HEI HEALTH EFFECTS INSTITUTE

Statistical Methods for Epidemiologic Studies of the Health Effects of Air Pollution

William Navidi, Duncan Thomas, Bryan Langholz, and Daniel Stram

University of Southern California School of Medicine, Los Angeles, CA

Includes the Commentary of the Institute's Health Review Committee

Research Report Number 86
May 1999

HEI HEALTH EFFECTS INSTITUTE

The Health Effects Institute, established in 1980, is an independent and unbiased source of information on the health effects of motor vehicle emissions. HEI studies all major pollutants, including regulated pollutants (such as carbon monoxide, ozone, nitrogen dioxide, and particulate matter), and unregulated pollutants (such as diesel engine exhaust, methanol, and aldehydes). To date, HEI has supported more than 200 projects at institutions in North America and Europe.

Typically, HEI receives half its funds from the U.S. Environmental Protection Agency and half from 28 manufacturers and marketers of motor vehicles and engines in the United States. Occasionally, funds from other public or private organizations either support special projects or provide resources for a portion of an HEI study. Regardless of funding sources, HEI exercises complete autonomy in setting its research priorities and in reaching its conclusions. An independent Board of Directors governs HEI. The Institute's Research and Review Committees serve complementary scientific purposes and draw distinguished scientists as members. The results of HEI-funded studies are made available as Research Reports, which contain both the Investigators' Report and the Review Committee's evaluation of the work's scientific quality and regulatory relevance.

HEI Statement

Synopsis of Research Report Number 86

Statistical Methods for Epidemiologic Studies of the Health Effects of Air Pollution

BACKGROUND

Using epidemiologic approaches to determine the health risks of exposure to air pollutants is challenging: it is difficult to measure exposure to the relatively low levels of pollutants to which people are generally exposed, and to find populations with different degrees of pollutant exposure but comparable exposure to potentially confounding factors. Furthermore, the frequently high correlation among different pollutants makes it difficult to identify the effect of a single agent. In the early 1990s the Health Effects Institute (HEI) supported an Environmental Epidemiology Planning Project to identify the methodological issues that needed to be addressed in future studies of the health effects of air pollutants (HEI Communications Number 3, 1994). HEI also funded research targeted to the development of epidemiologic methods in this area of air pollution research and, when possible, to the testing of these methods using appropriate data sets. HEI funded the study conducted by Dr. William Navidi and colleagues to develop statistical methods that would improve the estimates of the health effects air pollution obtained from epidemiologic studies. One important feature of this study was that the investigators were able to test the statistical models they developed using data from the University of Southern California Children's Health Study of the long-term effects of air pollutants on children.

STATISTICAL METHODS AND CRITIQUE

In this report, Navidi and colleagues discussed the development of three sophisticated statistical methods and their application to air pollution epidemiologic studies. First, they took a standard case-crossover design (an approach to assessing the effect of transient exposures on the risk of onset of an acute event) and introduced a bidirectional element where control data were obtained both before and after the health event of interest. Navidi and colleagues showed that their bidirectional case-crossover design method provided a better estimate than the standard case-crossover design and illustrated the potential bias in the latter approach. The use of the bidirectional case-crossover promises to be an advance over the use of the case-crossover design and offers promise for reducing bias in environmental epidemiologic studies. More investigation is needed, however, into possible problems, such as sensitivity to the mathematical model employed and "carryover effects"—the individual behaving differently after the health event.

Second, because measurement error (the difference between true and measured exposures) can have a substantial impact on the accuracy of estimated health effects, Navidi constructed a model to evaluate the reliability of two approaches to estimating cumulative exposure to air pollutants. The investigators imply that neither the direct nor the indirect approach to measuring exposure gave reliable answers when uncorrected measurement errors were present. This conclusion appears justified based on the simple model the investigators used. The investigators believe that more elaborate models may correct themselves and thus result in reliable estimates of cumulative exposure, but it is not clear how this will occur. Additional modeling and simulations are needed to make any definite conclusions. Certainly more basic research is needed to understand if the two approaches can be shown capable of estimating cumulative exposures and, if so, how well.

Third, the investigators adapted a multilevel analytic design to air pollution epidemiology. This design has the potential to combine individual and group level comparisons. Such designs, which have been used for many years in the social sciences and are now being applied in health effects fields, are undergoing intense development. Navidi has made an advance in this area by identifying the components needed to model the health effects of air pollution and by developing reasonable mathematical models. Further work is needed to determine the sensitivity of the investigators' multilevel model to model assumptions and bias before it can be recommended for general use.

COMMENTS

Navidi and his collaborators have made advances in statistical methodology. Their main contribution has been less in the development of new methods than in applying and extending methods in use in other areas—particularly the bidirectional case-crossover design and multilevel analytic design—to the air pollution health effects field. Some issues remain unresolved; applications of the investigators' methods to exposure problems that include careful model development and validation are needed to resolve these questions.

This Statement, prepared by the Health Effects Institute and approved by its Board of Directors, is a summary of a research project sponsored by HEI from March 1993. This study was conducted by Drs. William Navidi, Duncan Thomas, Bryan Langholz, and Daniel Stram of the University of Southern California, Los Angeles, CA. The following Research Report contains both the detailed Investigators' Report and a Commentary on the study prepared by the Institute's Health Review Committee.

Library of Congress Catalog Number for the HEI Research Report Series: WA 754 R432.
The paper in this publication meets the minimum standard requirements of the ANSI Standard Z39.48-1984 (Permanence of Paper) effective with Report Number 21, December 1988, and with Report Numbers 25, 26, 32, 51, and 65 Parts IV, VIII, and IX excepted. These excepted Reports are printed on acid-free coated paper.

TABLE OF CONTENTS

Research Report Number 86

Statistical Methods for Epidemiologic Studies of the Health Effects of Air Pollution

William Navidi, Duncan Thomas, Bryan Langholz, and Daniel Stram

(Continued on next page)

Statistical Methods for Epidemiologic Studies of the Health Effects of Air Pollution

William Navidi, Duncan Thomas, Bryan Langholz, and Daniel Stram

This Investigators' Report was produced with a different page layout than other Research Reports because of the incompatibility of the program used to create the equations with our desktop publishing software. To prevent errors being introduced in the process of recreating equations and to ensure that the mathematical characters set down by the authors maintained their precision and authenticity, we used camera-ready copy produced by the authors and had them stripped into the pages by our printer.

The Abstract begins on page 2.

* A list of abbreviations appears at the end of the Investigators' Report.

This Investigators' Report is one part of Health Effects Institute Research Report Number 86, which also includes a Commentary by the Health Review Committee, and an HEI Statement about the research project. Correspondence concerning the Investigators' Report may be addressed to Dr. William Navidi, Colorado School of Mines, Department of Mathematical and Computer Science, Golden, CO 80401.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award R824835 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

0 Abstract

We describe two statistical designs that can provide efficient estimates of the health effects of exposure to air pollutants in epidemiologic studies. We also evaluate the effects of measurement error in exposure assessment on the accuracy of estimated health effects.

The bidirectional case-crossover design is a variant of a method proposed by Maclure (1991). Our version of the method takes advantage of the fact that in epidemiologic studies involving environmental exposure, accurate information about past exposure is more readily available, and that levels of exposure are generally unaffected by the response of the subject. It differs from other case-crossover methods in that control information is assessed both before and after failure, thus avoiding confounding due to time trends in exposure.

The multilevel analytic design provides a method of combining estimates of health effects made on the individual level with those made at the group level. It has great potential value in situations where variations in exposure within groups may not be great enough to provide adequate power to detect health effects, as is often the case in air pollution studies where exposure levels are similar within a geographic community.

Measurement errors in exposure assessment can have substantial impact on the accuracy of estimated health effects. When the microenvironmental approach is used to estimate exposure, a standard error of 30% in estimating indoor/outdoor ratios can increase the standard error of a relative risk estimate by 50%, and introduce bias as well. Similar results hold when exposure is estimated with personal samplers. When the microenvironmental approach is used, errors in estimating indoor/outdoor ratios have more influence on the accuracy of risk estimation than do errors in estimating the time spent in microenvironments.

1 Introduction

Fundamental challenges common to many environmental epidemiological studies include (1) difficulties in measuring exposure, disease, and related variables, (2) the problem of finding populations with different degrees of exposure that are comparable with respect to potentially confounding factors, and (3) the difficulty of distinguishing effects at the individual level, often as a result of the spatially clustered nature of many environmental exposures.

We have addressed the methodological implications of these challenges in several ways. We studied an epidemiologic design that combines two approaches to studying exposure-disease associations, the analytic approach based on comparisons between individuals (for example, case-control and cohort studies), and the ecologic approach based on comparisons between groups. The analytic approach generally provides a stronger basis

for inference, in part because of freedom from between-group confounding and better quality data, but the ecologic approach is less susceptible to attenuation bias from measurement error and may provide greater variability in exposure. The design we propose entails the selection of a number of groups and enrollment of individuals within each group. Exposures, outcomes, confounders, and modifiers are to be assessed on each individual, but additional exposure data might be available on the groups. The analysis then combines both the individual-level and the group-level comparisons, with appropriate adjustments for exposure measurement errors, and tests for compatibility between the two levels of analysis, for example to determine whether the associations at the individual level can account for the differences in disease rates between groups. We also studied the implications for design efficiency in terms of trade-offs between numbers of groups, numbers of individuals, and the extent of the individual and group measurement protocols.

We have studied the impact of measurement error in assessing individual exposure to air pollutants on the accuracy of the estimation of health effects. We have considered the "microenvironmental" approach, the personal sampler (badge) approach, and an approach in which each individual's exposure was estimated to be the ambient outdoor level measured at a monitoring station. In the microenvironmental approach, the region occupied by an individual during the time period of interest is divided into a number of "microenvironments", e.g. home, school, office, car, outdoor areas, etc. To estimate an individual's cumulative exposure one estimates the concentrations of the pollutant of interest and the times spent by the individual in each microenvironment, then computes the appropriate time-weighted average. In the personal sampler approach, each individual wears a badge that traps or reacts with the pollutant to provide an estimate of the cumulative exposure. We have gained considerable insight into the association between measurement error in exposure assessment and bias in the estimation of health effects. This should be useful in the development of validation studies whose results can be used to improve the accuracy of these estimates.

We have also studied the case-crossover design (Maclure, 1991) for examining the acute effects of exposure to air pollutants. This is a design in which only cases are sampled, and their exposure at the time of their failure is compared with some estimate of their typical level of exposure. The definition of "failure" varies from study to study. For example, in the USC Children's Health Study, failure is defined as a day of absence from school. In studies of the relationship between particulate exposure and mortality, failure refers to death. Mittleman, Maclure, and Robins (1995) discuss Maclure's method and some variations on it. In all of the methods they describe, all control information is assessed prior to failure. Sampling all control information prior to failure makes the case-crossover design subject to bias from time trends in exposure. While time trends in exposure can be quite pronounced for environmental exposures, studying the effects of environmental exposures offers two compensating advantages. First, accurate information is often available about the levels of past exposure. Second, exposure after an event is unaffected by the occurrence of the event. We describe a modification of the case-crossover approach that makes use of both of these advantages. Our approach involves assessing control

information both before and after failure, which avoids confounding due to time trends.

We have applied some of these ideas to the University of Southern California Children's Health Study. This is an ongoing study of the health effects of air pollution in Southern California, in which 12 communities with different levels and types of pollution have been selected and more than 3500 school children have been enrolled in a ten-year cohort study. In this study, exposure assessment protocols involve a combination of ambient monitoring, microenvironmental sampling, personal sampler monitoring, and questionnaire data on time-activity and household characteristics. Outcomes of interest include several measurements of pulmonary function, as well as morbidity outcomes as reported on questionnaires and as indicated by school absences.

