Journal of Exposure Analysis and Environmental Epidemiology **(2003) 13, 294–317**

www.nature.com/jea

Using human activity data in exposure models: Analysis of discriminating factors

THOMAS McCURDY AND STEPHEN E. GRAHAM

Exposure Modeling Research Branch, Human Exposure and Atmospheric Sciences Division, National Exposure Research Laboratory/Office of Research and Development, US Environmental Protection Agency, North Carolina, USA

This paper tests factors thought to be important in explaining the choices people make in where they spend time. Three aggregate locations are analyzed: outdoors, indoors, and in-vehicles for two different sample groups: a year-long (longitudinal) sample of one individual and a cross-sectional sample of 169 individuals from the US Environmental Protection Agency's Consolidated Human Activity Database (CHAD). The cross-sectional sample consists of persons similar to the longitudinal subject in terms of age, work status, education, and residential type. The sample groups are remarkably similar in the time spent per day in the tested locations, although there are differences in participation rates: the percentage of days frequenting a particular location. Time spent outdoors exhibits the most relative variability of any location tested, with in-vehicle time being the next. The factors found to be most important in explaining daily time usage in both sample groups are: season of the year, season/temperature combinations, precipitation levels, and day-type (work/nonwork is the most distinct, but weekday/weekend is also significant). Season, season/temperature, and day-type are also important for explaining time spent indoors. None of the variables tested are consistent in explaining in-vehicle time in either the cross-sectional or longitudinal samples. Given these findings, we recommend that exposure modelers subdivide their population activity data into at least season/temperature, precipitation, and day-type "cohorts" as these factors are important discriminating variables affecting where people spend their time.
Journal of Exposure Analysis and Environmental Epidemiology (2003) **13,** 294–317. doi:10.1038/sj.jea.7500281

Keywords: *cross-sectional data, exposure modeling, human activity data, indoor time, longitudinal data, outdoor time, motor vehicle time.*

Introduction

All human exposure assessments of environmental contaminants require that human activity/location data in one form or another be conflated with environmental concentration estimates or measures. *Time series exposure models* of the type used by the US Environmental Protection Agency (EPA) requires that activity/location data be available on a daily basis and that the daily *sequence* of an individual's activities and locations be preserved. This type of information is called "activity pattern" data. To meet this need, EPA has developed the Consolidated Human Activity Database (CHAD) that combines in one file all of the publicly-available activity pattern data that exist in this country (McCurdy et al., 2000). A major problem with activity pattern data is that most of it is cross-sectional in nature, usually consisting of only 1 day of information from a single individual, although some studies in CHAD have up to 3 days of information per person. This contrasts with the practice of human exposure modeling, which is longitudinal in nature. Regulatory agencies are interested in understanding the health impacts of being exposed repeatedly to a contaminant over time: months, years, or even lifetimes. Thus, exposure modelers have to utilize cross-sectional activity pattern data to mimic longitudinal activity patterns, which themselves are unknown.

The overall approach used to address this problem is to use modeling "decision rules" to allocate cross-sectional human activity data into "cohorts" of similar people that hopefully minimizes intraindividual variability while maintaining inherent interindividual variability in human activities (Xue et al., 2003). However, there has been little testing of these decision rules due to the dearth of longitudinal human activity pattern studies available for comparison. In fact, there have been very few analyses undertaken of the decision rules themselves to determine if they are theoretically sound or if they actually can explain differences in how people spend their time. We attempt to do this in our paper, to determine if there are factors or attributes that are available in CHAD activity/location data that can explain differences in time patterns. We call these "discriminating factors".

Probably the most extensive analyses to date of possible discriminating factors have been undertaken by exposure

1. Address all correspondence to: Thomas McCurdy, EMRB/NERL, E205-02, RTP, NC 27711, USA. Tel.: +1-919-541-0782.
E-mail: mccurdy.thomas@epa.gov
Received 21 October 2002; accepted 21 March 2003.

assessors developing and using EPA's Office of Air Quality Planning and Standards (OAQPS) exposure models. These analyses are contained primarily in consultant reports or conference proceedings that are not widely available (Johnson et al., 1992, 1995). Factors evaluated in those studies include various meteorological/weather conditions relating to temperature, cloud cover, relative humidity, wind chill, and wind speed, as well as information on activity-specific breathing rates, season of the year, work and/or school day-types, and time-of-day. Using regression analyses, the analysts found that factors related to the time-of-day, work/school status, and season were good predictors of time/activity patterns (Johnson et al., 1992), as were gender, ethnicity, income level (Johnson et al., 1995). Weather variables, while statistically significant — probably primarily due to the large sample size of the analyses — explained very little variance in time spent in five general locations. Based upon these analyses, many of the factors noted above have been used to define population cohorts in OAQPS models designed to estimate exposures associated with alternative air quality standards. See Johnson (1995) and McCurdy (1995) for a discussion of these models.

Methods

In this paper, we explore further the issue of how best to define cohorts used for exposure modeling. We start theoretically by evaluating the literature from various disciplines regarding causal factors thought to be important in affecting human activities. This is summarized in tabular form in Table A1, which is divided into four categories reflecting the disciplines that investigate human activity from different perspectives. These disciplines are: physical activity epidemiology (including exercise), leisure-time studies, human activity (time budget) sociology, and exposure-related surveys of activities of importance. There is a varying amount of analytical support for each causal factor as a review of the "General Relationships" column of Table A1 indicates. There is a relatively strong relationship between activity choices and age, gender, seasons of the year, degree of urbanization, health status (particularly obesity), and "life stages" considerations. There is a mixed and complex relationship between activity and heredity, socioeconomic status (SES), aspects of which are education, occupation, and income; race/ethnicity; day-of-the-week; and geographic region. There is a minor effect overall of parental values on children/adolescent activity choices, but a major effect for some families.

Organizing the factors and considerations used in Table A1 into an "activity factors typology" indicates that 11 identifiable, but not independent, factors are thought to significantly affect an individual's choices regarding the type and amount of activity she/he engages in. They are:

External considerations

1. Season of the year, climate/weather
2. Day-of-the-week
3. Region of the country, degree of urbanization

Individual, intrinsic biological factors

4. Age, gender, growth and development stage
5. Race, ethnicity, genetic makeup/inheritance
6. Health considerations, disability, obesity, pregnancy

Individual, situational/social factors

7. SES, income, and/or education
8. Employment status, occupational type, vacation period
9. Activity lifestyle (e.g., an exerciser or a "couch potato")
10. Home type (single-family residence, apartment, etc.)
11. Life stage: marital status, family values and responsibilities

Only some of these factors are considered in most exposure modeling efforts. Usually, the population of interest is subdivided by age and gender (activity factors typology #4) and by season of the year (#1) factors. Some models consider daily weather (#1), day-of-the-week (#2), region of the country (#3, but only in a general sense), employment status (#8), and home-type characteristics (#10) (Johnson, 1995; McCurdy, 1995). There are only a few modeling efforts that focus on the health status (#6) of a modeled population that we could find; two are EPA's analyses of (1) "jogging asthmatics" exposed to sulfur dioxide (Biller et al., 1984) and (2) persons with pre-existing cardiovascular disease exposed to carbon monoxide (Johnson et al., 1999). Even then, activity patterns used in those assessments were not specific to the compromised health groups of interest, but were abstracted from similar age/gender activity data using nonhealth decision rules. We could find no time-series exposure modeling effort that disaggregated activities by race or ethnicity (#5), by SES (#7), by activity lifestyle" (#9), or by family life stages (#10). These intuitively are important factors in modeling what people do during their daily life, and not considering them in exposure modeling efforts is a shortcoming of our nascent discipline.

Can the items contained in the above typology indeed act as "discriminating factors" to explain differences in the amount of time spent in locations by different cohorts? If so, then they should be used to the extent possible in defining exposure modeling cohorts. We address this question by coevaluating the factors in two different activity databases: a year-long diary kept by a single individual, and a cross-sectional database obtained from CHAD for people similar to that individual in terms of age, work status, education, and residential type. We realize that only generalizations can be made from a longitudinal sample of one; however, there is a need for exposure modelers to understand how well the cross-sectional cohort-approach using CHAD diaries actually represents a "real" individual that the cohort is supposed to represent. Our literature review indicates that this has never been done before. A second paper in this series (Graham and McCurdy, 2003) expands the goal of evaluating the activity typology to the entire CHAD database, covering all ages and both genders.

Development of exposure-relevant locational metrics

In order to develop research hypotheses that can be tested, the area of interest was narrowed toward variables that are currently available in CHAD. The unit of analysis is an *exposure event*, which occurs when a human receptor inhabits a specific location having a uniform pollutant concentration for a specified time period (termed a "microenvironment"). During that event, the exposed person undertakes some type of activity at a specific physiological level. If the microenvironmental concentration, location, or activity/activity level changes, a new exposure event occurs. While a receptor's activity itself sometimes affects a microenvironmental concentration, such as cooking or vacuuming, in this paper we focus on the locational aspects of human activity data.

The longitudinal-diary subject used 124 location codes at least once over the diary year (and 86 activity codes), but only two to 16 unique location codes were used on any one day. The median number of unique codes used per day was 8. While this number seems low, in order to reduce subject burden and increase subject compliance, the time spent in rooms within a building was generally not coded. It was sufficient to note that the subject was "at work", and not that he was in his office rather than a colleague's, or in a hallway or restroom, etc. If the subject went to an office in another building, of course, it was coded as another office. Likewise, rooms within the home typically were not distinguished except for the bathroom and the basement, but even then use of individual room codes was inconsistent. For purposes of this paper, the home location includes all rooms within the house, even if more detail is noted in the diary.

The cross-sectional data are treated in the same manner: all rooms in most buildings are treated as a single aggregated location. (In fact, only aggregated location data are available for most of the studies contained in CHAD.) On this basis, the cross-sectional database of similar people (described below) had between two and 15 unique aggregated location codes per day, with the same median (i.e., 8) as the longitudinal sample. Thus, the number of *major* locations that are normally visited by an individual on a given day is not large, due in part to combining all rooms within a building into one locational code. Since not all of the studies in CHAD collect data for even the same set of aggregate locations, we had to first undertake a comprehensive analysis of the cross-sectional data set to identify the locational codes that are used universally in all CHAD studies. These "common-denominator" codes then become the locational indicators used later for statistical comparison of the activity information.

An initial attempt focused on seven indoor and five outdoor locations, with motor vehicles considered as a separate category due to their high air exchange rate and proximity to roadway pollutants.[2] Upon inspection, however, not all of these aggregated locational codes were used by all studies in CHAD. Problem locations were: work, office, yard (own), other's yard, and park. These problem locations subsequently affect the two locational categories that aggregate all other indoor and outdoor locations. Thus, the original 13 location categories were aggregated up to the following six categories:

- Residences (own & other's)
- All other indoor locations
- Total indoors (summation of the above)
- Motor vehicles (usually a privately owned-vehicle (POV), occasionally a bus or plane)
- Gas stations
- Outdoor locations (except for gas stations)

All of the comparative analyses discussed here are performed using them; the categories are similar to those used by Johnson et al. (1995) for their analyses.

Development of the cross-sectional sample

Given the information in Table A1, at the minimum, the cross-sectional sample should contain similar age/gender characteristics as the individual with longitudinal data. After that, what additional factors or considerations should be used to subdivide the available activity data into logical categories for rigorous comparison? The subject of the year-long activity pattern diary study is a Caucasian male aged 58 years at the start of the study with a post-graduate college education, who is employed full-time performing sedentary desk work. By most SES schema, he is middle-class with a moderate income. He lives and works in a rapidly growing Southeastern USA Metropolitan Statistical Area (MSA), and lives in a suburban single-family detached house with about a 25 min commute to work by automobile. He is married and an "empty-nester" (lifestyle characteristics). Therefore, for the cross-sectional comparison, all records available for similar individuals were extracted from CHAD. There were immediate difficulties in trying to obtain a sufficient sample size for the analysis as described below. It was not considered appropriate to relax the age/gender or

2. The indoor codes were: residence; other's residence; "work" (all locations); offices; malls/stores; restaurants/bars; and "all others". Outdoor locations were: residential yard; sidewalks/parking lots; parks/natural areas; gas stations; and "all others".

educational criteria significantly as they are known to affect quantitatively activities undertaken, particularly physical activity.[3] Therefore, to increase sample size but still maintain cohesiveness, the age/gender selection criteria were restricted to 57–60-year-old males (1 year on both sides). Within that cohort, we found that there was no effect of age on the number of events recorded in the diaries, educational level, or the physical activity level (PAI) of subjects in the final sample; thus, it appeared that relaxing the age range to increase the sample size was reasonable and probably would not significantly bias our results.

There are seven studies in CHAD containing activity information for that cohort of interest. Two of them are slightly different versions (for air and water) of the National Human Activity Pattern Survey (NHAPS), which is a national telephone random probability survey of the prior day's activities (Klepeis et al., 1996, 2001). Another is the California Air Resources Board (CARB) state-level random probability telephone survey of yesterday's activities (Wiley et al., 1991). Three studies in CHAD are random probability concurrent-diary studies for specific MSA's: Cincinnati (Johnson, 1989), Denver (Johnson, 1984; Akland et al., 1985), and Washington DC (Hartwell et al., 1984; Akland et al., 1985). The final study in CHAD that contains information on the cohort of interest is the diary panel study undertaken in Valdez, Alaska in 1990–1991 (Goldstein et al., 1992).

