiking	0	-	-	0	-	-	0	-	-

(continued)

Table 3-4 (Continued)

Abbreviated activity category	Construction (linear)			Construction (log-log)			Construction (average)		
	n[a]	GM[b]	GSD[c]	n	GM	GSD	n	GM	GSD
30 Jogging or bicycling	0	-	-	0	-	0	-	-	-
31 Taking a walk	0	-	-	0	-	0	-	-	-
32 Artistic creations, music, and hobbies	0	-	-	0	-	0	-	-	-
33 Other active leisure	0	-	-	0	-	0	-	-	-
34 Reading	0	-	-	0	-	0	-	-	-
35 Television or radio	1	6.89	-	1	7.45	-	1	7.17	-
36 Conversation and correspondence	0	-	-	0	-	-	0	-	-
37 Relaxing, reflecting, thinking (no visible activity)	6	7.80	1.31	6	8.52	1.23	6	8.17	1.27
38 Other passive leisure	2	8.81	1.38	2	8.91	1.34	2	8.86	1.36
39 Asthma attack	0	-	-	0	-	-	0	-	-
40 Other sudden illness or injury	0	-	-	0	-	-	0	-	-
41 Uncertain of applicable code	11	11.27	1.42	11	11.26	1.37	11	11.28	1.39
42 No entry in diary	1	8.95	-	1	6.82	-	1	7.88	-
43 Interview	0	-	-	0	-	-	0	-	-
44 Wakeup	11	5.66	2.79	11	7.40	1.77	11	6.77	1.95
45 Baby crying	0	-	-	0	-	-	0	-	-
80 Sitting or standing (including driving on job site)	299	12.93	1.48	299	12.75	1.45	299	12.86	1.46
81 Walking	481	14.20	1.54	481	14.14	1.50	481	14.20	1.50
82 Hand-carrying building materials or equipment	147	16.25	1.40	147	16.09	1.48	147	16.19	1.44
83 Working at trade (hammering, sawing, framing, etc.)	728	14.79	1.49	728	14.72	1.50	728	14.79	1.48

[a]Number of event EVR values.
[b]Geometric mean of EVR (liters $\cdot$ $m^{-1} \cdot m^{-2}$).
[c]Geometric standard deviation of EVR (dimensionless).

Four of the activities (recess and physical education, active sports and games outside school, taking a walk, and other active leisure) appear in the listings for at least two of the studies.

Activities associated with relatively small geometric means (given n $\geq$ 10) are listed below.

<u>Low EVR Activities</u>

Elementary school:

other indoor chores (Code 10)
sleeping (16)
other passive leisure (38)
uncertain of applicable code (41)

High school:

income-related work (2)
sleeping (16)
reading (34)
wakeup (44)

Outdoor worker:

laundry (9)
child care and child-centered activities (12)
sleeping (16)
cafe, bar, tearoom (23)
wakeup (44)

Construction worker:

all destination oriented travel (1)
wakeup (44)

Two of the low EVR activities (sleeping and wakeup) are common to listings for three of the four studies.

MICROENVIRONMENT CATEGORY

Five microenvironment categories were used to characterize diary entries for each of the four studies: indoors - residence, indoors - other, outdoors - near road, outdoors - other, and in vehicle. Table 3-5 provides geometric means and standard deviations for EVR values by microenvironment for the elementary school, high school, and outdoor worker studies, respectively. Table 3-6 provides similar statistics for the construction worker study according to calibration method. In each of the four studies, the outdoors - other microenvironment is the microenvironment associated with the largest geometric mean EVR value. The second largest geometric mean is associated with the in-vehicle microenvironment for the elementary school study, the indoors - other microenvironment for the high school study, the in-vehicle microenvironment for the outdoor worker study, and the outdoor - near road microenvironment for the construction worker study.

TIME OF DAY

Each event EVR value can be assigned to a clock hour (0 to 23) based on the time that the associated event began. For example, an event beginning at 4:23 pm can be assigned to clock hour 16. Table 3-7 presents geometric means and standard deviations for event EVR by clock hour for the elementary school, high school, and outdoor worker studies, respectively. Table 3-8 provides similar statistics for the construction worker study according to calibration method.

Note that there are few entries in these tables for the clock hours 0 to 5 because few events began during these hours. Sleeping events usually began during the hours 22, 23, and 24. In addition, the subjects of the construction worker study were typically monitored between wake-up and quitting time each day. There are few data entries in this database for evening events after work.

Table 3-5. Geometric means and standard deviations of event EVR values by microenvironment (elementary school, high school, and outdoor worker studies).

Microenvironment	Elementary school			High school			Outdoor worker		
	n[a]	GM[b]	GSD[c]	n	GM	GSD	n	GM	GSD
Indoors - residence	381	10.23	1.52	871	7.78	1.36	696	7.27	1.47
Indoors - other	188	11.33	1.66	471	9.17	1.39	284	7.62	1.53
Outdoors - near road	90	12.00	1.47	241	9.12	1.36	483	8.51	1.61
Outdoors - other	127	12.42	1.66	305	10.48	1.44	232	9.93	1.73
In vehicle	54	12.25	1.66	167	8.46	1.31	495	8.67	1.98

[a]Number of event EVR values.
[b]Geometric mean of EVR (liters $\cdot$ min^{-1} $\cdot$ m^{-2}).
[c]Geometric standard deviation of EVR (dimensionless).

Table 3-6. Geometric means and standard deviations of event EVR values by microenvironment (construction worker study).

Microenvironment	Construction (linear)			Construction (log-log)			Construction (average)		
	n[a]	GM[b]	GSD[c]	n	GM	GSD	n	GM	GSD
Indoors - residence	51	7.62	1.71	51	8.41	1.37	51	8.09	1.45
Indoors - other	10	8.02	1.71	10	8.50	1.44	10	8.29	1.55
Outdoors - near road	13	11.80	1.52	13	11.83	1.52	13	11.82	1.52
Outdoors - other	1,724	14.34	1.50	1,724	14.24	1.49	1,724	14.32	1.48
In vehicle	63	7.75	2.45	63	9.17	1.95	63	8.60	2.07

[a]Number of event EVR values.
[b]Geometric mean of EVR (liters $\cdot$ min^{-1} $\cdot$ m^{-2}).
[c]Geometric standard deviation of EVR (dimensionless).

Table 3-7. Geometric means and standard deviations of event EVR values by beginning clock hour (elementary school, high school, and outdoor worker studies).

Clock hour	Elementary school			High school			Outdoor worker		
	n[a]	GM[b]	GSD[c]	n	GM	GSD	n	GM	GSD
0	0	-	-	7	6.96	1.28	23	8.09	1.15
1	0	-	-	0	-	-	14	8.76	1.54
2	0	-	-	0	-	-	7	9.49	1.27
3	0	-	-	0	-	-	0	-	-
4	0	-	-	0	-	-	12	8.04	1.70
5	0	-	-	0	-	-	32	6.72	1.45
6	25	10.05	1.41	19	7.28	1.37	62	7.53	1.59
7	49	12.13	1.53	114	9.64	1.29	74	8.17	1.84
8	45	11.67	1.50	179	9.50	1.36	98	7.16	1.54
9	43	9.76	1.79	123	8.45	1.38	117	9.23	2.02
10	38	12.36	1.78	143	9.02	1.44	156	8.39	1.61
11	45	12.55	1.65	157	9.37	1.51	147	8.35	1.71
12	75	11.63	1.66	168	8.47	1.41	139	9.27	1.87
13	62	10.01	1.56	157	9.11	1.40	153	9.01	1.71
14	77	12.26	1.52	167	9.55	1.42	143	8.62	1.69
15	80	12.86	1.47	116	8.59	1.37	139	8.66	1.68
16	64	11.60	1.64	114	8.95	1.37	140	8.57	1.63
17	73	10.83	1.56	111	7.88	1.38	152	8.33	1.77
18	64	10.91	1.50	130	8.14	1.35	139	7.19	1.58
19	42	9.79	1.44	121	8.26	1.30	100	7.51	1.46
20	32	9.00	1.38	95	7.35	1.36	116	8.49	1.55
21	44	10.66	1.49	67	7.72	1.36	89	7.04	1.36
22	9	10.50	1.84	42	6.92	1.25	84	6.94	1.47
23	3	4.76	1.25	25	6.89	1.41	54	6.20	1.73

[a]Number of event EVR values.
[b]Geometric mean EVR (liters $\cdot$ min^{-1} $\cdot$ m^{-2}).
[c]Geometric standard deviation of EVR (dimensionless).

Table 3-8. Geometric means and standard deviations of event EVR values by beginning clock hour (construction worker study).

Clock hour	Construction (linear)			Construction (log-log)			Construction (average)		
	n[a]	GM[b]	GSD[c]	n	GM	GSD	n	GM	GSD
0	0	-	-	0	-	-	0	-	-
1	0	-	-	0	-	-	0	-	-
2	0	-	-	0	-	-	0	-	-
3	0	-	-	0	-	-	0	-	-
4	3	9.62	1.08	3	9.30	1.09	3	9.46	1.09
5	56	6.19	2.13	56	7.47	1.56	56	6.95	1.70
6	127	10.99	1.96	127	11.34	1.69	127	11.26	1.75
7	249	13.41	1.42	249	12.88	1.40	249	13.16	1.40
8	226	13.51	1.48	226	13.26	1.45	226	13.40	1.46
9	178	13.30	1.47	178	12.99	1.44	178	13.17	1.45
10	196	16.63	1.37	196	16.51	1.42	196	16.58	1.39
11	181	15.35	1.46	181	15.35	1.53	181	15.37	1.49
12	130	14.16	1.44	130	14.23	1.46	130	14.21	1.45
13	152	16.63	1.44	152	16.78	1.50	152	16.72	1.47
14	164	14.13	1.67	164	14.66	1.58	164	14.48	1.57
15	94	14.35	1.59	94	14.81	1.65	94	14.60	1.62
16	77	13.34	1.36	77	13.20	1.37	77	13.27	1.36
17	22	10.71	1.43	22	11.10	1.35	22	10.91	1.39
18	5	10.95	1.30	5	10.50	1.27	5	10.73	1.28
19	1	6.89	-	1	7.45	-	1	7.17	-
20	0	-	-	0	-	-	0	-	-
21		-	-	0	-	-	0	-	-
22	0	-	-	0	-	-	0	-	-
23	0	-	-	0	-	-	0	-	-

[a]Number of event EVR values.
[b]Geometric mean EVR (liters $\cdot$ min^{-1} $\cdot$ m^{-2}).
[c]Geometric standard deviation of EVR (dimensionless).

Listed below are the clock hours associated with the largest, second largest, second smallest, and smallest geometric means by study (n >= 10).

Study	Largest	Second largest	Second smallest	Smallest
Elementary school	15	11	9	20
High school	7	14	22	23
Outdoor workers	12	9	5	23
Construction workers	13	10	17	5

The clock hours associated with the largest and second largest geometric means include 7, 9, 10, 11, 12, 13, and 15. The clock hours associated with the smallest and second smallest geometric means include 5, 9, 17, 20, 22, and 23. In general, these results suggest that we can define a daily "active" period as including clock hours 7 through 16 and a daily "inactive" period as including clock hours 0 through 6 and 17 through 23.

DESCRIPTIVE STATISTICS FOR EVENT EVR VALUES BY DURATION

As indicated above, events are defined as beginning whenever a subject changed activity, microenvironment, or breathing rate and ending whenever a new event begins. Because study subjects reported times to the nearest minute, each event is associated with a duration expressed as an integer number of minutes. Table 3-9 provides descriptive statistics for event duration by study. The median duration length is 14 minutes for the elementary school study, 8 minutes for the high school study, 12 minutes for the outdoor worker study, and 2 minutes for the construction worker study. Note that the construction workers reported few events having durations greater than 60 minutes.

Table 3-9. Descriptive statistics for event durations by study.

Statistic	Value of statistic by study			
	Elementary school	High school	Outdoor workers	Construction workers
Number of values	874	2055	2190	1861
Arithmetic mean, min	29.1	18.9	27.3	5.4
Arithmetic std. dev., min	40.7	28.7	42.8	8.4
Skewness	4.3	3.4	3.8	4.2
Kurtosis	36.6	16.4	21.6	24.1
Minimum, min	1.0	1.0	1.0	1.0
25th percentile, min	5.0	3.0	4.0	1.0
50th percentile, min	14.0	8.0	12.0	2.0
75th percentile, min	38.0	23.0	32.0	6.0
90th percentile, min	71.5	48.0	71.0	13.0
95th percentile, min	103.0	70.0	113.0	20.0
98th percentile, min	161.0	109.9	162.0	33.5
99th percentile, min	193.3	155.4	217.5	42.4
99.5th percentile, min	220.4	182.0	262.0	55.8
Maximum, min	556.0	299.0	540.0	85.0

Table 3-10 provides geometric means and standard deviations for event durations classified by study and breathing rate category. Note that there are relatively few events associated with the sleeping category, as the majority of documented activities occurred between 5 a.m. and 6 p.m. (see Table 3-8).

With respect to the first three studies listed in Table 3-10, geometric means tend to be lower for the medium category than for the slow and fast categories. This pattern does not hold for the remaining study (construction workers). The geometric means for the slow and medium categories are essentially equal, whereas the geometric mean for the fast category is approximately 25 percent lower (2.1 minutes). Note that geometric means for the construction worker study are consistently smaller than the corresponding values for the other studies. The construction worker means are smaller because (1) a trained observer filled out the diary of each subject, (2) few non-work activities were recorded, and (3) construction work tends to consist of a relatively large number of short-term tasks.

To facilitate the statistical analyses described below, analysts defined seven duration categories:

DUR1: 0 to 5 minutes
DUR2: 6 to 10 minutes
DUR3: 11 to 20 minutes
DUR4: 21 to 30 minutes
DUR5: 31 to 45 minutes
DUR6: 46 to 60 minutes
DUR7: 61+ minutes.

Table 3-11 provides geometric means and standard deviations for event EVR values by duration category for the elementary school, high school, and outdoor worker studies, respectively. Table 3-12 provides similar statistics for the construction worker study according to calibration method.

Table 3-10. Descriptive statistics for event durations by study and breathing rate category.

Study	Breathing rate	Number of values	Geometric mean, min	Geometric std. dev.[a]
Elementary school	Sleeping	20	23.2	5.93
	Slow	474	14.2	3.70
	Medium	321	11.6	4.10
	Fast	59	13.0	3.58
	All	874	13.3	3.90
High school	Sleeping	42	5.8	5.59
	Slow	1442	9.0	3.83
	Medium	508	6.0	3.46
	Fast	55	8.4	4.06
	All	2055[b]	8.1	3.82
Outdoor workers	Sleeping	78	22.8	6.22
	Slow	1903	10.6	4.14
	Medium	194	9.6	4.60
	Fast	15	25.4	3.44
	All	2190	10.8	4.27
Construction workers	Sleeping	0	-	-
	Slow	733	2.8	2.92
	Medium	1014	3.0	2.70
	Fast	112	2.1	2.09
	All	1861[c]	2.9	2.77

[a]Dimensionless.
[b]Includes 8 values not classified by breathing rate.
[c]Includes 2 values not classified by breathing rate.

Table 3-11. Geometric means and standard deviations of event EVR values by duration category (elementary school, high school, and outdoor worker studies).

Duration class	Elementary school			High school			Outdoor worker		
	n[a]	GM[b]	GSD[c]	n	GM	GSD	n	GM	GSD
DUR1: 0 to 5 minutes	231	11.65	1.63	865	8.89	1.40	687	8.25	1.67
DUR2: 6 to 10 minutes	121	11.08	1.63	323	9.10	1.36	316	8.49	1.70
DUR3: 11 to 20 minutes	168	11.74	1.58	311	9.03	1.42	422	8.52	1.73
DUR4: 21 to 30 minutes	81	11.28	1.69	156	8.59	1.36	202	7.97	1.72
DUR5: 31 to 45 minutes	88	11.26	1.60	171	8.06	1.43	181	8.19	1.71
DUR6: 46 to 60 minutes	79	11.08	1.64	90	7.79	1.32	110	7.54	1.48
DUR7: 61+ minutes	106	10.51	1.58	139	7.05	1.35	272	7.32	1.55
All values	874	11.32	1.62	2055	8.66	1.40	2190	8.14	1.68

[a]Number of event EVR values.
[b]Geometric mean of EVR (liters $\cdot$ min^{-1} $\cdot$ m^{-2}).
[c]Geometric standard deviation of EVR (dimensionless).

Table 3-12. Geometric means and standard deviations of event EVR values by duration category (construction worker study).

Duration class	Construction (linear)			Construction (log-log)			Construction (average)		
	n[a]	GM[b]	GSD[c]	n	GM	GSD	n	GM	GSD
DUR1: 0 to 5 minutes	1,394	14.03	1.63	1,394	14.08	1.56	1,394	14.10	1.56
DUR2: 6 to 10 minutes	232	13.64	1.44	232	13.64	1.48	232	13.65	1.46
DUR3: 11 to 20 minutes	144	12.91	1.44	144	12.93	1.45	144	12.93	1.44
DUR4: 21 to 30 minutes	52	11.51	1.49	52	11.72	1.45	52	11.63	1.46
DUR5: 31 to 45 minutes	24	11.34	1.30	24	11.17	1.22	24	11.27	1.25
DUR6: 46 to 60 minutes	8	10.58	1.36	8	10.51	1.31	8	10.55	1.33
DUR7: 61+ minutes	7	8.84	1.33	7	9.08	1.30	7	8.97	1.31
All values	1,861	13.74	1.59	1,861	13.78	1.53	1,861	13.79	1.54

[a]Number of event EVR values.
[b]Geometric mean of EVR (liters · min^{-1} · m^{-2}).
[c]Geometric standard deviation of EVR (dimensionless).

In general, the geometric means and standard deviations display a weak tendency to decrease as duration increases. This result is consistent with the physiological principal that maximal sustainable EVR decreases as the duration of a task increases. If high EVR's are reduced as duration increases, the mean and the standard deviation should also be reduced. The pattern is weak because most of the reported events do not represent maximum exertion levels. At these lower exertion levels, EVR is relatively independent of task duration.

Section 4
MONTE CARLO MODELS FOR GENERATING EVENT EVR VALUES

DATABASE TYPES

A report by Johnson et al. (Appendix A) lists 10 time/activity databases that are appropriate for use in pNEM and similar exposure models. Table 4-1 presents a summary of the study that produced each database. The databases can be categorized according to the four general types presented in Table 4-2.

A Type 1 database will include data that can be used to define four breathing rate categories (sleeping, slow, medium, and fast) and five microenvironments (indoors - residence, indoors - other, outdoors - near road, outdoors - other, and in vehicle) for each event. In addition, the database will indicate the activity associated with each event, the gender of each subject, and the daily maximum temperature associated with each day of diary data. These data must be available for most waking hours. Four of the ten studies listed in Table 4-1 produced Type 1 databases: Cincinnati, elementary school, high school, and outdoor workers.

Type 2 data support the use of the five Type 1 microenvironments and four breathing rate categories. However, Type 2 data are limited to male construction workers and omit most non-work activities. Work activities are defined according to four special codes not used in any other diary study. Otherwise, Type 2 data are similar to Type 1 data. The construction worker study produced the only Type 2 database.

Type 3 data were acquired through the Denver and Washington personal monitoring studies. These data support the use of the five Type 1 microenvironments, but omit breathing rate data. Type 3 data are similar to Type 1 data in most other respects.

Table 4-1. Characteristics of studies associated with the 10 time/activity databases.

Database name	Reference number	Characteristics of subjects	Number of subject-days	Study calendar periods	Diary type	Diary time period	Breathing rates reported?
California - 11 and under	14	Children ages 1 to 11	1200	April 1989 - Feb. 1990	Retrospective	Midnight to midnight	No
California - 12 and over	15	Ages 12 to 94	1762	Oct. 1987 - July 1988	Retrospective	Midnight to midnight	No
Cincinnati	6	Ages 0 to 86	2800	March and August 1985	Real-time	Midnight to midnight	Yes
Denver	16	Ages 18 to 70	859	Nov. 1982 - Feb. 1983	Real-time	7 p.m to 7 p.m. (nominal)	No
Los Angeles - construction	9	Construction workers (ages 23 to 42)	19	July - Nov. 1991	Real-time[a]	Subject wakeup to subject returns home from work	Yes
Los Angeles - elem. school	7	Elementary school students, 10 to 12 years	58	Oct. 1989	Real-time[a]	Midnight to midnight	Yes
Los Angeles - high school	7	High school students, 13 to 17 years	66	Sept. and Oct. 1990	Real-time[a]	Midnight to midnight	Yes
Los Angeles - outdoor worker	8	Adult outdoor workers (ages 19 to 50)	60	Summer 1989	Real-time[a]	Midnight to midnight	Yes
Valdez	17	Ages 10 to 72	405	Nov. 1990 - Oct. 1991	Retrospective	Varying 24-h period	No
Washington	18	Ages 18 to 70	705	Nov. 1982 - Feb. 1983	Real-time	7 p.m. to 7 p.m. (nominal)	No

[a]Study employed the Cincinnati diary format.