2 Specific Aims

This study had three specific aims: (1) to develop efficient statistical designs for panel studies, (2) to develop methods for allowing for the effects of exposure measurement error, and (3) to develop methods for combining individual and aggregate level comparisons.

Our work on the bidirectional case-crossover design addressed the first aim by presenting a statistical design that has proven successful in analyzing acute health effects of exposure to air pollutants in a major epidemiologic study. We also addressed this aim in our work on the multilevel analytic design, by measuring the impact of design decisions involving trade-offs between numbers of groups, numbers of individuals, and the extent of the individual and group measurement protocols.

We addressed the second aim in two ways. We developed mathematical models for measurement error that allow us to compare the accuracy of the indirect or microenvironmental approach to exposure assessment with the direct or personal sampling approach. We also assessed the impact of these errors on a model that describes the effect of exposure on a health outcome.

We addressed the third aim by developing the multilevel analytic design. A basic feature of this design is that it can be applied at the individual level, by studying the within-group differences, or at the aggregate level, by studying the group averages. The within-groups and the between-groups estimates can be combined to yield a pooled estimator that is more efficient than either of the two estimates separately.

3 The Bidirectional Case-crossover Approach

3.1 Introduction

The case-crossover design (Maclure, 1991) is a design in which only cases are sampled, and their exposure at the time of their failure is compared with some estimate of their typical level of exposure. This design has the advantage that by comparing exposures within subjects, all time invariant confounders such as sex, race, diet, age (over short periods), and community of residence are automatically controlled for.

Data analysis in case-crossover studies is done by standard case-control methods. The basic principle is to estimate risk by comparing the exposure of the subject during a time interval just before failure (the case interval) with the exposure during one or more prior time periods (control intervals). Conditional logistic regression is generally used, with the Mantel-Haentzel approximation an option when the exposure measurement is binary.

In crossover experiments, it is common to randomize the order of treatment, so that in a clinical trial, for example, approximately equal numbers of subjects receive treatment followed by placebo as the other way around. Randomization prevents the order of treatment from being confounded with other risk factors. In observational studies this is of course impossible. The fixed sequence of exposures, combined with the fact that the case exposure is always sampled last, makes the case-crossover design subject to bias from time trends in exposure. Greenland (1996) points out that this is a form of selection bias, since it tends to cause controls to be systematically either more or less exposed than cases. In fact, it is the same sort of bias that could occur in a standard case-control study if the controls were selected earlier in time than the cases.

A number of variations of the case-crossover design have been proposed, involving different strategies for selecting control information. Mittleman, Maclure, and Robins (1995) describe methods in which one or more control intervals are selected at specified time lags before the event, which provide analogs to a one-to-one or many-to-one matched case-control study. Alternatively, one may estimate the proportion of time exposed during a longer interval (e.g., one year) before the event. This is the approach used by Maclure (1991), and referred to by Mittleman, Maclure, and Robins as the "Usual Frequency Approach." Marshall and Jackson (1993) describe a version of this approach in which exposure could be continuous rather than binary. In all of these approaches, secular trends in exposure are assumed to be absent, although Mittleman, Maclure, and Robins (1995) describe several methods for adjusting for cyclical trends. Feldman (1993ab) describes a method for estimating acute and transient risks associated with intermittent exposure. It is designed for binary exposure and for non-terminal outcomes (i.e., outcomes other than death). It does not automatically adjust for individual susceptibility factors, but instead requires that baseline risk be correctly modeled as a function of measured covariates.

Time trends can be particularly strong for environmental exposures such as air pollutants.

Thus standard case-crossover methods can be severely biased. However, studying the effects of environmental rather than behavioral exposures offers two advantages. First, fairly accurate information about past levels of exposure is often available. Second, levels of exposure are unaffected by failures of subjects. Therefore it is possible to determine at times post-failure what the level of exposure would have been had a subject not failed. The case-crossover approach can be adapted to make use of these advantages by assessing control information from times both before and after failure. We refer to this method as the *bidirectional* case-crossover method, since control information is assessed in both directions from the failure time. This allows consistent estimators of risk to be computed, regardless of time trends in exposure.

3.2 The Case-crossover Design for Environmental Exposure

3.2.1 Methods for School Absence Data

We have applied the bidirectional case-crossover design to study the relationship between school absence and pollution levels in the USC Children's Health Study (Peters, 1996). We describe an approach that is appropriate for this situation. For each subject i, let $t_1, ..., t_M$ be the set of days at risk, that is, the set of days that school is in session for that subject, and let X_{ij} be the exposure of subject i on day t_j. The quantity X_{ij} may be multivariate, containing levels of several pollutants at several lags, as well as values of potential confounders. Assume the log odds of absence for subject i on day t_j are given by

$$\log \frac{p_{ij}}{1 - p_{ij}} = \lambda_i + \beta X_{ij}. \tag{1}$$

where λ_i is the baseline level specific to subject i. Then

$$p_{ij} = \frac{e^{\lambda_i} e^{\beta X_{ij}}}{1 + e^{\lambda_i} e^{\beta X_{ij}}}. \tag{2}$$

We assume for simplicity that absences on distinct days are independent. While this is not strictly true, it is nearly so for days far apart in time. With a reasonably large amount of data, this assumption should be satisfactory. Assume that subject i is absent on exactly n_i days, and let A_i denote the set of days on which he is absent. It follows that the probability that a subject is absent precisely on these n_i days is

$$\mathrm{P}(A_i) = \left(\prod_{t_j \in A_i} p_{ij} \right) \left(\prod_{t_k \notin A_i} (1 - p_{ik}) \right). \tag{3}$$

Conditional on the number of absences being n_i, the probability that the absences occurred

precisely on those days in A_i is

$$\mathrm{P}(A_i|n_i) = \frac{\left(\prod_{t_j \in A_i} p_{ij}\right)\left(\prod_{t_k \notin A_i}(1-p_{ik})\right)}{\sum_{S \in D_{n_i}} \left(\prod_{t_j \in S} p_{ij}\right)\left(\prod_{t_k \notin S}(1-p_{ik})\right)}, \tag{4}$$

where D_{n_i} denotes the collection of all sets of n_i days. Substituting (2) into (4) yields

$$\mathrm{P}(A_i|n_i) = \frac{e^{\beta \sum_{t_j \in A_i} X_{ij}}}{\sum_{S \in D_{n_i}} e^{\beta \sum_{t_k \in S} X_{ik}}}. \tag{5}$$

Let $X_{A_i} = \sum_{t_j \in A_i} X_{ij}$, and for each subject i and each set of n_i days S, let $G_{iS} = \sum_{t_k \in S} X_{ik}$.

Then

$$\mathrm{P}(A_i|n_i) = \frac{e^{\beta X_{A_i}}}{\sum_{S \in D_{n_i}} e^{\beta G_{iS}}}. \tag{6}$$

The likelihood function is found by summing the logarithm of (6) over all subjects:

$$L(\beta) = \sum_i \left[\beta X_{A_i} - \log \sum_{S \in D_{n_i}} e^{\beta G_{iS}}\right]. \tag{7}$$

We estimate β by maximizing the right hand side of (7). The Fisher information is the negative of the second derivative of (7). In the case where β is univariate, this is given by

$$I(\beta) = \sum_i \frac{(\sum_{S \in D_{n_i}} G_{iS}^2\, e^{\beta X_{iS}})(\sum_{S \in D_{n_i}} e^{\beta X_{iS}}) - (\sum_{S \in D_{n_i}} G_{iS}\, e^{\beta X_{iS}})^2}{(\sum_{S \in D_{n_i}} e^{\beta X_{iS}})^2}. \tag{8}$$

The variance of the maximum likelihood estimate of β ($\hat{\beta}$) is estimated by substituting $\hat{\beta}$ for β in (8) and inverting.

The number of sets to be summed over in the denominator of (5) is $\frac{M!}{n_i!(M-n_i)!}$, where M is the total number of days at risk. This is too large for the sum to be computable. In practice, therefore, we select a few sets at random and sum over these. Results of Langholz and Goldstein (1997) imply that under this approach, equations (5)–(8) remain valid, with sums over the full collection of sets D_{n_i} replaced by sums over the randomly chosen subcollection of sets.

3.2.2 Methods for Mortality Data

In the case of mortality data, the probability that subject i fails at time t_j is given by the proportional hazards model

$$\Lambda_i(t_j) = \lambda_i \exp(\beta X_{ij}). \tag{9}$$

Conditional on having failed during the ascertainment period, the probability that subject i fails at time t_k is

$$p_{ik} = \frac{\exp(\beta X_{ik})}{\sum_{j=1}^{M} \exp(\beta X_{ij})}. \tag{10}$$

The log likelihood function can be derived in a manner analogous to that of the previous section. If the failure time of the ith subject is t_{k_i}, the log likelihood function is

$$L(\beta) = \sum_{i=1}^{n} \log p_{ik} = \sum_{i=1}^{n} \left[\beta X_{ik_i} - \left(\log \sum_{j=1}^{M} \exp(\beta X_{ij}) \right) \right]. \tag{11}$$

3.3 Simulations

We have conducted two simulation studies to evaluate the accuracy of this approach. In the first study, 1000 subjects were followed for 100 days. A univariate exposure X was generated, taking values in the repeated pattern $1, 1.5, 2, 2.5, 3, \ldots$. The true value β of the log relative risk associated with one unit of exposure was taken to be 1, and the baseline e^{λ} was taken to be 0.005 for each subject. In this way the mean number of days absent for each subject was 4.69. Table 1 presents the results. The column labeled "Number of Combinations" gives the number of sets of n_i days used to form sum in the denominator of (5). The column labeled "Mean" gives the average of the 1000 estimates of β, and the column labeled "SD" gives the standard deviation for that sample, which should be quite close to the standard error (SE) of $\hat{\beta}$. To assess the accuracy of the Fisher information estimate of the SE, The column labeled "Nominal SE" (Root Mean Square Standard Error) gives the square root of the average of the 1000 variance estimates computed by inverting the Fisher information number.

The results show that there is no appreciable bias, even when the denominator in (5) is estimated with a single control set. The standard deviation is reduced by increasing the number of control sets, but it is well estimated with the standard Fisher information estimate in all cases.

Our second simulation used actual exposure data collected for the Children's Health Study. We obtained data on concentrations of particulate matter 10 microns or less in diameter

Table 1: Simulation Results for the Bidirectional Case-crossover Design

True Value	Number of Combinations	Mean	SD	Nominal SE
1.000	1	1.006	0.0646	0.0635
1.000	10	1.001	0.0331	0.0332
1.000	100	1.001	0.0261	0.0266

(PM_{10}) for ten Southern California communities for each day from January 1 through December 30, 1994. (Several of the communities had some days with missing information; these days were ignored in the simulation.) For purposes of the simulation, we assumed there were 100,000 subjects at risk in each community on January 1. The probability that a subject in community i would fail on day j was given by (9). The exposure X_{ij} was taken to be the average daily concentration of PM_{10}, in units of $100\mu g/m^3$. We took the baseline hazard λ_i to be 10^{-5} for all subjects in all communities, and the true value of the log odds ratio β to be 0.1. This results in a true relative risk of about 1.1 per $100\mu g/m^3$, which is in good agreement with relative risks of mortality due to particulates as estimated by Schwartz (1994) and Schwartz and Dockery (1992).