There are 289 daily CHAD activity records for 57–60-year-old males in these seven studies; 195 Caucasians, 24 other specified ethnic/racial groups, and 70 (24% of the sample) unspecified ethnic/race categories. A preliminary analysis of race/ethnicity did not indicate any significant differences in the time spent in indoor and outdoor locations; therefore, race/ethnicity was dropped from further analysis. There was a problem with educational level, as it was missing for 31 person-days of data and there were not many respondents who were college graduates (37), had some grad school (10), or had a graduate/professional degree (28), three of the educational categories used in CHAD that relate to our longitudinal subject.[4] To match the longitudinal diary subject's educational level exactly would reduce the sample size to only 28 days, which is too small for meaningful comparative analyses. Because of this, the educational requirement was weakened to having at least a high school education. Some implications of this relaxation are noted below.

After these adjustments, a sample size of 245 person-days was obtained. However, not all of them were usable for comparisons of interest (see Table A2). Since CHAD data have a prescribed clock hour subdivision, making 24 events/day (d) the minimum number possible, diaries with fewer than 30 daily event records (25% more than the minimum possible) were thought to under-report activities. This figure was turned into an acceptance criterion, and 32 records were removed that had <30 events for a day. Outdoor workers or deliverymen (truck drivers, etc.) in the sample also were removed (15 person-days of data) due to their workplace location (mostly a motor vehicle) that would have distorted the time spent in that location compared with an indoor worker. Eight subjects were removed because they worked on weekends or at home on weekdays, unlike the longitudinal subject. An additional 11 person-days of data were removed because they were missing *both* educational and occupational data, precluding any type of SES analyses for them. Finally, there were 8 days of "grossly" incompatible activity and location data and 2 days of "unusual" diary data that were also removed from the analysis data set; see Table A2. The remaining 169 person-days of data constitute the cross-sectional data set. The two NHAPS studies provided 110 days of data (65% of the total), while CARB and Cincinnati each provided 26 additional days (15% each). The remainder (7 person-days) came from Denver and Washington, DC. None of the Valdez data survived the vetting process.

Analysis of the cross-sectional (C/S) sample by study[5]

A modeler generally combines data from all appropriate studies when defining a cohort for exposure assessment, but the practical impacts of combining activity pattern data from different studies usually are not evaluated. We perform such an evaluation. Since the three MSA-specific studies are very small, they are combined into a single category — called the "MSA studies" — and compared to the three recall studies treated separately. Thus, there are four study types used to determine if the study type itself affects locational choices.

For evaluating categorical variables, a contingency table approach and a χ^2 test statistic were utilized to determine if a "study effect" existed. More formally, the test, or null, hypothesis (H_0) is that there is no statistically significant difference (at $\alpha = 0.05$) among studies with respect to a particular activity typology characteristic; for example, educational level or day-type. The analyses indicated that

3. Physical activity (PA) is a broad term in the exercise physiology/nutrition literatures. It is any movement of the body produced by skeletal muscles that results in an expenditure of energy (Kohl et al., 1988). Exercise is one component of physical activity. PA basically is everything that a human does, including all energy expended *except* that needed for sleeping, basal metabolism, digestion, or growth (retained energy) and reproduction. A marker of daily PA is the physical activity index (PAI). It basically is the total energy expended for a day divided by the person's basal metabolism. See McCurdy (2000) for more information on these subjects.

4. Other educational categories used in CHAD are: none; some elementary; elementary school graduate; some high school; high school graduate; some college; missing; and "any," which includes all categories, including missing.

5. The PC-based Statgraphics® and SAS® statistical packages were used for all analyses in this paper. Most of the analyses were conducted using SAS version 8.2.

H_0 could not be rejected (alternatively, we could not demonstrate a relationship between a study and a factor) for the following factors:

- Age of the subject ($\chi^2 = 6.8$, 9 df, $p = 0.65$; df = degrees of freedom)
- Educational level of the subject ($\chi^2 = 9.7$, 8 df, $p = 0.28$); for the three recall surveys only(no educational data exists for the MSA studies)
- Day-of-the-week for the diary ($\chi^2 = 18.7$, 18 df, $p = 0.41$)
- Day-type[6] ($\chi^2 = 6.32$, 9 df, $p = 0.71$)
- Precipitation[7] level ($\chi^2 = 8.0$, 6 df, $p = 0.24$)

There were too many zero cells to test H_0 for the subject's type of residence, occupation, or the month of the activity data gathering.

With respect to observed continuous variables, a Kolmogorov–Smirnov (K–S) test was used to determine if the study made a difference in a variable's data distribution. The maximum daily temperature of the activity day was significantly different only for the CARB-MSA pair of studies ($D_N = 0.44$, $\chi^2 = 1.66$, $p = 0.008$). This is understandable because the three MSA studies were conducted in northern climates during nonsummer months (March, September, December–February), while the California study was year-round. The modeling implication of this is that if maximum daily temperature cutpoints are used to define a population cohort in an exposure modeling exercise — which is followed in some models (e.g., APEX, HAPEM, pNEM[8]) — then there will be a less-than-proportionate probability of picking a day from the various studies in CHAD for a particular temperature classification.

Evaluating the number of recorded activity–location events/d, the MSA studies ($\bar{X} = 48.1 \pm 9.0$; range = 33–69)[9] had significantly more events (at P-values < 0.001) than the CARB survey ($\bar{X} = 41.0 \pm 7.7$; 30–64), each of which in turn had more than the two NHAPS surveys, which could not be distinguished apart (NHAPS Air: $\bar{X} = 35.5 \pm 9.0$; 30–49 and NHAPS Water: $\bar{X} = 35.7 \pm 3.9$; 30–44). See Figure 1. H_0 could not be rejected for the NHAPS pair ($D_N = 0.24$, $\chi^2 = 1.22$; $p = 0.10$). Since the number of recorded events is particularly important for modeling the exposure and dose profile of a population, possible reasons for the differences among studies were explored. There were no significant differences among attributes of the studies, such as the respondent's state of residence, educational level, day-of-the-week, day-type, or temperature/precipitation level. By deduction, then, differences in the number of events recorded among the studies are likely due to the format used to gather activity information. Basically, the diary format provides more information than the recall format. Johnson et al. (1995) found the same result. While logical, this does not explain why the CARB study has significantly more recorded events than the two NHAPS studies ($D_N = 0.55$; $\chi^2 > 2.2$, $p < 0.001$) since all three were telephone recall surveys and used many of the same questions (Tsang and Klepeis, 1997).

One activity that is recorded in the MSA diary studies that is not generally mentioned in the telephone recall studies is time spent in a "travel transition," the time it takes to go from one location to another via a third location. Examples are (1) going from home to work via a car or other vehicle, and (2) going from an office to the car via a parking lot. Between 75 and 100% ($\bar{X} = 79.2\%$) of the people in the three MSA studies recorded an average of 7.2 travel transitions/d at 5.1 min per event (min/event). For the two NHAPS studies, only 5.6% of the respondents recorded a travel transition, and for those that did there was an average of 3.3 transitions/d at 5.0 min/event. Note that the time per transition was about the same, but the "participation rate" in NHAPS was a meager 7% of the MSA participant's, and the number of travel transitions/d was half that of the MSA

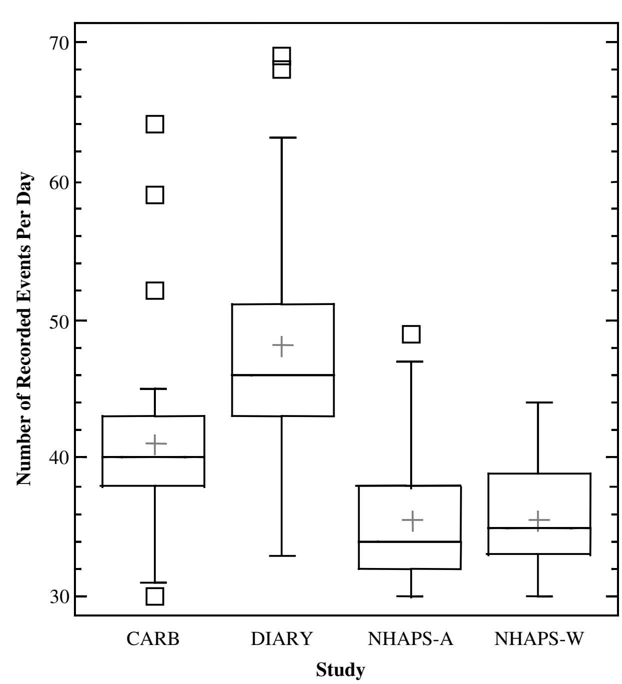

Figure 1. Box plot of the number of events recorded by study.

6. A categorization of the activity day into four day-types: "working" (income-producing) weekday; nonworking weekday; Saturday, and Sunday. Nine different day-types are used for the longitudinal data set, as described below, but the finer distinctions could not be made for the C/S data. Day-type subsequently was reduced to two categories: working weekdays and all-other days.
7. Precipitation is in inches of water measured; there was no distinction made in the data available to us between rain and snow events.
8. These are all OAQPS models. APEX stands for Air Pollution Exposure Model; HAPEM is the Hazardous Air Pollution Exposure Model; pNEM is the Probabilistic NAAQS Exposure Model.
9. The standard deviation (SD) of the sample is shown by the $\pm$ symbol. Subsequent citations to the range drop the word "range" and only show it as a set of figures separated by a dash; $\bar{X}$ = Mean.

studies. The CARB study was even less specific on this point: only one of the 26 CARB respondents in our sample (3.8%) mentioned even one travel transition in his diary. Thus, if multimodal exposures are of concern in a particular assessment, it is suggested that only activity data from one or more of the MSA-specific diary studies in CHAD should be used.

The results of these study comparisons within the cross-sectional data set indicate that there are some "structural" differences among the four studies and that they are associated with the data gathering process itself. The next step was to test if time spent in the six aggregated locations was different among the four studies. Using the two-sample K–S test at $\alpha = 0.05$, H_0 was not rejected for any of the indoor locations: a residence; a nonresidence; and indoors-total. Thus, the four studies are not different in the amount of time their participants spend in these locations. The H_0 of no-study differences in time spent outdoors can be rejected for all possible pairs of the studies ($D_N = 0.32$–0.38, $\chi^2 = 1.47$–1.77, $p = 0.004$–0.020). Participants in the NHAPS-Air survey reported the most time in min/d outdoors ($\bar{X} = 138 \pm 184$; 0–698) and NHAPS-Water the least ($\bar{X} = 94 \pm 127$; 0–475). The values for California ($\bar{X} = 107 \pm 122$; 0–325) and the MSAs ($\bar{X} = 131 \pm 177$; 0–777) fell in between. These differences are not necessarily limited to survey design differences. Much of the statistical difference can be attributed to the unequal distribution of day-type, season, and weather parameters among the studies, resulting in a "study-effect" being noticed here (ANOVA not shown). This is an unavoidable impact of combining data from different sources.

The H_0 of no-study differences in the amount of time spent per day in motor vehicles could only be rejected at $\alpha = 0.05$ for study-pairs involving the CARB California survey. The time spent in motor vehicles for participants in that study was much higher ($\bar{X} = 150 \pm 110$; 20–520) than any of the others (NHAPS-Water was the next highest at $\bar{X} = 95 \pm 105$; 0–655 min/d). There was no statistically significant difference among non-CARB studies: the lowest K–S test p-value for these study-pairs was 0.813. Finally, there were too many zero entries for time spent in gas stations to test for differences among the studies.

To summarize, the four studies differed in two of the tested locational categories: outdoors and motor vehicles (CARB study only). These differences among studies, even for highly aggregated locational categories, raise a cautionary note to users who combine data from them without fully assessing the impacts of doing so.

Evaluation of the activity factors typology in the cross-sectional sample

Statistics for time spent in the aggregated locations for habitués (people who occupy a particular location for >0 min/d) are depicted in Table 1 for both the cross-sectional and longitudinal samples. Besides the six aggregated categories, some data are provided for specialized locations, but these are only of real interest for the longitudinal analysis.

Averaging over all individuals, about 1224 min/d (85% of the day) is spent in indoor locations, but this varies from 570 to 1440 min/d (40–100% of the day) across the sample. Everyone in the cross-sectional sample spends some time indoors, as might be expected; we call this the participation rate in Table 1. Note that 100% of the sample is in a residence on their diary day, either their own or somebody else's. About 72% of the individuals spend some time outdoors and for them it is 160 min/d on average (about 11% of the day), with a range of between 3 and 777 min (<1–54% of the day). This is not the case for the longitudinal sample. The longitudinal sample person spends less time outdoors per day on average, 118 min/d, but is outside some time almost every day (98.6% of the days). In general, the two samples track each other fairly closely with respect to temporal averages, but less so for the participation rates. The longitudinal results in Table 1, and the K–S tests of differences in the longitudinal and cross-sectional samples, are discussed in more detail in the next section.