Table 4-2. Database types.

Type 1:	Time/activity data acquired using the "Cincinnati" diary. Type 1 data support the use of four breathing rate categories (sleeping, slow, medium, and fast) and five microenvironments (indoors - residence, indoors - other, outdoors - near road, outdoors - other, and in vehicle). Representative studies: Cincinnati, elementary school, high school, and outdoor workers. Studies providing regression data: elementary school, high school, and outdoor workers.
Type 2:	Time/activity data acquired using the Cincinnati diary. Type 2 data support the use of the five Type 1 microenvironments and four breathing rate categories. Data limited to male construction workers. Data omit most non-work activities. Representative study: construction workers. Study providing regression data: construction workers.
Type 3:	Time/activity data acquired using the "Denver/Washington" diary. Type 3 data support the use of the five Type 1 microenvironments. Breathing rate data are not available. Representative studies: Denver and Washington. Studies providing regression data: elementary school, high school, and outdoor workers.
Type 4:	Time/activity data acquired using other diary formats. Type 4 data support the use of four microenvironments: indoors - residence, indoors - other, outdoors, and in vehicle. Breathing rate data are not available. Representative studies: California - children, California - adults, and Valdez. Studies providing regression data: elementary school, high school, and outdoor workers.

Databases obtained from the remaining studies listed in Table 4-1 (California - children, California - adults, and Valdez) are classified as Type 4. These data are similar to Type 3 data except that they do not permit the use of the outdoors - near road microenvironment. Type 4 data support four microenvironments (indoors - residence, indoors - other, outdoors, and vehicle).

GENERAL PROCEDURE FOR DEVELOPMENT OF MONTE CARLO MODELS

ITAQS developed one or more Monte Carlo models for each database type. Each model was specific to one of the following four demographic groups: elementary students, high school students, outdoor workers, and construction workers. Each model consisted of an algorithm capable of generating an EVR for each event in a time/activity dataset of the specified database type and demographic group. Each algorithm predicted EVR as a function of six or more predictor variables which constituted a "predictor set."

Each predictor set was developed by first defining a candidate variable set for the database type and then performing stepwise linear regression analyses to determine which of the candidate variables are significant predictors of EVR for a particular demographic group. The regression analyses were performed on the Hackney/Linn databases, as these are the only databases available that provide a "measured" EVR value for each exposure event. Three of the Hackney/Linn databases (elementary school, high school, and outdoor workers) were used to develop demographic-specific Monte Carlo models for database Types 1, 3, and 4. The remaining database (construction workers) was used to develop a Monte Carlo model for Type 2 data. Note that the construction worker database is the only existing example of a Type 2 database.

The results of the regression analyses were used to estimate the coefficients of various terms in each Monte Carlo model. Each Monte Carlo model has the general form

$$\ln[EVR(i,j)] = b_0 + (b_1)[VAR_1(i,j)] + (b_2)[VAR_2(i,j)] + \ldots + (b_m)[VAR_m(i,j)] + e(i,j) \qquad (7)$$

where EVR(i,j) is the EVR value for event j associated with subject i; b_0, b_1, b_2, ..., b_m are constants; $VAR_1(i,j)$, $VAR_2(i,j)$, ..., $VAR_m(i,j)$ are the values of the predictor variables for the event; and e(i,j) is the residual. The next subsection provides a listing of the candidate variables sets used in each regression analysis. A later subsection discusses patterns identified by researchers in the regression residuals.

CANDIDATE VARIABLE GROUPS

Table 4-3 provides a complete list of the candidate variables defined for use in the regression analyses. With the exception of the continuous variable LGM (the natural logarithm of the geometric mean of all event EVR values associated with a subject), each of these variables is a binary "dummy" variable. A dummy variable equals one when specified conditions are met and equals zero under all other conditions. Among the variables listed in Table 4-3 are variables which indicate breathing rate, event duration, microenvironment, subject gender, time of day, day of the week, and temperature. Several variables classify activities according to level of exertion, work status (work/non-work), and travel status (travel/non-travel).

The best overall predictor variable listed in Table 4-3 is LGM. Analyses of variance performed on the four Hackney/Linn data sets indicated that inter-subject variability with respect to average EVR was a major source of variability in the event EVR values. LGM is an indicator of average subject EVR which can be related directly to ln(EVR), the dependent variable in Equation 7.

Table 4-3. Candidate variables used in stepwise linear regression analyses.

Variable	Explanation	Candidate variable group 1	Candidate variable group 2	Candidate variable group 3	Candidate variable group 4
LGM	Natural logarithm of geometric mean of event EVR values for individual subject	*	*	*	*
SLEEP	SLEEP=1 if breathing rate = sleeping, 0 otherwise	*	*		
SLOW	SLOW=1 if breathing rate = slow, 0 otherwise	*	*		
MEDIUM	MEDIUM=1 if breathing rate = medium, 0 otherwise	*	*		
FAST	FAST=1 if breathing rate = fast, 0 otherwise	*	*		
DUR1	DUR1=1 if duration ≤ 5 minutes, 0 otherwise	*	*	*	*
DUR2	DUR2=1 if 6 ≤ duration ≤ 10 minutes, 0 otherwise	*	*	*	*
DUR3	DUR3=1 if 11 ≤ duration ≤ 20 minutes, 0 otherwise	*	*	*	*
DUR4	DUR4=1 if 21 ≤ duration ≤ 30 minutes, 0 otherwise	*	*	*	*
DUR5	DUR5=1 if 31 ≤ duration ≤ 45 minutes, 0 otherwise	*	*	*	*
DUR6	DUR6=1 if 46 ≤ duration ≤ 60 minutes, 0 otherwise	*	*	*	*
DUR7	DUR7=1 if duration > 60 minutes, 0 otherwise	*	*	*	*
INDOOR	INDOOR = 1 if event occurs in an indoor microenvironment, 0 otherwise	*	*	*	*
OUTDOOR	OUTDOOR = 1 if event occurs in an outdoor microenvironment, 0 otherwise	*	*	*	*
OUTOTHER	OUTOTHER = 1 if event occurs in the outdoors - other microenvironment, 0 otherwise	*	*	*	
VEH	VEH = 1 if event occurs in a vehicle microenvironment, 0 otherwise	*	*	*	*
MALE	MALE = 1 if subject is male, 0 otherwise	*	*	*	*
WEEKDAY	WEEKDAY = 1 if event occurs on a weekday, 0 otherwise	*		*	*
HIGHTOWK	HIGHTOWK = 1 if daily maximum temperature exceeds 79°F and event occurs in outdoor microenvironment and activity code = 2 (work), 0 otherwise	*		*	*
DAYACT	DAYACT = 1 if event begins between 7:00 a.m. and 4:59 p.m., 0 otherwise	*	*	*	*
TRAVEL	TRAVEL = 1 if activity code is 1 (travel), 0 otherwise	*	*	*	*
HIGHACT	HIGHACT = 1 if activity code is 11, 27, 28, 29, 30, 31, or 33; 0 otherwise	*		*	*
LOWACT	LOWACT = 1 if activity code is 10, 12, 16, 23, 34, 35, 37, or 44; 0 otherwise	*	*	*	*
WORK	WORK = 1 if activity code = 2 (work), 0 otherwise	*	*	*	*
CONWORK	CONWORK = 1 if activity code = 81, 82, or 83; 0 otherwise		*	*	*
ACT80	ACT80 = 1 if activity code 80, 0 otherwise		*		
ACT81	ACT81 = 1 if activity code 81, 0 otherwise		*		
ACT82	ACT82 = 1 if activity code 82, 0 otherwise		*		
ACT83	ACT83 = 1 if activity code 83, 0 otherwise		*		

Table 4-3 specifies a group of candidate variables to be included in the regression analyses for each database type. For example, the candidate variable group for Type 1 data contains the 24 variables marked by asterisks under the column heading "Group 1." The results of exploratory data analyses suggested that these variables were promising candidates for predicting EVR in Type 1 databases. Note that this group contains all but five of the variables listed in Table 4-3. The omitted variables (CONWORK, ACT80, ACT81, ACT82, and ACT83) can only be defined in terms of data from the construction worker study.

The candidate variable group for Type 2 data contains 25 variables: 20 of the Type 1 candidate variables (including LGM), 4 variables relating to construction work activities (ACT80, ACT81, ACT82, and ACT83), and 1 variable indicating a work activity likely to increase ventilation rate (CONWORK). The WEEKDAY variable was not included in the candidate variable group for Type 2 data because all of the construction worker data relate to weekday activities.

The candidate variable group for Type 3 data is identical to the Type 1 group, except that it omits the four variables associated with breathing rate (BR13, BR14, BR15, and BR16). As indicated above, Type 3 data do not include breathing rate information.

The Type 4 group omits one additional variable (OUTOTHER). This variable indicates that an event occurred in the outdoor - other microenvironment. Type 4 data do not distinguish between the outdoor - near road and outdoor - other microenvironments.

RESULTS OF STEPWISE LINEAR REGRESSION ANALYSES

Tables 4-4 though 4-7 present the results of the stepwise linear regression analyses in which the dependent variable is ln(EVR), i.e., the natural logarithm of event EVR. The results are organized according to the database used in the analysis. Entries in

Table 4-4. Results of stepwise linear regression analyses performed on elementary school data set.

Candidate variable set	Selected variable[a]	Regression coefficient	p value	Cumulative R^2
1	Constant	-0.08174	0.1544	0.0000
	LGM	0.98606	0.0000	0.6600
	OUTDOOR	0.12156	0.0000	0.6812
	FAST	0.16111	0.0000	0.6939
	DAYACT	0.07188	0.0001	0.7002
	SLEEP	-0.17393	0.0021	0.7037
	WEEKDAY	0.04674	0.0062	0.7063
	HIGHACT	0.05962	0.0159	0.7083
3,4	Constant	-0.04230	0.5242	0.0000
	LGM	0.99012	0.0000	0.6600
	OUTDOOR	0.07338	0.0268	0.6812
	DAYACT	0.08146	0.0000	0.6892
	HIGHACT	0.09849	0.0000	0.6951
	WEEKDAY	0.05709	0.0012	0.6983
	DUR7	-0.06056	0.0203	0.7002
	INDOOR	-0.06199	0.0403	0.7016

[a]HIGHTOWK not applicable to this data set.

Table 4-5. Results of stepwise linear regression analyses performed on high school data set.

Candidate variable set	Selected variable[a]	Regression coefficient	p value	Cumulative R^2
1	Constant	0.16385	0.0158	0.0000
	LGM	0.91365	0.0000	0.3063
	OUTOTHER	0.11198	0.0000	0.3505
	HIGHACT	0.15447	0.0000	0.3833
	SLOW	-0.07989	0.0000	0.4038
	DAYACT	0.08175	0.0000	0.4209
	DUR7	-0.13637	0.0000	0.4325
	LOWACT	-0.08749	0.0000	0.4408
	WEEKDAY	0.05873	0.0000	0.4462
	DUR5	-0.09184	0.0000	0.4503
	FAST	0.14685	0.0001	0.4545
	DUR6	-0.10629	0.0001	0.4582
	VEH	-0.04650	0.0255	0.4596
3	Constant	0.01146	0.8620	0.0000
	LGM	0.94621	0.0000	0.3063
	OUTOTHER	0.12257	0.0000	0.3505
	HIGHACT	0.20361	0.0000	0.3833
	DAYACT	0.08162	0.0000	0.4018
	DUR7	-0.13371	0.0000	0.4146
	WEEKDAY	0.06418	0.0000	0.4229
	LOWACT	-0.06447	0.0007	0.4292
	DUR5	-0.09104	0.0000	0.4346
	DUR6	-0.10098	0.0003	0.4387
	VEH	-0.08339	0.0002	0.4409
	TRAVEL	0.04644	0.0011	0.4439
4	Constant	-0.06451	0.3748	0.0000
	LGM	0.94794	0.0000	0.3063
	OUTDOOR	0.15598	0.0000	0.3495
	HIGHACT	0.20379	0.0000	0.3848
	DAYACT	0.08090	0.0000	0.4011
	DUR7	-0.13089	0.0000	0.4122
	WEEKDAY	0.06476	0.0000	0.4198
	DUR5	-0.09361	0.0000	0.4257
	LOWACT	-0.06752	0.0004	0.4307
	DUR6	-0.10000	0.0003	0.4343
	INDOOR	0.07342	0.0016	0.4360
	TRAVEL	0.03528	0.0236	0.4374

[a]HIGHTOWK not applicable to this data set.

Table 4-6. Results of stepwise linear regression analyses performed on outdoor worker data set.

Candidate variable set	Selected variable	Regression coefficient	p value	Cumulative R^2
1	Constant	0.05679	0.4033	0.0000
	LGM	0.96278	0.0000	0.3020
	INDOOR	-0.09539	0.0000	0.3264
	FAST	0.72697	0.0000	0.3479
	DAYACT	0.10270	0.0000	0.3582
	HIGHACT	0.24594	0.0000	0.3673
	MEDIUM	0.11492	0.0002	0.3719
	HIGHTOWK	0.35876	0.0002	0.3755
	LOWACT	-0.06100	0.0135	0.3779
	DUR7	-0.06875	0.0121	0.3794
	DUR6	-0.08170	0.0427	0.3806
3,4	Constant	0.06180	0.3672	0.0000
	LGM	0.96665	0.0000	0.3020
	INDOOR	-0.09738	0.0000	0.3264
	HIGHACT	0.37953	0.0000	0.3465
	DAYACT	0.10104	0.0000	0.3568
	HIGHTOWK	0.36557	0.0002	0.3605
	LOWACT	-0.06711	0.0070	0.3633
	DUR7	-0.06996	0.0116	0.3649
	DUR6	-0.08770	0.0313	0.3662

Table 4-7. Results of stepwise linear regression analyses applied to the construction worker data set (average values)

Candidate variable set	Selected variable	Regression coefficient	p value	Cumulative R^2
2	Constant	-0.38586	0.0023	0.0000
	LGM	0.96795	0.0000	0.1730
	OUTDOOR	0.26600	0.0000	0.2570
	DAYACT	0.20823	0.0000	0.2751
	CONWORK	0.07459	0.0029	0.2873
	SLOW	-0.05756	0.0075	0.2906
	ACT82	0.07266	0.0230	0.2926

each table indicate the candidate variable set used in a particular regression analysis, the variables selected by the regression procedure as being significant predictors of ln(EVR), the regression coefficient associated with each variable, the p value associated with the coefficient, and the cumulative R^2 value that resulted when the variable was added to the regression equation.

Table 4-4 provides the results of the regression analyses performed on the elementary school database. The regression analysis of the Candidate Variable Group 1 selected seven variables for the regression equation (LGM, OUTDOOR, FAST, DAYACT, SLEEP, WEEKDAY, and HIGHACT). The cumulative R^2 values for all seven variables is 0.7083. This value indicates that the regression equation explains 70.83 percent of the variation in ln(EVR).

The regression analyses performed on Candidate Variable Groups 3 and 4 selected a different set of seven variables: LGM, OUTDOOR, DAYACT, HIGHACT, WEEKDAY, DUR7, and INDOOR. Five of these variables match the variables selected for Candidate Group 1 (LGM, OUTDOOR, DAYACT, HIGHACT, and WEEKDAY). As SLEEP and FAST were not included in Candidate Groups 3 and 4, these two variables could not be selected by the regression procedure. The cumulative R^2 value for all seven variables is 0.7016.

Table 4-5 presents regression results for the high school database. The regression procedure selected 12 variables from Candidate Variable Set 1. The first three variables to be selected were LGM, OUTOTHER, and HIGHACT. These three variables had a cumulative R^2 value of 0.3833. The cumulative R^2 value for all 12 variables is 0.4596.

The regression procedure selected 11 variables from Candidate Variable Set 3. The first three variables were again LGM, OUTOTHER, and HIGHACT (cumulative R^2 = 0.3833). The cumulative R^2 for all 11 variables is 0.4439. Eleven variables were also selected from Candidate Variable Group 4 (cumulative R^2 = 0.4374). The first three selected variables were LGM, OUTDOOR, and HIGHACT; the cumulative R^2 for these variables equals 0.3848.

Regression results for the outdoor worker database are presented in Table 4-6. Ten variables were selected from Candidate Variable Set 1 (cumulative R^2 = 0.3806). The first three selected variables were LGM, INDOOR, and FAST. The same eight variables were selected from Candidate Variable Groups 3 and 4 (cumulative R^2 = 0.3662). The first three selected variables were LGM, INDOOR, and HIGHACT .

Table 4-7 provides results from the regression analyses of the construction worker database. Only Candidate Variable Group 2 was used in these analyses. The regression procedure selected six variables: LGM, OUTDOOR, DAYACT, CONWORK, SLOW, and ACT82. The cumulative R^2 value for these six variables is 0.2926.

A comparison of the results presented in Tables 4-4 through 4-7 finds that 11 variables are frequently among the selected predictor variables: LGM, OUTDOOR,

INDOOR, DAYACT, HIGHACT, LOWACT, WEEKDAY, SLOW, HIGH, DUR6, and DUR7. Each of these variables appears in at least two tables. LGM appears in all tables and is always the first variable selected. LGM always contributed at least 0.170 to the cumulative R^2 value. In Table 4-4, adding LGM increased the R^2 value by 0.660.

These results suggest that variables associated with average subject EVR (LGM), microenvironment (OUTDOOR, INDOOR), daytime activities (DAYACT), the exertion level of activities (HIGHACT, LOWACT), day of week (WEEKDAY), breathing rate (SLOW, HIGH) and duration of activity (DUR6, DUR7) are useful in predicting event EVR. Note that the duration variables tended to be relatively insignificant predictors. Adding DUR6 or DUR7 to the regression equation never increased the cumulative R^2 value by more than 0.015.

THE DISTRIBUTION OF REGRESSION RESIDUALS

Each regression analysis produced a set of residual values, one for each EVR value. Researchers performed a series of exploratory data analyses in which they attempted to find patterns in the residuals that could be used to characterize random effects in the Monte Carlo approach. Statistical analysis of the residuals indicated that (1) the standard deviation of the residuals varied significantly from subject to subject and (2) the distribution of the subject-specific standard deviations was approximately lognormal.

Based on these findings, researchers assumed that the residual term in Equation 7 could be represented by a normally distributed random variable with mean equal to zero and standard deviation equal to SDRES. The value of SDRES was assumed to vary with subject and to be lognormally distributed among subjects; i.e., the natural logarithm of SDRES [LSDRES = ln(SDRES)] is normally distributed with mean = MU and standard deviation = SIGMA. The values of MU and SIGMA were specific to the data set undergoing the regression analysis.

Consistent with these assumptions, analysts performed the following statistical analysis of the residuals obtained from each regression analysis:

1. Classify residuals by subject.
2. Calculate the standard deviation of residuals associated with each subject (SDRES).
3. Calculate a value of LSDRES for each subject where LSDRES is the natural logarithm of each SDRES value obtained in Step 2.
4. Calculate the mean (MU) and the standard deviation (SIGMA) of the LSDRES values determined for all subjects.

Table 4-8 lists the values of MU and SIGMA determined in Step 4.

THE DISTRIBUTION OF LGM VALUES

As indicated above, researchers found that the LGM variable was the single best predictor of ln(EVR) in each regression analysis. An analysis of the LGM values associated with the subjects in each of the Hackney/Linn studies indicated that the distribution of LGM values for each study was approximately normal. The parameters of these normal distributions were estimated by the following procedure.

1. Classify the event EVR values by subject.
2. Calculate ln(EVR) for each event.
3. Calculate the mean of the ln(EVR) values associated with each subject (LGM).
4. Calculate the arithmetic mean and arithmetic standard deviation of the LGM values determined for all subjects in the database.

The arithmetic mean and standard deviation values determined in Step 4 are listed in Table 4-8.

Table 4-8. Distribution of LGM and LSDRES values.

Database	Number of subjects	Parameters of normal distribution fit to subject LGM values[a]		Wilk-Shapiro[b] statistic for LGM	Regression model producing residuals	Parameters of normal distribution fit to subject LSDRES values[c]		Wilk-Shapiro statistic for LSDRES
		Arithmetic mean	Arithmetic standard deviation			Mu	Sigma	
Elementary school	16	2.3629	0.4324	0.9630	1 3,4	-1.6068 -1.5736	0.4450 0.4026	0.9540 0.9363
High school	19	2.1621	0.1890	0.9597	1 3 4	-1.4662 -1.4438 -1.4391	0.2997 0.2977 0.3047	0.9845 0.9716 0.9742
Outdoor workers	20	2.0740	0.2918	0.9602	1 3,4	-1.1335 -1.1139	0.4338 0.4229	0.9704 0.9706
Construction workers	19	2.6286	0.1940	0.9621	2	-1.2307	0.3834	0.9607

[a]LGM is the natural logarithm of the geometric mean of the event EVR values associated with one subject.
[b]An indicator of normality (1.0 = normal distribution).
[c]LSDRES is the natural logarithm of the standard deviation of the regression residuals associated with one subject.