Figure 1 shows a plot of the daily average levels of PM_{10} for Mira Loma, California, one of the ten communities. The time trends in this town are representative of most of the others, and of particle pollution in Southern California generally. Levels are highest in the winter and in the summer. The winter peak is related to cloud cover, stagnation, and primary emissions, while the summer peak is a result of photochemistry and emissions.

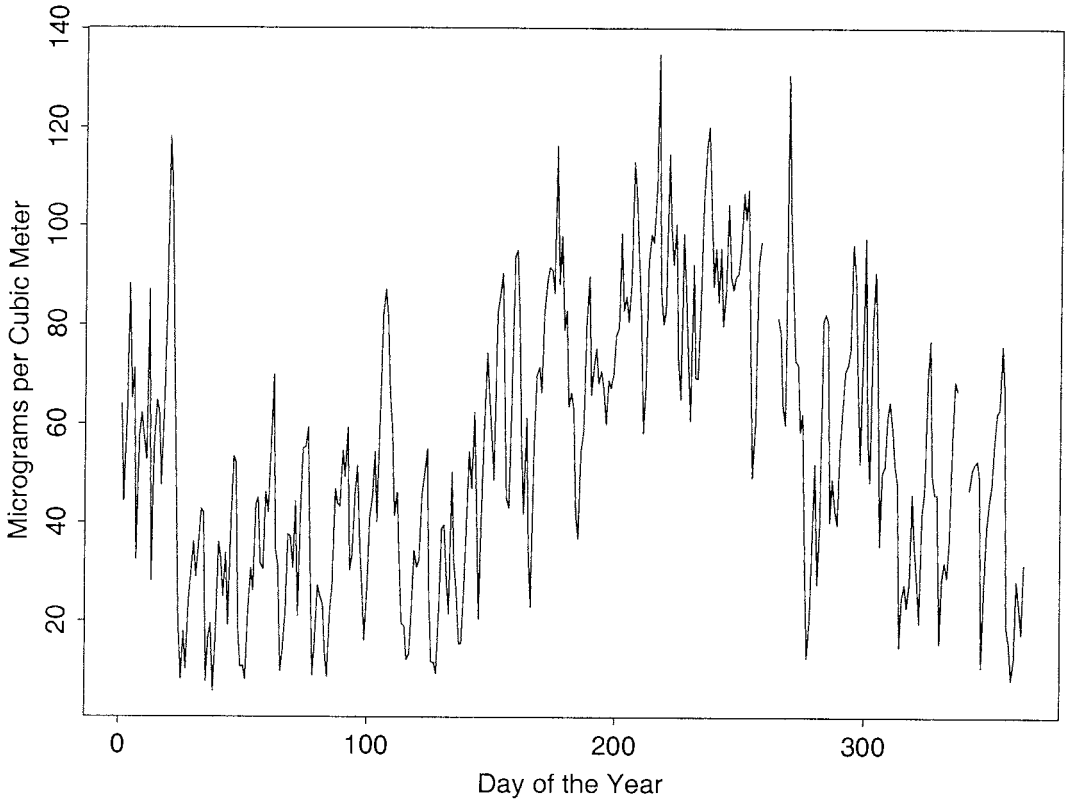

Figure 1: PM_{10} levels in Mira Loma, California, 1994.

We generated 1000 artificial data sets. Each data set was constructed by generating, for each community i and each day j, an observation from a Poisson distribution with mean $N_{ij}\lambda_i e^{\beta X_{ij}}$ where N_{ij} represents the number of subjects at risk in community i on the jth day. For each data set, we computed five different estimates of β. Two of them, $\hat{\beta}_{lag14}$ and $\hat{\beta}_{lag28}$, are conventional case-crossover estimates, each using a single control day for each subject 14 or 28 days before failure, respectively. $\hat{\beta}_{1-1}$ is the bidirectional case-crossover estimate with a single control day for each subject randomly chosen from among the 363 control days. $\hat{\beta}_{TH}$ is a "total history" version of the case-crossover estimator, in which all days prior to failure are used as controls. Finally, $\hat{\beta}_{BI}$ is the bidirectional case-crossover estimator using the full stratum of 363 control days for each subject. For each data set, we computed each of the five estimators of β along with its estimated asymptotic variance. Thus we had 1000 replications of each estimator, along with 1000 replications of its estimated asymptotic variance.

The third column of Table 2 presents, for each estimator, the average of the 1000 replications across the simulated data sets. It can be seen that $\hat{\beta}_{1-1}$ and $\hat{\beta}_{BI}$, the two bidirectional case-crossover estimators, are nearly unbiased, while the others, which use only information prior to failure, are considerably biased. The fourth column presents for each estimator the sample standard deviation of the 1000 replications, which should be a good approximation to the standard error. As expected, estimators using more control information have smaller standard errors. In fact, the standard error of $\hat{\beta}_{1-1}$ is greater than that of $\hat{\beta}_{BI}$ by a factor very close to $\sqrt{2}$, which is in accordance with the rule of thumb for case-control studies that the relative efficiency of a case-control study with M controls per case is $M/(M+1)$ compared to a cohort study. The fifth column of Table 2 presents, for each estimator, the square root of the average of the 1000 replications of its asymptotic variance. The estimated asymptotic variance is seen to be nearly unbiased for the true variance for each of the estimators.

We can obtain some insight into the nature of the bias for $\hat{\beta}_{lag14}$ and $\hat{\beta}_{lag28}$ by examining the time trends in the data more closely. For the lag 14 data, only those cases who fail later than January 14 are considered in the analysis. Their exposures at failure are compared with their exposures 14 days earlier. In all 10 communities taken together, 50.1% of the days after January 14 had lower levels than their corresponding control day. On average, exposure on days after January 14 was 0.08 μg/m^3 lower than on the corresponding control day. Thus there is a slight tendency for control days to have greater exposure. When applied to the 1,000,000 subjects in the simulation, it causes the effect of exposure to be underestimated, as shown in the first line of Table 2.

The corresponding results for the lag 28 data are that 50.3% of days after January 28 had lower levels of exposure than their corresponding control days, with an average difference of 0.26 μg/m^3. The tendency for control days to have greater exposure is more pronounced than in the lag 14 data, and the effect of exposure is more severely underestimated.

Table 2: Comparison of Standard Case-crossover and Bidirectional Case-crossover Estimators

Estimator	True Value	Mean	SD	Nominal SE
$\hat{\beta}_{lag14}$	0.1000	0.0625	0.1707	0.1730
$\hat{\beta}_{lag28}$	0.1000	-0.0220	0.1708	0.1701
$\hat{\beta}_{1-1}$	0.1000	0.0993	0.1330	0.1412
$\hat{\beta}_{TH}$	0.1000	0.4894	0.0710	0.0856
$\hat{\beta}_{BI}$	0.1000	0.0995	0.0934	0.0988

The true value of β is 0.1. $\hat{\beta}_{lag14}$ and $\hat{\beta}_{lag28}$ are conventional case-crossover estimates, each using a single control day for each subject 14 or 28 days before failure, respectively. $\hat{\beta}_{1-1}$ is the bidirectional case-crossover estimate with a single control day for each subject randomly chosen from among the 363 control days. $\hat{\beta}_{TH}$ is a "total history" version of the case-crossover estimator, in which all days prior to failure are used as controls. $\hat{\beta}_{BI}$ is the bidirectional case-crossover estimator using the full stratum of 363 control days for each subject.

3.4 Application to USC Children's Health Study

During the calendar year 1994, a pilot school absence monitoring study was conducted at the University of Southern California. School absences for approximately 3600 Southern California children were recorded. The school absence monitoring study was part of the Children's Health Study, a large longitudinal study of the health effects of air pollution on children. We describe here an analysis in which the bidirectional case-crossover design was used to estimate acute effects of exposure to PM_{10} on the risk of absence from school.

The probability of absence for subject i on day j was modeled with the logistic model (1). The vector of covariates X_{ij} contained the peak PM_{10} level on the previous day, maximum temperature and humidity lagged one and two days, as well as indicators for the month and for the day of the week. This model ignores the fact that absences are autocorrelated, that is, if a subject is absent on a given day, the probability of absence the next day is greatly increased. In the analysis presented here, this was compensated for by treating only the first day of each period of absence as a failure, and discarding subsequent consecutive absences. Days on which the subject attended school were used as control days. Other approaches are possible. For example, one could include indicator functions for absence on the previous day, previous two days, etc., and interactions between these indicators and other covariates. In addition, since there is likely to be a waiting time of some duration between the end of one period of absence and the beginning of the next, one might include a function of the time since the last absence as a covariate.

Risks were estimated using the procedure described in section 3.3. The denominator in (6) was approximated by using a single set of control days as well as the case set. Results of

the analysis are presented in Tables 3 through 6.

Table 3 presents relative risks of absence in relation to peak ambient levels of O_3, PM_{10}, and NO_2 both one and two days prior to the start of the absence. No results are statistically significant, although peak O_3 lagged one day appears somewhat inversely related to absence risk.

Table 3: Relative Risk of Absence in Relation to Recent Exposure to Air Pollutants

Pollutant	Relative Risk per Unit Exposure	95 % C.I.
Peak O_3 lag 1	0.9778	0.9560–1.0001
Peak O_3 lag 2	1.0026	0.9813–1.0244
Peak PM_{10} lag 1	1.0082	0.9987–1.0018
Peak PM_{10} lag 2	0.9917	0.9819–1.0017
Peak NO_2 lag 1	0.9464	0.9273–1.0150
Peak NO_2 lag 2	1.0474	1.0242–1.0711

There were 2486 subjects for whom absences were reported. Adjusted for temperature and humidity lagged one and two days. Units are 10 ppb for O_3 and NO_2, 10 μg / m^3 for PM_{10}.

Table 4 presents relative risks of absence due to respiratory illness. No statistically significant results are observed. The sample size is much smaller than that for the analysis presented in Table 3, as is reflected in the increased width of the confidence intervals.

Table 4: Relative Risk of Absence Due to Respiratory Illness in Relation to Recent Exposure to Air Pollutants

Pollutant	Relative Risk per Unit Exposure	95 % C.I.
Peak O_3 lag 1	1.0295	0.9432–1.1238
Peak O_3 lag 2	0.9696	0.8970–1.0482
Peak PM_{10} lag 1	0.9832	0.9399–1.0285
Peak PM_{10} lag 2	0.9851	0.9488–1.0227
Peak NO_2 lag 1	1.0248	0.9457–1.1105
Peak NO_2 lag 2	0.9939	0.9076–1.0885

There were 499 subjects who reported an absence due to respiratory illness. Adjusted for temperature and humidity lagged one and two days. Units are 10 ppb for O_3 and NO_2, 10 μg / m^3 for PM_{10}.

Tables 5 and 6 present relative risks of absence due to respiratory illness for subjects with wheeze or with asthma, respectively. No statistically significant results are observed for wheezers, but peak NO_2 lagged two days is associated with a statistically significant

increase in risk for asthmatic subjects, and the risk associated with NO_2 lagged one day is elevated as well.

Table 5: Relative Risk of Absence Due to Respiratory Illness in Relation to Recent Exposure to Air Pollutants: Subjects With Wheeze Only

Pollutant	Relative Risk per Unit Exposure	95 % C.I.
Peak O_3 lag 1	0.9033	0.4552–1.7925
Peak O_3 lag 2	1.3696	0.7389–2.5385
Peak PM_{10} lag 1	0.8152	0.5275–1.2597
Peak PM_{10} lag 2	0.9210	0.5897–1.4384
Peak NO_2 lag 1	1.4720	0.9290–2.3324
Peak NO_2 lag 2	0.9117	0.5571–1.4921

There were 59 subjects with wheeze who reported an absence due to respiratory illness. Adjusted for temperature and humidity lagged one and two days. Units are 10 ppb for O_3 and NO_2, 10 μg / m^3 for PM_{10}.