For those factors that could be tested in the cross-sectional sample, we explored if individual items in the typology affected the time spent per day in the six locations using two-sample K–S tests. Table 2 depicts those factors having a statistically significant relationship ($\alpha = 0.05$) with the time spent outdoors for habitués in the cross-sectional sample; Table 3 contains significant results for the total-indoors and motor vehicle microenvironments. As can be seen in Table 2, many of the typology factors relate significantly to the time spent in the outdoor location. Season of the year, particularly winter, is important, using either months or combined monthly metrics; see Figure 2. The month entry in Table 2 needs some explanation. The first entry, 1:5–1:10, means that each of the comparisons for month 1 (January) with month 5 (May) through month 10 (October), by pairs, are significantly different with respect to time spent outdoors; the K–S statistics for individual month-pairs vary within the ranges shown. Thus, January and May are statistically different with respect to time spent outdoors, as are January and June (1:6); January and July (1:7), etc. The same is true for month-pairs involving September and February (9:2), September and March (9:3), and so forth through the September and June (9:6) pair. The September and July (9:7) pair is not statistically different with respect to time spent outdoors, and thus is not shown in Table 2, but the September and August (9:8), September and November (9:11), and September and December (9:12) pairs are statistically different. The rest of Table 2 is read the same way, as is Table 3. Pairs not shown — for any of the factors listed in Table 2 — are not statistically different for the tested metric.

Table 1. Daily time in minutes spent in aggregated locational categories (habitués only).

Category	Cross-sectional sample ($n=169$)						K–S test[d]				Longitudinal sample ($n=368$)					
	No. of days	Mean (min)	SD[a] (min)	COV[b] (%)	Range (min)	Participation[c] (%)	D_N	χ^2	p	Reject H_0	No. of days	Mean (min)	SD[a] (min)	COV[b] (%)	Range (min)	Participation[c] (%)
Indoor locations																
Residences (own and others)	169	903	224	24.8	85–1440	100	0.13	1.38	0.045	Yes	350	889	249	28.0	18–1434	95.1
All other indoor locations	149	363	239	65.7	3–925	88.2	0.26	2.56	<0.001	Yes	306	461	248	53.8	5–1235	83.2
Total indoors	169	1224	179	14.6	570–1440	100	0.17	1.79	0.003	Yes	368	1228	158	12.7	491–1434	100.0
Motor vehicles	159	107	93	87.0	3–655	94.1	0.16	1.67	0.008	Yes	326	106	99	93.4	7–682	88.3
Gas stations	7	16	9	59.3	5–30	4.1	0.54	1.37	0.048	Yes	68	7	5	69.0	2–23	18.4
Outdoor locations (not gas station)	122	160	161	101	3–777	72.2	0.16	1.52	0.017	Yes	365	118	122	103	6–655	98.4

[a]Standard deviation.
[b]Coefficient of variation (SD/Mean * 100).
[c]Participation rates or percent of sample days in the study spending some time (>0 min/day) in the location (habitués/total sample days).
[d]The two-sample Kolmogorov-Smirnov (K–S) test H_0 is that the distribution of variable 1 is the same as variable 2, using a χ^2 test statistic at $\alpha=0.050$.

When individual days were tested for their impact on the amount of time spent outdoors by habitués, the K–S statistics showed no impact, but when days were combined into weekday and weekend categories, the importance of day-of-the-week on locational decisions is significant. "Day-type", another day-of-the-week metric, relates even more closely to the time spent outdoors; it basically distinguishes among working (paid workday) and nonworking weekdays, Saturday, and Sunday. As seen in Table 2, no significant differences can be found in the time spent outdoors among nonworking weekdays and weekend days (i.e., between 2:3 and 2:4 under day-type). The highest χ^2 statistics are obtained when a distinction is made only between working weekdays and all other days (1:2+) — see the "recoded day-type" metric — possibly due to the larger sample size (n) for each of these day-types. Since the longitudinal sample subject did not do paid work on weekends, we could not analyze this factor further.

Precipitation category in Table 2 (none *versus* $>0.5''$) negatively relates to the time spent outdoors, as expected (0:2 metrics). This is one of the surrogate measures of weather that was tested; the other is daily temperature. Two types of temperature metrics were used: a two-category metric that disaggregated temperature into two classes by season (four combined temperature/season classes). This approach mimics that long used by OAQPS in its exposure modeling work (McCurdy, 1995).

The other temperature metric is a continuous variable, and so it could not be tested via a K–S approach. It was used in a series of regression equations to determine if it — along with certain other variables — could predict habitué time spent outdoors. To determine if the basic assumptions for regression analyses were met, all variables of interest were tested for normality using a K–S test of parameters of the distribution at $\alpha=0.05$. If not, alternative distributions were tested. Time spent outdoors is log-normally distributed in the cohort; temperature and PAI are normally distributed. None of the standard distributions fit any of the other variables in this dataset, so they are used nontransformed in the regression analyses. It must be recognized, however, that having non-normally distributed independent variables violates one of the assumptions of multiple regression; thus, these results should be viewed with some caution.[10]

The overall conclusion from the regression analyses is that the most robust equation (i.e., lowest standard error (SE_{Est}) and having significant independent variables at $\alpha=0.05$) for "explaining" time spent in outdoor locations contains the following independent variables: temperature, precipitation,

10. A dimensional assessment of the *maximum* possible impact on the results predicted by Eq. (1) indicates that non-normally distributed variables account for 20–26% of the variance in ln(OUT). The actual impact would be much less than that in practice.

Table 2. Statistical analyses of factors in the activity typology for the time spent outdoors by habitués.

Activity factor	Tested metric	Tested pair[a]	K–S test D_N	χ^2	p	Notes[b]
External considerations						
1. Time of year	Season	1:2	0.49	1.90	0.002	1 = Winter (Nov–Feb); 2 = Summer (Jul–Sep); 3 = Others
		1:3	0.37	1.68	0.007	
	Month	1:5–1:10	0.65–1.00	1.58–1.90	0.002–0.013	1 = Jan; 2 = Feb; ...; 12 = Dec
		9:2–9:6	0.73–0.90	1.44–1.89	0.006–0.033	
		9:8, 9:11, 9:12	0.65–1.00	1.40–1.55	0.016–0.04	
		7:3, 7:12	0.61–0.67	1.41–1.50	0.023–0.037	
Weather	Precipitation	0:2	0.58	1.95	0.001	0 = None; 1 = Trace (< 0.5 in.); 2 = > = 0.5 in.
	Temperature	–	–	–	–	This continuous variable is tested in regression.
	Temperature/season	C:H	0.68	1.93	0.001	C = < 55°F, Oct to Jun; NC = > 54°F, Oct to Jun; H = > 83°F, Jul to Sep; NH = < 84°F, Jul to Sep.
		C:NC	0.32	1.43	0.033	
		C:NH	0.47	1.73	0.005	
		NH:NC	0.35	1.44	0.032	
2. Day of the week	Days	All	Cannot reject H_o for any day pair			1 = Mon; 2 = Tue; …; 7 = Sun.
	Weekday/weekend	1:2	0.28	1.45	0.029	1 = Weekday; 2 = Weekend.
	Day-type	1:2	0.44	2.00	< 0.001	1 = Weekday, working (paid work); 2 = Weekday, not working (paid work); 3 = Saturday; 4 = Sunday.
		1:3	0.57	1.90	< 0.001	
		1:4	0.48	1.84	0.002	
	Recoded day-type	1:3+	0.51	2.35	< 0.001	1 = Weekday, working; 3+ = Sat and Sun.
	Recoded day-type	1:2+	0.44	2.42	< 0.001	1 = Weekday, working; 2+ = all other day types.
3 Region	Region	All	Cannot reject H_o for any region pair			1 = Northern & Mountains; 2 = Southwest & Great Plains; 3 = Southeast & West Coast.
Intrinsic biological factors						
4. Age	Age	all	Cannot reject H_o for any age pair			Range = 57–60.
5. Race	Black vs White	all	Cannot reject difference using a t-test			—
6. Health considerations	—	—	—	—	—	Not tested due to lack of data.
Situational and social factors						
7. SES/income/ education	Education levels	12:19	0.61	2.15	< 0.001	12 = High school; 14 = Some college; 16 = College graduate; 18 = Some graduate school; 19 = graduate or professional degree.
		14:19	0.61	1.67	0.008	
		16:19	0.59	1.70	0.008	
8. Employment status	—	—	—	—	—	Not tested due to lack of data.
9. Activity lifestyle	PAI	—	—	—	—	This continuous variable is tested in regression.
10. Home type	Residence	All	Cannot reject H_o for any home type pair			1 = Single family house; 2 = All other houses; 3 = Apartment
11. Family role/life stage	—	—	—	—	—	Not tested due to lack of data.

[a] Only pairs assigned statistical significance ($P < 0.050$) are shown.
[b] The note applies to all tested pairs within the tested metric.

Table 3. Statistical analyses of factors in the activity typology for time spent indoors and in motor vehicles.

Activity factor	Tested metric	Tested pair[a]	K–S test D_N	χ^2	p	Notes[b]
Indoors						
1. Time of year	Season	1:2	0.37	1.68	0.007	1 = Winter (Nov-Feb); 2 = Summer (Jul–Sep); 3 = Others.
		1:3	0.26	1.44	0.032	
Weather	Precipitation	all	Cannot reject H_o for any precipitation pair			0 = None; 1 = Trace (<0.5 in.); 2 = > = 0.5 in.
	Temperature/season	C:NH	0.40	1.66	0.008	C = < 55° F, Oct to Jun; NC = > 54° F, Oct to Jun; H = > 83° F, Jul to Sep; NH = < 84° F, Jul to Sep.
		NH:NC	0.32	1.52	0.020	
2. Day of the week	Day type	1:2	0.40	2.14	<0.001	1 = Weekday, working (paid work); 2 = Weekday, not working (paid work); 3 = Saturday; 4 = Sunday.
		1:3	0.54	1.99	0.001	
		1:4	0.39	1.74	0.005	
	Recoded day-type	1:2	0.41	2.65	<0.001	1 = Weekday, working; 2 = all other day types.
Motor vehicle						
1. Time of year	Season	2:3	0.34	1.66	0.008	1 = Winter (Nov–Feb); 2 = Summer (Jul–Sep); 3 = Others.
Weather	Precipitation	all	Cannot reject H_o for any precipitation pair			0 = None; 1 = Trace (<0.5 in.); 2 = > = 0.5 in.
	Temperature/season	C:H	0.44	1.46	0.028	C = < 55° F, Oct to Jun; NC = > 54° F, Oct to Jun; H = > 83° F, Jul to Sep; NH = < 84° F, Jul to Sep.
		H:NC	0.47	1.65	0.009	
2. Day of the week	Day type	all	Cannot reject H_0 for any day pair			1 = Weekday, working (paid work); 2 = Weekday, not working (paid work); 3 = Saturday; 4 = Sunday.
	Recoded day type	1:2	Cannot reject H_0			1 = Weekday, working; 2+ = all other day types.

[a]Only pairs assigned statistical significance ($p<0.050$) are shown.
[b]The note applies to all tested pairs within the tested metric.

PAI, and the dichotomous day-type (the original four-part day-type metric also works well in regression analyses). On the other hand, for the indoor, nonresidential indoor, and motor vehicle locations the only significant independent variable is day-type for any metric formulation (results not shown).

The most robust equation ($R_a^2 = 0.38$, $SE_{Est} = 0.98$) for time spent outdoors by habitués follows; p-values for the regression coefficients are depicted in brackets:

$$\ln(\mathrm{OUT}) = \underset{[0.469]}{-0.53+} \quad \underset{[<0.001]}{0.02\times\mathrm{TEMP}} \quad \underset{[0.006]}{-0.39\times\mathrm{PREC}} \quad \underset{[<0.001]}{+1.07\times\mathrm{DAYT}} \quad \underset{[<0.001]}{+1.18\times\mathrm{PAI}} \tag{1}$$

where ln(OUT) is the natural logarithm of the time spent outdoors (min/d); TEMP the temperature (°F); PREC the precipitation (0 = none; 1 = trace to <0.5″; 2≥0.5″); DAYT the day-type (1 = work weekday; 2 = other) and PAI the physical activity index (unitless, median of 10 simulations).

If the PAI independent variable is left out of the equation, the predictive power of the regression equation is diminished ($R_a^2 = 0.29$, $SE_{Est} = 1.05$), thus:

$$\ln(\mathrm{OUT}) = \underset{[<0.001]}{+1.90} \quad \underset{[<0.001]}{+0.02\times\mathrm{TEMP}} \quad \underset{[0.006]}{-0.40\times\mathrm{PREC}} \quad \underset{[<0.001]}{+0.88\times\mathrm{DAYT}} \tag{2}$$

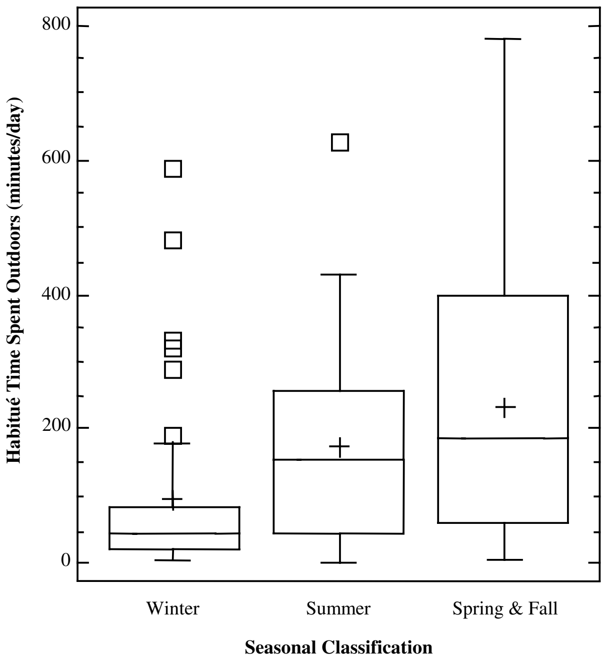

Figure 2. Box plot of habitué time spent outdoors by season.