ALGORITHM FOR EXECUTING THE MONTE CARLO MODEL

Table 4-9 presents the algorithm used to execute the Monte Carlo model. When this algorithm is applied to an appropriate input database, it generates a sequence of event EVR values for each subject in the database. The EVR value generated for each individual event is determined by:

- The values of the specified predictor variables;
- The regression coefficient associated with each predictor variable;
- An LGM value randomly selected from a normal distribution;
- A residual value selected from a subject-specific normal distribution.

The model coefficients are common to all subjects and are obtained from the regression analyses. Tables 4-4 through 4-7 lists the coefficients determined from the regression analyses performed on each of the four Hackney/Linn studies.

A value for LGM is selected for each subject from a normal distribution common to all subjects. Table 4-8 lists the arithmetic mean and standard deviation of LGM values associated with each of the Hackney/Linn studies.

A residual value is determined for each event. The residual is randomly selected from a normal distribution (mean = 0) which is subject-specific. The standard deviation for this normal distribution is randomly selected from a lognormal distribution that applies to all subjects. Table 4-8 lists the parameters (MU and SIGMA) for the lognormal distribution associated with the residuals determined from the application of each regression model to each of the four Hackney/Linn studies.

Section 5 presents the results of initial efforts to validate this algorithm.

Table 4-9. Algorithm used to execute the Monte Carlo model for generating event-specific values of equivalent ventilation rate.

1. User specifies

    MEANLGM: mean of LGM values
    SDLGM: standard deviation of LGM values
    MU: mean of LSDRES values
    SIGMA: standard deviation of LSDRES values
    b_0: constant
    b_m: coefficient for variable VAR_m

    These values are specific to the database (i.e., elementary school, high school, outdoor workers, or construction workers). Denote the value of b_m for variable LGM as b_1.

2. Go to first/next subject i.

3. Calculate LGM for subject i:

    LGM(i) = MEANLGM + (SDLGM)[Z1(i)]

    Z1(i): randomly selected value from unit normal distribution (normal distribution with mean = 0 and standard deviation = 1).

4. Calculate RESSIGMA for subject i.

    LSDRES(i) = MU + (SIGMA)[Z2(i)]

    RESSIGMA(i) = Exp[LSDRES(i)]

    Z2(i): randomly selected value from unit normal distribution.

5. Go to first/next event associated with subject i.

6. Calculate residual value for event j of subject i.

    RES(i,j) = [RESSIGMA(i)][Z(i,j)]

    Z(i,j): randomly selected value from unit normal distribution.

(continued)

Table 4-9 (Continued)

7. Calculate LEVR for event j of subject i:

$$LEVR(i,j) = b_0 + (b_1)[LGM(j)] + (b_2)[VAR_2(i,j)] + (b_3)[VAR_3(i,j)] + \ldots + (b_m)[VAR_m(i,j)] + RES(i,j)$$

8. Calculate EVR for event j of subject i:

$$EVR(i,j) = Exp[LEVR(i,j)]$$

9. Write EVR(i,j) to output file.

10. If last event of subject i, go to Step 2. If not, go to Step 5.

Section 5
VALIDATION OF MONTE CARLO MODELS

At the time of this report (March 1994), the four Hackney/Linn databases provided the only means of validating the Monte Carlo approach described in the last section. These were the only databases available that included high quality time/activity data together with EVR values determined from heart rate measurements. This section summarizes the results of initial efforts to validate the Monte Carlo approach using these four databases.

APPLICATION OF THE ALGORITHM TO THE HACKNEY/LINN DATABASES

Table 4-9 in Section 4 presents an algorithm which is able to generate an EVR value for each event in a time/activity database, given that the database is one of the four types listed in Table 4-2. Three of the Hackney/Linn studies (elementary school, high school, and outdoor workers) produced Type 1 databases. The remaining study (construction workers) produced a Type 2 database. Consequently, the application of the algorithm to these databases should provide an indication of model performance with respect to Type 1 and Type 2 databases.

It should be noted that the model is expected to provide better results when applied to Type 1 and Type 2 databases than when applied to Type 3 and Type 4 databases. Types 1 and 2 include event-specific data on breathing rate (classified as sleeping, slow, medium, and fast) which are used by the model to improve its predictions. Types 3 and 4 do not include breathing rate data.

The algorithm was applied to the elementary school database in the following manner. Researchers used the regression results listed in Table 4-4 for Candidate Variable Set 1 to determine the set of predictor variables, the coefficient of each variable, and the constant. The selected predictor variables were LGM, OUTDOOR, FAST, DAYACT, SLEEP, WEEKDAY, and HIGHACT. The constant was -0.082, the coefficient for LGM was 0.986, the coefficient for OUTDOOR was 0.122, and so on. The resulting EVR generator equation was

$$\begin{aligned}\ln(EVR) = -0.082 &+ (0.986)(LGM) + (0.122)(OUTDOOR) + \\ &(0.161)(FAST) + (0.072)(DAYACT) + (-0.174)(SLEEP) \\ &+ (0.047)(WEEKDAY) + (0.060)(HIGHACT) + e.\end{aligned}$$

This equation was applied to each event listed in the elementary school database. The values of OUTDOOR, FAST, DAYACT, SLEEP, WEEKDAY, and HIGHACT for each event were determined by diary entries associated with the event. The value of LGM was constant for each of the 16 subjects, but was allowed to vary among subjects. The LGM value for each subject was randomly selected from a normal distribution with mean = 2.3629 and standard deviation = 0.4324, the normal distribution specified in Table 4-8 for elementary school students.

The value of e was selected from a normal distribution with mean = 0 and standard deviation = SDRES. The value of SDRES was constant for each subject. Subject-specific SDRES values were selected from a lognormal distribution defined by the parameters MU = -1.6068 and SIGMA = 0.4450. These parameter values were also obtained from Table 4-8 (elementary school -- regression model 1).

Table 5-1 provides descriptive statistics for the event EVR values generated by three applications (runs) of the model to the elementary school database. The results vary from run to run because of the random elements incorporated into the Monte Carlo algorithm. Table 5-1 also presents the average of the three runs and descriptive statistics for the observed event EVR values. A comparison of the three-run model averages with the corresponding observed statistics indicates good agreement (less than a 10 percent difference) with respect to arithmetic mean, standard deviation, and percentiles up to the 99th percentile. The model underestimates the 99.5th percentile (36.32 $l \cdot min^{-1} \cdot m^{-2}$ versus 48.18 $l \cdot min^{-1} \cdot m^{-2}$) and the maximum value (52.70 $l \cdot min^{-1} \cdot m^{-2}$ versus 86.04 $l \cdot min^{-1} \cdot m^{-2}$).

Table 5-1. Descriptive statistics for modeled and observed event EVR values (elementary school database).

Statistic[a]	Modeled data				Observed data
	Run 1	Run 2	Run 3	Average of three statistics	
Number of event EVR values	870	870	870	870	870
Arithmetic mean	13.57	11.94	12.84	12.78	12.45
Arithmetic std. dev.	7.11	5.95	6.18	6.41	6.53
Skewness[b]	1.67	1.05	1.44	1.39	3.71
Kurtosis[b]	4.78	1.14	3.67	3.20	30.71
Minimum	4.07	2.03	2.57	2.89	2.80
25th percentile	8.44	7.64	9.20	8.43	8.64
50th percentile	11.63	10.10	11.65	11.13	11.22
75th percentile	16.81	15.37	15.48	15.89	15.21
90th percentile	23.55	20.31	20.30	21.39	19.72
95th percentile	27.79	23.26	24.16	25.07	21.98
98th percentile	33.15	27.36	32.13	30.88	27.36
99th percentile	36.03	30.19	35.86	34.03	30.52
99.5th percentile	40.27	31.63	37.05	36.32	48.18
Maximum	67.05	41.23	49.81	52.70	86.04

[a]Units are liters · min^{-1} · m^{-2} unless otherwise indicated.
[b]Dimensionless.

This analysis was repeated for the high school, outdoor worker, and construction worker databases. The EVR generator equation for the high school database included a constant (0.16385) and 12 variables. The first grouping in Table 4-5 lists these variables and the associated coefficients. For example, the table indicates that OUTOTHER is one of the variables and that its coefficient is 0.11198. Consistent with Table 4-8, LGM values for the high school database were selected from a normal distribution with mean equal to 2.1621 and standard deviation equal to 0.1890. MU was set equal to -1.4662; SIGMA was 0.2997.

Table 5-2 presents descriptive statistics for three applications of the algorithm to the high school database, averages of these statistics, and descriptive statistics for the observed EVR values. The modeled and observed data compare favorably with respect to the mean, standard deviation, and percentiles up to the 99th percentile. The model underestimates the 99.5th percentile (21.28 $l \cdot min^{-1} \cdot m^{-2}$ versus 28.81 $l \cdot min^{-1} \cdot m^{-2}$) and the maximum value (31.61 $l \cdot min^{-1} \cdot m^{-2}$ versus 48.67 $l \cdot min^{-1} \cdot m^{-2}$).

The EVR generator equation for the outdoor worker database included a constant (0.05679) and the 12 variables listed in the first grouping in Table 4-6. Consistent with Table 4-8, LGM values for the outdoor worker databse were selected from a normal distribution with mean equal to 2.0740 and standard deviation equal to 0.2918. MU and SIGMA were set equal to -1.1335 and 0.4338, respectively.

Table 5-3 presents descriptive statistics for three applications of the algorithm to the outdoor worker database, averages of these statistics, and descriptive statistics for the observed EVR values. The modeled and observed data compare favorably with respect to the mean and percentiles up to the 90th percentile. The model underestimates the standard deviation (5.49 $l \cdot min^{-1} \cdot m^{-2}$ versus 10.42 $l \cdot min^{-1} \cdot m^{-2}$) and the percentiles above the 90th percentile.

Table 5-2. Descriptive statistics for modeled and observed event EVR values (high school database).

Statistic[a]	Modeled data				Observed data
	Run 1	Run 2	Run 3	Average of three statistics	
Number of event EVR values	2055	2055	2055	2055	2055
Arithmetic mean	9.30	9.55	9.34	9.40	9.21
Arithmetic std. dev.	2.82	3.31	2.99	3.04	3.75
Skewness[b]	1.00	2.59	0.92	1.50	3.30
Kurtosis[b]	2.13	15.53	1.24	6.30	22.07
Minimum	2.47	3.31	3.36	3.05	3.73
25th percentile	7.37	7.52	7.21	7.37	6.96
50th percentile	8.95	8.99	8.85	8.93	8.41
75th percentile	10.78	10.89	11.05	10.91	10.59
90th percentile	12.71	13.15	13.21	13.02	13.27
95th percentile	14.58	14.98	14.81	14.79	15.51
98th percentile	16.97	17.42	17.33	17.24	18.25
99th percentile	18.50	21.67	18.42	19.53	20.80
99.5th percentile	19.21	24.68	19.94	21.28	28.81
Maximum	27.66	43.02	24.14	31.61	48.67

[a]Units are liters $\cdot$ min^{-1} $\cdot$ m^{-2} unless otherwise indicated.
[b]Dimensionless.

Table 5-3. Descriptive statistics for modeled and observed event EVR values (outdoor worker database).

Statistic[a]	Modeled data				Observed data
	Run 1	Run 2	Run 3	Average of three statistics	
Number of event EVR values	2190	2190	2190	2190	2190
Arithmetic mean	9.28	8.38	9.69	9.12	9.90
Arithmetic std. dev.	5.42	4.40	6.65	5.49	10.42
Skewness[b]	2.37	2.33	7.46	4.05	5.65
Kurtosis[b]	10.69	9.89	113.72	44.77	39.45
Minimum	1.28	1.67	0.84	1.26	2.25
25th percentile	5.76	5.54	6.12	5.81	6.03
50th percentile	8.28	7.33	8.49	8.03	7.71
75th percentile	11.36	9.91	11.62	10.96	9.52
90th percentile	15.22	13.64	15.53	14.80	13.01
95th percentile	18.54	16.90	18.97	18.14	22.86
98th percentile	24.62	20.62	24.71	23.32	44.91
99th percentile	31.78	24.97	29.29	28.68	65.07
99.5th percentile	35.86	27.79	38.79	34.15	80.24
Maximum	60.25	47.53	137.17	81.65	119.47

[a]Units are liters $\cdot$ min^{-1} $\cdot$ m^{-2} unless otherwise indicated.
[b]Dimensionless.

Analysts hypothesized that the relatively poor matchup between modeled and observed EVR values in the upper percentiles was caused by an anomaly in the observed data. As indicated in Table 2-12, the subject associated with the highest average observed EVR (Subject No. 1299) is also the subject who contributed the largest number of individual EVR values (178 values = 8 percent of the total) to the database. Consequently, a disproportionately large number of the high observed EVR values are associated with this one subject. When the algorithm is applied to this subject, the algorithm tends to produce EVR values that are lower than those observed for the subject. In this situation, the algorithm will tend to produce fewer high EVR values than the number observed.

To test this hypothesis, analysts recalculated the descriptive statistics in Table 5-3, omitting Subject No. 1299. Table 5-4 presents the results. Omitting Subject No. 1299 produced a significant reduction in the observed values for percentiles above the 90th percentile. However, these new observed percentile values are still larger than the corresponding modeled values. For example, the observed 99th percentile value is 65.07 $l \cdot min^{-1} \cdot m^{-2}$ with Subject No. 1299 and 48.43 $l \cdot min^{-1} \cdot m^{-2}$ without Subject No. 1299. The three-run average for the modeled 99th percentile is 28.68 $l \cdot min^{-1} \cdot m^{-2}$ with Subject No. 1299 and 29.20 $l \cdot min^{-1} \cdot m^{-2}$ without Subject No. 1299. Subject No. 1299 is not the complete explanation for the difference in modeled and observed results in the upper percentiles.

The EVR generator equation for the <u>construction worker database</u> included a constant (-0.038586) and the six variables listed in in Table 4-7 (LGM, OUTDOOR, DAYACT, CONWORK, SLOW, and ACT82). Consistent with Table 4-8, LGM values for the construction worker database were selected from a normal distribution with mean equal to 2.6286 and standard deviation equal to 0.1940. MU and SIGMA were set equal to -1.2307 and 0.3834, respectively.

Table 5-4. Descriptive statistics for modeled and observed event EVR values (outdoor worker database -- Subject No. 1299 omitted).

Statistic[a]	Modeled data				Observed data
	Run 1	Run 2	Run 3	Average of three statistics	
Number of event EVR values	2012	2012	2012	2012	2012
Arithmetic mean	9.45	8.42	10.06	9.31	9.10
Arithmetic std. dev.	5.51	4.53	6.79	5.61	8.35
Skewness[b]	2.38	2.28	7.51	4.06	6.51
Kurtosis[b]	10.54	9.34	112.36	44.08	55.54
Minimum	1.28	1.67	0.84	1.26	2.25
25th percentile	5.91	5.46	6.44	5.94	5.87
50th percentile	8.37	7.33	8.83	8.18	7.56
75th percentile	11.53	9.98	11.95	11.15	9.35
90th percentile	15.49	13.88	15.85	15.07	12.11
95th percentile	18.85	17.36	19.54	18.58	17.30
98th percentile	25.96	21.51	25.54	24.34	34.92
99th percentile	32.95	25.13	29.51	29.20	48.43
99.5th percentile	36.70	28.43	40.11	35.08	74.58
Maximum	60.25	47.53	137.17	81.65	118.51

[a]Units are liters $\cdot$ min^{-1} $\cdot$ m^{-2} unless otherwise indicated.
[b]Dimensionless.

Table 5-5 presents descriptive statistics for three applications of the algorithm to the construction worker database, averages of these statistics, and descriptive statistics for the observed EVR values. The modeled and observed data compare very favorably with respect to the mean, standard deviation, and percentiles up to the 99.5th percentile. The model overestimated the maximum value (94.82 $l \cdot min^{-2} \cdot m^{-1}$ versus 62.38 $l \cdot min^{-2} \cdot m^{-1}$).

In general, the performance of the algorithm was rated "good" with respect to the elementary school, high school, and construction worker databases. Performance was rated as "fair to good" with respect to the outdoor worker database.

Table 5-5. Descriptive statistics for modeled and observed event EVR values (construction worker database).

Statistic[a]	Modeled data				Observed data
	Run 1	Run 2	Run 3	Average of three statistics	
Number of event EVR values	1861	1861	1861	1861	1861
Arithmetic mean	15.81	15.86	15.38	15.68	15.06
Arithmetic std. dev.	6.45	9.00	6.11	7.19	6.45
Skewness[b]	1.91	4.30	1.77	2.66	1.42
Kurtosis[b]	7.90	36.66	10.26	18.27	4.22
Minimum	4.46	2.66	3.12	3.41	1.50
25th percentile	11.67	10.77	11.22	11.22	10.69
50th percentile	14.78	14.28	14.68	14.58	13.85
75th percentile	18.56	18.84	18.36	18.59	18.18
90th percentile	23.04	24.11	22.61	23.25	23.47
95th percentile	27.17	28.97	26.09	27.41	27.03
98th percentile	33.88	37.96	30.59	34.14	31.71
99th percentile	38.59	47.20	34.65	40.15	36.41
99.5th percentile	43.69	58.26	38.15	46.70	42.04
Maximum	70.29	133.00	81.16	94.82	62.38

[a]Units are liters $\cdot$ $min^{-1} \cdot m^{-2}$ unless otherwise indicated.
[b]Dimensionless.

Section 6
SUMMARY AND RECOMMENDATIONS

The EPA has developed the pNEM methodology as a means of evaluating current and proposed NAAQS. The pNEM approach provides exposure estimates for defined population groups based on activity data specific to each group. Similar to other exposure models, pNEM characterizes each exposure event by time period and pollutant concentration. Unlike most other exposure models, pNEM also characterizes each exposure event by EVR, defined as ventilation rate divided by body surface area.

This report presents a series of algorithms that can be used by pNEM and similar models to estimate EVR values by exposure event. Each algorithm is optimized for use with one of four specific data types (Table 4-2). The algorithms were developed through an analysis of data reported by a research team directed by Jack Hackney and William Linn. The Hackney/Linn team conducted four studies in Los Angeles to obtain ventilation rate data representative of typical daily activities.

- Elementary school students
- High school students
- Outdoor workers
- Construction workers.

The heart rate of each study subject was continuously monitored as the subject documented his or her activities in a special diary. Separate clinical trials were conducted in which the heart rate and ventilation rate of each subject were measured simultaneously. These measurements were used to develop a "calibration curve" for each subject relating heart rate to ventilation rate.

Existing calibration curves for three studies (elementary school, high school, and outdoor workers) were used to convert one-minute heart rate measurements into one-minute ventilation rates. The ventilation rate values were in turn divided by the subject's estimated body surface area to produce one-minute EVR values.

Hackney and Linn did not provide calibration curves for the construction worker study. An analysis of the applicable calibration data collected during the construction worker study produced two alternative sets of calibration curves, one based on a linear relationship (Model A) and the other on a log-log relationship (Model C). Each set of calibration curves was used to convert one-minute heart rate values into corresponding one-minute ventilation rates. The results were averaged to produce a final set of one-minute ventilation rates. The values were then divided by body surface area to produce a database listing one-minute EVR values for the construction workers.

Algorithms for predicting EVR were developed by applying a four-step procedure to each of the one-minute EVR databases. In <u>Step 1</u>, ITAQS processed each one-minute EVR database to produce a special "event EVR file." Each file provided a sequence of exposure events keyed to the activities documented by each subject. The listing for each event included the average EVR for the event and the values of 20 variables which were considered likely to influence EVR values.

In <u>Step 2</u>, ITAQS prepared tables of descriptive statistics for event EVR values which had been categorized by breathing rate, activity, microenvironment, time of day, and event duration. These statistics provided an initial means for identifying factors to be considered in developing the EVR prediction algorithms. These factors were compiled into sets of candidate variables, each set specific to a particular database type (Table 4-3).

In Step 3, ITAQS developed one or more Monte Carlo models for each database type. Each model was specific to one of the following four demographic groups: elementary students, high school students, outdoor workers, and construction workers. Each model consisted of an algorithm capable of generating an EVR for each event in a time/activity data set of the specified database type and demographic group. Each algorithm predicted EVR as a function of six or more predictor variables which constituted a predictor set.

Each predictor set was developed by first defining a candidate variable set for the database type and then performing stepwise linear regression analyses to determine which of the candidate variables were significant predictors of EVR for a particular demographic group. The regression analyses were performed on the Hackney/Linn databases, as these were the only databases available which provided a "measured" EVR value for each exposure event. The results of the regression analyses determined the variables to be included in the predictor set and the coefficients of various terms in the associated Monte Carlo model (Equation 7).