Table 6: Relative Risk of Absence Due to Respiratory Illness in Relation to Recent Exposure to Air Pollutants: Subjects With Asthma Only

Pollutant	Relative Risk per Unit Exposure	95 % C.I.
Peak O_3 lag 1	0.7211	0.4904–1.0603
Peak O_3 lag 2	1.2892	0.8739–1.9018
Peak PM_{10} lag 1	0.9266	0.7933–1.0824
Peak PM_{10} lag 2	0.8251	0.6653–1.0232
Peak NO_2 lag 1	1.1096	0.8564–1.4377
Peak NO_2 lag 2	1.5034	1.0291–2.1962

There were 58 subjects with asthma who reported an absence due to respiratory illness. Adjusted for temperature and humidity lagged one and two days. Units are 10 ppb for O_3 and NO_2, 10 μg / m^3 for PM_{10}.

4 The Multilevel Analytic Design

4.1 Introduction

Epidemiologists recognize two basic strategies for looking at the association between an exposure and a disease: ecologic studies, in which disease rates in groups of individuals are related to the average exposure rates in these groups, and analytic studies, in which individuals' disease outcomes are related to their own exposure values. Cohort studies and case-control studies are examples of the latter type. The epidemiologic literature is full of examples of discrepancies between the conclusions of the two types of studies. In a classic example, Durkheim found suicide rates in provinces of Western Europe to be highly correlated with the proportion of Protestants. Regression analyses of these rates produced an estimate of the rate ratio for Protestants relative to Catholics of 7.5, compared with a value of 2 estimated on an individual basis (Selvin, 1958). Similarly, numerous associations between cancer rates and mean consumption of various dietary factors have been found in ecologic correlation studies, but establishing such associations at an individual level has proven more elusive (Prentice and Sheppard, 1990).

The resolution of such paradoxes usually turns on three issues: between-group confounding, measurement error, and restricted variability. Between-group confounding refers to a characteristic of groups that is not accounted for in the model but is the real risk factor. In the suicide example, such a factor might be the alienation felt by Catholics in predominantly Protestant provinces. This is the essential explanation of the "ecologic fallacy", in which spurious ecological associations may be caused by a tendency for the individuals in the higher exposure groups who get the disease not to have been exposed themselves but rather to have gotten the disease as a result of some other group characteristic. Exposure measurement error has different effects on the two types of studies, generally biasing associations at the individual level towards zero, but not at the aggregate level. Finally, studies conducted within a single group may have a restricted range of variation in exposure and hence limited power. Thus, in the diet example, the positive associations at the ecologic level might be explained by some confounding variable such as race that is not accounted for in the analysis, whereas the lack of association at the individual level might be due to dilution of a real effect by measurement error or by restricted variability in diet within racial groups.

Each of these designs has advantages and disadvantages. The main advantage of the ecologic design is cost, but the fact that there is often greater variation in exposure between groups than within groups is another advantage. On the other hand, ecologic studies typically suffer from between-group confounding (partly because groups will be more heterogeneous with respect to confounders than members of groups and partly because data on confounders are unavailable) and the exposure data are usually of poor quality (e.g., food disappearance rates rather than mean intake rates). Analytic studies are more readily controlled for confounding factors, and have better quality data, but may

suffer from the effects of measurement error and restricted variability.

Multilevel designs, in which both individual and aggregate level variables are used to explain individual outcomes, represent an attempt to capture the desirable features of both approaches. Von Korff et al., (1992) describe many ways in which multilevel analyses have been applied to problems in the social sciences, and recommend their application to epidemiology.

The design we will discuss here is a hybrid design that we shall call the "multilevel analytic design". Key to this design is an analysis that exploits both levels of comparison. Exposure and confounder data are assembled on individuals, to have the best possible quality. The resulting exposure-response relations can then be tested for compatibility with the between-group differences in rates, and if compatible, the two analyses can be pooled for greater power. In particular, this allows one to assess how much of the differences in disease rates between groups can be explained by differences in the distribution of risk factors.

We next provide some details about the basic design and its analysis. In the following section, we describe how the effects of measurement error may be incorporated. We then address the issue of design optimization, and provide an example with a simulation study. Finally, we describe an application to the USC Children's Health Study.

4.2 The Multilevel Analytic Design and Its Analysis

The design begins with a selection of a number of groups $g = 1, ..., G$, which might be defined by geographic areas (as in a study of air pollution), ethnicity (as in a study of diet), or any other factor for which group identifying data are readily available. Within each group, individuals $i = 1, ..., I_g$ are selected. (For notational simplicity, we set $I_g \equiv I$). Data on outcomes y_{gi}, exposures x_{gi}, and confounders v_{gi} are collected on each individual, and in addition, certain characteristics of the group X_g may also be collected. For example, in an air pollution study, individual exposure information might comprise personal exposure estimates (e.g., ozone badges), microenvironmental sampling (e.g., in homes, schools, cars, outdoors), or individual exposure modifying factors such as proportion of time spent outdoors or characteristics of their homes (air conditioning, presence of a smoker, heating and cooking sources, etc.). Group exposure characteristics might include estimates of the ambient levels from area monitoring. The specifics of the outcomes (continuous or binary, cross-sectional or longitudinal) and the sampling plan for individuals (survey, cohort, or case-control) will vary from study to study, but are not germane to the issues discussed here.

For conceptual and notational simplicity, we will assume that the outcome, exposure, and confounder are all univariate and continuous, and that the individuals in each group are chosen by simple random sampling. We also assume that the quantities of interest are

linearly related, that is,

$$y_{gi} = \alpha_g + \beta x_{gi} + \gamma v_{gi} + \varepsilon_{gi} \tag{12}$$

where α_g is the baseline outcome for group g and the ε_{gi} are independent random variables with $\mathrm{E}(\varepsilon_{gi}) = 0$, $\mathrm{Var}(\varepsilon_{gi}) = \sigma^2$. Interest centers on the estimation of β, the exposure effect.

The baseline effects α_g may be considered fixed or random. Considering them random may be appropriate when the groups on which data are collected are randomly chosen from a larger population of groups. If effects are fixed, then one parameter for each group is included in the model. If effects are random, then it is necessary only to include enough parameters to characterize their distribution. Usually this requires only two parameters, a mean and a variance. Therefore it is advantageous to treat effects as random when appropriate, so as to decrease the number of parameters in the model, resulting in more accurate estimation. The true exposures x_{gi} and the confounders v_{gi} may also be considered either fixed or random. Considering them random is appropriate, for example, when individual exposure is considered to be distributed around a function of ambient exposure and one or more exposure modifying variables, as will be described in section 4.4. In what follows we will consider α_g, x_{gi}, and v_{gi} to be random, and we make the following assumptions:

1. The random variables α_1, ..., α_G are independent and identically distributed (e.g., the groups are selected by simple random sampling).
2. The group baseline effects α_g are independent of both x_{gi} and v_{gi}.

In general, the true exposures x_{gi} will be unknown, and will be estimated by measured values, as discussed in the section 4.4. For the remainder of this section, in order to study the performance of our method under ideal circumstances, we will ignore the effect of measurement error, effectively assuming the true exposures to be known. We will also assume that the true values of the confounder v_{gi} are known, although measurement error in v_{gi} can bias the estimator of β (Greenland, 1980).

Eq. (12) can be used to estimate β, and is appropriate when the α_g's are considered fixed. When the α_g are independent random variables with $\mathrm{E}(\alpha_g) = \alpha$, $\mathrm{Var}(\alpha_g) = \tau^2$, an estimator with smaller variance is obtained using the equation

$$y_{gi} = \alpha + \beta x_{gi} + \gamma v_{gi} + \eta_{gi}. \tag{13}$$

The error η_{gi} is equal to $\alpha_g - \alpha + \varepsilon_{gi}$. The covariance matrix of $\boldsymbol{\eta}$ can be described as follows: Let $\rho = \tau^2/(\sigma^2 + \tau^2)$. Define $\boldsymbol{\Sigma} = (1-\rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}^T$, where $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ is an I-dimensional column of 1s. Define $\boldsymbol{\Sigma}_{BIG}$ to be the $GI \times GI$ block diagonal matrix, consisting of G identical blocks of the matrix $\boldsymbol{\Sigma}$. Then the covariance matrix of $\boldsymbol{\eta}$ is equal to $(\sigma^2 + \tau^2)\boldsymbol{\Sigma}_{BIG}$. The matrix $\boldsymbol{\Sigma}^{-1}$ is equal to $a\mathbf{I} + b\mathbf{1}\mathbf{1}^T$, where $a = 1/(1-\rho)$, and $b = -\rho/\{[(I-1)\rho + 1](1-\rho)\}$. Thus $\boldsymbol{\Sigma}_{BIG}^{-1}$ is a block diagonal matrix with each block equal to $\boldsymbol{\Sigma}^{-1}$.

If ρ is known, the parameters α, β, and γ can be estimated by weighted least squares. If σ^2 and τ^2 are unknown, the parameters can be estimated by a two-stage procedure. In the first stage, only within groups differences are used. This is accomplished by using Eq. (12) to estimate the parameters $\alpha_1, ..., \alpha_G, \beta, \gamma$ by ordinary least squares. Denote by $\hat{\beta}_1$ the estimate of β obtained from this first-stage regression, and by $\hat{\sigma}^2$ the usual mean square residual estimate of error variance. The second stage regression involves only the between groups differences. The regression equation is obtained from Eq. (13) by averaging over i:

$$\bar{y}_{g.} = \alpha + \beta \bar{x}_{g.} + \gamma \bar{v}_{g.} + \bar{\eta}_{g.}. \tag{14}$$

The variables $\bar{\eta}_{g.}$ are independent with mean 0 and variance $\tau^2 + \sigma^2/I$. Denote by $\hat{\beta}_2$ the ordinary least squares estimate of β from Eq. (14). The mean square residual is an estimate of $\tau^2 + \sigma^2/I$, which can be combined with $\hat{\sigma}^2$ to yield an estimate of τ^2. The estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are uncorrelated. Let $\mathbf{X}_1$ and $\mathbf{X}_2$ denote the design matrices from the first and second stages, respectively. Then $\mathrm{Var}(\hat{\beta}_1)$ and $\mathrm{Var}(\hat{\beta}_2)$ are estimated with appropriate elements from the diagonals of the matrices $(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\hat{\sigma}^2$ and $(\mathbf{X}_2^T\mathbf{X}_2)^{-1}(\hat{\sigma}^2/I + \hat{\tau}^2)$. The two-stage procedure is completed by computing the variance weighted average of $\hat{\beta}_1$ and $\hat{\beta}_2$ to obtain the estimator

$$\hat{\beta}_{pooled} = \frac{\mathrm{Var}(\hat{\beta}_2)\hat{\beta}_1 + \mathrm{Var}(\hat{\beta}_1)\hat{\beta}_2}{\mathrm{Var}(\hat{\beta}_1) + \mathrm{Var}(\hat{\beta}_2)}. \tag{15}$$

The relationship between weighted least squares and the two-stage procedure is given by the following:

Theorem: Let $\hat{\sigma}^2$, $\hat{\tau}^2$ be the estimators of σ^2, τ^2 from the two-stage procedure. Then $\hat{\beta}_{pooled}$ is the weighted least squares estimate of β when $\rho = \hat{\tau}^2/(\hat{\sigma}^2 + \hat{\tau}^2)$.