Notice that removing PAI results in the constant term becoming statistically significant, which is not desired. It is done for comparison with the longitudinal sample later.

Returning to Table 2, although there are noticeable differences in the time spent outdoors by region, the H_o of no regional differences cannot be rejected at $p<0.05$. Possibly the small sample sizes do not allow definitive conclusions to be drawn concerning this variable. The same is true for age, since there is very little age variability in the sample used. Educational level is the only item in the SES/Income category that could be tested, but the cross-sectional sample only contains high school graduates or above. It appears that educational level can account for differences in the time spent outdoors for persons with post-undergraduate college education — their time spent outdoors is about one-third that of people with lower educational levels (61 *versus* 176–209 min/d) — but not for other possible educational levels. Education is not highly correlated with any other aggregated location category either (not shown). Home-type does not affect outdoor time nor is it significantly correlated with any other location. Finally, due to lack of information, we could not test whether health considerations or family role/life stage affects the time spent in the outdoor — or any other — location.

Table 3 indicates that fewer of the activity typology factors significantly affect time spent indoors or in a motor vehicle. Season and temperature classes have some impact on these locations, but precipitation does not. Day-type metrics affect time spent indoors but not in-vehicle time.

Most of the activity typology factors are weakly to moderately related with time spent in outdoor locations and to the other aggregate locations that were tested. Analyses of the entire CHAD database provide even more support for this conclusion (Graham and McCurdy, 2003). Exposure modelers should consider these factors, particularly age, gender, season of the year, temperature, precipitation, and day-type (or at least a weekday/weekend breakdown) when choosing activity patterns for modeling purposes. While this seems to be an obvious conclusion, there has been limited testing of these factors with respect to their use in framing exposure modeling efforts; the Johnson et al. citations (1992, 1995) are exceptions to this finding. Many of the factors found to be important here are considered in the APEX, HAPEM, pNEM, and more recently, the Stochastic Human Exposure and Dose Simulation (SHEDS) exposure models, but in a different formulation. The authors know of no other models that do so.

Development of the longitudinal sample

The longitudinal sample consists of 368 days of activity data (out of 369 days[11]) from a male who was aged between 57 and 58 years at the time of the survey. The diary used in the study is depicted in Figure 3. It is a revision of the one used in the Cincinnati study discussed earlier. Every time there was a change in any item on the form (location, activity, activity level, the opening or closing of windows, smoking situation, combustion status, or solvent usage), the subject was instructed to fill out a new diary page. Compliance by the motivated subject, the first author, was excellent; there was less than 0.5% missing data for any item on the form. In addition to the diary data, the subject for each day recorded daily maximum temperature, if it rained or snowed, and if he had to alter any activity that day due to sickness, injury, or some other specified reason.

The paper diary data were put into a Lotus 1-2-3® file, and a number of quality control checks were made to ensure that the data were entered correctly. Activity start/stop times were checked to make sure they were not juxtaposed. Activity and location codes were checked to ascertain that they were within the range of usable CHAD codes, and a number of logic checks were made to ensure that the coded activity was not obviously incompatible with a coded location. Finally, coded total daily time was checked to ascertain that it equaled 1440 min each day. If any of these checks uncovered a problem for a day, *every* entry for that day was compared to the filled-out diary and corrected if needed. Finally, some day-types of interest were 100% reviewed even if they passed all of the above checks. These included: weekdays-not-at-

11. The subject lost a diary about 20.00 h on day #358, February 14, 2000, in snow covering an airport parking lot as he was loading luggage into a car. The "penalty" for this noncompliance was carrying the diary for 3 additional days, which is why $n=368$ rather than 365.

Date	
Day of Week	
Max Temp.	° F
Rain or snow?	Y N
Alter any activities because of ill health?	Y N

Thanks for your participation. EPA needs this information to understand where/how people are exposed to environmental pollutants. We greatly need information on what people do over many days, as these data are non-existent. So please, STICK WITH IT AS LONG AS YOU CAN!

Thank you.

Time	am	pm
Location		
Activities		
If you breathed hard or broke a sweat, how many minutes did you?		
IF YOU ARE INDOORS:		
Any window(s) open?	Y	N
A smoker present?	Y	N
Combustion activities[1]	List:	
Solvent usage[2]	List:	

1. Activities of interest include:
   cooking with a gas burner/oven
   heating with a kerosene heater
   using a fireplace
   using a gas stove to heat the house

2. Solvents of interest include:
   insect repellents
   spray or other paints
   floor or furniture wax or polishes
   hair sprays/cooking sprays
   deodorants

If this is the last activity in the day, go to the "day sheet" and start at midnight.
Fill out a new set of diary pages for each day (midnight-to-midnight).

Figure 3. Cover sheet and an event-specific diary page used in the longitudinal study.

work, vacation days, work-related trip days, sick/injury days, and "regional disaster" days when all highway travel was officially banned in the subject's MSA (*Hurricane Floyd:* 1 day; a major snowstorm: 4 days).

The day-types originally used in the longitudinal study are depicted in Table A3. For comparison with the cross-sectional (C/S) analyses, day-type 1 is the same for both samples, as is nonvacation Saturday and Sunday (#s 7 and 8 in the table). C/S day-type 2: weekday, not at work is roughly equivalent to day-types 3-5 in Table A3, but there is no C/S equivalency for longitudinal day-types 2, 6, 9, and 10. While "real", these day-types, constituting 35 days (9.5% of the sample), cannot be modeled using CHAD or any other known activity database since activity data normally are gathered only when subjects are at home and can be contacted. We combine some of the day-types, and ignore "work-related trip days" and "regional disaster days", when considering day-types to be compatible with the cross-sectional sample. However, from an exposure modeling perspective these day-types should not be ignored in future activity data-gathering surveys.

Summary descriptive statistical data for the longitudinal habitué sample are depicted in Table 1 for the same locations used in the cross-sectional sample. While the mean time spent in the six aggregated locations is similar for the two samples, as are the COVs and ranges, K–S tests indicate that their distributions are statistically different at $\alpha = 0.05$ for all six categories. In addition, there are some large disparities in participation rates (number of days in a location ÷ total sample days) between the two samples, particularly for gas stations, outdoor locations, and outdoors-own yard (a "specialized" location).

Descriptive statistics for the time spent in outdoor, indoor, and in-vehicle locations are depicted in Tables 4–6 for the longitudinal sample for those days with >0 min time in a location. The data are disaggregated by day-types (four categories), precipitation category (three categories), and temperature/season classes (four categories). As mentioned, the temperature-by-season class approach is that used by OAQPS in its exposure modeling work, based upon analyses of weather effects on human locational decisions undertaken by Johnson et al. (1992, 1995).

The four classes are:

- temperature $<55°$F , October – June [C]
- temperature $\geq 55°$F, October – June [NC]
- temperature $<84°$F, July – September [NH]
- temperature $\geq 84°$F, July – September [H]

There are fairly large differences in all of the descriptive statistics shown in the outdoor and in-vehicle locations (Tables 4 and 6), respectively. There is much less difference shown in the indoor locations (Table 5). This corresponds to the cross-sectional sample findings.

Also shown in the table are the distributions that best "fit" the data[12] using the K–S test; four distributions were evaluated: normal (N), lognormal (L), exponential (E), and Weibull (W). The distributions are listed in the order of best fit. If none of the four distributions fit the data, it is designated "nf." In general, the lognormal and exponential fit the data best.

Evaluating the activity factors typology in the longitudinal sample

Not many of the items in the activity factors typology can be evaluated with the longitudinal sample, since many of them are keyed to subject characteristics that do not vary in the longitudinal study. We were able to test for the day-type, temperature, and precipitation factors, using the two-sample K–S test discussed above. With respect to time spent outdoors (Table 4), the K–S test indicates that all the possible day-type pairs — there are six of them — except the Weekend/Holiday and Vacation pair (the 2:3 pair) can be rejected at $\alpha = 0.05$. This indicates that those two day-types can safely be combined without loss of discrimination for this parameter, but that the other day-types should be accounted for. For the three possible precipitation pairs, the only one that can be rejected is the No/Measurable precipitation (0:2), just as was determined for the cross-sectional sample. None *versus* trace and trace *versus* measurable precipitation cannot be distinguished for the time spent outdoors by the longitudinal sample. With respect to the six possible temperature-by-season categories, only two pairs can be marginally distinguished: H:NC ($D_N = 0.20$; $\chi^2 = 1.36$; $p = 0.049$) and NH:NC ($D_N = 0.27$; $\chi^2 = 1.40$; $p = 0.041$). When day-type is considered, there are no longer any significant differences in the K–S test, and we attribute the differences to a disproportionate amount of vacation days in the NC grouping, which makes for a high mean/standard deviation for time spent outdoors for that season/temperature class. See also Figure 4.

Information for indoor locations appears in Table 5. With respect to day-type, more time is spent indoors when sick than for any other type of day, as expected (and it is spent at home, data not shown). The least amount of time spent indoors is on vacation, another expected result. All of the day-type pairs can be distinguished by the K–S test, except the Sick/Workday pair (0:1), and we believe that this is due to sick time being, in essence, substituted for work time. Precipitation class affects time spent indoors for at least the "Measurable" category, with 1274 min/d, and this category can be distinguished from the no-precipitation class using the K–S test, but not from the "Trace" amount class. This result is the same as was found for the precipitation impact on outdoor time. Indoor and outdoor results are very different with respect to K–S testing of the temperature/season categories. For indoor time, three pairs have statistically different distributions: H:NH during the "summer" season and the two "cross-season" pairs: H:NC and NH:NC. The reasons for these differences are the "mirror image" of those noted above for time spent outdoors.

These findings probably are most efficiently depicted by bar charts that show the interaction of day-type with season/temperature and precipitation for time spent outdoors per day. This is shown in Figures 4 and 5 for nonzero time days only. There is an interaction in structure by both weather factors, and sometimes the differences for the various joint sets are significant. For instance, time spent outdoors per day is less on workdays (day-type 4 in Figure 4) than any other day-type, with hardly any season/temperature variability. This is expected. The most time spent outdoors is during vacation (#2) and weekend/holiday (#3) days, and of these day-types the maximum corresponds to the NH (d) and NC (c) days. This again is expected. With respect to Figure 5, more time is spent outdoors on vacation and during weekend/holiday days when it is not appreciably raining than on other day-types, another logical finding.

Also shown on the two figures are codes above individual bars that represent the other cases that have significantly different distributions of the time spent outdoors than the case depicted (using the K–S test). However, we evaluated these differences only for cases within the same day-type

12. Actually, the H_0 is that the data fit the stated distribution using a χ^2 approximation to a large-sample K–S statistic at $=0.05$.

Table 4. Time spent (min/d) in outdoor locations by day-type, precipitation, and temperature categories for longitudinal subject.

Parameter	Distribution[a]	n[b]	Mean (min)	Std Dev (min)	COV (%)	Range (min)
Day[c]						
0	E, N, L	11	88	72	81	6-251
1	L, E	205	54	54	99	3-325
2	nf	108	220	143	65	8-655
3	N, W, L	31	192	107	56	3-424
Precipitation[d]						
0	E	283	126	126	100	4-655
1	N, L, E	11	186	170	91	6-482
2	E, L	69	77	83	108	3-372
Temperature[e]						
C	E, L	72	90	87	97	3-398
NC	nf	200	139	141	102	4-655
NH	L, E	30	92	115	125	3-482
H	E, L	61	99	77	78	5-327

[a]N = normal; L = lognormal; E = Exponential; W = Weibull; nf = none fit. Order implies best fit.
[b]In addition to the data shown here (habitués only) there were a few days where time spent outdoors was zero (0); for day $n = 4$; precipitation $n = 5$; temperature $n = 5$.
[c]0 = Sick; 1 = Work Week; 2 = Weekend/Holiday; 3 = Vacation.
[d]0 = none; 1 = trace; 2 = measurable.
[e]C = October to June $< 55°F$; NC = October to June $> = 55°F$; NH = July to September $< 84°F$; H = July to September $> = 84°F$.

Table 5. Time spent (min/d) in indoor locations by day-type, precipitation, and temperature categories for longitudinal subject.[a]

Parameter	Distribution	n[b]	Mean (min)	Std Dev (min)	COV (%)	Range (min)
Day						
0	L, N, E	11	1310	86	7	1170–1434
1	nf	209	1293	77	6	772–1381
2	nf	108	1173	168	14	596–1432
3	N, L	31	962	201	21	491–1335
Precipitation						
0	nf	287	1219	162	13	491–1418
1	N, E	11	1181	178	15	874–1377
2	nf	70	1274	126	10	788–1434
Temperature						
C	nf	72	1255	145	12	570–1390
NC	nf	204	1208	166	14	491–1434
NH	nf	31	1264	130	10	910–1372
H	nf	61	1248	147	12	596–1395

[a]See Table 4 for relevant notes.
[b]There were no data containing zero (0) time spent indoors.