The best overall predictor variable was found to be LGM, the natural logarithm of the geometric mean of all event EVR values associated with a subject. Statistical analysis of the LGM values indicated that the distribution of LGM values was approximately normal.

In addition to LGM, the regression analyses suggested that variables associated with microenvironment, daytime activities, the exertion level of activities, day of week, breathing rate, and duration of activity were generally useful in predicting event EVR. It should be noted that the duration variables were among the least significant of these predictors.

Each regression analysis produced a set of residual values, one for each EVR value. Statistical analysis of the residuals indicated that (1) the standard deviation of the residuals varied significantly from subject to subject and (2) the distribution of the subject-specific standard deviations was approximately lognormal.

Table 4-9 presents the general algorithm included in each Monte Carlo model. When this algorithm is applied to an appropriate database, it generates a sequence of event EVR values for each subject in the database. The EVR value generated for each individual event is determined by the values of the specified predictor variables, the regression coefficient associated with each predictor variable, an LGM value randomly selected from a study-specific normal distribution, and a residual value selected from a subject-specific normal distribution. Because the algorithm uses Monte Carlo techniques to produce EVR estimates, each application of an algorithm to a particular time/activity database will produce a different sequence of exposure estimates if the initial random number selected is different.

In Step 4, ITAQS performed an initial validation of the Monte Carlo approach by applying the EVR-generator algorithm to each of the four Hackney/Linn databases. Each application produced a distribution of event EVR values that could be compared with the distribution of "measured" values. For three of the databases (elementary school, high school, and construction workers), the modeled and observed distributions compared favorably with respect to mean, standard deviation, and percentiles up to the 99th or 99.5th percentiles. At higher percentiles, the algorithm tended to underestimate EVR for the elementary and high school databases and over estimate EVR for the construction worker database.

The algorithm did not perform as well for the outdoor worker database. In this case, the model significantly underestimated the standard deviation and the percentiles above the 90th percentile. Analysts found that the differences between modeled and observed distributions were significantly reduced when a particular atypical subject was removed from the analysis. This subject did not account for all of the difference between the modeled and observed results, however.

In general, the performance of the algorithm was rated "good" with respect to the elementary school, high school, and construction worker databases. Performance was rated "fair to good" with respect to the outdoor worker study.

At this stage in its development, the EVR algorithm presented here is already far superior to the algorithm currently installed in pNEM. The new algorithm includes the following advances over the existing pNEM algorithm:

1. Greater specificity: Distinct versions of the new algorithm are available for each of four demographic groups (elementary students, high school students, outdoor workers, and construction workers). The existing pNEM algorithm distinguishes between only two demographic groups (children and adults).

2. Greater prediction power: The new algorithm estimates EVR values as a function of at least six different variables (including breathing rate) which appear in the time/activity data. The existing pNEM algorithm considers only one variable (breathing rate).

3. Partitioning of variability: The new algorithm explicitly accounts for the effects of within-subject variability and between-subject variability. The existing pNEM algorithm does not separate these two sources of variability.

4. Applicability to incomplete databases: Special versions of the new algorithm are available for application to Type 3 databases and Type 4 databases. By definition, these databases do not include data on breathing rate categories (sleeping, slow, medium, and fast). The existing pNEM algorithm requires that the input time/activity data include breathing rate categories.

The new EVR algorithm was constructed to make use of the data items included in the four Hackney/Linn databases while being applicable to the ten time/activity databases listed in Table 4-1. Additional versions of the model should be constructed as new databases become available. EPA is currently developing a national time/activity database using retrospective diaries administered by telephone. ITAQS recommends that API support additional research to develop a version of the

EVR algorithm applicable to this database when it becomes available for use in pNEM.

Researchers typically use the log-linear relationship (Equation 4) as the basis for the calibration curve to convert minute heart rate (MINHR) to minute ventilation rate (MINVR). This equation does not always produce reasonable MINVR values for large MINHR values, particularly when the equation is applied outside the range of calibration measurements. For example, ITAQS found that the log-linear relationship did not provide reasonable estimates of MINVR when applied to the construction worker database.

ITAQS recommends that additional research be conducted to identify a defensible general model that can be used to predict MINVR from MINHR. This general model would be expressible as a mathematical function containing two or three coefficients whose values would be estimated by fits to subject-specific calibration data. William Adams is collecting experimental data that may provide useful calibration data for this effort. In addition, researchers should reanalyze the calibration data previously collected by the Hackney/Linn research team. Note that no attempt has been made to date to fit curves other than Equation 4 to the data obtained from the elementary school, high school, and outdoor worker studies. Other functional forms may provide better fits to the calibration data collected by these studies.

Section 7
REFERENCES

1. Johnson, T., J. Capel, E. Olaguer, and L. Wijnberg. 1993. *Estimation of Ozone Exposures Experienced by Urban Residents Using a Probabilistic Version of NEM.* U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

2. Johnson, T., J. Capel, E. Olaguer, and L. Wijnberg. 1992. *Estimation of Ozone Exposures Experienced by Residents of the ROMNET Domain Using a Probabilistic Version of NEM.* U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

3. Johnson, T., J. Capel, R. Paul, and L. Wijnberg. 1992. *Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Denver Residents Using a Probabilistic Version of NEM.* U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

4. McDonnell, W. F., D. H. Horstman, and M. J. Hazucha. 1983. Pulmonary Effects of Ozone Exposure During Exercise: Dose Response Characteristics. *Journal of Applied Physiology.* 54:1345.

5. Johnson, T., L. Wijnberg, and J. Capel. 1990. *Overview of Research Relating to Population Exposure to Air Pollution.* API Publication No. DR201. American Petroleum Institute, Washington D.C.

6. Johnson, T. 1987. *A Study of Human Activity Patterns in Cincinnati, Ohio.* Electric Power Research Institute, Palo Alto, California.

7. Spier, C. E., D. E. Little, S. C. Trim, T. R. Johnson, W. S. Linn, and J. D. Hackney. 1992. Activity Patterns in Elementary and High School Students Exposed to Oxidant Pollution. *Journal of Exposure Analysis and Environmental Epidemiology.* 2:277-293.

8. Shamoo, D. A., T. R. Johnson, S. C. Trim, D. E. Little, W. S. Linn, and J. D. Hackney. 1991. Activity Patterns in a Panel of Outdoor Workers Exposed to Oxidant Pollution. *Journal of Exposure Analysis and Environmental Epidemiology.* 1:423-438.

9. Linn, W. S., C. E. Spier, and J. D. Hackney. 1993. Activity Patterns in Ozone-Exposed Construction Workers. *Journal of Occupational Medicine and Toxicology*. 2:1-14.

10. Raizenne, M. E., and J. D. Spengler. 1988. Dosimetric Model of Acute Health Effects of Ozone and Acid Aerosols in Children. *Atmospheric Ozone Research and its Policy Implications, Proceedings of the Third US-Dutch International Symposium*. Nijmegan, the Netherlands. pp. 322-329.

11. Astrand, P. and K. Rodahl. 1977. *Textbook of Work Physiology*. McGraw-Hill, New York.

12. DuBois, D. and E. F. DuBois. 1916. A Formula to Estimate the Approximate Surface Area If Height and Weight Be Known. *Archives of Internal Medicine*. 17, 863-871.

13. Anderson, E. et al. 1985. *Development of Statistical Distributions or Ranges of Standard Factors Used in Exposure Assessments*. Report No. GCA-TR-CH-85-01. GCA Corporation, Chapel Hill, North Carolina.

14. Wiley, J. A. et al. 1991. *Study of Children's Activity Patterns*, Research Division, California Air Resources Board, Sacramento, California.

15. Wiley, J. A. et al. 1991. *Activity Patterns of California Residents*, Research Division, California Air Resources Board, Sacramento, California.

16. Johnson, T. 1984. *A Study of Personal Exposure to Carbon Monoxide in Denver, Colorado*. EPA-600/54-84-014, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

17. Goldstein, B. D. et al. 1992. *Valdez Air Health Study*. Alyeska Pipeline Service Company, Anchorage, Alaska.

18. Hartwell, T. D. et al. 1984. *Study of Carbon Monoxide Exposure of Residents of Washington, D. C., and Denver, Colorado*. EPA-600/54-84-031, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina.

APPENDIX A

A COMPARISON OF TEN TIME/ACTIVITY DATABASES: EFFECTS OF GEOGRAPHIC LOCATION, TEMPERATURE, DEMOGRAPHIC GROUP, AND DIARY RECALL METHOD

Appendix A reprinted with the permission of the Air & Waste Management Association from *Tropospheric Ozone: Nonattainment and Design Value Issues*, (ed. J. J. Vostal) TR-23, 1993, pp. 255-276.

A Comparison of Ten Time/Activity Databases: Effects of Geographic Location, Temperature, Demographic Group, and Diary Recall Method

Ted Johnson, Michael McCoy, James E. Capel, and Louis Wijnberg
IT Air Quality Services
3710 University Drive, Suite 201
Durham, North Carolina 27707

Dr. Will Ollison
American Petroleum Institute
1220 L Street, N.W.
Washington, D.C. 20005

ABSTRACT

A number of computer-based exposure models have been developed to realistically simulate the movements of people through zones of varying air quality. Many of these models employ time/activity databases obtained from surveys in which subjects documented their activities in diaries. These surveys differ with respect to geographic location, local climate, demographics of the sampled population, and diary recall approach (real-time versus retrospective). This paper presents results of a preliminary set of statistical analyses performed on ten time/activity databases to determine the significance of these and other factors. The databases represent the activity patterns of residents of five cities (Cincinnati, OH; Denver, CO; Washington, DC; Los Angeles, CA; and Valdez, AK) and of the entire state of California. Results of the analyses suggest that major divisions of the day (sleeping, work, school), student status, and season are generally more important in determining the time spent in selected microenvironments at various breathing rates than retirement status, school commuting periods, and midmorning time periods. The databases associated with the Los Angeles and Cincinnati studies were found to contain the highest number of activities per hour. Each of these databases was obtained through the use of a real-time activity diary. The Valdez database contained the lowest number of diary entries per hour. A 24-hour recall (retrospective) diary was used in this study.

INTRODUCTION

A number of computer-based models have been developed for estimating the exposure of human populations to air pollution. Examples include the exposure models commonly referred to by the acronyms pNEM[1], HAPEM[2], SHAPE[3], SIMSYS[4], PAQM[5], and EPEM[6]. Each of these models attempts to realistically simulate the movements of people through zones of varying air quality. Many of these models employ time/activity databases obtained from surveys in which subjects documented their activities in diaries. These surveys differ with respect to geographic location, local climate, demographics of the sampled population, and diary recall approach (real-time versus retrospective). This paper presents results of a preliminary series of statistical analyses performed on a variety of time/activity databases to determine the significance of these and other factors. The work was performed by IT Air Quality Services (ITAQS) with funding provided by the American Petroleum Institute (API). Other comparisons of time/activity databases can be found in reports by Johnson et al.[7] and by Schwab.[8]

THE TIME/ACTIVITY BASES

A survey was conducted to identify and evaluate all time/activity databases developed since 1980. Researchers identified 12 databases for which adequate descriptive material was available and completed a three-page evaluation form for each database. The evaluation form provided information on the time period included in each diary, the data items included in the diary, the data items included in any supplementary questionnaires, the diary protocol (real-time, prospective, or retrospective), the diary time resolution, and the inclusion of three data items of current interest to the researchers (time spent in current residence, trips made by non-workers, and trips made by students). Ten databases were selected as promising candidates for an in-depth statistical analysis. Table I lists these databases.

Eight of the databases relate to five individual urban areas: Cincinnati, Ohio; Denver, Colorado; Los Angeles, California; Valdez, Alaska; and Washington, D.C. The remaining two databases relate to the entire State of California. In the discussion that follows, each database will be identified according to the associated geographical area. When a geographical area is associated with more than one database, each database will be further distinguished according to the sampled population (e.g., Los Angeles - outdoor workers).

California

The California Air Resources Board conducted two state-wide time/activity studies[9,10] to provide a large pool of activity pattern data suitable for use in estimating environmental exposures. The first study, referred to hereafter as the "California - high school/adult" study, was conducted between October 1987 and July 1988. During the study, interviewers collected one day of activity data from each of 1762 California residents over the age of 11. The second study ("California - children") was conducted from April 1989 through February 1990. The study gathered one day of activity data from each of 1200 children ages 11 and under. Both studies employed retrospective telephone interviews to obtain a record of each subject's activities during the preceding day.

Cincinnati

The Cincinnati Activity Diary Study[11] was conducted by the Electric Power Research Institute during March and August 1985 to provide an extensive database for evaluating human exposure to air pollution. The sampled population included all residents of a three-county area in and around Cincinnati, Ohio. Each subject recorded his or her activities over a three-day period in a real-time diary and completed a detailed background questionnaire. The 487 March subjects provided 1401 subject-days of diary data; the 486 August subjects provided 1399 subject-days. Activity diary data collected during the Cincinnati study have been used by the U.S. Environmental Protection Agency (EPA) in various applications[1,12,13] of the pNEM/CO and pNEM/O3 exposure models.

Denver and Washington

The U.S. Environmental Protection Agency conducted studies of adults (18 to 70 years) in Denver, Colorado, and Washington, D.C., during the winter of 1982 - 1983 for the purpose of collecting representative data on personal exposure to carbon monoxide. In the Denver study[14], each of 454 subjects carried a personal exposure monitor (PEM) and a real-time time/activity diary for two 24-hour periods. Each subject also provided a breath CO sample at the end of each monitored period and completed a detailed background questionnaire. The Washington study[15] employed a similar protocol to obtain data for a single 24-hour period from each of 908 subjects. Activity diary data from these two studies have been used in conjunction with data from the Cincinnati study in applications of EPA's pNEM/CO exposure model[12].

Los Angeles

Between 1989 and 1991, a research team headed by Dr. Jack Hackney conducted four activity diary studies in Los Angeles with funding provided by the American Petroleum Institute. The first of these, the "Los Angeles - outdoor worker study", was conducted during the summer of 1989. Each of 20 outdoor workers wore a heart rate monitor for a three-day period during which he recorded his activities in a real-time diary identical to that used in the Cincinnati study.[16]

In October of 1989, the outdoor worker study was expanded to include 20 healthy elementary school children. During this phase of the Los Angeles study, referred to here as the "Los Angeles - elementary school" study, each child wore a heart-rate monitor for two or three days and recorded his or her activities in the real-time Cincinnati diary. Approximately 58 subject-days of activity data were collected.[17]

A third phase of the Los Angeles study (the "Los Angeles - high school" study) was conducted during September and October 1990. During this phase, 66 subject-days of real-time activity data were collected from 19 students between the ages 13 and 17 using the Cincinnati diary.[18]

The Hackney research team conducted a fourth study in Los Angeles between July and November 1991. Each of 19 construction workers between the ages of 23 and 42 wore a heart rate monitor during a typical work day. The Cincinnati diary was used to record each subject's activities during this period. The study protocol differed from the other Los Angeles studies in that each diary was completed by a trained observer rather than by the subject. The observer monitored each subject's activities visually and by two-way radio. This approach produced unusually detailed diaries of high accuracy.[19]

Valdez

The Valdez Air Health Study[20] was undertaken by the Alyeska Pipeline Service Company in response to concerns expressed by the citizens of Valdez, Alaska, regarding their potential exposure to certain volatile organic compounds (VOCs). Between November 1990 and October 1991, 405 subjects aged 10 to 72 years were interviewed and requested to report their daily activities for an earlier 24-hour period. In addition to the activity data, researchers collected extensive data on personal exposures to VOC's, ambient air quality, and meteorological conditions.

PREPARATION OF DATABASES FOR ANALYSIS

A complete time/activity database was obtained for each of the ten studies listed in Table I. These databases differed greatly with respect to format, documented time periods, data items included, coding conventions, and other factors. To facilitate the statistical comparison of the ten databases, researchers developed a common data format and coding scheme which could be applied to all ten databases. The result was a set of ten simplified "person-hour" databases.

Each record in one of the person-hour databases pertained to a single person-hour of time/activity data. Each person-hour began on a clock hour (midnight, 1 a.m., 2 a.m., etc.) and ended 60 minutes later. To permit sorting by demographic characteristics, each person-hour record included entries assigning the subject to a demographic group (Table II) and indicating whether or not the subject's residence had an attached garage, a central or window air conditioner, or a gas stove. Each person-hour was also classified as to season (summer or not summer), daytype (weekday or weekend), and maximum temperature of the calendar day associated with the person-hour.

In most of the time/activity studies, subjects provided diary records for continuous 24-hour periods which included sleep time. The subjects of the Los Angeles - construction worker study omitted most of their activities which occurred outside working hours, including sleep time. As no attempt was made by researchers to estimate activities during these periods, the person-hour database developed for this study is biased toward daytime work-related activities.

The subjects of the other three Los Angeles studies provided better coverage of their daily activities, routinely reporting evening activities through 11 pm. The diary entries for these subjects frequently omit the actual time that the subject retired, however. In these cases, researchers with IT Air Quality Services made reasonable assumptions as to the probable bedtime of each subject and added this information to the associated person-hour database. As a result of this minor data enhancement effort, most of the subjects of the time/activity studies listed in Table I were represented in the person-hour databases by one or more continuous 24-hour time periods which included sleep time.

Five of the person-hour databases were obtained from studies which used an activity diary originally developed for the Cincinnati study. The Cincinnati diary provides data on both the microenvironment and the breathing rate associated with each activity documented by a subject. Table II lists five microenvironment and five breathing rate categories which have been applied to the Cincinnati data in various applications of the pNEM exposure model. Based on these categories, analysts defined a set of 25 distinct combinations of microenvironment and breathing rate. Each person-hour record associated with a study that used the Cincinnati diary included 25 values for MIN(m,b) and 25 values for ACT(m,b), where MIN(m,b) equaled the number of minutes associated with microenvironment m and breathing rate b and ACT(m,b) equaled the number of distinct activities associated with the m,b combination.

Breathing rate data were not available for the remaining studies. Consequently, the person-hour databases developed for these studies included entries for MIN(m) and ACT(m) where each variable was classified as to microenvironment but not as to breathing rate. The set of five microenvironments listed in Table II was employed in developing each of the Denver, Washington, and Los Angeles person-hour databases. As the diary location codes used in the two California studies and in the Valdez study did not permit analysts to differentiate between outdoor activities conducted near roads (Category 3) and other outdoor activities (Category 4), these two categories were combined into a general outdoor category (designated "3/4") in these databases.

Table III presents a preliminary set of 22 binary or "dummy" variables which were defined to facilitate statistical analyses of the person-hour databases. The value of each variable was set equal to 1 when the indicated condition(s) were met and equal to zero in all other cases. It should be noted that some of the variables were assigned the same value for all records in a particular database. For example, OUTWORK (i.e., the subject is a member of demographic group 6, "outdoor workers") was set equal to 1 for all records associated with the Los Angeles - outdoor worker study. It should also be noted that the air conditioning variables ACRES and ACPOT were excluded from five databases (California - 11 and under, California - 12 and over, Los Angeles - elementary students, Los Angeles - outdoor workers, Washington, and Valdez) because data concerning these variables were unavailable or difficult to interpret.

STATISTICAL ANALYSIS OF THE PERSON-HOUR DATABASES

Table IV provides arithmetic means for the MIN(m,b) and ACT(m,b) for the five studies which employed the Cincinnati diary. Table V lists mean values of MIN(m) and ACT(m) for the Denver and Washington studies for each of five defined microenvironments. Table VI contains the means for the two California studies and for the Valdez study. In this table, the means are listed for four microenvironments. As previously discussed, the coding convention used in these studies does not permit analysts to distinguish between activities associated with the "outdoors - near road" microenvironment and those associated with "outdoors - other." Consequently, these categories were collapsed into the single "outdoors" microenvironment.

According to the mean MIN statistics, all databases indicate that the majority of time is spent in Microenvironments No. 1 (indoors - residence) and No. 2 (indoors - other). The total time spent in the two outdoor microenvironments is higher for the Los Angeles databases than for the other databases. This pattern is probably the result of 1) limiting the subjects of the Los Angeles studies to young people and outdoor workers and 2) omitting nighttime hours from the sampled time periods in the Los Angeles- construction worker study. It is interesting to note that the twin Denver and Washington studies produce almost identical results with respect to the mean values of MIN (Table V).

Consistent with expectations, the apportionment of time among microenvironments and breathing rates varies with time of day. The general pattern for Cincinnati subjects can be observed in Figure 1, which presents mean statistics (minutes per hour) by time of day. Time spent in the indoors - other microenvironment (No. 2) peaks in the early afternoon. Time spent in the outdoor and in-vehicle microenvironments (Nos. 3, 4, and 5) is highest between the hours of 3:00 p.m. and 8:00 p.m. With respect to breathing rate categories, subjects are most likely to report time at the moderate (No. 3) and fast (No. 4) levels during afternoon hours.