Corollary: If the errors η_{gi} are normally distributed, then the MLE of β satisfies Eq. (15), with $\hat{\beta}_{MLE}$ substituted for $\hat{\beta}_{pooled}$ and $\mathrm{Var}(\hat{\beta}_1)$ and $\mathrm{Var}(\hat{\beta}_2)$ evaluated at $\hat{\beta}_{MLE}$.

Proofs of these claims are provided in section 4.7. The corollary suggests that Eq. (15), if iterated, will converge to the MLE.

4.3 Potential Sources of Bias

As with any model, departures from model assumptions can lead to bias. This is especially true with ecologic analyses, for which many different types of bias have been documented. In the case where exposure is binary, Brenner et al., (1992) discuss bias that can result from nondifferential exposure misclassification in ecologic studies. This type of bias can occur in both linear and log-linear models, and is always away from the null.

Richardson et al., (1987) discuss biases resulting when the outcome is related to a non-linear function of exposure. They also discuss bias that results when baseline risks

within groups are related to mean exposures within groups. Greenland and Robins (1994) discuss bias resulting from nonlinearity and nonadditivity of effects. In particular, they show how interactions on the individual level may not be appropriately represented by the corresponding ecologic level interaction. That is, if the individual outcome y is related to individual level variables x_1 and x_2 by the relation

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2, \tag{16}$$

and if μ_y, μ_{x_1}, μ_{x_2} are the mean levels in a given group, then the ecologic model

$$\mu_y = \beta_0 + \beta_1 \mu_{x_1} + \beta_2 \mu_{x_2} + \beta_3 \mu_{x_1} \mu_{x_2} \tag{17}$$

may give biased results.

While our model, like others, is subject to bias when its assumptions are violated, it should be relatively free from certain common ecologic biases due to the fact that exposure variables and confounders are assessed at the individual level. Therefore in the example given in equations (16) and (17), if we define μ_{12} to be the mean level of the interaction term $x_1 x_2$ in a particular group, our ecologic model would be

$$\mu_y = \beta_0 + \beta_1 \mu_{x_1} + \beta_2 \mu_{x_2} + \beta_3 \mu_{12}, \tag{18}$$

which does not suffer from nonadditivity bias.

With regard to bias resulting from a relationship between baseline risk and mean exposure across groups, this will not occur if our assumption of random selection of groups is met. If this assumption is violated, however, ecologic bias may result.

4.4 Allowance for Exposure Measurement Error

In many circumstances, it may not be feasible to obtain complete and error-free data on all individuals and hence some variables will only be available for some (randomly selected) subset of individuals. For example, in a dietary study, one might wish to validate the use of a food frequency questionnaire in the entire group by repeated 7-day records. In an air pollution study, it might be feasible to obtain personal monitoring or microenvironmental sampling data on only a sample, but questionnaire data on individual modifying factors might be available on the entire group. Optimization of the design would typically entail trade-offs between the number of groups and the number of individuals in the main study and in validation substudies, and the extent of the measurement protocols, subject to constraints on the total costs. These design issues will be discussed further below. In this section, we will focus on the effect of exposure measurement error. To simplify matters, we will ignore confounding. The extension of our results to the situation in which imperfectly

measured confounders are included is mathematically straightforward, but considerably more complex notationally.

We make a distinction in our analysis between two types of measurement error. The first type, known as the "Berkson" error model (Armstrong, 1990), applies when individuals are assigned their group average exposures. The second type, known as the "classical" error model, applies when the assigned exposure is a random variable whose expected value is the true exposure.

Let x_{gi} denote the unobservable true exposure for individual i in group g and let z_{gi} indicate the measured value (e.g., from personal monitoring). The classical error model assumes that the measured values are randomly distributed around the true value with the property that $\mathrm{E}(z_{gi}|x_{gi}) = x_{gi}$. As is well known (see Thomas, Stram, and Dwyer (1993) for a review), the classical error model produces a bias towards the null, essentially because the measured exposures are overdispersed ($\mathrm{Var}(z_{gi}) = \mathrm{Var}(x_{gi}) + \mathrm{Var}(z_{gi}|x_{gi}) > \mathrm{Var}(x_{gi})$). Thus if $\mathrm{Var}(x_{gi}) = \phi_g^2$ and $\mathrm{Var}(z_{gi}|x_{gi}) = \omega^2$, the regression on z_{gi} produces a slope estimate $\hat{\beta}$ that has expectation $c_g = \phi_g^2/(\phi_g^2 + \omega^2)$ times the expectation of the slope of the regression on the x_{gi}. This suggests a simple correction for measurement error if these variances are equal and known. First fit the naive regression on z_{gi} and then correct the estimated slope coefficient by dividing it by c_g (Rosner, Willett, and Spiegelman, 1989). For more complex situations, for example if the variances differ between groups, a useful strategy is to replace the z_{gi}'s by $\hat{x}_{gi} = \mathrm{E}(x_{gi}|z_{gi}) = c_g z_{gi} + (1 - c_g)\mathrm{E}(x_{gi})$ and then use these $\hat{x}_{gi}$'s as if they were the true exposures in the regression.

The Berkson error model assumes instead that the true exposures x_{gi} of individuals are distributed around their group estimates X_g with the property that $\mathrm{E}(x_{gi}|X_g) = X_g$. Thus, in an air pollution study with no personal monitoring, we might assume that individuals' exposures are randomly distributed around the ambient levels for their communities. A consequence of this assumption is that, at least for linear exposure-response models, the regression on the measured values provides unbiased estimates of the true slope. If $y_{gi} = \alpha_g + \beta x_{gi} + \varepsilon$, then

$$\begin{aligned} \mathrm{E}(y_{gi}|X_g) &= \alpha_g + \beta\mathrm{E}(x_{gi}|X_g) + \mathrm{E}(\varepsilon|X_g) \\ &= \alpha_g + \beta X_g. \end{aligned}$$

Thus, Berkson error produces no bias towards the null for linear models.

Typically, it would not be feasible to obtain true exposure data on any individuals. Rather, a surrogate variable w would be obtained on everybody and higher quality measurements z only on a sample. The measurements are assumed to be unbiased in the classical error sense and might be replicated T times. In this case, it will not be possible to use the z's directly in modeling y because they are available on too few subjects, but they could be used to build a model for the relationship between z and w which could then be used for imputing $\hat{x}$-values in the first stage regression. The surrogate variable w might be a simpler measure of x (such as a food frequency questionnaire) or it might be a personal modifier of

a group exposure characteristic X (for example, percent time spend outdoors in an air pollution study could modify the ambient pollution level).

To give a concrete example of this imputation procedure, assume that at times $t = 1, 2, ..., T$ we have measurements of a group exposure characteristic X_{gt} for each group, and for a subset of individuals we have an exposure modifying variable w_{git} and an exposure measurement z_{git}. We assume that X and w are assessed without error, and z has a classical error structure in relation to true exposure x. For the sake of simplicity, we will consider that the replicate measurements have been averaged over time, and we will drop the subscript t. We assume the following relationships:

$$x_{gi} \sim N(X_g + \delta_0 + \delta_1 w_{gi}, \phi^2) \tag{19}$$
$$z_{gi} \sim N(x_{gi}, \omega^2) \tag{20}$$

We assume that ω^2 is known from other studies or from another set of replicate measurements, but that δ_0, δ_1, and ϕ^2 are unknown. Combining Eqs. (19) and (20) yields

$$z_{gi} \sim N(X_g + \delta_0 + \delta_1 w_{gi}, \phi^2 + \omega^2) \tag{21}$$

from which we can obtain unbiased estimates of δ_0, δ_1, and ϕ^2 (since ω^2 is known). We then estimate x_{gi} as $\hat{x}_{gi} = X_g + \hat{\delta}_0 + \hat{\delta}_1 w_{gi}$, which is an unbiased estimator of $\mathrm{E}(x_{gi}|X_g, w_{gi})$ since $\hat{\delta}_0$ and $\hat{\delta}_1$ are unbiased.

Allowing for measurement error complicates the two-stage procedure for estimating the parameter β as follows. We assume

$$\alpha_g \sim N(\alpha, \tau^2) \tag{22}$$
$$y_{gi} \sim N(\alpha_g + \beta x_{gi}, \sigma^2) \tag{23}$$

where α, τ^2, and σ^2 are unknown. Since the x_{gi} are also unknown, we replace them with their estimates $\hat{x}_{gi}$ when fitting the model. The first stage model is thus:

$$y_{gi} = \alpha_g + \beta \hat{x}_{gi} + \varepsilon_{gi}, \tag{24}$$

where the ε_{gi} are independent normal random variables with $\mathrm{E}(\varepsilon_{gi}) = 0$ and $\mathrm{Var}(\varepsilon_{gi}) = \sigma^2 + \beta^2 \mathrm{E}[(x_{gi} - \hat{x}_{gi})^2 | X_g, w_{gi}]$.

Since $\mathrm{Var}(\varepsilon_{gi})$ depends on g, i, and on the unknown parameter β, the model Eq. (24) may be fit by iteratively reweighted least squares (IRLS). In order to use this procedure, we must first express the weights $\mathrm{Var}(\varepsilon_{gi})$ in usable form. Let $\hat{V}(\hat{\delta}_0)$, $\hat{V}(\hat{\delta}_1)$, and $\hat{C}(\hat{\delta}_0, \hat{\delta}_1)$, be the estimates of $\mathrm{Var}(\hat{\delta}_0 | X_g, \{w_{gi}\})$, $\mathrm{Var}(\hat{\delta}_1 | X_g, \{w_{gi}\})$, and $\mathrm{Cov}(\hat{\delta}_0, \hat{\delta}_1 | X_g, \{w_{gi}\})$, respectively, calculated from regression model Eq. (21). Since $\hat{x}_{gi} \approx \mathrm{E}(x_{gi}|X_g, w_{gi})$, $\mathrm{Var}(\varepsilon_{gi})$ can be approximated by $V^* = \sigma^2 + \beta^2 \mathrm{Var}(x_{gi}|X_g, w_{gi}) \approx \sigma^2 + \beta^2 W_{gi}$ where $W_{gi} = \hat{V}(\hat{\delta}_0) + 2w_{gi}\hat{C}(\hat{\delta}_0, \hat{\delta}_1) + w_{gi}^2 \hat{V}(\hat{\delta}_1) + \hat{\phi}^2$. The IRLS procedure is then conducted as follows. Set $\hat{\sigma}^2$ and $\hat{\beta}$ to arbitrary initial values, then fit model Eq. (24) by weighted least

squares regression using weights V^{*-1}. This produces an updated estimate $\hat{\beta}^{(1)}$ of β, and fitted values $\hat{y}_{gi}^{(1)}$. To obtain an updated estimate of σ^2, take the average of the values $(y_{gi} - \hat{y}_{gi}^{(1)})^2 - (\hat{\beta}^{(1)})^2 W_{gi}$ and then repeat the entire process.

Let $\hat{\alpha}_g$, $\hat{\beta}_1$, and $\hat{\sigma}^2$ be the estimates of α_g, β, and σ^2 obtained from the first-stage model Eq. (24). Let $\mathbf{W}$ be the diagonal matrix whose diagonal elements are $\hat{\sigma}^2 + \hat{\beta}_1 W_{gi}$. Let $\mathbf{X}$ be the design matrix corresponding to the first-stage model Eq. (24). The usual estimate of the covariance matrix of $\hat{\alpha}_g$, $\hat{\beta}_1$ is $(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}$, which is accurate when the substudy sample is reasonably large (see simulation results below).