"column" or within the same season/temperature "row"; in other words, we did not use the K–S test for cases on the diagonal. As one example, the 4a case (outdoor time on workdays during the cold period [a]) is significantly different than the 2a and 3a cases: vacation time and weekend/holiday day-types in the October–June $< 55°F$ class. The absence of any day-type four codes above the 4a bar means that time spent outdoors on workdays is not significantly affected by season/temperature classes. This, too is a logical finding. All of the "super-scripted" values noted in Figures 4 and 5 should be read in the manner just described; remember, diagonal cases were not evaluated for statistical difference.

The in-vehicle data are depicted in Table 6. The mean time spent per day in a motor vehicle is significantly more on vacation days than for any other day-type. A K–S test of motor vehicle time indicates that (1) it is significantly different with respect to day-type for all possible combina-

Table 6. Time spent in minutes in automobiles by day-type, precipitation, and temperature categories for longitudinal subject.[a]

Parameter	Distribution	n^b	Mean (min)	Std Dev (min)	COV (%)	Range (min)
Day						
0	N, E	6	71	60	85	10–168
1	nf	208	92	44	47	41–479
2	E	73	68	99	145	7–654
3	N, L	28	303	155	51	54–682
Precipitation						
0	nf	257	107	103	97	8–682
1	E	7	113	77	68	50–280
2	nf	62	100	84	84	7–424
Temperature						
C	nf	63	107	119	111	7–682
NC	nf	178	109	92	84	7–519
NH	nf	28	96	80	84	8–424
H	nf	57	99	109	111	13–654

[a]See Table 4 for relevant notes.
[b]In addition to the data shown here (habitues only) there were a few days where time spent in autos was zero (0); for day $n=44$; precipitation $n=42$; temperature $n=42$.

tions except sick–workday and sick–weekend/holiday pairs, and (2) it is not significantly different for any of the possible precipitation or temperature pairs. Thus, weather conditions do not affect the longitudinal subject's time spent per day in a motor vehicle, but this may be a peculiarity of the subject himself. Weather conditions may very well affect the time spent in motor vehicles for persons with more discretionary time or those not on a "fixed" time schedule, such as the elderly, unemployed individuals, and retirees.

An analysis of variance was undertaken for outdoor data in the longitudinal data set; results of this analysis appear in Table 7. For this assessment, only a two-class precipitation variable was used: none + trace and measurable. All other variables are defined as above, and they all are statistically significant in explaining total model variance, itself accounting for about 56% of total sample variance in the time spent outdoors per day. It should be noted that day-type accounts for most of variance explained and that there is interaction between day-type and (1) temperature and (2) precipitation that accounts for more variance than temperature or precipitation alone. This finding suggests that using a combination of some of the original typology factors relating to day-of-the-week and weather considerations may be important in developing more relevant cohorts for exposure modeling purposes.

A regression analysis of the longitudinal time spent outdoors, similar to that carried out for the cross-sectional sample, was undertaken using the same variable definitions. PAI data were not available for the longitudinal study subject. The most robust (non-PAI) equation is

$$\ln(\mathrm{OUT}) = \underset{[<0.001]}{+1.64} \quad \underset{[0.015]}{+0.01\times\mathrm{TEMP}} \quad \underset{[<0.001]}{-0.30\times\mathrm{PREC}} \quad \underset{[<0.001]}{+1.49\times\mathrm{DAYT}} \tag{3}$$

These regression parameters and statistics are almost identical to those of the cross-sectional sample. The longitudinal equation is marginally better at explaining variance in outdoor time ($R_a^2=0.40$, $SE_{Est}=0.94$), probably because a single individual is involved instead of disparate individuals with different lifestyles, etc. On the other hand, it is remarkable that the equations are so similar. This hints that day-to-day variability within an individual approaches that found between kindred individuals of like age/gender/education/work status.

Finally, we evaluated the day-to-day correlation structure that exists in the longitudinal sample for time spent outdoors. We looked at all-possible day-of-week combinations, such as Monday/Tuesday, Monday/Wednesday, and so forth, making 21 unique pairs total. We did this for all day-types as well as specific day-types, such as workdays and non-workdays. For all day-types, one-third of the possible daily pairs are significantly correlated: the highest r is 0.56 ($p<0.001$) for Saturday/Sunday. The six other significant correlations follow:

- Monday/Tuesday: $r=0.43$ ($p=0.001$)
- Monday/Thursday: $r=0.28$ ($p=0.042$)

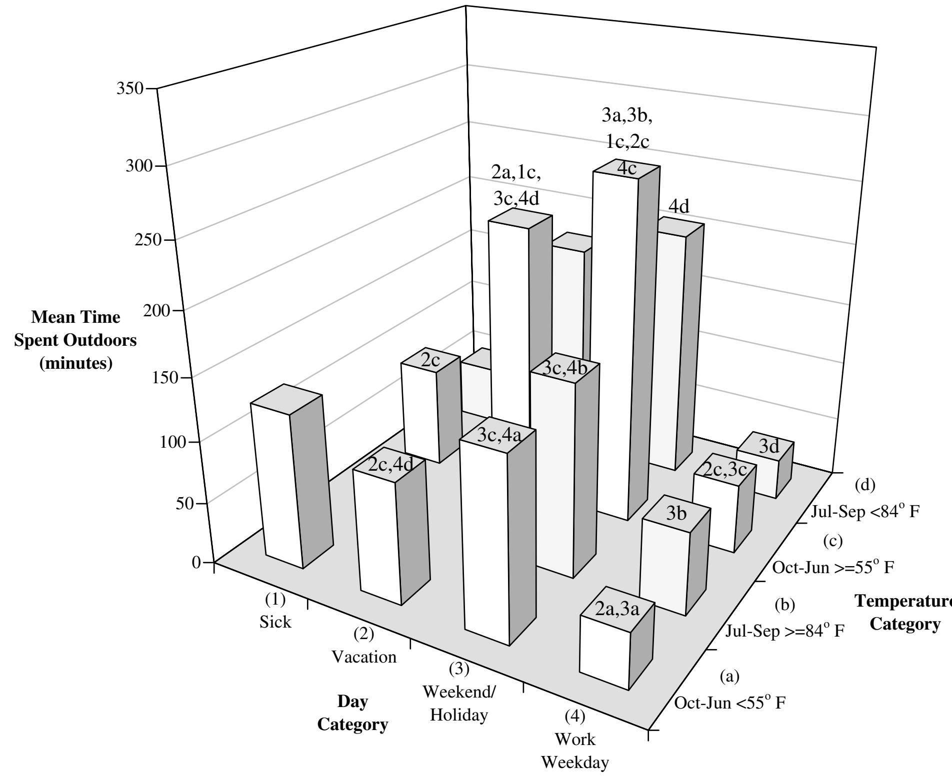

Figure 4. Mean time spent outdoors for longitudinal subject by day and temperature categories. Pairwise significant differences in distribution (K–S test) are indicated by number–letter pattern.

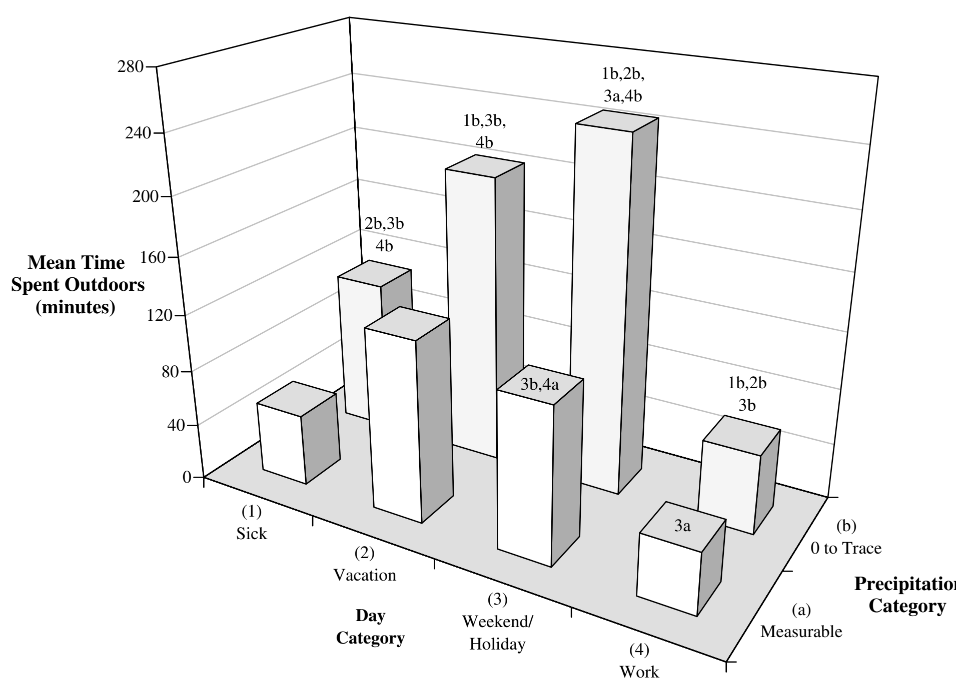

Figure 5. Mean time spent outdoors for longitudinal subject by day and precipitation categories. Pairwise significant differences in distribution (K–S test) are indicated by number–letter pattern.

Table 7. Analysis of variance of time spent outdoors for longitudinal subject.

Source	DF	SS	MS	F	p	R^2
Model	17	2,982,677	175,452	25.4	<0.0001	0.558
Error	341	2,358,987	6,918			
Corrected total	358	5,341,663				
Type III SS for source categories [a]						
Day	3	514,033	171,344	24.8	<0.0001	
Temperature	3	92,912	30,970	4.5	0.0042	
Precipitation [b]	1	84,544	84,545	12.2	0.0005	
Day * temperature	7	291,390	41,627	6	<0.0001	
Day * precipitation	3	169,576	56,526	8.2	<0.0001	

[a]Day and temperature categories included all levels noted in Table 4.
[b]For the precipitation parameter there were 2 levels; "no precipitation" was combined with trace level since there was no significant difference between the two (0 to trace level); measurable amounts constituted the second level.

- Tuesday/Wednesday: $r=0.49$ ($p<0.001$)
- Tuesday/Thursday: $r=0.41$ ($p=0.002$)
- Tuesday/Friday: $r=0.47$ ($p<0.001$)
- Wednesday/Friday: $r=0.36$ ($p=0.008$)

Therefore, while some adjacent day-pairs are significantly correlated with respect to the time spent outdoors by day, others are not. None of the r's are particularly high, even though they are significantly different than 0.0 at $\alpha=0.05$. The correlations are worse when the data are subdivided by day-type. For nonworkdays, only three of the 21 possible day pairs are significantly related — and one is Saturday/Sunday, which is identical to the data used above. With respect to workdays, the only significant pair out of the 20 possible — Saturday/Sunday drops out — is Wednesday/Friday ($r=0.54$; $P=0.001$), but this could very well be an artifact.[13]

We also looked at day-pair correlations for the time spent indoors by day-type, but there are very few significant r's among the possible day-pairs. There were none for indoors for workdays and four for nonworkdays (out of 21). We think that the main reason for these results is that there is not much daily variability in time spent in aggregate locations for a single individual; hence, there is not much variance to "work with" for the correlations. This problem is exacerbated when the data are divided into common day-type categories, where you also run into small sample size problems. The analyses indicate that high day-to-day correlations in the time spent in even aggregated locations by an individual do not exist in general, contrary to what might be expected. While this may be a peculiar finding of the individual being analyzed, it is consistent with other correlation analyses seen in the literature (Schwab et al., 1990, 1991, 1992; MacIntosh, 2001; Xue et al., 2003).

13. With $\alpha=0.05$, you could expect one out of 20 correlations to be significant by chance, even if there is no actual correlation.

Summary and Conclusions

An activity-factor typology was developed based upon factors thought by various social science disciplines to affect human activity choices. Some of these factors are used in a few exposure modeling efforts, but the practice is not standardized. Each of the explanatory factors that could be evaluated was statistically tested against location data contained in two samples of data: (1) a year-long longitudinal pattern from one individual, and (2) a cross-sectional data set of similar individuals selected from CHAD. The tests used aggregated (common denominator) locations for the most part: time spent per day in the outdoors, indoors, and in-vehicles. More detailed locational analyses were not possible due to lack of consistency in coding locations in the various studies comprising the CHAD database.

Many of the factors in the activity typology could not be tested. Data simply are not available in CHAD for all of these factors for all of the studies. These are called "structural differences" in the studies. Combining a study that contains data for a particular factor with one that does not obviously biases downward aggregated statistics developed for the set as a whole. It also makes it difficult to obtain a large and representative sample size for exposure modeling of exposures experienced in specific locations, such as alongside-the-road and in parking lots. Making a transition between travel locations was used as an example of a specific location with data from only a few surveys.

On the other hand, in the sample evaluated, there were no significant differences among the CHAD studies for many of the activity factors deemed important in affecting human choices. Included were age, educational level, day-of-the week of the diary, type-of-day (weekend/weekday), and precipitation amount. Thus, there is a mixed message here, and it is an important one: a modeler has to be very careful in combining CHAD studies when developing an input file of human activity data.