Table VII provides a listing for each person-hour database of the average number of activities (all microenvironments) occurring per person-hour within eight specified time periods. Hereafter, statistics of this kind will be referred to as "activity rates." The Table VII tabulation provides a means for comparing the relative level of detail provided by the various databases adjusted for time of day. According to the activity rate indicators presented in Table VII, most of the databases show less activity during nighttime hours, particularly the hours 0 to 5, and more activity during daylight hours. The pattern observed in the Cincinnati database is typical. The activity rate averages 1.1 activities per person-hour for the 0 - 5 time period, increases to 2.1 or 2.2 for the early and mid-morning time periods, rises to a value between 2.3 and 2.5 for the late morning and afternoon time periods, falls back to 2.1 for early evening, and then decreases to 1.6 for the late evening hours before midnight.

The differences among the databases with respect to activity rates are likely the result of differences in the sampled populations and in the data collection protocols. Although the relative significance of these factors is difficult to determine, the results in Table VII suggest that data collection protocol is particularly important. The five databases with the highest activity rate values (the four Los Angeles databases and Cincinnati) were all obtained through the use of real-time diaries based on the Cincinnati diary format. The overall activity rate of each of these studies equals or exceeds 1.9 activities per person-hour.

The Valdez database had the lowest overall activity rate (1.4). The Valdez data were obtained through the use of a retrospective diary administered by telephone. The remaining two real-time databases (Denver and Washington) and the remaining two retrospective databases (the California studies) share a common overall activity rate, 1.6 activities per person-hour. The results would seem to suggest that the combination of real-time data collection and the use of the Cincinnati diary yields higher activity rates. As demonstrated by the very high overall activity rate of the Los Angeles - construction database (10.3), the use of a trained observer appears to significantly increase the level of detail documented in each activity diary.

A preliminary series of stepwise linear regression analyses were performed on the person-hour databases. In each analysis, the dependent variable was either MIN(m,b) or MIN(m). The set of independent (or explanatory) variables in each regression analysis included all or most of the 22 variables listed in Table 3. Each regression analysis permitted forward and backward stepping. The F statistic for determining variable inclusion was set at 0.15. Under this generous inclusion criterion, the majority of the regression equations contained at least three explanatory variables.

The stepwise linear regression analyses were performed on six person-hour databases: Cincinnati, Denver, Washington, Valdez, aggregate California, and aggregate Los Angeles. The aggregate California database was formed by combining the two individual databases associating with California. In a similar way, the aggregate Los Angeles database was formed by combining the four individual databases associated with Los Angeles. The two aggregate databases were similar to the other four databases in that each database included a mix of demographic groups, facilitating the use of the demographic-group dummy variables in the regression analysis of each database. This mix was not available in the individual California and Los Angeles databases.

Tables VIII through XIII provide results of these analyses by database, by microenvironment, and by breathing rate (when applicable). The table entry for each regression analysis lists the three most significant explanatory variables and the regression coefficient of each of these variables. As each coefficient value has the units of minutes, each coefficient can be interpreted as the average number of minutes that will be added or subtracted to the hourly total for the indicated combination of microenvironment and breathing rate when the variable takes a value of 1. For example, the entry in Table VIII for Microenvironment No. 1 (indoors - residence) and Breathing Rate No. 2 (slow - awake) lists -30.3 as the coefficient for the NIGHT variable. This result can be interpreted as indicating that the time spent during a person-hour in this combination of microenvironment and breathing rate will be reduced by 30.3 minutes, on average, when the person-hour falls within the NIGHT time period (hours 0 to 5).

The following variables appear at least nine times in the listings of the most significant variables (first, second, and third) in Tables VIII through XIII: NIGHT (59 listings), SUNUP (20), WORKTIME (19), SCHTIME (17), MIDAFT (11), SCHOOL (11), LUNCH (9), SUMMER (9), and OUTWKR (9). The following variables do not appear more than four times: MIDMORN (0), SCHCOM (0), RETIRE (0), SUMHOT (1), ACRES (1), WORKER (1), WINCOLD (2), ACPOT (3), SUNDOWN (3), and WINWARM (4). With respect to the variable listed as most significant (first) in each listing, the following variables appear at least four times: NIGHT (34), WORKTIME (11), SCHTIME (6), and SCHOOL (4). The following variables never appear as first variable: SUMHOT, ACPOT, SUNUP, MIDMORN, MIDAFT, SCHCOM, WORKER, and RETIRE. These results suggest that variables relating to major divisions of the day (e.g., NIGHT, SCHTIME), student status (SCHOOL), and season (SUMMER) are generally more useful in predicting activities than variables relating to temperature (e.g., SUMHOT), midmorning time periods (MIDMORN), school commuting time periods (e.g., SCHCOM), and general work status (e.g., WORKER, RETIRE). The infrequent occurrence in the tables of the air conditioning variables (ACRES and ACPOT) is difficult to interpret because these variables were not present in many of the databases which were analyzed.

CONCLUSIONS

The American Petroleum Institute is conducting an on-going research program to evaluate the various databases which have been developed for use in exposure modeling studies. The program will also sponsor original research to fill in data gaps as they are identified. The results summarized above represent the preliminary findings of research efforts within this program to evaluate and compare the various time/activity databases which have been developed for use in estimating population exposure to air pollution. Additional analyses of these databases are planned, and new time/activity databases will be analyzed as they become available.

The preliminary results presented here suggest that the data collection protocol can have a significant effect on the level of detail in the collected data. The combined use of observers and the real-time Cincinnati diary in the Los Angeles - construction study appears to have increased the level of detail (as quantified by the activity rate statistic) by at least a factor of four. The use of the real-time Cincinnati diary appears to have contributed to the relatively high levels of detail observed in the other Los Angeles databases and in the Cincinnati database. The use of a retrospective diary in the Valdez study may have contributed to the relatively low level of detail observed in that database.

The results of an initial set of stepwise regression analyses suggest that factors related to major divisions of the day (particularly time periods associated with sleeping, work, and school), student status, and season are relatively good predictors of time/activity patterns. Factors related to retirement status, school commuting periods, and midmorning time periods are relatively poor predictors. It should be noted, however, that these results are dependent on the set of candidate variables available for selection into each regression model. The set of 22 variables defined here may not adequately represent some of the proposed explanatory factors such as temperature. Other potentially important explanatory factors, such as precipitation, have been omitted completely. Additional analyses should be performed to establish a definitive set of variables for predicting time/activity patterns.

REFERENCES

1. T. Johnson, J. Capel, E. Olaguer, and L. Wijnberg, Estimation of Ozone Exposures Experienced by Urban Residents Using a Probabilistic Version of NEM, prepared by IT Air Quality Services for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 1992.

2. T.R. Johnson, J.E. Capel, R.A. Paul and E.P Olaguer, A Comparison of the Hazardous Air Pollutant Exposure Model (HAPEM) and the Human Exposure Model (HEM), prepared by International Technology Air Quality Services for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, September 1991.

3. W. Ott, D. Mage, and L. Wallace, "Validation of the Simulation of Human Activity and Pollutant Exposure (SHAPE) Model Using Paired Days from the Denver, Colorado, Carbon Monoxide Field Study," Atmospheric Environment, 22:1 (1988).

4. P.B. Ryan, "An Overview of Human Exposure Modeling," Toxicology and Industrial Health, 1:4, pp.453 - 474 (1991).

5. S.R. Hayes, "Estimating the Effect of Being Indoors on Total Personal Exposure to Outdoor Air Pollution," JAPCA, 39(11):1453 (1989).

6. T. Johnson, L. Wijnberg, J. Capel, and J. Vostal, "The Use of Activity Diary Data to Estimate the Probability of Exposure to Air Pollution," Proceedings of the 1991 International Specialty Conference on Tropospheric Ozone and the Environment II, Air and Waste Management Association, 1991.

7. T. Johnson, L. Wijnberg, and J. Capel, Review and Evaluation of New Research Relating to Population Exposure to Air Pollution, prepared by PEI Associates, Inc., for the American Petroleum Institute, Washington, D.C., April 1990.

8. M. Schwab, "Time Activity Patterns in Air Pollution Exposure: A Review of Methods, Available Information, Documented Relationships, and New Directions for Data Collection," Workshop on Measuring, Understanding, and Predicting Exposures in the 21st Century, International Society of Exposure Analysis, Atlanta, November 1991.

9. J.A. Wiley et al., Activity Patterns of California Residents, Research Division, California Air Resources Board, Sacramento, California, May 1991.

10. J.A. Wiley et al., Study of Children's Activity Patterns, Research Division, California Air Resources Board, Sacramento, California, September, 1991.

11. T.R. Johnson, A Study of Human Activity Patterns in Cincinnati, Ohio, Electric Power Research Institute, Palo Alto, November 1987.

12. T. Johnson, J. Capel, L. Wijnberg, and T. McCurdy, "Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Denver Residents Using a Probabilistic Version of NEM," paper no. 92-145.01, 85th Annual Meeting of the Air and Waste Management Association, Kansas City, Missouri, 1992.

13. T. Johnson, J. Capel, E. Olaguer, and L. Wijnberg, "Estimation of Ozone Exposures Experienced by Residents of the ROMNET Domain Using a Probabilistic Version of NEM," prepared by IT Air Quality Services for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 1992.

14. T.R. Johnson, A Study of Personal Exposure to Carbon Monoxide in Denver, Colorado, EPA-600/54-84-014, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, March 1984.

15. T.D. Hartwell et al., Study of Carbon Monoxide Exposure of Residents of Washington, D.C., and Denver, Colorado, EPA-600/54-84-031, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, March 1984.

16. J.D. Hackney et al., Development of Improved Methods to Measure Effective Doses of Ozone, prepared by Environmental Health Service for the American Petroleum Institute, Washington, D.C., November 1989.

17. J.D. Hackney et al., Activity Patterns in Children, prepared by Environmental Health Service for the American Petroleum Institute, Washington, D.C., January 1990.

18. J.D. Hackney et al., Activity Patterns in High School Students, prepared by Environmental Health Service for the American Petroleum Institute, Washington, D.C., January 1991.

19. J.D. Hackney et al., Activity Patterns in Ozone-Exposed Construction Workers, prepared by Environmental Health Service for the American Petroleum Institute, Washington, D.C., February 1992.

20. B.D. Goldstein et al., Valdez Air Health Study, Alyeska Pipeline Service Company, Anchorage, Alaska, June 1992.

Table I. Characteristics of studies associated with the 10 time/activity databases.

Database name	Reference number	Characteristics of subjects	Number of subject-days	Study calendar periods	Diary type	Diary time period	Breathing rates reported?
California - 11 and under	10	Children ages 1 to 11	1200	April 1989 - Feb. 1990	Retrospective	Midnight to midnight	No
California - 12 and over	9	Ages 12 to 94	1762	Oct. 1987 - July 1988	Retrospective	Midnight to midnight	No
Cincinnati	11	Ages 0 to 86	2800	March and August 1985	Real-time	Midnight to midnight	Yes
Denver	14	Ages 18 to 70	859	Nov. 1982 - Feb. 1983	Real-time	7 p.m. to 7 p.m. (nominal)	No
Los Angeles - construction	19	Construction workers (ages 23-42)	19	July - Nov. 1991	Real-time[a]	Wakeup to subject returns home from work	Yes
Los Angeles - elem. school	17	Elementary school students, 10 to 12 years	58	Oct. 1989	Real-time[a]	Midnight to midnight	Yes
Los Angeles - high school	18	High school students, 13 to 17 years	66	Sept. and Oct. 1990	Real-time[a]	Midnight to midnight	Yes
Los Angeles - outdoor worker	16	Adult outdoor workers (ages 19-50)	60	Summer 1989	Real-time[a]	Midnight to midnight	Yes
Valdez	20	Ages 10 to 72	405	Nov. 1990 - Oct. 1991	Retrospective	Varying 24-h period	No
Washington	15	Ages 18 to 70	705	Nov. 1982 - Feb. 1983	Real-time	7 p.m. to 7 p.m. (nominal)	No

[a]Study employed the Cincinnati diary format.

Table II. Codes used to identify demographic groups, microenvironment categories, and breathing rate categories.

Classification variable	Category	Code
Demographic group (DG)	Children 0 to 5 years	1
	Children 6 to 13 years	2
	Children 14 to 18 years	3
	Workers with low probability of outdoor work	4
	Workers with moderate probability of outdoor work	5
	Workers with high probability of outdoor work	6
	Nonworking adults under 35 years	7
	Nonworking adults 35 to 54 years	8
	Nonworking adults 55+ years	9
Microenvironment (ME)	Indoors - residence	1
	Indoors - other	2
	Outdoors - near road	3
	Outdoors - other	4
	In vehicle	5
Breathing rate (BR)	Slow - sleeping	1
	Slow - awake	2
	Moderate	3
	Fast	4

Table III. Dummy variables created for stepwise linear regression analysis.

Variable type	Variable name	Conditions for value = 1
Season/temperature/AC	SUMMER	season = summer
	SUMHOT	(season = summer) and (temperature ≥ 84° F.)
	WINWARM	(season = not summer) and (temperature ≥ 55°F.)
	WINCOLD	(season = not summer) and (temperature < 55° F.)
	ACRES	residential AC = yes
	ACPOT	(residential AC = yes) and (temperature > 75°F.)
Time of day/day of week	NIGHT	(hour < 6) or (hour > 22)
	SUNUP	hour = 6, 7, or 8
	MIDMORN	hour = 9 or 10
	LUNCH	hour = 11, 12, or 13
	MIDAFT	hour 14, 15, or 16
	SUNDOWN	hour = 17 or 18
	SCHCOM	(DG = 2 or 3) and (day type = weekday) and (hour = 6, 7, 8 ,14, 15, or 16) and (season = not summer)
	WORKCOM	(DG = 4, 5, or 6) and (day type = weekday) and (hour = 6, 7, 8, 16, 17, or 18)
	SCHTIME	(DG = 2 or 3) and (day type = weekday) and (7 < hour < 16) and (season = not summer)
	WORKTIME	(DG = 4, 5, or 6) and (day type = weekday) and (7 < hour < 19)
	WEEKEND	day type = weekend
Demographic group	PRESCH	DG = 1
	SCHOOL	DG = 2 or 3
	WORKER	DG = 4 or 5
	OUTWKR	DG = 6
	RETIRE	DG = 9

Notes

temperature: maximum daily temperature
AC: Air conditioning system
DG: demographic group (see Table II)

Table IV. Arithmetic means of MIN(m,b) and ACT(m,b) statistics for Cincinnati and Los Angeles databases.

		Arithmetic means by database									
		Cincinnati		Los Angeles - outdoor worker		Los Angeles - elem. school		Los Angeles - high school		Los Angeles - construction	
ME[a]	BR[b]	MIN[c]	ACT[d]	MIN	ACT	MIN	ACT	MIN	ACT	MIN	ACT
1	Any	42.3	1.10	36.7	1.02	33.7	1.06	40.3	1.40	2.6	0.16
	1	22.2	0.41	22.0	0.40	18.6	0.34	24.9	0.46	0.0	0.00
	2	19.2	0.65	13.7	0.57	10.3	0.47	13.3	0.79	2.5	0.14
	3	0.8	0.03	1.0	0.04	4.6	0.24	1.9	0.15	0.1	0.01
	4	0.1	0.00	0.0	0.00	0.3	0.02	0.1	0.01	0.0	0.00
2	Any	9.5	0.27	7.5	0.29	6.0	0.31	11.1	0.53	0.7	0.05
	1	0.1	0.00	0.4	0.01	0.0	0.00	1.5	0.03	0.0	0.00
	2	8.5	0.24	6.5	0.26	3.3	0.16	7.1	0.36	0.5	0.03
	3	0.8	0.03	0.6	0.02	2.2	0.12	2.2	0.13	0.2	0.01
	4	0.1	0.00	0.0	0.00	0.4	0.02	0.3	0.02	0.0	0.00
3	Any	2.1	0.20	6.5	0.47	4.6	0.20	2.1	0.22	0.7	0.06
	1	0.1	0.00	0.0	0.00	0.4	0.01	0.0	0.00	0.0	0.0
	2	0.9	0.11	5.5	0.39	2.0	0.07	0.9	0.11	0.6	0.04
	3	1.1	0.08	0.9	0.07	1.7	0.09	1.1	0.11	0.1	0.01
	4	0.1	0.01	0.1	0.00	0.6	0.02	0.1	0.01	0.0	0.0
4	Any	2.9	0.17	3.5	0.21	13.6	0.45	3.9	0.35	50.8	9.84
	1	0.1	0.00	0.0	0.00	4.2	0.08	0.0	0.00	0.0	0.0
	2	1.5	0.11	2.7	0.16	4.9	0.19	1.6	0.16	17.5	3.83
	3	1.1	0.06	0.5	0.04	3.6	0.14	1.8	0.15	31.2	5.24
	4	0.2	0.01	0.2	0.01	0.9	0.03	0.5	0.03	2.0	0.76
5	Any	3.1	0.19	5.8	0.38	2.1	0.10	2.6	0.16	5.2	0.22
	1	0.0	0.00	0.0	0.00	0.0	0.00	0.1	0.00	0.0	0.0
	2	3.1	0.19	5.4	0.36	1.5	0.08	2.1	0.13	4.3	0.18
	3	0.0	0.00	0.4	0.02	0.5	0.02	0.4	0.02	0.6	0.03
	4	0.0	0.00	0.0	0.00	0.0	0.00	0.0	0.00	0.2	0.0
Any	Any		1.93		2.36		2.12		2.67		10.32

[a]Microenvironment (see Table II).
[b]Breathing rate (see Table II).
[c]Mean number of minutes per person-hour.
[d]Mean number of activities per person-hour.

Table V. Arithmetic means of MIN(m) and ACT(m) statistics for Denver and Washington databases.

Microenvironment	Arithmetic means by database			
	Denver		Washington	
	MIN[a]	ACT[b]	MIN	ACT
Indoors - residence	43.1	1.02	42.3	1.00
Indoors - other	11.2	0.30	11.9	0.30
Outdoors - near road	1.0	0.06	1.2	0.07
Outdoors - other	0.6	0.03	0.4	0.02
In vehicle	4.0	0.22	4.3	0.20
All		1.63		1.59

[a]Mean number of minutes per person-hour.
[b]Mean number of activities per person-hour.

Table VI. Arithmetic means of MIN(m) and ACT(m) statistics for California and Valdez databases.

Microenvironment	Arithmetic means by database					
	California - 11 and under		California - 12 and over		Valdez	
	MIN[a]	ACT[b]	MIN	ACT	MIN	ACT
Indoors - residence	45.3	1.11	41.2	1.00	39.8	0.79
Indoors - other	5.9	0.17	10.8	0.28	11.6	0.28
Outdoors	6.2	0.22	4.1	0.13	4.3	0.12
In vehicle	2.5	0.14	3.9	0.19	4.3	0.17
All		1.64		1.60		1.36

[a]Mean number of minutes per person-hour.
[b]Mean number of activities per person-hour.

Table VII. Arithmetic means for the number of activities per person-hour according to time period.

Database	Mean number of activities per person-hour according to time period								
	0-5	6-8	9-10	11-13	14-16	17-18	19-21	22-23	All
California - 11 and under	1.0	1.8	1.8	2.0	2.0	2.1	1.9	1.2	1.6
California - 12 and over	1.1	1.8	1.7	1.9	1.8	2.0	1.7	1.4	1.6
Cincinnati	1.1	2.1	2.2	2.4	2.3	2.5	2.1	1.6	1.9
Denver	1.1	1.9	1.9	2.0	1.9	2.1	1.7	1.5	1.6
Los Angeles - construction	2.9	13.1	12.6	9.6	8.6	2.2	2.0	a	10.3
Los Angeles - elem. school	1.0	3.0	2.8	2.8	2.7	2.9	2.2	1.2	2.1
Los Angeles - high school	1.0	3.1	3.8	3.9	3.4	3.4	3.4	1.7	2.7
Los Angeles - outdoor workers	1.2	2.3	3.1	3.2	3.1	3.1	2.4	2.0	2.4
Valdez	1.0	1.4	1.4	1.6	1.5	1.6	1.4	1.2	1.4
Washington	1.1	1.8	1.7	1.9	1.8	1.9	1.7	1.5	1.6

[a]No data obtained for this time period.

Table VIII. Results of stepwise linear regression analyses performed on the Cincinnati person-hour database (62,671 person-hours).