The second-stage model is obtained from Eq. (24) by averaging over i and by replacing α_g with its mean α to obtain

$$\bar{y}_{g\cdot} = \alpha + \beta \hat{x}_{g\cdot} + \eta_g \tag{25}$$

where $\hat{x}_{g\cdot} = I^{-1} \sum_{i=1}^{I} \hat{x}_{gi}$, and η_g are independent random variables with $\mathrm{E}(\eta_g) = 0$ and

$$\mathrm{Var}(\eta_g) = \tau^2 + \sigma^2/I + \beta^2 \mathrm{E}[(\bar{x}_{g\cdot} - \hat{x}_{g\cdot})^2 | X_g, \{w_{gi}\}]. \tag{26}$$

The expectation on the right hand side of Eq. (26) may be estimated with $\bar{W}_{g\cdot}$, the average of the W_{gi} described above.

The second-stage model may also be fitted by IRLS, updating β as in the first-stage regression, and updating the sum $\tau^2 + \sigma^2/I$ as a single quantity. After convergence, a separate estimate of τ^2 can be obtained by combining the estimate of $\tau^2 + \sigma^2/I$ with the estimate of σ^2 from the first-stage regression. Let $\hat{\beta}_2$ be the estimate of β from the second-stage regression. The covariance matrix of $\hat{\alpha}$ and $\hat{\beta}_2$ can be obtained in a manner analogous to the method described above for the first-stage model.

If the true weights were known and did not depend on β, the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ would be uncorrelated. When variables are normally distributed, they are nearly uncorrelated when the weights are estimated as well (simulation results, not reported). For many applications in environmental epidemiology, it is more appropriate to assume that true exposures are nonnegative and lognormally distributed and that measurement errors are lognormally distributed and multiplicative. Furthermore, the individual exposure modifiers w_{gi} might also be assumed to act multiplicatively on the group means X_g. All this can be accomplished without any new theory simply by redefining x, X, and z to be the logarithms of their respective quantities.

4.5 Design Optimization

At the design stage, the epidemiologist needs to consider the trade-off between the number of groups and the number of subjects per group, the selection of the specific groups to be included, the number of subjects in the main study versus the number in the validation

sample, and the number and complexity of measurements to be made on each sample. These are important issues that have been given only limited attention in the context of analytic studies and none in the context of ecologic or hybrid designs. For analytic studies, Greenland (1988) and Spiegelman and Gray (1991) have considered the trade-offs between numbers of subjects in the main and validation studies and have provided explicit formulae for determining the optimal design where it is planned to use measurement error adjustment methods in the analysis like those described above. Rosner and Willett (1988) considered the trade-off between numbers of subjects and numbers of replicate measurements in a validation study.

For linear models with a continuous normally distributed outcome, ignoring confounding at the individual and group levels, measurement error, and assuming exposure is assessed only at the group level, the power of the study can be computed as a function of four quantities: the number of groups G, the number of subjects I sampled in each group, the true R^2 between group mean exposure and group mean outcome, and the ratio $VR = V_W/V_B$, where V_W and V_B are outcome variances within and between groups, respectively. Given these quantities, we compute $R_*^2 = IV_BR^2/(V_W + IV_B)$, the squared correlation between group mean exposure and the average outcome among the individuals sampled from the group. The quantity R_*^2 is less than R^2, because the sample mean outcome rather than the true group mean outcome is used. The power to detect a nonzero R^2 is calculated by using Fisher's transformation of R_*^2.

Table 7 illustrates the results for a variety of choices of the model and design parameters. It is clear that the power is much more strongly influenced by the number of groups than by the number of subjects. For a logistic model for binary outcomes, the power also depends on the overall disease frequency, but the same basic result emerges — the power of the aggregate analysis depends much more strongly on the number of groups than on the number of subjects per group.

Table 8 provides similar power calculations for testing a partial R^2 for the individual regression after removing group effects, again using Fisher's transformation. The power of these analyses depends only on the total number of individuals and it is clear that with sample sizes in the thousands, there is adequate power for detecting very small correlations. However, it is important to note that these are the correlations with the measured exposures, which could be severely attenuated by measurement error.

To provide further guidance for the design of the USC Children's Health study, we undertook a limited simulation study. For this purpose, we varied the number of groups G, the number of subjects per group in the main study I, and the number of subjects per group in the exposure substudy S. Relationships amongst the variables were as given in Eqs. (19–23), with the ambient levels X_g, X_{gt} and individual modifiers w_{gi}, w_{git} being normally distributed. For each choice of design parameters, 1000 replicate data sets were simulated and analyzed using the methods described above. We tabulated the bias and variance of the parameter estimates from the individual level regressions (with and without adjustment for measurement error), the ecologic regression, and the proposed pooled

Table 7: Statistical Power of Between-groups Comparisons

		R^2 for Between-groups Regression					
		$R^2 = 0.1$		$R^2 = 0.3$		$R^2 = 0.5$	
G	$G \times I$	$VR = 10$	$VR = 100$	$VR = 10$	$VR = 100$	$VR = 10$	$VR = 100$
5	1000	0.12	0.10	0.21	0.18	0.33	0.28
	4000	0.12	0.11	0.22	0.20	0.34	0.32
10	1000	0.21	0.15	0.46	0.32	0.72	0.54
	4000	0.21	0.19	0.48	0.43	0.75	0.68
20	1000	0.34	0.20	0.75	0.45	0.96	0.73
	4000	0.37	0.29	0.80	0.68	0.97	0.92
40	1000	0.52	0.23	0.94	0.54	1.00	0.84
	4000	0.60	0.41	0.97	0.86	1.00	0.99

G is the number of groups, and I is the number of individuals per group. $VR = V_W/V_B$, where V_W is the within-groups outcome variance and V_B is the between-groups outcome variance.

combination of the two regressions.

The design parameters were chosen to approximate those being considered for the USC Children's Health study, and the model parameters were then adjusted to illustrate a hypothetical situation in which the two approaches to estimation would be roughly equally informative. The true parameter values for the simulation are given in a footnote to Table 9. Table 9 illustrates the effect of modifying the design parameters under the constraint that the total number of measurements $G(I + SM)$ be fixed at 3000. Under the assumptions of the simulation, measurement error is minimized when one measurement is taken per individual in the substudy. Therefore we set $T = 1$. All parameter estimates appear to be nearly unbiased. The columns labeled "sample SD" give the sample standard

Table 8: Statistical Power of Within Groups Comparisons

$G \times I$	$R^2 = 0.001$	$R^2 = 0.005$	$R^2 = 0.01$	$R^2 = 0.05$
1000	0.26	0.72	0.94	1.00
2000	0.41	0.94	1.00	1.00
4000	0.64	0.99	1.00	1.00

Table 9: Standard Errors of Parameter Estimates in the Presence of Measurement Error

			$SE(\hat{\beta}_1)$		$SE(\hat{\beta}_2)$		$SE(\hat{\beta}_{pooled})$	
			Sample	Nominal	Sample	Nominal	Sample	Nominal
G	I	S	SE	SE	SE	SE	SE	SE
6	200	300	0.19	0.15	0.32	0.30	0.15	0.13
12	100	150	0.19	0.15	0.19	0.19	0.13	0.11
24	50	75	0.19	0.15	0.14	0.14	0.11	0.10
48	25	38	0.18	0.15	0.11	0.11	0.092	0.086
6	400	100	0.24	0.11	0.30	0.30	0.19	0.10
12	200	50	0.25	0.11	0.18	0.18	0.16	0.090
24	100	25	0.24	0.11	0.12	0.12	0.12	0.080
48	50	13	0.23	0.11	0.092	0.092	0.10	0.069

G is the number of groups, I is the number of individuals per group in the main study, and S is the number of individuals per group in the exposure validation substudy.
1000 data sets were generated for each value of G, I, and S.
Exposures measured subject to error; measurement error variance estimated in substudy.
True parameter values are $\phi^2 = 1.0$, $\omega^2 = 25.0$, $\tau^2 = 1.0$, $\sigma^2 = 25.0$. In addition, $\text{Var}(X_g) = 4.0$, $\text{Var}(X_{gt}|X_g) = 0.25$, $\text{Var}(w_{gi}) = 1.0$, and $\text{Var}(w_{git}|w_{gi}) = 0.25$.

deviation of the 1000 parameter estimates. The columns labeled "Nominal SE" give the square root of the average of the 1000 conditional variance estimates obtained from the covariance matrix estimator $(X^T W^{-1} X)^{-1}$. Each block of the table shows the effect of varying the number of groups and the number of subjects per group. In agreement with Tables 7 and 8, the efficiency of $\hat{\beta}_2$ improves rapidly as the number of groups increases, whereas $\hat{\beta}_1$ depends only on the total number of subjects. Comparing the two blocks illustrates the trade off between the number of subjects in the main study and validation substudy. The second-stage estimator is relatively insensitive to this parameter, while the first-stage estimator is improved by having a larger proportion in the validation study.

4.6 Application to the USC Children's Health Study

In January 1992, the California Air Resources Board (CARB) awarded a contract to the University of Southern California to initiate a 10-year cohort study of the health effects of air pollution in Southern California. The study enrolled a cohort of about 3500 school children from 12 communities selected so as to represent a variety of types and levels of air pollution that are represented in the basin. The primary focus of the study is on the effects of chronic exposure to one-hour peak ozone (O_3), but particulates (PM_{10}), nitrogen dioxide (NO_2), acids (H^+), and other pollutants are also being measured. Health outcomes

measured annually include various lung function tests, symptoms reported by questionnaire, and absences abstracted from school records.

4.6.1 Community Selection

Some preliminary power calculations based on assumed values for true effects indicated that for studying a single pollutant, it would be necessary to have at least 10 groups for power to be adequate. We carried out further calculations along similar lines to assess the prospects for doing multivariate analyses of two or more pollutants and concluded that it would be possible, provided groups could be selected in such a way that the correlations in pollutant levels across groups were not too large. Thus, the optimal choice would have to take account of the actual levels of exposure to each of the pollutants we wished to assess.

Fortunately, extensive data were available on the four highest priority pollutants from the CARB's monitoring program. Year-round average levels for the period 1986-1990 were obtained from 86 monitoring stations scattered across Southern California. (For some pollutants, notably acids, the values had to be interpolated from other stations on an inverse-distance weighted basis). Our initial selection of sites was based on the intuitive notions that (1) we wished to maximize the dispersion of each of the pollutants, and (2) we wished to represent as many combinations of high and low levels of each pollutant as possible.

For each pollutant, we calculated the mean level over the 86 communities, then for each community, we converted the pollution levels to standard units. Each community was assigned a "profile" by recording it as either above (+) or below (−) the mean level for each pollutant. For a design based on all four pollutants, there were thus $2^4 = 16$ possible profiles, of which demographically suitable examples could be found for 7 of them. Within each profile, we then selected from one to three communities whose sum of squared standardized pollution levels were large. Table 10 describes the characteristics of the communities that we judged to be the most suitable on this basis, under the constraint that we could afford to study no more than twelve. This selection process differs from the one described above in that the groups were not randomly chosen. Thus the group effects must be considered fixed rather than random.