The time spent per day outdoors exhibited the greatest relative variability of the locations investigated, with in-vehicle time being next. With respect to factors in the typology, the cross-sectional sample indicates that time spent outdoors is significantly affected by seasons (and months), precipitation (none *versus* measurable amounts only), temperature, seasonal/temperature combinations, day-type (various coding schemes tested), education, and relative physical activity level. Factors not found to be important are age (likely due to the narrow age range tested), race, and type of residence. Statistical testing indicates that only season, season/temperature class, and day-type significantly affect the time spent indoors. With respect to in-vehicle time, there were few significant activity factors that affected this location: one season-pair and two season/temperature classes. The remaining typology factors were not important distinguishing variables or they could not be tested for total indoor or in-vehicle time.

The impact of the activity typology factors on time allocation in the longitudinal sample was also tested to the extent possible. The findings for this sample were quite similar to those for the cross-sectional sample. For outdoor time, most day-type categories provided statistically different time per day estimates, as did no *versus* measurable precipitation levels, and some season/temperature classes. Only day-type significantly affected the amount of time spent in a vehicle for the longitudinal sample.

A series of multiple regressions of the time spent outdoors were undertaken for the two samples. The structure of the best-performing cross-sectional equation was then used for the longitudinal sample. Estimated parameters of the two, admittedly modest, regressions were remarkably similar in both cases. Statistically significant explanatory variables found in both samples were temperature, precipitation, and day-type. We recommend that exposure modelers use these factors to subdivide their activity days into bins or "cohorts" from which to sample human activity data, as they are important discriminating variables affecting people's choices in spending time in aggregate locations.

Acknowledgments

We thank the two anonymous reviewers of this paper for helpful suggestions regarding its content and format. The authors also thank two of our EPA colleagues, Dr.s Janet Burke and Alan Huber for a detailed review. Responding to all of the issues and comments that arose hopefully improved the clarity, presentation, and applicability of this paper.

Disclaimer

The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. It has been subjected to Agency review and approved for publication. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use.

References

Aarino M., Winter T., Kujala U.M., and Kaprio J. Familial aggregation of leisure-time physical activity—a three generation study. *Int J Sports Med* 1997: 18: 549–556.

Aaron D.J., Dearwater S.R., Anderson R., Olsen T., Driska A.M., and LaPorte R.E. Physical activity and the initiation of high-risk health behaviors in adolescents. *Med Sci Sports Exer* 1995: 27: 1639–1645.

Aaron D.J., Kriska A.M., Dearwater S.R., Anderson R.L., Olsen T.L., Cauley J.A., and LaPorte R.E. The epidemiology of leisure physical activity in an adolescent population. *Med Sci Sports Exer* 1993: 25: 847–853.

Adair J.H., and Spengler J.D. Assessing activity patterns for air pollution exposure research. In: Starks T.H. (Ed.) *Research Planning Conference on Human Activity Patterns*. US Environmental Protection Agency, Las Vegas, 1989: pp. 6–1—6–19, EPA-600/4-89-004.

Ainsworth B.E., Irwin M.L., Addy C.L., White M.C., and Stolarczyk L.M. Moderate physical activity patterns of minority women: the Cross-Cultural Activity Participation Study. *J Women's Health Gender-Based Med* 1999: 8: 805–813.

Ainsworth B.E., Richardson M.T., Jacobs D.R. Jr., and Leon A.S. Gender differences in physical activity. *Women Sport Phys Activity* 1993: 2: 1–16.

Akland G.G., Hartwell T.D., Johnson T.R., and Whitmore R.W. Measuring human exposure to carbon monoxide in Washington, D.C. and Denver, Colorado during the Winter of 1982–83. *Environ Sci Tech* 1985: 19: 911–918.

Andersen L.B. Changes in physical activity are reflected in changes in fitness during late adlescence. *J Sports Med Phys Fitness* 1994: 34: 390–397.

Andersen L.B., and Haraldsóttir J. Tracking of cardiovascular disease risk factors including maximal oxygen uptake and physical activity from late teenage to adulthood: an 8 year follow-up study. *J Intern Med* 1993: 234: 309–315.

Anderssen N., Jacobs D.R. Jr., Sidney S., Bild D.E., Sternfeld B., Slattery M.L., and Hannan P. Change and secular trends in physical activity patterns in young adults: a seven-year longitudinal follow-up in the Coronary Artery Risk Development in Young Adults Study (CARDIA). *Am J Epidem* 1996: 143: 351–362.

Atkins S., Stratton G., Dugdill L., and Reilly T. The free-living physical activity of schoolchildren: a longitudinal study. in: Armstong N., Kirby B., Welsman J. (Eds.). *Children and Exercise XIX*. E & FN Spon, London, 1997: pp. 145–150.

Baranowski T., and de Moor C. How many days was that? Intra-individual variability and physical activity assessment. *Res Q Exer Sports* 2000: 71: 74–78.

Baranowski T., Smith M., Thompson W.O., Baranowski J., Hebert D., and de Moor C. Intraindividual variability and reliability in a 7-day exercise record. *Med Sci Sports Exerc* 1999: 31: 1619–1622.

Beunen G., and Thomis M. Genetic determinants of sports participation and daily physical activity. *Int J Obes* 1999: 23 (Supp. 3): S55–S63.

Bijnen F.C.H., Feskens E.J.M., Caspersen C.J., Mosterd W.L., and Kromhout D. Age, period and cohort effects on physical activity among elderly men during 10 years of follow-up: the Zutphen elderly study. *J Gerontol Med Sci* 1998: 53A: M235–M241.

Biller W.F., Thomas H.C. Jr., Stoeckenius T.E., and Paul R.A. "Estimation of short-term sulfur dioxide population exposures."

Paper presented at the 77th Annual Meeting of the Air Pollution Control Association, San Francisco, June, 1984.

Black A.E., Coward W.A., Cole T.J., and Prentice A.M. Human energy expenditure in affluent societies: an analysis of 574 doubly-labelled water measurements. *Eur J Clin Nutr* 1996: 50: 72–92.

Bouton C.V., Lichtenbelt W.D.V., and Westerterp K.R. Body mass index and daily physical activity in anorexia nervosa. *Med Sci Sports Exerc* 1996: 28: 967–973.

Brooks C. Adult physical activity behavior: a trend analysis. *J Clin Epidemiol* 1988: 41: 385–392.

Bullen B.A., Reed R.B., and Mayer J. Physical activity of obese and nonobese adolescent girls appraised by motion picture sampling. *Am J Clin Nutr* 1964: 14: 211–224.

Burchfiel C.M., Sharp D.S., Curb J.D., Rodriguez B.L., Hwang L.J., Marcus E.B., and Yano K. Physical activity and incidence of diabetes: the Honolulu Heart Program. *Am J Epidemiol* 1995: 141: 360–368.

Burton N.W., and Turrell G. Occupation, hours worked, and leisure-time physical activity. *Prev Med* 2000: 31: 673–681.

Butcher J. Longitudinal analysis of adolescent girls' participation in physical activity. *Sociol Sport J* 1985: 2: 130–143.

Centers for Disease Control. Sex-, age-, and region-specific prevalence of sedentary lifestyle in selected states in 1985 — the Behavioral Risk Factor Surveillance System. *MMWR* 1987: 36: 195–204.

Centers for Disease Control and Prevention. Health-risk behaviors among our nation's youth: United States, 1992. *Vital Health Statistics*, Series 10, No. 192. US Department of Health and Human Services, Hyattsville MD, 1995.

Centers for Disease Control and Prevention. *Physical Activity and Health*. US Department of Health and Human Services, Washington, DC, 1996.

Centers for Disease Control and Prevention. *Physical Activity and Health: A Report of the Surgeon General*. US Department of Health and Human Services, Washington, DC, 2000a. (www.cdc.gov/nccdphp/sgr.html).

Centers for Disease Control and Prevention. Prevalence of leisure-time physical activity among overweight adults — United States 1998. *MMWR*. 2000b: 49: 326–330.

Chapin F.S Jr. *Human Activity Patterns in the City*. John Wiley & Sons, New York, 1974.

Cooper A.R., Page A., Fox K.R., and Misson J. Physical activity patterns in normal, overweight and obese individuals using minute-by-minute accelerometry. *Euro J Clin Nutr* 2000: 54: 887–894.

Crespo C.J., Smit E., Carter-Pokras O., and Andersen R. Acculturation and leisure-time physical inactivity in Mexican American adults: results from NHANES III, 1988-1994. *Am J Public Health* 2001: 91: 1254–1257.

Cunningham D.A., Montoye H.J., Metzner H.L., and Keller J.B. Physical activity at work and active leisure as related to occupation. *Med Sci Sports* 1969: 1: 165–170.

Dempsey J.M., Kimiecik J.C., and Horn T.S. Parental influence on children's moderate to vigorous physical activity participation: a expectancy-value approach. *Pediatr Exer Sci* 1993: 5: 151–167.

Dovey S.M., Reeder A.I., and Chalmers D.J. Continuity and change in sporting and leisure time physical activities during adolescence. *Br J Sports Med* 1998: 32: 53–57.

Eaton C.B., Feldman H., Reynes J., Lasater T., Assaf A.R., and Carleton R.A. Predicting physical activity change in men and women in two New England communities. *Am J Prev Med* 1993: 9: 209–219.

Echols S.L., MacIntosh D.L., Hammerstrom K.A., and Ryan P.B. Temporal variability of microenvironmental time budgets in Maryland. *J Expos Anal Environ Epidem* 1999: 9: 502–512.

Echols S.L., MacIntosh D.L., and Ryan P.B. Temporal patterns of activities potentially related to pesticide exposure. *J Expos Anal Environ Epidem* 2001: 11: 389–397.

Farrow A., Taylor H., and Golding J. Time in the home by different family members. *Environ Tech* 1997: 18: 605–614.

Felton G.M., and Parsons M.A. Factors influencing physical activity in average-weight and overweight young women. *J Comm Health Nurs* 1994: 11: 109–119.

Folsom A.R., Cook T.C., Sprafka J.M., Burke G.L., Norsted S.W., and Jacobs D.R. Jr. Differences in leisure-time physical activity levels between blacks and whites in population-based samples: the Minnesota Heart Study. *J Behav Med* 1991: 14: 1–9.

Ford E.S., Merritt R.K., Heath G.W., Powell K.E., Washburn R.A., Kriska A., and Haile G. Physical activity behaviors in lower and higher socioeconomic populations. *Am J Epidemiol* 1991: 133: 1246–1256.

Fortier M.D., Katzmarkky P.T., Malina R.M., and Bouchard C. Seven-year stability of physical activity and musculoskeletal fitness in the Canadian population. *Med Sci Sports Exerc* 2001: 33: 1905–1911.

Freeman N.C.G., Lioy P.J., Pellizzari E., Zelon H., Thomas K., Clayton A., and Quackenboss J. Responses to the Region 5 NHEXAS time/activity study. *J Expes Anal. Environ Epidemiol* 1999: 9: 414–426.

Goldstein B.R., Tardiff G., Hoffnagle G., and Kester R. *Valdez Health Study: Summary Report*. Alyeska Pipeline Service Company, Anchorage, AK, 1992.

Goodchild M., and Jannelle D. *The City Around the Clock: Space–Time Patterns of Urban Ecological Structure*. Department of Geography, University of Western Ontario, London, Ontario, 1984.

Goran M.I., Carpenter W.H., McGloin A., Johnson R., Hardin J.M., and Weinsier R.L. Energy expenditure in children of lean and obese parents. *Am J Physiol* 1995: 268: E917–E924.

Goran M.I., Gower B.A., Nagy T.R., and Johnson R.K. Developmental changes in energy expenditure and physical activity in children: evidence for a decline in physical activity in girls before puberty. *Pediatrics* 1998: 101: 887–891.

Graham S.E., and McCurdy T. (2003). Developing meaningful cohorts for human exposure models. *J Expos Anal Environ Epidemiol*, accepted.

Hartwell T.D., Clayton C.A., Michie R.M., Whitmore R.W., Zelon H.S., Jones S.M., and Whitehurst D.A. *A Study of Carbon Monoxide Exposure of Residents of Washington, D.C. and Denver, Colorado*. US Environmental Protection Agency, Research Triangle Park, NC, 1984, EPA-600/S4-84-031.

Heath G.W., Pratt M., Warren C.W., and Kann L. Physical activity patterns in American high school students. *Arch Pediatr Adolesc Med* 1994: 148: 1131–1136.

Henderson K.A., Bialeschki M.D., Shaw S.M., and Freysinger V.J. *A Leisure of One's Own*. Venture Publishing. State College, PA, 1989.

Hovell M.F., Sallis J.F., Kolody B., and McKenzie T.L. Children's physical activity choices: a developmental analysis of gender, intensity levels, and time. *Pediatr Exerc Sci* 1999: 11: 158–168.

Johnson T. *A Study of Personal Exposure to Carbon Monoxide in Denver, Colorado*. US Environmental Protection Agency, Research Triangle Park NC, 1984, EPA 600/S4-84-014.

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

Johnson T. *Human Activity Patterns in Cincinnati, Ohio*. Electric Power Research Institute, Palo Alto, CA, 1989.

Johnson T. Recent advances in the estimation of population exposure to mobile source pollutants. *J Expos Anal Environ Epidemiol* 1995: 5: 551–571.

Johnson T., Capel J., and McCoy M. *An Analysis of Ten Time/Activity Databases: The Effects of Selected Facvtors on the Apportionment of Time Among Microenvironments and Breathing Rate Categories*. IT Corporation, Durham, NC, 1995.