		Regression model		Most significant variables					
				First		Second		Third	
ME[a]	BR[b]	Variables in model	R^2	Variable	Coef.[c]	Variable	Coef.	Variable	Coef.
1	Any	19	0.312	NIGHT	14.1	WORKTIME	-17.8	SCHTIME	-27.5
	1	16	0.622	NIGHT	45.4	SUNUP	27.2	PRESCH	7.1
	2	20	0.247	NIGHT	-30.3	WORKTIME	-13.6	SUNUP	-17.1
	3	15	0.014	NIGHT	-0.8	SUNUP	-0.3	PRESCH	0.1
	4	11	0.002	WINCOLD	0.1	NIGHT	-0.1	PRESCH	0.0
2	Any	18	0.290	WORKTIME	19.1	SCHTIME	32.5	NIGHT	-5.2
	1	11	0.003	PRESCH	0.3	OUTWKR	0.3	NIGHT	0.1
	2	17	0.285	WORKTIME	18.4	SCHTIME	31.9	NIGHT	-4.6
	3	14	0.013	WORKTIME	1.0	NIGHT	-0.6	WORKER	0.5
	4	11	0.004	SCHTIME	0.6	SUMMER	-0.1	NIGHT	-0.1
3	Any	19	0.048	NIGHT	-2.6	SUNUP	-1.9	SCHOOL	1.5
	1	7	0.001	SCHOOL	0.0	ACPOT	0.0	SUMHOT	0.0
	2	19	0.019	NIGHT	-1.2	SUMMER	0.6	SUNUP	-1.0
	3	14	0.031	NIGHT	-1.3	SCHOOL	1.4	SUNUP	-0.9
	4	13	0.007	SCHOOL	0.3	NIGHT	-0.1	SUNDOWN	0.3
4	Any	18	0.061	NIGHT	-3.4	ACPOT	1.4	SUNUP	-2.5
	1	10	0.004	SCHOOL	0.2	NIGHT	0.2	SUMMER	0.1
	2	18	0.037	NIGHT	-2.2	ACPOT	1.1	SUNUP	-1.8
	3	15	0.029	NIGHT	-1.2	SUNUP	-0.8	SUMMER	0.6
	4	15	0.012	SCHOOL	0.5	NIGHT	-0.2	SUNUP	-0.2
5	Any	18	0.062	NIGHT	-2.8	WORKCOM	3.6	SUNUP	-1.7
	1	6	0.002	PRESCH	0.1	WEEKEND	0.0	NIGHT	0.0
	2	18	0.063	NIGHT	-2.8	WORKCOM	3.6	SUNUP	-1.7
	3	2	0.000	ACRES	0.0	WORKTIME	0.0		
	4	0							

[a]Microenvironment (see Table II). [b]Breathing rate (see Table II). [c]Coefficient, minutes.

Table IX. Results of stepwise linear regression analyses performed on the aggregate Los Angeles person-hour database (2,727 person-hours).

		Regression model		Most significant variables					
				First		Second		Third	
ME[a]	BR[b]	Variables in model	R^2	Variable	Coef.[c]	Variable	Coef.	Variable	Coef.
1	Any	13	0.292	NIGHT	11.4	SCHTIME	-27.1	WORKTIME	-16.8
	1	12	0.484	NIGHT	38.6	SUNUP	21.7	WORKCOM	-10.2
	2	11	0.168	NIGHT	-22.4	SCHTIME	-15.8	SUNUP	-14.3
	3	8	0.049	NIGHT	-4.0	SCHOOL	2.1	SCHTIME	-3.7
	4	4	0.007	SUNDOWN	0.4	SCHOOL	0.1	WEEKEND	0.1
2	Any	9	0.244	SCHTIME	32.4	NIGHT	-4.6	LUNCH	5.0
	1	4	0.016	NIGHT	1.4	WEEKEND	1.2	SUNUP	0.7
	2	8	0.204	SCHTIME	22.7	NIGHT	-4.6	SUMMER	4.3
	3	8	0.092	SCHTIME	7.4	NIGHT	-1.4	SCHOOL	0.5
	4	4	0.015	SCHTIME	1.1	SCHOOL	0.2	MIDAFT	-0.3
3	Any	9	0.072	NIGHT	-3.4	SUMMER	5.0	MIDAFT	4.8
	1	2	0.004	NIGHT	0.3	WEEKEND	0.2		
	2	8	0.075	SUMMER	4.2	NIGHT	-3.0	MIDAFT	2.7
	3	7	0.021	NIGHT	-1.4	SCHOOL	1.3	WORKTIME	1.1
	4	4	0.008	WINWARM	0.2	MIDAFT	0.3	NIGHT	-0.2
4	Any	7	0.119	WORKTIME	16.5	SUMMER	-14.3	OUTWKR	9.6
	1	6	0.037	WEEKEND	1.8	NIGHT	2.3	OUTWKR	1.5
	2	9	0.076	WORKTIME	6.2	WINWARM	7.5	OUTWKR	6.8
	3	6	0.100	WORKTIME	7.3	NIGHT	-2.2	WINWARM	6.6
	4	4	0.009	NIGHT	0.7	LUNCH	0.4	WINWARM	0.3
5	Any	10	0.085	NIGHT	-2.4	OUTWKR	2.2	SUNUP	-1.9
	1	2	0.002	LUNCH	0.1	WEEKEND	0.1		
	2	10	0.080	OUTWKR	2.5	NIGHT	-1.7	MIDAFT	3.0
	3	8	0.013	NIGHT	-0.7	SUNUP	-0.8	SUNDOWN	0.4
	4	1	0.002	WORKCOM	0.1				

[a]Microenvironment (see Table II). [b]Breathing rate (see Table II). [c]Coefficient, minutes.

Table X. Results of stepwise linear regression analyses performed on the Denver person-hour database (19,203 person-hours).

	Regression model		Most significant variables					
			First		Second		Third	
ME[a]	Variables in model	R^2	Variable	Coef.[b]	Variable	Coef.	Variable	Coef.
1	15	0.359	NIGHT	7.1	WORKTIME	-14.2	MIDAFT	-23.4
2	15	0.281	WORKTIME	15.0	NIGHT	-4.1	LUNCH	16.8
3	11	0.024	NIGHT	-0.3	LUNCH	2.0	MIDAFT	2.0
4	14	0.030	SCHTIME	13.1	NIGHT	-0.4	OUTWKR	1.3
5	13	0.078	NIGHT	-2.3	MIDAFT	5.6	LUNCH	4.3

[a]Microenvironment (see Table II).
[b]Coefficient, minutes.

Table XI. Results of stepwise linear regression analyses performed on the Washington person-hour database (16,345 person-hours).

	Regression model		Most significant variables					
			First		Second		Third	
ME[a]	Variables in model	R^2	Variable	Coef.[b]	Variable	Coef.	Variable	Coef.
1	13	0.485	WORKTIME	-26.2	NIGHT	5.7	MIDAFT	-16.0
2	14	0.446	WORKTIME	26.6	NIGHT	-3.1	SUNUP	-5.4
3	10	0.027	NIGHT	-0.3	SCHTIME	6.4	LUNCH	1.8
4	9	0.016	OUTWKR	1.0	NIGHT	-0.2	LUNCH	0.8
5	12	0.085	NIGHT	-1.9	WORKCOM	5.4	OUTWKR	3.8

[a]Microenvironment (see Table II).
[b]Coefficient, minutes.

Table XII. Results of stepwise linear regression analyses performed on the aggregate California person-hour database (70,938 person-hours).

ME[a]	Regression model		Most significant variables					
	Variables in model	R^2	First		Second		Third	
			Variable	Coef.[b]	Variable	Coef.	Variable	Coef.
1	17	0.327	NIGHT	7.0	WORKTIME	-15.6	SCHTIME	-21.1
2	16	0.249	WORKTIME	18.1	SCHTIME	24.0	NIGHT	-3.3
3/4	17	0.104	NIGHT	-1.9	MIDAFT	8.9	LUNCH	9.2
5	14	0.057	NIGHT	-1.9	MIDAFT	3.2	MIDAFT	3.6

[a]Microenvironment (see Table II).
[b]Coefficient, minutes.

Table XIII. Results of stepwise linear regression analyses performed on the Valdez person-hour database (9,715 person-hours).

ME[a]	Regression model		Most significant variables					
			First		Second		Third	
	Variables in model	R^2	Variable	Coef.[b]	Variable	Coef.	Variable	Coef.
1	16	0.309	NIGHT	14.3	WORKTIME	-13.9	SCHTIME	-27.6
2	14	0.244	WORKTIME	13.1	SCHTIME	29.3	NIGHT	-8.0
3/4	9	0.057	NIGHT	-4.0	SUNUP	-2.5	WINCOLD	-1.5
5	12	0.038	NIGHT	-2.6	SUNUP	-1.0	SUMMER	1.8

[a]Microenvironment (see Table II).
[b]Coefficient, minutes.

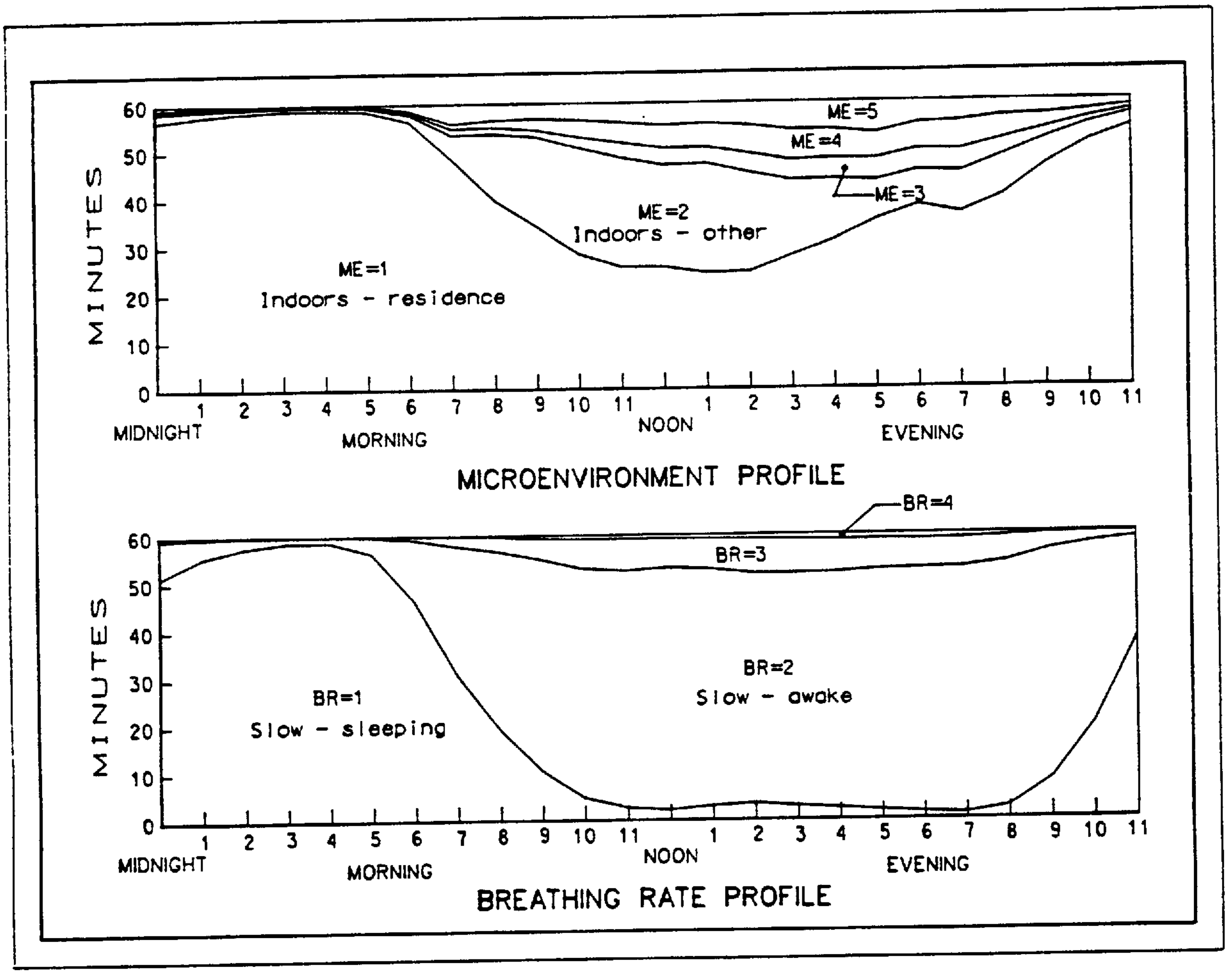

Figure 1. Mean number of minutes per hour spent by Cincinnati subjects in indicated microenvironment (ME) and breathing rate (BR) categories by time of day.

APPENDIX B

AN ALGORITHM FOR DETERMINING MAXIMUM SUSTAINABLE VENTILATION RATE ACCORDING TO GENDER, AGE, AND EXERCISE DURATION

AN ALGORITHM FOR DETERMINING MAXIMUM SUSTAINABLE VENTILATION RATE ACCORDING TO GENDER, AGE, AND EXERCISE DURATION

Ted R. Johnson
IT Air Quality Services
3710 University Drive
Durham, North Carolina 27707

William C. Adams, Ph.D.
Department of Physical Education
University of California at Davis
Davis, California 95616

Abstract

In estimating human exposure to air pollution, researchers frequently employ time/activity data to develop typical exposure event sequences for individuals or demographic types. In typical applications, each event in one of these sequences assigns a person to a geographic location and to a microenvironment for a specified time interval. In applications of the pNEM exposure model by the U.S. Environmental Protection Agency, each event also specifies an equivalent ventilation rate (EVR) for the time interval, where EVR is defined as ventilation rate divided by body surface area. Through the use of a Monte Carlo process, values for EVR are randomly selected from distributions specific to the level of exertion (low, medium, high) associated with the event. An algorithm has been developed which provides an upper limit for the average EVR that can be sustained over time intervals from 1 minute to nine hours. This algorithm is based on an empirically-derived relationship between maximal aerobic power, oxygen uptake rate, and exercise duration. This paper describes the algorithm and provides tables listing algorithm parameters by gender and age.

Key Words

Air pollution

Human exposure

Ventilation rate

Algorithms

INTRODUCTION

During the past 15 years, a number of researchers have developed computer-based models for simulating the exposure of human populations to air pollution. These models are frequently used to estimate exposures under various regulatory control scenarios. For example, the U. S. Environmental Protection Agency (EPA) has developed an exposure model specifically to evaluate current and proposed National Ambient Air Quality Standards (NAAQS). A recent version of this model incorporating stochastic features (the probabilistic NAAQS Exposure Model or pNEM) has been used to estimate the exposures of urban populations to ozone (1,2) and carbon monoxide (3).

The pNEM approach provides exposure estimates for defined population groups based on activity data specific to each group. Similar to other exposure models, pNEM characterizes each exposure by time period and pollutant concentration. Unlike most other exposure models, pNEM also characterizes each exposure by a measure of respiration, the equivalent ventilation rate (EVR). EVR is defined as ventilation rate (liters per minute) divided by body surface area (square meters). Clinical research by EPA suggests that EVR exhibits less inter-person variability than ventilation rate (in liters per minute) for a given level of exertion (4).

In pNEM, EVR values are allowed to vary according to specified lognormal distributions with the constraint that EVR must not exceed a specified upper limit. The limit is assumed to vary with age, gender, and activity duration. This paper describes an algorithm which can be used to determine appropriate values for these EVR limits. A preliminary version of this algorithm was described in an unpublished report by Johnson et al. (5).

THE ESTIMATION OF EVR IN PNEM

In a typical application of pNEM, the population of a designated study area is divided into homogenous population groups called cohorts. Using a special file containing diary-derived time/activity data, the model generates a distinct activity pattern for each cohort. The pattern consists of a sequence of "exposure events." Each event assigns the cohort to a specific environmental setting and to a specific breathing rate category. The environmental setting is characterized by geographic location (e.g., within 5 km of a fixed-site monitor) and a microenvironment (e.g., outdoors near a road). Through the use of ambient monitoring data related to the geographic location and a mass balance model appropriate for the microenvironment, pNEM provides an estimate of the pollutant concentration associated with each exposure event.

In addition to the environmental setting, each exposure event assigns the cohort to one of four breathing rate categories: sleeping, slow, medium, and fast. These categories were selected to permit use of time/activity data collected during the Cincinnati Activity Diary Study (6). During this study, over 900 subjects completed three-day time/activity diaries which employed these breathing categories to characterize the exertion level associated with each activity.

During the summer of 1989, a research team directed by Dr. Jack Hackney conducted a study in Los Angeles to obtain ventilation rate data representative of typical daily activities (7). During this study, the heart rates of 36 subjects were continuously monitored over a three-day period. Each subject documented his or her activities in an activity diary similar to that used in the Cincinnati study. Separate clinical trials were conducted in which the heart rate and ventilation rate of each subject were measured simultaneously. Analysts used these measurements to develop a relationship relating heart rate to ventilation rate for each subject.

The heart rate data for each subject were reported as one-minute averages, referred to as minute heart rates (MINHR). Johnson et al. (5) analyzed the MINHR data for each subject and noted a number of outliers. Following the recommendations of the Hackney research team, they excluded MINHR below 30 beats/min and above a limit defined as 220 beats/min minus age in years. The upper limit was based on a recommendation by McArdle (8). The remaining "valid" MINHR values were converted to minute EVR values by a two-step process. MINHR values were first converted to minute ventilation rates using subject-specific relationships derived from the clinical trials. The resulting ventilation rates were then divided by body surface area to yield minute EVR values. Body surface areas were calculated by a formula developed by Dubois and Dubois (9).

Activity diary data were available for the period during which each subject's heart rate was monitored. The minute EVR values were combined with the activity diary responses associated with each minute, permitting each "valid" EVR value to be categorized according to the diary breathing rate categories (sleeping, slow, medium, and fast).

Researchers aggregated the minute EVR values to create a second data file containing one record for each exposure event. An exposure event began whenever a new activity diary entry was made or when a new clock hour began. For example, an activity that a diary indicated as lasting from 1:15 to 3:23 would yield three exposure events, 1:15 to 2:00, 2:00 to 3:00, and 3:00 to 3:23. The pNEM methodology treats activities in this same manner, so that hourly-average exposures can be calculated by aggregating exposure events over clock hours.

Researchers averaged the EVR values for each event in which at least 75 percent of the minutes were represented by valid EVR values. EVR was treated as a missing value for the remaining events. This procedure produced EVR values for 3700 exposure events of one to 60 minutes. These event EVR values were then grouped by breathing rate category and by age group (children versus adults). Researchers identified 71 outliers (less than 2 percent of the event EVR values) by inspection of probability plots of the event EVR values in each group. After removal of these outliers, researchers found that a two-parameter lognormal distribution provided a good fit to the EVR values in each group. Table 1 lists the geometric mean (GM) and geometric standard deviation (GSD) of each fitted distribution.

In a typical pNEM analysis, each exposure event can be characterized using the breathing rate and age/gender categories listed in Table 1. Consequently, an EVR value can be assigned to each exposure event by randomly selecting values from the appropriate lognormal distribution in Table 1. Because lognormal distributions are not bounded with respect to large values, the selection process must provide some method for identifying "impossible" EVR values. Ideally, this method would provide an upper EVR limit based on the characteristics of the simulated cohort and the duration of the event. If a selected EVR value exceeds the specified limit, the limit value would be substituted for the selected value. This procedure can be stated formally as follows.

TABLE 1. PARAMETER VALUES OF LOGNORMAL DISTRIBUTIONS FIT TO GROUPED EVENT EVR VALUES

Age group	Breathing rate	Number of values		Parameter values of fitted lognormal distribution	
		Fitted	Omitted	Geometric mean[a]	Geometric standard deviation
Children	Sleeping	33	0	8.1	1.60
	Slow	577	21	10.0	1.46
	Medium	423	8	12.3	1.44
	Fast	79	0	14.8	1.62
Adults	Sleeping	88	3	5.4	1.22
	Slow	2169	29	7.1	1.36
	Medium	237	10	8.6	1.34
	Fast	23	0	18.9	1.92
		3629	71		

[a]Liters/min per m^2.

1. Select appropriate values for geometric mean (GM) and geometric standard deviation (GSD) from Table 1 based on the breathing rate category of the event (sleeping, slow, medium, or fast) and the age group of the cohort (children or adults).

2. Determine EVR using the lognormal generating function

$$EVR = (GM)(GSD)^z$$

where z is randomly selected from a normal distribution with mean equal to zero and standard deviation equal to 1.

3. Determine the maximum value of EVR that can be sustained by the members of the cohort for the specified event duration. This value is designated EVRLIM. Compare EVR with EVRLIM. If EVR $\leq$ EVRLIM, assign EVR to the event. If EVR > EVRLIM, assign EVRLIM to the event.

In pNEM analyses, the value of EVRLIM is estimated through the use of an algorithm based on the concept of maximal aerobic power.

THE CONCEPT OF MAXIMAL AEROBIC POWER

Respiration provides oxygen to accomplish work and removes the resulting carbon dioxide. Pulmonary ventilation rate (V_E) is defined as the volume of air exhaled per minute. V_E increases with increasing work rate up to a maximal level which can exceed 200 liters/min in extreme cases. The rate of oxygen uptake (VO_2) by the lungs in liters/min also increases with increasing work rate, although the ratio of V_E to VO_2 is not constant. At rest and during moderately heavy work, this ratio (often termed the "ventilatory equivalent") is typically between 20 and 25 for adults. At maximal work rates, the ratio may increase to a value between 30 and 35 (10).