In order to compare alternative designs based on different selections of priority pollutants, we then carried out a further simulation study, based purely on the second stage ecologic regression, but allowing the actual pollutant levels to differ from the measured values subject to a covariance structure estimated from the observed data. Table 11 summarizes the results of this simulation, which led us to the conclusion that, if all four pollutants had health effects, then the optimal design would need to be based on all four. This design appears to have adequate power for detecting differences in mean FEV_1 of about 3-5% between the high and low communities for each of the four pollutants in multivariate analysis, assuming that one-third of the variance in FEV_1 is explained by variation in the

Table 10: Characteristics of the Communities Selected for the University of Southern California Children's Health Study

Community	Profile	Annual Mean Level				Demographic Characteristics		
		O_3	PM_{10}	NO_2	H^+	% White	% Age 5-18	People/Room
Glendora	++++	109.2	67.0	39.1	2.93	89	20	0.50
Upland	++++	92.0	75.6	44.6	3.09	75	22	0.58
Mira Loma	+++−	95.1	84.9	32.8	1.05	68	24	0.62
Riverside	+++−	95.1	84.9	32.8	1.05	77	22	0.51
Perris	++−+	81.3	60.1	15.4	1.99	73	23	0.62
Lancaster	++−+	70.8	47.0	13.2	3.16	81	21	0.54
Lake Gregory	+−++	98.8	38.3	23.6	2.26	95	22	0.51
Alpine	+−−−	80.5	37.4	16.7	1.18	93	19	0.48
North Long Beach	−+++	45.2	49.5	44.8	2.43	58	18	0.59
Santa Maria	−−−−	30.2	28.0	7.7	0.91	66	20	0.61
Santa Barbara	−−−−	30.4	31.0	10.4	0.91	84	19	0.50
Lompoc	−−−−	34.8	30.0	1.6	0.91	72	21	0.58

In the "Profile" column, a "+" signifies that the pollution level is above the mean level of the 86 communities considered, a "−" signifies that the pollution is below that level. O_3 and NO_2 are measured in ppb on a mass basis. PM_{10} is measured in $\mu g/m^3$. H^+ is measured in ppb on a molar basis. The communities of Glendora, Perris, and Santa Barbara elected not to participate in the study, and were replaced by San Dimas, Lake Elsinore, and Atascadero, whose pollution characteristics were similar.

pollutants, and that O_3 and PM_{10} each contribute twice as much to the health effect as do NO_2 and H^+. Alternative designs that ignore one or more of these pollutants (with the same total number of communities) may slightly increase the variability of the pollutants of primary interest, which normally would be expected to yield an improvement in power. However, they also substantially weaken the power for controlling the confounding effect of the omitted pollutants and therefore in most instances reduce the power for the effects of interest in a multivariate analysis.

To determine whether we could significantly improve our selection of communities under the four-pollutant design, we conducted a final simulation along similar lines, starting with the choice given in Table 10 and in a stepwise fashion considered replacing each of the twelve communities by each of the remaining candidates. This led to the conclusion that, under an optimality criterion that maximized the sum of the powers for the four pollutants, it was theoretically possible to improve the design further by changing 5 of the 12 sites. This alternative choice attained better overall power by substantially reducing the correlations among the exposure variables. However, it did so at the expense of

Table 11: Comparison of Power to Detect Effects of Four Priority Pollutants from Alternative Choices of Sites

Community Selection Based On	Pollutants Included in Model	G	Power			
			O_3	PM_{10}	NO_2	H^+
O_3 PM_{10} H^+	O_3 PM_{10} NO_2 H^+	12	0.88	0.66	0.23	0.28
O_3 PM_{10} NO_2	O_3 PM_{10} NO_2 H^+	12	0.89	0.89	0.40	0.30
O_3 PM_{10} NO_2 H^+	O_3 PM_{10} NO_2 H^+	12	0.78	0.87	0.58	0.77
Random	O_3 PM_{10} NO_2 H^+	12	0.34	0.61	0.34	0.37
O_3 PM_{10}	O_3 PM_{10} NO_2 H^+	8	0.48	0.33	0.14	0.14
O_3 PM_{10} NO_2	O_3 PM_{10} NO_2 H^+	6	0.07	0.08	0.06	0.06
O_3 PM_{10} H^+	O_3 PM_{10} NO_2 H^+	6	0.13	0.10	0.07	0.07
Original	O_3 PM_{10}	5	0.22	0.23		

G is the number of groups in the study.

substantially reducing the variance of each exposure. Since we were unsure of the validity of the correlation estimates because many of the entries were based on interpolation, and since the overall improvement in power was modest, we decided to retain our original selection. Essentially, we judged that the primary objective of the study was to maximize the overall power to detect any air pollution effect and that the separation of the effects of particular pollutants was of only secondary importance, after having demonstrated an overall effect. We therefore felt that it was more important to maximize the variance in exposures than to minimize their covariances.

4.6.2 Exposure Modeling

The measurement protocol entails a combination of ambient monitoring, personal monitoring, microenvironmental sampling, and questionnaire assessment of personal modifying factors. Ambient data is routinely collected by the CARB for each of the communities, and provides long-term average levels throughout the study as well as historically. The questionnaire is administered to all subjects annually and includes items on residence history, usual indoor and outdoor times and activities and household characteristics (smoking by family members, air conditioning and heating, air exchange, sources of indoor pollution, etc.). Protocols for personal monitoring and microenvironmental sampling are still under discussion, and for this reason, we have not yet had the opportunity to bring the full force of our methods to bear on this study. We

describe below our analysis plan.

The goal of the analysis will be to combine these various data sources in such a way as to provide estimates of individual and group mean exposures for the first- and second-stage regressions described above, including estimates of measurement error distributions for adjustment purposes. The actual form of the models to be used is still under development, since it depends on the protocols adopted for the determination of personal exposure. To illustrate the general approach, we make some simplifying assumptions that will be remedied in our final analyses.

First, we assume that the relevant exposure variable is the long-term arithmetic mean (i.e, the time-weighted average, TWA). We also assume that ambient levels, true personal exposures, and measurement errors are lognormally distributed. Finally, we assume that the ratio of personal exposures to ambient levels is described by a multiplicative factor that depends loglinearly on the personal modifying factors. The basic relationships are thus as described in Eqs. (19)–(24), except for the additional complexities introduced by the lognormal assumptions. Using the estimates from this model, we can compute for each subject in the main study the TWA, $\mathrm{E}(e^{x_{gi}}|X_g, w_{gi})$, for use in the first-stage regression, together with the average over all subjects of these TWAs for use in the second-stage regression. Whether it will be possible to assess exposure effects at an individual level in a longitudinal analysis will depend primarily on the ability of the exposure model to accurately predict personal exposures. We have been successful in assessing exposure effects in the cross-sectional phase of the study, using lifetime exposures inferred from historical ambient measurements and residential histories. Even if it is not possible to assess exposure-response relations at an individual level, however, the use of average TWAs rather than X_g in the second stage should lead to more reliable estimates, because communities with different exposure patterns are likely to differ substantially in modifying factors such as use of air conditioning and proportion of time spent outdoors, because of major differences in climate across Southern California.

4.6.3 Estimation of Health Effects

We present an illustration of the multilevel procedure using the pulmonary function data in the USC Children's Health Study. More than 3500 children from 12 Southern California communities underwent spirometric assessment of pulmonary function. We determine the relationship between the measured FEV_1 and lifetime exposure to O_3. Lifetime exposure was assessed by compiling a residential history on each subject, then consulting historical records to determine the ambient levels of pollutants during each month of the subject's life.

The model used to analyze the data adjusted for exposure to the pollutants PM_{10} and NO_2, and for age, sex, race, height, and weight. The results were $\hat{\beta}_1 = -1.87 \pm 0.91$, $\hat{\beta}_2 = -0.16 \pm 0.97$, and $\hat{\beta}_{pooled} = -1.07 \pm 0.66$.

We were fortunate in the USC Children's Health study that, due to considerable levels of migration in Southern California, there was a large degree of variation within each community in the lifetime exposure to the pollutants under study. Therefore the individual level analysis provided a good level of efficiency for studying the effects of lifetime exposure, so we did not combine it with ecologic analysis in the Phase II report of this study (Peters, 1996). As described above, we expect the multilevel analytic design to play a considerable role in the longitudinal phase of the study.

4.7 Proofs

Proof of Theorem: We prove the theorem for a more general case with an arbitrary number of confounders. The model is

$$y_{gi} = \alpha + \beta x_{gi} + \mathbf{V}\gamma + \eta_{gi}$$

where $\mathbf{V}$ is a matrix of confounder variables. Let $\hat{\sigma}^2$ and $\hat{\tau}^2$ be estimates of σ^2 and τ^2, and let $\hat{\mathbf{\Sigma}}_{BIG}$ be the corresponding estimate of $\mathbf{\Sigma}_{BIG}$.

Assume without loss of generality that $\hat{\mathbf{\Sigma}}_{BIG}^{-1/2}\mathbf{x}$ is orthogonal to $\hat{\mathbf{\Sigma}}_{BIG}^{-1/2}\mathbf{V}$. Otherwise, replace $\mathbf{x}$ with $\mathbf{x} - \mathbf{V}(\mathbf{V}^T\hat{\mathbf{\Sigma}}_{BIG}^{-1}\mathbf{V})^{-1}\mathbf{V}^T\hat{\mathbf{\Sigma}}_{BIG}^{-1}\mathbf{x}$ and reparameterize the confounders γ. The weighted least squares estimates of α and β are the values minimizing

$$(\mathbf{y} - \alpha - \beta\mathbf{x})^T\hat{\mathbf{\Sigma}}_{BIG}^{-1}(\mathbf{y} - \alpha - \beta\mathbf{x}) \tag{27}$$

which is equal to

$$\sum_{g=1}^{G}(\mathbf{y}_g - \alpha - \beta\mathbf{x}_g)^T\hat{\mathbf{\Sigma}}^{-1}(\mathbf{y}_g - \alpha - \beta\mathbf{x}_g) \tag{28}$$

where $\mathbf{y}_g = (y_{g1}, ..., y_{gI})^T$, and $\mathbf{x}_g$ is defined similarly.

Let $\rho = \hat{\tau}^2/(\hat{\sigma}^2 + \hat{\tau}^2)$, $a = 1/(1-\rho)$, and $b = -\rho/(1-\rho)(1+(I-1)\rho)$. Substitute $(a\mathbf{I} + b\mathbf{1}\mathbf{1}^T)/(\hat{\sigma}^2 + \hat{\tau}^2)$ for $\hat{\mathbf{\Sigma}}^{-1}$ in Eq. (28) to obtain

$$\frac{1}{\hat{\sigma}^2 + \hat{\tau}^2}\sum_{g=1}^{G}\left[a\sum_{i=1}^{I}(y_{gi} - \alpha - \beta x_{gi})^2 + bI^2(\bar{y}_{g.} - \alpha - \beta\bar{x}_{g.})^2\right]. \tag{29}$$

Algebraic manipulation yields

$$\frac{a}{\hat{\sigma}^2 + \hat{\tau}^2}\sum_{g=1}^{G}\sum_{i=1}^{I}[y_{gi} - \bar{y}_{g.} - \beta(x_{gi} - \bar{x}_{g.})]^2 + \frac{Ia + I^2b}{\hat{\sigma}^2 + \hat{\tau}^2}\sum_{g=1}^{G}[\bar{y}_{g.} - \alpha - \beta\bar{x}_{g.}]^2. \tag{30}$$

Substitute appropriate expressions for a and b in terms of $\hat{\sigma}^2$ and $\hat{\tau}^2$ to obtain

$$\frac{1}{\hat{\sigma}^2}\sum_{g=1}^{G}\sum_{i=1}^{I}[y_{gi} - \bar{y}_{g.} - \beta(x_{gi} - \bar{x}_{g.})]^2 + \frac{1}{\hat{\sigma}^2/I + \hat{\tau}^2}\sum_{g=1}^{G}[\bar{y}_{g.} - \alpha - \beta\bar{x}_{g.}]^2. \tag{31}$$

Differentiating with respect to α and β and setting partial derivatives equal to 0 shows that the value of β minimizing Eq. (31) is

$$\hat{\beta}_{WLS} = \frac{\sum_g \sum_i (x_{gi} - \bar{x}_{g.})(y_{gi} - \bar{y}_{g.})/\hat{\sigma}^2 + \sum_g (\bar{x}_{g.} - \bar{x}_{..})(\bar{y}_{g.} - \bar{y}_{..})/(\hat{\sigma}^2/I + \hat{\tau}^2)}{\sum_g \sum_i (x_{gi} - \bar{x}_{g.})^2/\hat{\sigma}^2 + \sum_g (\bar{x}_{g.} - \bar{x}_{..})^2/(\hat{\sigma}^2/I + \hat{\tau}^2)}. \tag{32}$$

Let $\hat{V}_{\hat{\beta}_1} = \hat{\sigma}^2/\sum_g \sum_i (x_{gi} - \bar{x}_{g.})^2$ be the estimate of $\hat{\beta}_1$ and let $\hat{V}_{\hat{\beta}_2} = (\hat{\sigma}^2/I + \hat{\tau}^2)/\sum_g (\bar{x}_{g.} - \bar{x}_{..})^2$ be the estimate of $\hat{\beta}_2$ from the first- and second-stage models, respectively.