Johnson T., McCoy M., Capel J.E., Wijnberg L., and Ollison W. A Comparison of Ten Time/Activity Databases: Effects of Geographic Location, Temperature, Demographics Group, and Diary Recall Method. Paper presented at AWMA International Conference on Tropospheric Ozone Nonattainment and Design Value Issues, October 1992.

Johnson T., Mihlan G., LaPointe J., Fletcher K., and Capel J. *Estimation of Carbon Monoxide Exposures and Associated Carboxyhemoglobin Levels in Denver Residents Using pNEM/CO (Version 2.0)*. TRJ Environmental, Chapel Hill, NC, 1999.

Kann L., Warren W., Collins J.L., Ross J., Collins B., and Kolbe L.J. Results from the National School- based Youth Risk Behavior Survey and progress toward achieving related health objectives for the nation. *Public Health Rep* 1993: 108(Suppl. 2): 47–55.

Kann L., Warren C.W., Harris W.A., Collins J.L., Williams B.I., Ross J.G., and Kolbe L.J. Youth Risk Behavior Surveillance — United States, 1995. *J School Health* 1996: 66: 365–377.

Kelder S.H., Perry C.L., Klepp K.-I., and Lytle L.L. Longitudinal tracking of adolescent smoking, physical activity, and food choice behaviors. *Am J Public Health* 1994: 84: 1121–1126.

Kelder S.H., Perry C.L., Peters R.J. Jr., Lytle L.L., and Klepp K.-I. Gender differences in the Class of 1989. Study The school component of the Minnesota Heart Health Program. *J Health Educ* 1995: 26(Suppl): 36–44.

Kimiecik J.C., and Horn T.S. Parental beliefs and children's moderate-to-vigorous physical activity. *Res Q Exerc Sport* 1988: 69: 163–175.

Kimiecik J.C., Horn T.S., and Shurin C.S. Relationships among children's beliefs, perceptions of their parents' beliefs, and their moderate-to-vigorous physical activity. *Res Q Exerc Sport* 1996: 67: 324–336.

Klepeis N.E., Nelson W.C., Ott W.R., Robinson J.P., Tsang A.M., Switzer P., Behar J.V., Hern S.C., and Engelmann W.H. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *J Expos Anal Environ Epidemiol.* 2001: 11: 231–252.

Klepeis N., Tsang A., and Behar J.V. *Analysis of the National Human Activity Pattern Survey (NHAPS) Respondents from a Standpoint of Exposure Assessment*. National Exposure Research Laboratory, US Environmental Protection Agency, Las Vegas, NV, 1996, EPA-600-R-96/074.

Kohl H.W., LaPorte R.E., and Blair S.N. Physical activity and cancer. *Sports Med* 1988: 6: 222–237.

Lauderdale D.S., Fabsitz R., Meyer J.M., Sholinsky P., Ramakrishnan V., and Goldberg J. Familial determinants of moderate and intense physical activity: a twin study. *Med Sci Sports Exerc* 1997: 29: 1062–1068.

Lee M.M., Whittemore A.S., and Jung D.L. Reliability of recalled physical activity, cigarette smoking, and alcohol consumption. *Ann Epidemiol* 1992: 2: 705–714.

Leech J.A., Nelson W.C., Burnett R.T., Aaron S., and Raizenne M.E. It's about time: A comparison of Canadian and American time-activity patterns. *J Expos Anal Environ Epidemiol* 2002: 12: 427–432.

Leslie E., Fotheringham M.J., Owen N., and Bauman A. Age-related differences in physical activity levels of young adults. *Med Sci Sports Exer* 2001: 33: 255–258.

Levin S., Jacobs D.R. Jr., Ainsworth B.E., Richardson M.T., and Leon A.S. Intra-individual variation and estimates of usual physical activity. *Ann Epidemiol* 1999: 9: 481–488.

MacIntosh D.L. *Refinements to EPA/NERL's Aggregate SHEDS-Pesticide Model.* EPA report, RTP, NC, 2001.

Maffeis C., Pinelli L., Zaffanello M., Schena F., Iacumin P., and Schutz Y. Daily energy expenditure in free-living conditions in obese and non-obese children: Comparison of doubly labelled water ($^2H_2{}^{18}O$) method and heart-rate monitoring. *Int J Obes* 1995: 19: 671–677.

Matthews C.E., Hebert J.R., Freedson P.S., Stanek III E.J., Merriam P.A., Ebbeling C.B., and Ockene I.S. Sources of variance in daily physical activity levels in Seasonal Variation of Blood Cholesterol Study. *Am J Epidemiol* 2001: 153: 987–995.

McCrory M.A., Kim H.-R., Wright N.C., Lovelady C.A., Aitkens S., and Kilmer D.D. Energy expenditure, physical activity, and body composition of ambulatory adults with hereditary neuromuscular disease. *Am J Clin Nutr* 1998: 67: 1162–1169.

McCurdy T. Estimating human exposure to selected motor vehicle pollutants using the NEM series of models: Lessons to be learned. *J Expos Anal Environ Epidemiol* 1995: 5: 533–550.

McCurdy T. Conceptual basis for multi-route intake dose modeling using an energy expenditure approach. *J Expos Anal Environ Epidemiol* 2000: 10: 86–97.

McCurdy T., Glen G., Smith L., and Lakkadi Y. The National Exposure Research Laboratory's Consolidated Human Activity Database. *J Expos Anal Environ Epidemiol* 2000: 10: 566–578.

Moore M.E., Pond J., and Korslund M.K. Energy expenditure of pre-adolescent girls. *J Am Diet Assoc* 1966: 49: 409–412.

Ng A.V., and Kent-Braun Quantification of lower physical activity in persons with multiple sclerosis. *Med Sci Sports Exerc* 1997: 4: 517–523.

Pate R.R., Baranowski T., Dowda M., and Trost S.G. Tracking of physical activity in young children. *Med Sci Sports Exerc* 1996: 28: 92–96.

Perusse L., Tremblay A., Leblanc C., and Bouchard C. Genetic and environmental influences on level of habitual physical activity and exercise participation. *Am J Epidemiol* 1989: 129: 1012–1022.

Potvin L., Gauvin L., and Nguyen N.M. Prevalence of stages of change for physical activity in rural, suburban, and inner-city communities. *J Community Health* 1997: 22: 1–13.

Prentice A.M., Black A.E., Coward W.A., and Cole T.J. Energy expenditure in overweight and obese adults in affluent societies: an analysis of 319 doubly-labelled water measurements. *Euro J Clin Nutr* 1996: 50: 93–97.

Quackenboss J.J., and Lebowitz M.D. The utility of time/activity data for exposure assessment: summary of procedures and research needs. Paper presented at the 82nd Annual Meeting of the Air and Waste Management Association; June 1989.

Ransdell L.B., and Wells C.L. Physical activity in urban white, African-American, and Mexican-American women. *Med Sci Sports Exerc* 1998: 30: 1608–1615.

Ribeyre J., Fellmann N., Montaurier C., Delaitre M., Vernet J., Coudert J., and Vermorel M. Daily energy expenditure and its main components as measured by whole-body indirect calorimetry in athletic and non-athletic adolescents. *Br J Nutr* 2000: 83: 355–362.

Rieper H., Karst H., Noack R., and Johnsen D. Intra- and inter-individual variations in energy expenditure of 14–15-year-old schoolgirls as determined by indirect calorimetry. *Br J Nutr* 1993: 69: 29–36.

Robinson J.P. *How Americans Use Time*. Praeger Publishers, New York, 1977.

Robinson J.P., and Godbey G. *Time for Life*. The Pennsylvania State University Press, University Park, PA, 1999.

Sallis J.F., Berry C.C., Broyles S.L., McKenzie T.L., and Nader P.R. Variability and tracking of physical activity over 2 yr in young children. *Med Sci Sports Exerc* 1995: 27: 1042–1049.

Sallis J.F., McKenzie L.T., Elder J.P., Hoy P.L., Galati T., Berry C.C., Zive M.M., and Nader P.R. Sex and ethnic differences in children's physical activity: Discrepancies between self-report and objective measures. *Pediatr Exerc Sci* 1998: 10: 277–284.

Sallis J.F., Patterson T.L., Buono M.J., Atkins C.J., and Nader P.R. Aggregation of physical activity habits in Mexican-American and Anglo families. *J Behav Med* 1988: 11: 31–41.

Sallis J.F., Prochaska J.J., and Taylor W.C. A review of correlates of physical activity of children and adolescents. *Med Sci Sports Exerc* 2000: 32: 963–975.

Schoeller D.A., and Fjeld C.R. Human energy metabolism: what have we learned from the doubly labeled water method? *Ann Rev Nutr* 1991: 11: 355–373.

Schoeller D.A., and Hnilicka J.M. Reliability of the doubly labeled water method for the measurement of total daily energy expenditure in free-living subjects. *J Nutr* 1996: 126: 348S–354S.

Schwab M., Colomé S.D., Spengler J.D., Ryan P.B., and Billick I.H. Activity patterns applied to pollutant exposure assessment: data from a personal monitoring study in Los Angeles. *Toxicolus Ind Health* 1990: 6: 517–532.

Schwab M., McDermott A., and Spengler J.D. Using longitudinal data to understand children's activity patterns in an exposure context: data from the Kanawha County health study. *Environ Int* 1992: 18: 173–189.

Schwab M., Terblanche A.P.S., and Spengler J.D. Self-reported exertion levels on time/activity diaries: application to exposure assessment. *J Expox Anal Environ Epidemiol* 1991: 1: 339–356.

Shephard R.J., Jéquier J.-C., Lavallée H., LaBarre R., and Rajic M. Habitual physical activity: Effects of sex, milieu, season and required activity. *J Sports Med* 1980: 20: 55–66.

Simons-Morton B.G., McKenzie T.J., Stone E., Mitchell P., Osganian V., Strikmiller P.K., Ehlinger S., Cribb P., and Nader P.R. Physical activity in a multiethnic population of third graders in four states. *Am J Pub Health* 1997: 87: 45–50.

Sternfeld B., Sidney S., Jacobs D.R. Jr., Sadler M.C., Haskell W.L., and Schreiner P.J. Seven-year changes in physical fitness, physical activity, and lipid profile in the CARDIA Study. *Ann Epidemiol* 1999: 9: 25–33.

Treuth M.S., Butte N.F., and Wong W.W. Effects of familial predisposition to obesity on energy expenditure in multiethnic prepubertal girls. *Am J Clin Nutr* 2000: 71: 893–900.

Trost S.G., Pate R.R., Freedson P.S., Sallis J.F., and Taylor W.C. Using objective physical activity measures with youth: how many days of monitoring are needed? *Med Sci Sports Exerc* 2000: 32: 426–431.

Tsang A.M., and Klepeis N.E. *Descriptive Statistics Tables from a Detailed Analysis of the National Human Activity Pattern Survey (NHAPS) Data*. US Environmental Protection Agency, Las Vegas, NV, 1996, EPA/600/R-96/148.

Tsang A.M., and Klepeis N.E. *Three Telephone Surveys of Human Activities in California: The 1992-94 National Human Activity Pattern Survey; The 1987–88 California Activity Pattern Survey of Adults and Teenagers; and the 1989-90 California Activity Pattern Survey of Children*. US Environmental Protection Agency, Las Vegas, NV, 1997.

Uitenbroek D.G. Seasonal variation in leisure time physical activity. *Med Sci Sports Exerc* 1993: 25: 755–760.

Van Mechelen W., and Kemper H.C.G. Habitual physical activity in longitudinal perspective, in: Kemper H.C.G. (Ed.). *The Amsterdam Growth Study: A Longitudinal Analysis of Health, Fitness, and Lifestyle*. Human Kinetics, Champaign, IL, 1995, pp. 135–158.

Weinsier R.L. et al. Energy expenditure and free-living physical activity in black and white women: comparison before and after weight loss. *Am J Clin Nutr* 2000: 71: 1138–1146.

Welle S., Forbes G.B., Statt M., Barnard R.R., and Amatruda J.M. Energy expenditure under free-living conditions in normal-weight and overweight women. *Am J Clin Nutr* 1992: 55: 14–21.

Wiley J.A., Robinson J.P., Piazza T., Garrett K., Cirksena K., Cheng Y.-T., and Martin G. *Activity Patterns of California Residents*. Survey Research Center, University of California, Berkeley, CA, 1991.

Xue J., McCurdy T., Spengler J., and Özkaynak H. Intra- and inter-individual activity considerations in human exposure modeling. *J Expos Anal Environ Epidemiol,* 2003: submitted.

Zuzanek J., and Smale B.J.A. Life-cycle variations in across-the-week allocation of time to selected daily activities. *Society Leisure* 1992: 15: 559–586.