A person's maximum V_E (V_{Emax}) is determined by his or her maximum oxygen uptake rate (VO_{2max}) and the V_E/VO_2 ratio in effect under maximum oxygen uptake conditions (MAXRATIO) such that

$$V_{Emax} = (VO_{2max})(MAXRATIO). \tag{1}$$

VO_{2max} and MAXRATIO are functions of age, gender, and training, among other factors. For example, children may have MAXRATIO values as high as 40 (10).

As work rate increases, energy is provided primarily by aerobic (oxygen-based) processes up to the point of VO_{2max}, referred to as the point of maximal aerobic power (MAP). The additional energy required for higher work rates is provided by anaerobic processes. Consequently, the work rate where VO_{2max} is reached is less than a person's maximum work rate. Because V_{Emax} occurs at VO_{2max}, the discussion that follows will be concerned to some degree with the effect of age and gender on VO_{2max}.

EFFECT OF DURATION ON AVERAGE WORK RATE

Astrand and Rodahl (10) state that most individuals cannot maintain a work rate equal to 100 percent of MAP (i.e., a work rate where VO_2 equals VO_{2max}) for more than about five minutes. As the duration of work increases, there is a progressive decrease in the average VO_2 level that can be maintained. Astrand and Rodahl also state that a VO_2 level equal to 50 percent of VO_{2max} cannot be maintained for a whole working day.

Erb (11) has developed estimates of the percentage of "maximum work capacity" that can be maintained by young and middle-aged adults for durations of one to nine hours (middle column, Table 2). These values -- which apply to normally active, non-trained adults -- appear to be very nearly equivalent to the percentage of VO_{2max} ($PCTVO_{2max}$) that can be maintained and are so labeled in Table 2. According to Erb, a person can maintain 64 percent of VO_{2max} for one hour and 33 percent of VO_{2max} for nine hours without straining.

The following expression provides a close fit to the estimates presented in Table 2.

$$PCTVO_{2max}(t) = 121.2 - (14.0)[\ln(t)]. \quad (2)$$

Note that t is duration in minutes and ln indicates the natural logarithm.

As a means of establishing an upper limit on the EVR values simulated in pNEM analyses, it is desirable to have a continuous relationship between duration and the value of $PCTVO_{2max}$ that can be maintained. To be consistent with the statement by Astrand and Rodahl that 100 percent of VO_{2max} can be maintained for up to five minutes, the relationship should indicate that $PCTVO_{2max}$ equals approximately 100 percent when t = 5 minutes. Erb's relationship (Equation 2), which plots as a straight line on semi-log graph paper (Figure 1), meets this criterion if the linear relationship is extended to 5 minutes. Thus a reasonable model for work rate would assume that $PCTVO_{2max}$ = 100 percent for 0 to 5 minutes and that Equation 2 applies to longer durations.

PREDICTION OF VENTILATION RATE FROM WORK RATE

Astrand and Rodahl provide a graph showing average values for VO_{2max} by age and gender based on measurements of 350 subjects aged 4 to 64 (Figure 9-11, p. 319). In another graph, Astrand and Rodahl present data for the same 350 subjects in terms of VO_{2max} per kg

TABLE 2. APPROXIMATE PERCENTAGE OF MAXIMUM VO_2 THAT CAN BE MAINTAINED FOR GIVEN WORK DURATIONS

Work duration		Percentage of maximum VO_2 that can be maintained	
Hours	Minutes	Erb estimate	Adjusted estimate
1	60	64	75
2	120	54	68
3	180	48	64
4	240	44	61
5	300	41	59
6	360	39	57
7	420	37	55
8	480	35	54
9	540	33	53

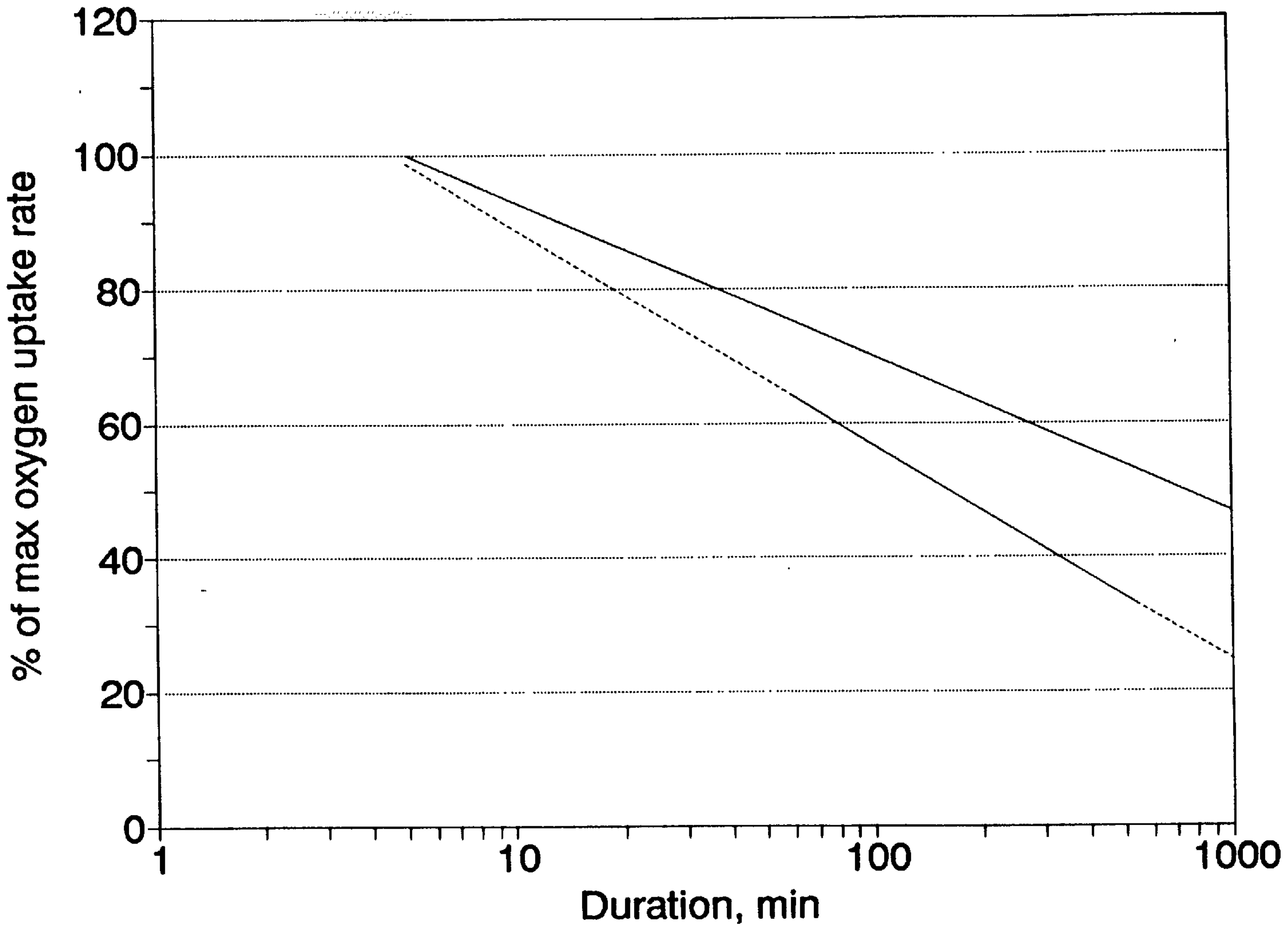

Figure 1. Percentage of maximum oxygen uptake rate that can be maintained versus activity duration. Lower solid line is formula proposed by Erb. Dashed lines are extrapolations of formula to durations less than 60 minutes and greater than 540 minutes. Upper solid line is alternative relationship proposed for regularly exercising persons motivated to attain high exertion levels.

of body weight (Figure 9-13, p. 321). They state that all subjects were healthy and moderately well trained; however, none of the subjects was an athlete. Astrand and Rodahl acknowledge that these subjects were "selected," but state that it is almost impossible to present results for "normal" subjects because of the difficulty in defining normal.

Astrand and Rodahl present data showing that VO_{2max}, expressed as liters/min, increases with age up to 20 years. Beyond this age, there is a gradual decline so that a 65-year-old individual attains only about 70 percent of his peak VO_{2max} at 25 years. Before the age of 10, there is no significant difference in VO_{2max} between girls and boys; from ages 11 to 17 years, the VO_{2max} of boys is 11 percent to 37 percent higher than that of girls. During the adult years the average difference in VO_{2max} between women and men amounts to 25 to 30 percent. Related to body weight, the gender difference after puberty is 15 to 20 percent. Top athletes in endurance events have a VO_{2max} that is nearly twice as high as the VO_{2max} of the average person.

The research findings presented by Astrand and Rodahl can be used to estimate VO_{2max} by gender and age. Tables 3 and 4 document the procedure. The first column in each table indicates age in years. The next column lists mean weight values for each age group obtained from Anderson et al. (12). The third column lists the average body surface area (BSA) for the age group. Values for ages 2 through 17 are median values obtained from Anderson et al. (11); the remaining values were determined by the following formula developed by Dubois and Dubois (9):

$$BSA = (7.182E\text{-}3)(H)^{0.725}\,(W)^{0.425} \qquad (3)$$

TABLE 3. ALGORITHM PARAMETER VALUES FOR MALES

Age, years	Weight, kg	Body surface area (BSA), m^2	Ratio of VO_{2max} to weight, ml/min per kg[a]	VO_{2max}, liters/min	Ratio of V_E to VO_2	
					MAXRATIO[b]	SUBRATIO[c]
2	13.4	0.60	44	0.59	25.3	21.5[d]
3	15.5	0.66	46	0.71	33.8	28.8[d]
4	17.6	0.73	48	0.84	39.3	33.5[d]
5	19.7	0.79	50	0.99	41.1	35.0
6	22.8	0.87	52	1.19	39.5	33.0
7	24.8	0.94	54	1.34	38.0	31.0
8	28.0	1.00	56	1.57	36.5	29.0
9	30.7	1.07	57	1.75	35.0	28.0
10	36.2	1.18	57	2.06	35.5	27.0
11	39.7	1.23	58	2.30	34.5	26.0
12	44.1	1.34	58	2.56	33.5	25.5
13	49.5	1.47	58	2.87	32.0	24.0
14	56.4	1.61	57	3.22	32.0	23.0
15	61.2	1.70	57	3.49	32.0	22.5
16	66.5	1.76	55	3.66	31.0	22.0
17	66.7	1.80	53	3.53	32.8	21.5
18-24	73.7	1.90	50	3.69	32.5	21.0
25-34	78.7	1.95	44	3.46	33.2	21.0
35-44	80.8	1.97	38	3.07	33.2	22.3
45-54	81.0	1.96	35	2.84	32.4	23.5
55-64	78.8	1.93	32	2.52	31.3	23.5
65-74	74.8	1.88	29	2.17	31.1	23.5

[a]VO_{2max} = VO_2 at maximum aerobic power.
[b]MAXRATIO = V_E/VO_2 at $VO_2 = VO_{2max}$.
[c]SUBRATIO = V_E/VO_2 at $VO_2 = (0.65)(VO_{2max})$.
[d]Estimated as SUBRATIO = (0.85)(MAXRATIO).

TABLE 4. ALGORITHM PARAMETER VALUES FOR FEMALES

Age, years	Weight, kg	Body surface area (BSA), m^2	Ratio of VO_{2max} to weight, ml/min per kg[a]	VO_{2max}, liters/min	Ratio of V_E to VO_2	
					MAXRATIO[b]	SUBRATIO[c]
2	12.8	0.58	43	0.55	23.6	19.5[d]
3	14.8	0.65	44	0.65	30.8	25.5[d]
4	16.8	0.71	46	0.77	36.4	30.1[d]
5	19.4	0.78	47	0.91	37.5	31.0
6	21.9	0.84	50	1.10	38.0	30.5
7	24.6	0.92	52	1.28	38.5	30.0
8	27.5	1.00	53	1.46	39.0	30.0
9	31.7	1.06	52	1.65	38.5	29.0
10	35.7	1.17	51	1.82	37.0	28.0
11	41.4	1.30	50	2.07	36.0	26.5
12	46.1	1.40	49	2.26	35.0	26.0
13	50.9	1.48	47	2.39	33.5	25.0
14	54.3	1.55	46	2.50	33.0	25.0
15	55.0	1.57	46	2.53	34.2	24.5
16	57.8	1.60	45	2.60	33.5	25.0
17	59.6	1.63	44	2.62	32.5	25.0
18-24	60.6	1.65	41	2.49	32.9	25.0
25-34	64.2	1.69	37	2.38	31.5	25.0
35-44	67.1	1.71	33	2.21	30.0	26.5
45-54	67.9	1.71	30	2.04	30.4	28.0
55-64	67.9	1.70	27	1.83	31.1	28.0
65-74	66.6	1.69	24	1.60	32.6	28.0

[a]VO_{2max} = VO_2 at maximum aerobic power.
[b]MAXRATIO = V_E/VO_2 at $VO_2 = VO_{2max}$.
[c]SUBRATIO = V_E/VO_2 at $VO_2 = (0.65)(VO_{2max})$.
[d]Estimated as SUBRATIO = (0.83)(MAXRATIO).

where H is mean height in centimeters and W is mean weight in kilograms. Mean height and weight values were obtained from Frisancho (13).

The next column in each table lists values for the VO_{2max}-to-weight ratio. The ratios for ages 6 to 64 were taken directly from the curves for males and females provided by Åstrand and Rodahl in Figure 9-13 of the Textbook of Work Physiology. The ratios for ages 0 to 5 and 65 to 74 were obtained by extending the male and female curves tangentially in both directions.

The VO_{2max}-to-weight ratios were multiplied by the corresponding mean body weights to determine the VO_{2max} values for each age group. For example, the VO_{2max}-to-weight ratio for males aged 16 (55 ml/min per kg) was multiplied by the mean body weight for males aged 16 (66.5 kg) to yield an estimate of 3.66 liters/min for this group.

Tables 3 and 4 also list values for MAXRATIO -- the ratio of V_E to VO_2 at VO_{2max} by age. Values for ages 5 through 17 were taken directly from the curves for males and females provided in Figure 7-10 (p. 230) of the Textbook of Work Physiology. The ratios for children under 5 and adults 18 and above were determined by extending the male and female curves for V_{Emax} and VO_{2max} tangentially in both directions.

The V_E-to-VO_2 ratios listed in Tables 3 and 4 under the heading MAXRATIO represent those achieved during VO_{2max} conditions (i.e., $VO_2 = VO_{2max}$). The ratio of V_E to VO_2 is known to be lower at lower levels of exertion. Astrand and Rodahl demonstrate this phenomenon by providing V_E/VO_2 versus age curves (5 to 26 years) for males and females exercising at VO_{2max} and at submaximal aerobic power (60 to 70 percent VO_{2max}). The V_E/VO_2 values at submaximal aerobic power are typically 70 to 80 percent of the corresponding V_E/VO_2 values at VO_{2max}. The submaximal V_E/VO_2 values are listed in Tables 3 and 4 under the heading SUBRATIO. Children under 5 were assigned a value calculated as 85 percent of the MAXRATIO. Values for adults 26 years and above were determined by increasing the subratio by 5 percent for age range 35 to 44 years and by 10 percent thereafter, according to differences in submaximal V_E/VO_2 ratios observed by Adams et al. (14).

Data cited by McArdle et al. (8) can be used to determine the general shape of the relationship between V_E/VO_2 and VO_2 when VO_2 is expressed a percentage of VO_{2max}. McArdle et al. provide a scatter plot displaying V_E plotted against VO_2 data as VO_2 is increased from approximately 1.0 liter/min to VO_{2max} at 4.0 liters/min. These data were transformed into a scatter plot showing V_E/VO_2 plotted against VO_2 expressed as a percentage of VO_{2max}. The V_E/VO_2 values varied from 21.6 at VO_2 = 1.0 liter/min (25 percent of VO_{2max}) to 37.7 at VO_2 = 4.0 liters/min (100 percent of VO_{2max}). The V_E/VO_2 values were relatively constant from VO_2 = 35 percent VO_{2max} to 65 percent VO_{2max}. From VO_2 = 65 percent VO_{2max} to $VO_2 = VO_{2max}$, V_E/VO_2 increased in an approximately linear fashion.

Let SUBRATIO indicate V_E/VO_2 for VO_2 = 65 percent VO_{2max}. The relationship between V_E/VO_2 and VO_2 observed above can be described mathematically by the following equations.

$VO_2 < (0.65)(VO_{2max})$:

$$V_E/VO_2 = \text{SUBRATIO} \tag{4}$$

$VO_2 \geq (0.65)(VO_{2max})$:

$$V_E/VO_2 = \text{SUBRATIO} + (\text{MAXRATIO} - \text{SUBRATIO})(\text{PCTVO}_{2max} - 65)/35 \tag{5}$$

If these equations are generally applicable, then the algorithm presented in Table 5 should produce reasonable estimates for the upper limit of EVR (EVRLIM) when age, gender, and activity duration are specified. Note that this algorithm reflects its data sources in that it is applicable to average persons in ordinary work situations.

SUGGESTED ADJUSTMENTS TO THE ALGORITHM

The EVRLIM values produced by the algorithm in Table 5 are assumed to represent the EVR limit applicable to an average person who is working under conditions in which he or she is not straining. A person of average genetic characteristics who exercises regularly and who is motivated to reach a high exertion rate may attain EVR values exceeding the EVRLIM value specified by the algorithm. For these individuals, the algorithm can be adjusted to apply to these exceptional persons by 1) recalibrating the Erb relationship and 2) increasing the VO_{2max} values.

The Erb equation (Equation 2) characterizes the relationship between $PCTVO_{2max}$ and duration for workers who are not straining. Experimental data suggest that young and middle-aged adults who jog or bicycle routinely attain higher VO_2 levels than those indicated by the Erb equation. For example, data obtained from Pollock et al. (15) and from Lieber et al. (16) indicate that a motivated recreational jogger, cyclist, or swimmer may exercise at 75 percent VO_{2max} for t = 60 minutes. The Erb equation indicates that a worker can maintain 63.9 percent VO_{2max} for t = 60 minutes.

TABLE 5. ALGORITHM A: METHOD FOR ESTIMATING UPPER LIMIT OF EQUIVALENT VENTILATION RATE (EVRLIM) FOR AVERAGE PERSONS IN ORDINARY WORK SITUATIONS

1. Identify age and gender of population group of interest (example: males aged 16). Find entry for group in Table 3 or 4. Note values for VO_{2max}, V_E/VO_2 at VO_{2max} (MAXRATIO), V_E/VO_2 at 65 percent VO_{2max} (SUBRATIO), and body surface area (BSA).

2. Specify t, the duration of the activity in minutes.

3. Let $PCTVO_{2max}$ indicate the percentage of VO_{2max} that can be maintained for a specified duration. If $t \leq 5$ minutes, $PCTVO_{2max} = 100$ percent. If $t > 5$ minutes, use Equation 2 to estimate $PCTVO_{2max}$.

4. Define RATIO = V_E/VO_2. If $PCTVO_{2max} < 65$ percent, RATIO = SUBRATIO. Otherwise,

   $$RATIO = SUBRATIO + (MAXRATIO - SUBRATIO)(PCTVO_{2max} - 65)/35$$

5. Calculate the upper limit for ventilation rate by the equation

   $$VELIM = (VO_{2max})(PCTVO_{2max})(RATIO)/100.$$

6. Calculate the upper limit for EVR by the equation

   $$EVRLIM = (VELIM)/(BSA).$$

The Erb equation has the general form

$$PCTVO_{2max} = a_1 - (a_2)[\ln(t)] \tag{6}$$

where a_1 and a_2 are constants. If one assumes that Erb's formula underestimates $PCTVO_{2max}$ for a given value of t, the equation can be adjusted to provide the "correct" result by changing a_1 and/or a_2 without changing the general logarithmic relationship expressed by the formula.

At present, there is no reason to change the requirement that $PCTVO_{2max} = 100$ at t = 5 minutes. If one assumes that $PCTVO_{2max} = 75$ percent at t = 60 minutes, one obtains the following simultaneous equations.

$$100 = a_1 - (a_2)[\ln(5)] \tag{7}$$

$$75 = a_1 - (a_2)[\ln(60)] \tag{8}$$

The solution is $a_1 = 116.19$ and $a_2 = 10.06$. Thus the "adjusted" Erb equation is

$$PCTVO_{2max} = 116.19 - (10.06)[\ln(t)]. \tag{9}$$

The right-hand column of Table 2 lists estimates of $PCTVO_{2max}$ for activity durations between 1 and 9 hours based on Equation 9. The estimate for 9 hours (53 percent) is slightly higher than the limit for a day's work specified by Astrand and Rodahl (50 percent).

The values of VO_{2max} listed in Tables 3 and 4 were obtained from Astrand and Rodahl. These values are likely to represent average individuals who do not exercise regularly. Research by Lieber et al. suggests that average individuals may increase the VO_{2max} values by as much as 28 percent with regular intense exercise (16). Wilmore and Costill (17) report a VO_{2max} increase of 15 to 20 percent as more typical for the average adult who was sedentary prior to training, and who had trained regularly for several months at 75 percent of his or her VO_{2max}.

Table 6 presents an alternative algorithm (Algorithm B) for estimating EVRLIM values which is appropriate for active persons who are motivated to reach high sustained rates of exertion during exercise or work. The algorithm incorporates Equation 9 as a means of estimating $PCTVO_{2max}$. Consistent with the reported increases in VO_{2max} noted above, Algorithm B multiplies the VO_{2max} values listed in Tables 3 and 4 by 1.20.

Figure 2 presents EVRLIM plotted against activity duration as estimated by each algorithm for males aged 16. For durations of 5 minutes or less, Algorithm B yields a constant EVRLIM value of 77.4 liters/min per square meter. This value is 20 percent higher than the corresponding Algorithm A estimate of 64.5 liters/min per square meter. At 60 minutes, the Algorithm B estimate (46 liters/min per square meter) is 59 percent higher than the Algorithm A estimate (29 liters/min per square meter). At 9 hours (540 minutes), the Algorithm B estimate is 92 percent higher than the Algorithm A estimate (29 versus 15 liters/min per square meter). In each case, the absolute difference between the two estimates falls between 12 and 14 liters/min per square meter.

CONSISTENCY OF ESTIMATES WITH EXPERIMENTAL DATA

The two algorithms described above provide estimates of VO_{2max}, V_E/VO_2, V_{Emax}, and EVRLIM for ages 2 through 74 (both genders) applicable to activity durations from 1 to 540 minutes. The question naturally arises as to whether these estimates are consistent with experimental data presented in the scientific literature. Experimental VO_{2max} data, often accompanied with V_{Emax} data, are abundant in the literature, although most of the data represent young and middle-aged adult males. For example, Adams et al. (14) studied a

TABLE 6. ALGORITHM B: METHOD FOR ESTIMATING UPPER LIMIT OF EQUIVALENT VENTILATION RATE (EVRLIM) FOR REGULARLY EXERCISING PERSONS WHO ARE MOTIVATED TO ATTAIN HIGH EXERTION LEVELS (E.G., JOGGERS)

1. Identify age and gender of population group of interest (example: males aged 16). Find entry for group in Table 3 or 4. Note values for VO_{2max}, V_E/VO_2 at VO_{2max} (MAXRATIO), V_E/VO_2 at 65 percent MAP (SUBRATIO), and body surface area (BSA).

2. Specify t, the duration of the activity in minutes.

3. Let $PCTVO_{2max}$ indicate the percentage of VO_{2max} that can be maintained for a specified duration. If $t \leq 5$ minutes, $PCTVO_{2max} = 100$ percent. If $t < 5$ minutes, use Equation 9 to estimate $PCTVO_{2max}$.

4. Define RATIO = V_E/VO_2. If $PCTVO_{2max} < 65$ percent, RATIO = SUBRATIO. Otherwise,

   $$RATIO = SUBRATIO + (MAXRATIO - SUBRATIO)(PCTVO_{2max} - 65)/35$$

5. Calculate the upper limit for ventilation rate by the equation

   $$VELIM = (1.2)(VO_{2max})(PCTVO_{2max})(RATIO)/100.$$

6. Calculate the upper limit for EVR by the equation

   $$EVRLIM = (VELIM)/(BSA).$$

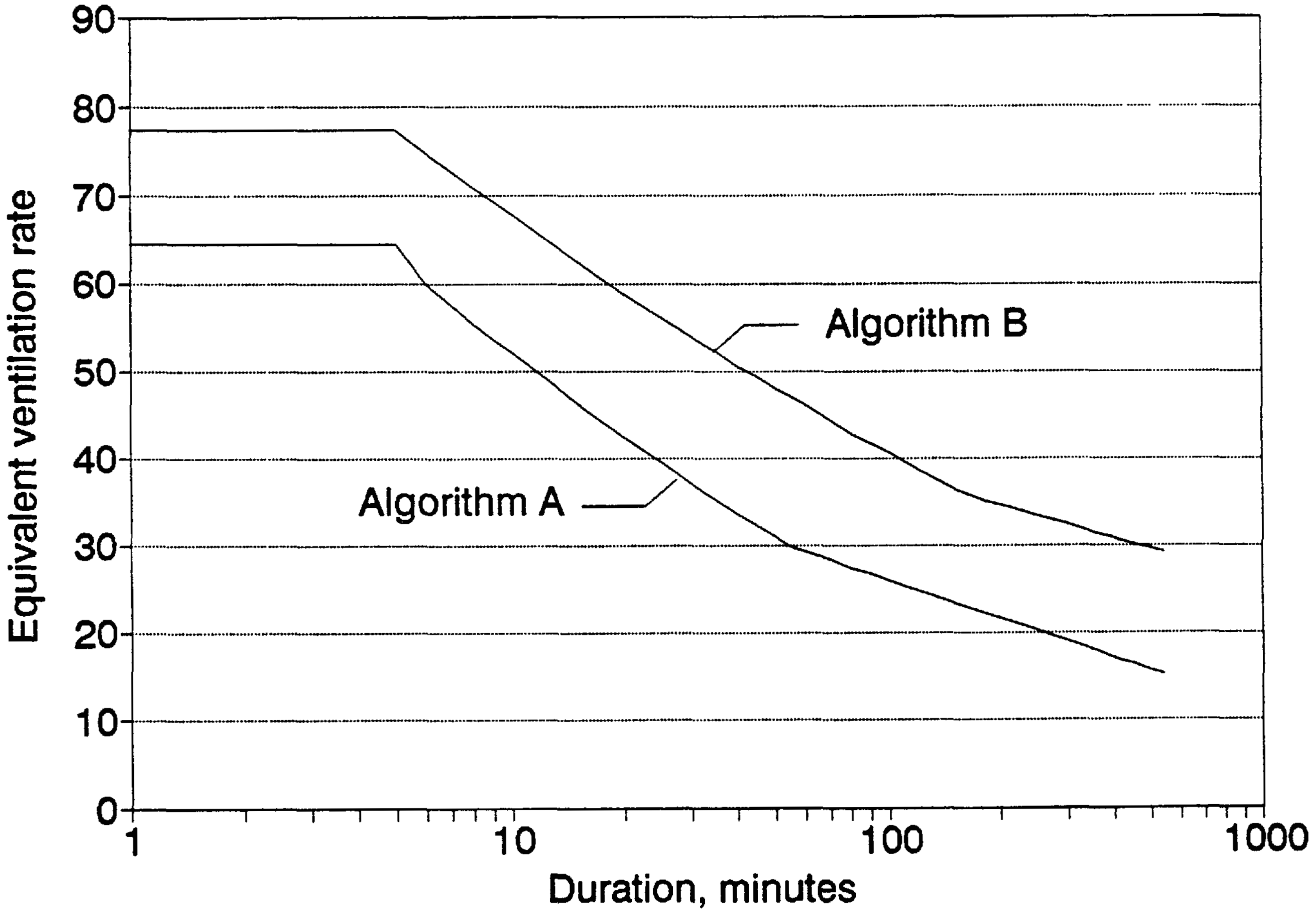

Figure 2. EVRLIM versus activity duration for males aged 16.

group of 80 sedentary males, ages 30 to 70, who were divided by decade into 4 groups. Their V_E/VO_2 values at VO_{2max} were very similar to the values given in Table 3. Adams et al. report a V_{Emax} of 78.8 liters/min at VO_{2max} for sedentary males 59 years of age. Dividing this value by a BSA of 1.945 square meters yields an EVR at VO_{2max} of 40.5 liters/min per square meter. The corresponding Algorithm A estimate for males 55-to-64 years is 40.9 liters/min per square meter.

Adams and Lauritzen (18) report that active young adult males and females may achieve VO_2/V_E values near 40 at VO_{2max}. The corresponding MAXRATIO values in Tables 3 and 4 are approximately 15 percent lower (32.8 and 32.9 for males and females, respectively). The EVRLIM values reported in this study for VO_{2max} were 67.0 liters/min per square meter for females and 82.5 liters/min per square meter for males. Algorithm B provides an EVRLIM estimate of 60 liters/min per square meter for active females at VO_{2max}; the estimate for males is 76 liters/min per square meter. These estimates are 8 to 10 percent lower than the experimental values.

Experimental data for upper limit EVR at durations longer than several minutes are almost exclusively related to VO_{2max} tests in athletic populations. The few studies which have focused on active non-athletes tend to support the assumption in Algorithm B that regularly exercising people can exercise at 75 percent VO_{2max} for 1 hour. Messineo and Adams (19) studied a group of young adult females (21.5 years) who regularly participated in aerobic exercise. The subjects completed one hour of continuous exercise on a cycle ergonomic at 68 percent of VO_{2max} while exposed to 0.30 ppm ozone. The average EVR during the test was 30.3 liters/min per square meter. Algorithm B yields a one-hour EVRLIM estimate of 37 liters/min per square meter for active, motivated females aged 18 to 24. The lower experimental EVR value is consistent with expectations, as the subjects of the study were directed to exercise at a high, but not maximal, sustained work rate.

In another non-maximal (but very sustained) exercise study by Adams (20), a group of sedentary young adult females exercised continuously on a cycle ergometer at 71 percent of VO_{2max} for one hour while exposed to filtered air (and on another occasion while exposed to 0.20 ppm ozone). The average EVR was 28.6 liters/min per square meter. Algorithm A provides a one-hour EVRLIM value of 27 liters/min per square meter for typical females 18 to 24 years of age.

In another heavy (but non-maximal) exercise study, Folinsbee et al. (21) reported an average EVR of 21.3 liters/min per square meter for a group of normally active young males engaged in 6.6 hours of exercise during an eight-hour period. The experimental conditions in the Folinsbee study represent a cross between Algorithm A and B in that the subjects were normally active persons (Algorithm A) who were directed to work at a level requiring a high degree of motivation (Algorithm B). Consequently, it is not unexpected that the experimental result (21.3) is bracketed by the estimates obtained from Algorithm A (15.3) and Algorithm B (27.4) when applied to males 18 to 24 years exercising for 6.6 hours. A close matchup between measured and estimated values can be obtained from Algorithm B if the 20 percent increase in VO_{2max} applicable to regular exercisers is eliminated from Step 5 (Table 6). With this appropriate modification, Algorithm B yields a 6.6 hour EVRLIM of 22.8 liters/min per square meter for males 18 to 24 years.

It is important to understand that competitive athletes can sustain much higher EVRs for prolonged time periods than can even highly motivated non-athletes. For example, Adams (22) reported that a highly trained long distance runner could run at 78 percent of VO_{2max} for 2.4 hours at an average V_E of 120 liters/min. Dividing this value by the runner's BSA (1.80 square meter) yields an EVR of 66.7 liters/min per square meter. Algorithm B provides an estimate of 33.0 liters/min per square meter for the EVRLIM of a highly motivated non-athlete (18 to 24 years) exercising continuously for 2.4 hours.

RECOMMENDATIONS FOR FUTURE RESEARCH

Two algorithms have been proposed for estimating the upper limit of EVR for a specified activity duration. Algorithm A applies to typical individuals engaged in ordinary work and recreational activities. Algorithm B applies to regularly exercising individuals engaged in activities in which they are motivated to reach high work rates (e.g., a member of an aerobic exercise class). Estimates from these algorithms can be used to determine "real-world" limits on the EVR values generated by pNEM and similar population exposure models.

Additional experimental data on VO_{2max}, VO_2/V_E, and EVR are needed to further refine and validate the algorithms. A review of the scientific literature indicates that most currently available experimental data relate to young adults exercising at or near VO_{2max} over short time intervals (five minutes or less). In addition, there are a few studies which report data for mostly-male subjects exercising at submaximal work rates for one-hour periods. The Folinsbee study (21) is one of the very few studies which provide data representing typical persons exercising at high work rates for multi-hour time periods. The Folinsbee subjects were all young adult males.

In general, the experimental data relate to trained or regularly exercising young male adults exercising for periods of one hour or less. Additional research is needed with respect to 1) males exercising for periods greater than one hour and 2) children, young adult females, and older adults exercising for all time periods. The research should distinguish between typical persons and those who exercise regularly. In addition, the experimental design of each study should provide for the quantification of a "motivational effect", that is, the incremental increase in EVRLIM that occurs when subjects are motivated to reach high work rates during exercise.

REFERENCES

1. T. Johnson, J. Capel, E. Olaguer, and L. Wijnberg, Estimation of Ozone Exposures Experienced by Urban Residents Using a Probabilistic Version of NEM (EPA Contract No. 68-DO-0062, 1993).

2. T. Johnson, J. Capel, E. Olaguer, and L. Wijnberg, Estimation of Ozone Exposures Experienced by Residents of the ROMNET Domain Using a Probabilistic Version of NEM (EPA Contract No. 68-DO-0062, 1992).

3. T. Johnson, J. Capel, R. Paul, and L. Wijnberg, Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Denver Residents Using a Probabilistic Version of NEM (EPA Contract No. 68-DO-0062, 1992).

4. W. F. McDonnell, D. H. Horstman, and M. J. Hazucha, "Pulmonary Effects of Ozone Exposure During Exercise: Dose-Response Characteristics," Journal of Applied Physiology, 54, 1345 (1983).

5. T. Johnson, L. Wijnberg, and J. Capel, Review and Evaluation of New Research Relating to Population Exposure to Air Pollution--Executive Summary, American Petroleum Institute, Washington, D.C., 1990.

6. T. R. Johnson, A Study of Human Activity Patterns in Cincinnati, Ohio, Electric Power Research Institute, Palo Alto, California, 1987.

7. S. C. Trim, Environmental Health Service, Rancho Los Amigos Medical Center, Inc., Downey, California, 1990, personal communication.

8. W. D. McArdle, F. I. Katch, and V. L. Katch, Exercise Physiology: Energy, Nutrition, and Human Performance, Lea and Febiger, Malvern, Pennsylvania, 1991.

9. D. DuBois and E. F. DuBois, "A Formula to Estimate the Approximate Surface Area If Height and Weight Be Known," Archives of Internal Medicine, 17, 863-871 (1916).

10. P. Astrand and K. Rodahl, Textbook of Work Physiology, McGraw-Hill, New York, 1977.

11. B. D. Erb, "Applying Work Physiology to Occupational Medicine," Occupational Health Safety, 50, 20-24 (1981).

12. E. Anderson, N. Browne, S. Duletsky, J. Ramig, and T. Warn, Development of Statistical Distributions or Ranges of Standard Factors Used in Exposure Assessments (EPA Contract No. 68-01-6775, 1985).

13. A. R. Frisancho, "New Standards of Weight and Body Composition by Frame Size and Height for Assessment of Nutritional Status of Adults and the Elderly," American Journal of Clinical Nutrition, 40: 808-819 (1984).

14. W. C. Adams, M. M. McHenry, and E. M. Bernauer, "Multistage Treadmill Walking Performance and Associated Cardiorespiratory Responses of Middle-Aged Men," Clinical Science, 42: 355-370 (1972).

15. M. L. Pollock, J. Dimmick, H. S. Miller, Jr., Z. Kendrick, and A. C. Linnerud, "Effect of Mode of Training on Cardiovascular Function and Body Composition of Adult Men," Medicine and Science in Sports and Exercise, 7: 139-145 (1975).

16. D. C. Lieber, R. L. Lieber, and W. C. Adams, "Effects of Run-Training and Swim-Training at Similar Absolute Intensities on Treadmill VO_{2max}," Medicine and Science in Sports and Exercise, 21: 655-661 (1989).

17. J. H. Wilmore and D. L. Costill, Training for Sport and Activity: The Physiological Basis of the Conditioning Process, Dubuque, Iowa, William C. Brown, 1988.

18. W. C. Adams and S. K. Lauritzen, "Ozone Inhalation Effects Consequent to Continuous Exercise in Females: Comparison to Males," Journal of Applied Physiology, 59:1601-1606, 1985.

19. W. C. Adams and T.D. Messiness, "Ozone Inhalation Effects in Females Varying Widely in Lung Size: Comparison to Males," Journal of Applied Physiology, 69:96-103, 1990.

20. W. C. Adams, The Effect of Heavy Sustained Exercise in Combination with Low Levels of Ozone Concentration in Inducing Acute Pulmonary Function Impairment in Humans -- Final Report to California Air Resources Board, (CARB Contract A1-158-33, 1984, pp. 86-110).

21. L. J. Folinsbee, W. F. McDonnell, and D. H. Horstman, "Pulmonary Function and Symptom Responses After 6.6 hours Exposure to 0.12 ppm Ozone with Moderate Exercise," Journal of the Air Pollution Control Association, 38:28-35, 1988.

22. W. C. Adams, "Influence of Exercise Mode and Selected Ambient Conditions on Skin Temperature." Annals of the New York Academy of Science, 301:110-127,1977.

APPENDIX C

DESCRIPTIONS OF VARIABLES IN EVENT-AVERAGED EQUIVALENT VENTILATION RATE DATA BASE

PID - Person/subject identification number

Month - 1-12 (January-December)

Day - 1-31 (as appropriate)

Year - 89, 90, or 91 (equals 1989, 1990, and 1991)

Time - Starting time for the event

Demogrp - Demographic group:
1 Children 0-5 years
2 Children 6-13 years
3 Children 14-18 years
4 Workers with low probability of outdoor work
5 Workers with moderate probability of outdoor work
6 Workers with high probability of outdoor work
7 Nonworking adults under 35 years
8 Nonworking adults 35-54 years
9 Nonworking adults 55+ years

Gender -
1 Male
2 Female

Race -
1 White
2 Black
3 Asian
4 Other
9 Unknown/Missing

Income - Income status:
1 Less than $10,000
2 Greater than or equal to $10,000 and less than $25,000
3 Greater than or equal to $25,000 and less than $50,000
4 Greater than or equal to $50,000
9 Unknown/Missing

Garage - Attached garage on household:
0 NO
1 YES
9 Unknown/Missing

AC - Air conditioner in the household:
0 NO
1 YES
9 Unknown/Missing

Gasstove - Gas stove in household:
0 NO
1 YES
9 Unknown/Missing

Season - Season of the year:
0 Winter-months 1-5 and 9-12
1 Summer-months 6-8

Daytype -
1 Weekday
2 Weekend

Temp - Maximum temperature in degrees Fahrenheit

Activity -
1 All destination oriented travel
2 Income-related work
3 Day-care
4 Kindergarten-12th grade
5 College or trade school
6 Adult education and special training
7 Homework
8 Meal preparation and cleanup
9 laundry
10 Other indoor chores
11 Yard work and outdoor chores
12 Child care and child-centered activities
13 Errands and shopping
14 Personal care outside home (doctor, hair dresser, etc.)
15 Eating
16 Sleeping
17 Other personal needs
18 Religious activities
19 Meetings of clubs, organizations, committees, etc.)
20 Other collective participation
21 Spectator sports events
22 Movies concerts, and other entertainment events outside home
23 Cafe, bar, tearoom
24 Museums and exhibitions
25 Parties and receptions
26 Visiting with friends
27 Recess and physical education
28 Active sports and games outside school, including exercises and aerobics
29 Hunting, fishing, hiking
30 Jogging or bicycling
31 Taking a walk
32 Artistic creations, music, and hobbies
33 Other active leisure
34 Reading

35 Television or radio
36 Conversation and correspondence
37 Relaxing, reflecting, thinking (no visible activity)
38 Other passive leisure
39 asthma attack
40 Other sudden illness or injury
41 Uncertain of applicable code
42 No entry in diary
43 Interview
44 Wakeup
45 Baby crying

Income-related work (construction workers only)

80 Sitting or standing (including driving on job site)
81 Walking
82 Hand-carrying building materials or equipment
83 Working at trade (hammering, sawing, framing, etc.)

Microenv - Microenvironment:
1 Indoors residence
2 Indoors other
3 Outdoors near road
4 Outdoors other
5 In vehicle

Breathrate - Breathing rate:
13 Sleeping
14 Slow
15 Moderate
16 Fast

Duration - Duration of event in minutes

Avgevr - Average equivalent ventilation rate in liters/min/m^2

168pp

0495.75C1P

American Petroleum Institute
1220 L Street, Northwest
Washington, D.C. 20005

Related API Publications ...

DR 201, *Overview of Research Relating to Population Exposure to Air Pollution. September 1990*

DR 223, *Review and Evaluation of Research Relating to Population Exposure to Air Pollution, April 1990*

DR 75, *Development of Improved Methods to Measure Effective Doses of Ozone, November 1994*

To order, call API Publications Department (202) 682-8375

Order No. 841-46170