Appendix

Table A1. Categorical variables used to discriminate among factors affecting human activity participation rates

Important independent variables[a]	General relationship with activity	Example citation(s)[b]
A. Physical activity (PA)		
Age and gender (growth & developmental factors are included in this category also)	Except for infants, PA↓ with age; time, frequency, and intensity of exercise is lower in ♀ than ♂ especially post-puberty	Ainsworth et al. (1993); Butcher (1985); CDC (2000a, b); Dovey et al. (1998); Goran et al. (1998); Heath et al. (1994); Hovell et al. (1999); Kelder et al. (1995); Leslie et al. (2001); Sallis et al. (1995, 2000); Shephard (1980)
Familial characteristics: Genetics (heredity):	Some find that inter-generational PA patterns are associated, and some do not; heritability coeff.'s vary between 0.3 and 0.8; twins are more alike than other siblings	Aarnio et al. (1997); Beunen and Thomis (1999); Goran et al. (1995); Lauderdale et al. (1997); McCrory et al. (1998); Pérusse et al. (1989)
"Milieu": (environment)	Parental values weakly influence children's PA	Dempsey et al. (1993); Kimiecik and Horn (1988); Kimiecik et al. (1996)
	Strong relationship between parental PA participation and children's	Moore et al. (1966)
Educational level	Moderately proportional to *leisure-time* PA	Brooks (1988); CDC (2000b); Crespo et al. (2001)
Income	Weak positive relationship	Brooks (1988); Crespo et al. (2001)

Table A1. (*continued*)

Employment: occupational category	*Leisure-time* PA ↑ for higher "status" occup., but PA ↓ as hours worked per week; work-PA ↑ for blue-collar jobs	Ainsworth et al. (1999); Burton and Terrell (2000); Cunningham et al. (1969)
Social-economic status (not otherwise distinguished)	PA ↑ with SES	Black et al. (1996); Ford et al. (1991)
Race/ethnicity	Weak-to-moderate relationship by races and PA; white people generally are more active than black people, although that finding is not universal; conclusions regarding other races are even more equivocal	Ainsworth et al. (1999); CDC (2000b); Felton and Parsons (1994); Folsom et al. (1991); Randsdell and Wells (1998); Sallis et al. (1988, 1998); Treuth et al. (2000); Weinsier et al. (2000);
Season of the year	PA ↑ in warmer months	Shephard et al. (1980); Uitenbroek (1993)
Degree of urbanization (rural, suburban, urban)	Strong positive relationship between PA & "structural/environmental aspects of comm."	Potvin et al. (1997); Shephard et al. (1980)
Region of the country	PA varies by state	CDC (1987, 2000a); Simons-Morton et al. (1997)
Day of the week	Complex	CDC (2000b); Ribeyre et al. (2000)
Health status, including obesity (BMI) and pregnancy[c]	PA has an inverse relationship with BMI, selected obesity measures, diabetes, multiple sclerosis; and hospital confinement	Bouton et al. (1996); Bullen et al. (1964); Burchfiel et al. (1995); CDC (2000a); Cooper et al. (2000); Maffeis et al. (1995); Ng and Kent-Braun (1997); Prentice et al. (1996); Welle et al. (1992)
Longitudinal stability[d] (from a baseline over a long period); in the exercise physiology/epidemiology literature this is called "tracking"	Weak-to-moderate relationship in general (for numerous age groups even over many years), and it is stronger for vigorous activities (where measured)	Anderssen et al. (1996); Atkins et al. (1997); Bijnen et al. (1998); Hovell et al. (1999); Kelder et al. (1994)
	30–35% remain in lowest (sedentary) and highest (exercisers) quintiles over 7 years (random would be ~20%)	Fortier et al. (2001)
	A 4-year study of middle-age people found: 60% sedentary both times; 12% were "PA-maintainers;" 16% ↑ PA, and 12% ↓ PA	Eaton et al. (1993)
	Rank cc.'s for PA over two 10-year periods (and for the entire per.) ≈0.35–0.40	Lee et al. (1992)
	PA (h week^{-1}) cc's not significant over an 8-year period	Andersen and Haraldsdóttir (1993)
	PA decreases in late adolescence ≈(16–19) for most individuals	Andersen (1994)
	Intra-individual COV ≈9–11% via "objective" measures & 39% by self-report	Levin et al. (1999)
	Sources of variability in PA over a year period (using 15 daily surveys), using multiple measures: Intra-individual: 50–60% Inter-individual: 20–30% Seasonal differences: 6% Day-of-week: 15%	Matthews et al. (2001)
	Over a 3-year period, a PA Index had 1–3 year r_s's ≈0.5–0.6	Pate et al. (1996)
	34% had stable PA over a 7-year period (vs. 23%↓ & 43%↓, which was expected, as PA generally ↓ with aging in adults)	Sternfeld et al. (1999)
	The mean within-subject COV (4 times over a year) for low-EE activities ≈4–7% and 9–14% for high-EE activities	Rieper et al. (1993)
	Within-subject COV due to physical activity alone ≈7%; individual studies show 3–9%	Schoeller and Fjeld (1991); Schoeller and Hnilicka (1996)
	Cross-sectionally, weekly TDEE is higher & varies less in ath. (4.3–6.2% by sex) than in non-ath. (5.1–6.7%); Ex. EE is much higher & varies less in ath. (7.0–16.4%) than in non-ath. (15.0–31.5%)	Ribeyre et al. (2000)

Table A1. (*continued*)

	Moderate tracking over 3–5 years (IPC≈0.32–0.58), but it drops off at 10–15 years.	Van Mechelen and Kemper (1995)
Longitudinal stability (repeated: days per week in *vigorous* events)	*Adolescents:*	
	0: 16%	Rework of CDC (1992, 1996) data; Kann et al. (1993, 1996); see also Aaron et al. (1993, 1995)
	1–2: 20% 3–4: 22% 5–6: 21% 7: 21%	
	Adults:	
	0: 25% ♀ 3% ♂ 1–2: 21% ♀ 14% ♂ 3+: 54% ♀ 86% ♂ (5+ = 56%)	Rework of CDC (1992, 1996) data
Longitudinal stability (misc. heart rate measures, METS/d, or min/d in exercise)	Generally, need between 2 and 7 days of HR data to capture variability in children (various ages) with an rcc of 0.80; only need 1–2 h to get same reliability for HR_{Rest}. For adult ♀ teachers, need 2 weeks of 7-day diaries for a 0.8 rcc.	Baranowski and de Moor (2000); Baranowski et al. (1999)
	For 80% reliability of PA, need 8–9 days of objective monitoring in adolescents 12+, but only need 4–5 in school age children	Trost et al. (2000)
	For 80% annual reliability of **total** PA, need 7–10 days for & 14–21 days for ♀; for leisure-time PA, need 21–28 days	Matthews et al. (2001)
B. Leisure-time activity		
"Life cycle" stage/role[e]	Highly complex for these general characteristics (listed for the sake of completeness)	Henderson et al. (1989); Zuzanek and Smale (1992)
C. General human activities		
"Preconditioning factors": age, sex; life-cycle, esp. child-rearing responsibilities; health status; occupation; income; home type; seasons; day-of-week; vacation status; region (climate)	Highly complex (as above)	Chapin (1974); Goodchild and Jannelle (1984); Farrow et al. (1997); Freeman et al. (1999); Johnson (1995); Johnson et al. (1992, 1995); Klepeis et al. (1996, 2001); Leech et al. (2002); Robinson (1977); Robinson and Godbey (1999); Tsang and Klepeis (1996,1997)
D. Important factors shown to affect activities in environmental-exposure studies		
Weekend/weekday, season	Format: 3 consecutive 24-h diaries 2 times per year for ~600 children using 15 min blocks. No stat. testing results presented.	Adair and Spengler (1989)
Age, season, day-of-week, gender, race, urbanization, education, income	Format: up to 7 24-h diaries 6 times per year for 79 people >6. Time spent at home varies among individuals (demographic characteristics), day-of-the-week, and seasons.	Echols et al. (1999)
Results are focused on pesticide exposures; variables as above	As above	Echols et al. (2001)
Age, day-of-week	*NHEXAS format:* 7 consecutive 24-h diaries for 249 people using 1-h blocks. Daily autocorrelation for some, but not all, activities (episodic vs. regular behaviors).	Freeman et al. (1999)
Weekday/Sat./Sun., age, gender, employment status, commuting time	Markov chain simulation of Cincinnati activity data ($n\approx3000$)	Johnson et al. (1987)
Season, age, working status	*Format:* highly aggregated summary questions (h/d); $n\approx1100$	Quackenboss and Lebowitz (1989)
Meteorological factors	*Format:* detailed development and analyses of numerous meteorological factors using 3 studies ($n\approx4200$)	Johnson et al. (1996)
Age, gender, occupational, winter seasonal, day-of-week diff. found (stat. significant).	*Format:* 2 consecutive 24 h data? times per year for 1283 adults using 30-min block diaries. Within-person variability is large for some location/activity events	Schwab et al. (1990)

Table A1. (*continued*)

Age, school-days *vs*. nonschool days	*Format:* 14 consecutive 24-h data 2 times per year for 62–72 children. Large within- & between-individual variability in travel & outdoor locations, esp. for a subsample of the group	Schwab et al. (1992)

♂ = Male; ♀ = female.
[a]Usually shown to be statistically significant in quantitative analyses, but often is used simply as a qualitative descriptor of organized data (without being tested for statistical difference).
[b]This is not a comprehensive compilation of available sources, but is a sample of important and/or early sources of information on the topic. For the most part, exercise epidemiology studies are not cited, nor are "exercise-intervention" studies. Studies focused on "validating" (evaluating) objective or subjective (questionnaire) measuring instruments also are not used here. "*In vitro*" (experimental) studies are also not reported, only "*in vivo*" or "free-ranging" studies are of interest.
[c]Psychological factors, such as depression and "social physique anxiety" also affect PA participation rates, but these are ignored here. There are a number of studies that look at PA as the independent variable in chronic disease formation (particularly chronic heart conditions or disease); since this reverses the causal order of interest here, they too are ignored.
[d]Studies that obtained longitudinal data using a multi- year retrospective (recall) survey instrument are not included here.
[e]Life cycle is defined as a composite variable functionally combining biological characteristics (age, gender, health, physical vigor) with social role characteristics (marital status, presence of children, occupational pattern) and social-psychological orientations/motivational structures (Zuzanek and Smale, (1992). Five dimensions of life cycle are generally emphasized: age, gender, martial status, presence of children, and employment status.

Abbreviations used:

ath.	Athlete	NCHS	National Center for Health Statistics
BMI	Body mass index	occup.	Occupation(s)
cc	Correlation coefficient	PA	Physical activity
CDC	Centers for Disease Control	per.	Period(s)
coeff	Coefficient	pop.	Population
COV	Coefficient of variation (s.d./mean)	rcc	Reliability correlation coefficient
diff.	Different (difference)	Sat.	Saturday
EE	Energy expenditure	S.D.	Standard deviation
Ex.	Exercise	stat.	Statistical, statistics
esp.	Especially	Sun.	Sunday
IPC	Interperiod correlation coefficient	TDEE	Total daily energy expenditure

Table A2. Development of the sample of 57–60-year-old males for the comparative activity analyses.

Item	Number of records involved	Study[a]
Total number of CHAD records for the age/gender group	245	California: 33; Cincinnati: 44; Denver: 6; NHAPS-A: 65; NHAPS-W: 87; Valdez: 1; DC: 9
Person-days with <30 records	32	California: 4; NHAPS-A: 8; NHAPS-W: 19; Valdez: 1
Outdoor workers or deliverymen	15	California: 2; Cincinnati: 4; NHAPS-A: 7; NHAPS-W: 2
Unemployed worker on a weekday	1	NHAPS-W: 1
Person on a boat all day	1	NHAPS-A: 1
Works at home (and it was a weekday)	8	California: 1; Cincinnati: 3; NHAPS-A: 1; NHAPS-W: 2; DC: 1
Missing both the education & occupation codes	11	Cincinnati: 8; Denver: 1; NHAPS-A: 1; DC: 1
Unsuitable coding[b]	8	Cincinnati: 3; Denver: 2; DC: 3
The analyzed sample	169	California: 26; Cincinnati: 26 (from 10 individuals); Denver: 3 (from 2 individuals); NHAPS-A: 47; NHAPS-W: 63; DC: 4

[a]Studies used in the comparative analyses are all contained in CHAD; they are described in McCurdy et al. (2000).
[b]More than 4 h with missing or blatantly incompatible location and activity codes (e.g., 4 h of child care in a garage; no sleep (all activities inside the home were coded as: "general household activities," and similar problems). Activity coding problems played a part in this winnowing procedure because eventually this sample will be used to estimate dose received from an exposure, and that is greatly affected by activity-level, itself directly related to the activity being undertaken. See McCurdy (2000) for an in-depth discussion of this point.

Table A3. Day-types used for the longitudinal activity diary subject.

Day-type	Number of days by type	Percent of category day[a]	Number of altered-activity days	Per cent of day-type days
Weekdays				
1. At the workplace	209	78.9	11	5.3
2. Work-related trip (Mexico City)	3	1.1	3	100.0
3. Sick or injured; not at work	11	4.2	11	100.0
4. Regional disaster: workplace closed	5	1.9	1	20.0
5. Day off of work, but stayed in the region[b]	17	6.4	1	5.9
6. Day off of work, not in the region[c]	19	7.2	0	0.0
Total weekdays	264	100.0	26	9.8
Weekends				
7. Saturday, nonvacation	46	44.2	4	8.7
8. Sunday, nonvacation	45	43.3	6	13.3
9. Sunday, work-related trip	1	1.0	0	0.0
10. Not in the region (Saturday or Sunday)[c]	12	11.5	0	0.0
Total weekend days	104	100.0	10	9.6
Total study days	368		36	9.8

[a]The categories are weekdays (71.7% of the days in the study) and weekends (28.3% of all days).

[b]These days are like a weekend with no vacation-type of travel. They include both annual leave (vacation) days (6) and federal holidays (11).

[c]On vacation generally; the day may be spent on travel or not. One day was spent on hurricane disaster relief work; it definitely was not a vacation day.