Multiplying numerator and denominator of Eq. (32) by $\hat{V}_{\hat{\beta}_1}\hat{V}_{\hat{\beta}_2}$ shows that

$$\hat{\beta}_{WLS} = \frac{\hat{V}_{\hat{\beta}_2}\hat{\beta}_1 + \hat{V}_{\hat{\beta}_1}\hat{\beta}_2}{\hat{V}_{\hat{\beta}_1} + \hat{V}_{\hat{\beta}_2}} = \hat{\beta}_{pooled}.$$

Proof of Corollary: If $\eta \sim N(0, \mathbf{\Sigma}_{BIG}^{-1})$, then the MLEs of α, β, σ, and τ are the values minimizing

$$L(\alpha, \beta, \sigma, \tau) = \log\left(\det(\mathbf{\Sigma}_{BIG})\right) + (\mathbf{y} - \alpha - \beta\mathbf{x})^T \mathbf{\Sigma}_{BIG}^{-1} (\mathbf{y} - \alpha - \beta\mathbf{x}). \tag{33}$$

Let $\nu^2 = \sigma^2/I + \tau^2$. The value of $\det(\mathbf{\Sigma}_{BIG})$ is $[(\sigma^2)^{I-1}\nu^2]^G$. Now

$$\begin{aligned} L(\alpha, \beta, \sigma, \tau) \;=\; L(\alpha, \beta, \sigma, \nu) \;&=\; G(I-1)\log(\sigma^2) + \tfrac{1}{\sigma^2}\textstyle\sum_{g=1}^{G}\sum_{i=1}^{I}[y_{gi} - \bar{y}_{g.} - \beta(x_{gi} - \bar{x}_{g.})]^2 \\ &+\; G\log(\sigma^2) + \tfrac{1}{\nu^2}\textstyle\sum_{g=1}^{G}[\bar{y}_{g.} - \alpha - \beta\bar{x}_{g.}]^2. \end{aligned} \tag{34}$$

For any given value of β, the values of α, β, σ, and ν minimizing L are

$$\begin{aligned} \hat{\alpha} &= \bar{y}_{..} - \beta\bar{x}_{..} \\ \hat{\sigma}^2 &= \frac{\sum_{g=1}^{G}\sum_{i=1}^{I}[y_{gi} - \bar{y}_{g.} - \beta(x_{gi} - \bar{x}_{g.})]^2}{G(I-1)} \\ \hat{\nu}^2 &= \frac{\sum_{g=1}^{G}[\bar{y}_{g.} - \bar{y}_{..} - \beta(\bar{x}_{g.} - \bar{x}_{..})]^2}{G}. \end{aligned}$$

For given values of σ^2, ν^2, it follows from the theorem that the value of β minimizing L is

$$\hat{\beta} = \frac{\sum_g \sum_i (x_{gi} - \bar{x}_{g.})(y_{gi} - \bar{y}_{g.})/\sigma^2 + \sum_g (\bar{x}_{g.} - \bar{x}_{..})(\bar{y}_{g.} - \bar{y}_{..})/\nu^2}{\sum_g \sum_i (x_{gi} - \bar{x}_{g.})^2/\sigma^2 + \sum_g (\bar{x}_{g.} - \bar{x}_{..})^2/\nu^2},$$

which is Eq. (15).

5 Measurement Error in Air Pollution Exposure Assessment

5.1 Introduction

In epidemiologic studies, cumulative exposure during a period of time is generally thought of as the integral of the function that expresses the concentration of the substance to which the subject is exposed in terms of time. This can be written

$$E = \int_0^T C(t)dt. \tag{35}$$

Exposure assessment methods attempt to estimate the integral on the right hand side of (35).

We considered the assessment of cumulative exposure to an air pollutant over a fixed time interval. We studied two classes of methods, the indirect or microenvironmental approach (Fugas, 1975; Duan, 1982), and the direct or personal sampling approach (e.g. Koutrakis et al., 1991, 1993). Using a simple model for true exposure, we compared the accuracy of exposure assessments under each method under a variety of assumptions about the magnitudes of the measurement errors involved. We also assessed the impact of these errors on a hypothetical model that describes the effect of exposure on a health outcome.

5.2 The Microenvironmental Method

In the microenvironmental method, the region occupied by an individual during the time period of interest is divided into a number of microenvironments, e.g. home, school, office, car, outdoor areas, etc. The cumulative exposure is expressed in terms of the concentrations of the pollutant of interest and the lengths of time spent in each microenvironment. Let $M_1, ..., M_n$ represent a complete list of the microenvironments occupied by an individual, let t_m be the length of time during which microenvironment M_m is occupied, and let C_m be the average concentration of the pollutant of interest in M_m during the time that M_m is occupied. Then

$$E = \sum_{m=1}^{n} C_m t_m. \tag{36}$$

Estimation of E involves specifying the microenvironments, then estimating the concentrations C_m and the times t_m. Estimates of the C_m are generally based on measured ambient levels in outdoor locations, indoor levels, and indoor emission strengths. Estimates of the t_m are generally based on questionnaires or diary information. There are several exposure models available to compute the estimates of C_m and t_m. They can be classified

as statistical models, physical models, and physical-stochastic models (Sexton and Ryan, 1988; Ryan, 1991). Statistical models are based on statistical analyses of actual exposures and factors thought to influence them. Physical and physical-stochastic models are based on physical laws. Examples include the SHAPE model (Ott, 1984; Ott et al., 1988), the NEM and pNEM models (Johnson and Paul, 1983; McCurdy and Johnson, 1989), the SIMSYS model (Ryan et al., 1986), the NEM-SAI (Hayes 1989), and the REHEX model (Winer et al., 1989; Lurmann et al., 1989, Lurmann and Colome 1991). Some models, such as pNEM and REHEX, also estimate the dose received by the subjects as the product of the concentration and the ventilation rate. Dose may be a more relevant metric than exposure for epidemiologic research and ultimately all exposure models may need to address this parameter. However, the current analysis is limited to exposure in order to reduce the complexity of the analysis and because errors in assignment of ventilation rates are likely to be large.

In most models, the microenvironmental concentration is expressed as the product of the ambient level in an outdoor location and a proportionality factor (e.g. indoor/outdoor ratio), plus a term representing the contribution from non-ambient sources in the microenvironment. Thus, the microenvironmental concentration can be expressed in the form

$$C_m = f_m A + S_m, \tag{37}$$

where A is the ambient level, f_m is an estimated constant of proportionality expressing microenvironmental concentration as a fraction of the ambient level, and S_m is the concentration from non-ambient sources.

Efforts to validate microenvironmentally based methods of exposure assessment have been quite limited. There is some evidence that these models tend to estimate mean population exposure well, but are rather inaccurate in describing the tails of the distribution (Ott et al., 1988). No physical model has been adequately evaluated for ozone (Strock et al., 1985; Johnson et al., 1990; Liu et al., 1993; Liu et al., 1995, McCurdy 1994).

5.3 The Personal Sampling Method

Integrated personal sampling devices (badges) are instruments worn by individuals that attempt to measure the exposure of the individual to a pollutant. Continuous instruments, which can provide 1-minute average concentrations for up to 48 hours of sampling, are available for personal sampling of some pollutants, such as carbon monoxide. Development of continuous ozone monitors has begun (Penrose et al., 1994; Topham et al., 1992). Most personal sampling devices, however, can only measure the cumulative exposure over the period that they are worn. Both passive and active (pumped) badges are manufactured. Passive badges consist of a diffusion barrier followed by an impregnated filter with which the depositing pollutant reacts. Recently developed passive ozone badges include the Koutrakis sampler (Koutrakis et al., 1991, 1993; Liu et al., 1994), the DGA ozone badge

(Grosjean and Hishani, 1992), and the Kanno-Yanagisawa ozone badge (Kanno et al., 1992). Passive badges are also used for sampling NO_2 (Palmes et al., 1976) and formaldehyde (Geisling et al., 1982). Most personal samplers for particulates are active, involving the pumping of air through a filter that traps the particles. The filter is weighed both before and after sampling to determine the concentration. Avol et al. (1989) evaluated a pumped ozone badge that uses a solid monitoring reagent. The reagent develops a visible color change that is quantifiable with acetone extraction and spectrophotometric analysis. They concluded that this sampler was appropriate only for exposure periods of a few hours or less. Geyh et al. (1994) have reported on an active ozone sampler that is suitable for sampling periods up to 48 hours and is more accurate and precise than the Koutrakis passive sampler.

Integrated personal sampling devices are subject to several types of measurement error. Probably most serious for passive badges is sensitivity to wind velocity, since wind blowing into the badge effectively reduces the length of the diffusion chamber, increasing the deposit rate. Readings can vary according to badge placement. Placement near the nose is probably most accurate, but not necessarily most practical. Finally interference with other pollutants can occur. Personal samplers are probably less accurate for ozone than for other pollutants, both because personal ozone sampling technology is new, and because ozone is more reactive than most other pollutants.

5.4 Modeling the Microenvironmental and Personal Sampler Methods

The microenvironmental and personal sampling methods differ with respect to two principles. First, in the microenvironmental method, concentrations are measured with a model chosen by the experimenter. Thus many different exposure metrics can be handled. The method may be used to estimate peak exposure or proportion of time above a certain concentration, as well as cumulative exposure. In contrast, integrated personal sampling devices can only measure cumulative exposure for most pollutants. The second difference, which is quite important, is that while the microenvironmental method requires specifications of microenvironments and estimation of time durations, both of which are subject to error, the personal sampler always spends the correct amount of time in each microenvironment. Thus in terms of equation (36), the personal sampler is subject to error only in the measurement of the C_m, while the microenvironmental method is subject to measurement errors in the t_m as well.

To make a more detailed comparison of the two procedures, we model the measurement error in each, as well as the true exposure process. In practice there may be other sources of error, such as non-response bias, that might affect the accuracy of these procedures. We do not consider these here. We model a study involving I hypothetical communities with J subjects in each community, and we assume that each subject spends time in K

microenvironments. Let A_i be the true ambient ozone level in community i. For each microenvironment $m = 1, 2, ..., K$, let f_{ijm} be the concentration in microenvironment m to which subject j in community i is exposed, expressed as a fraction of the ambient level A_i, and let t_{ijm} be the duration of time spent in microenvironment m by subject j in community i. We ignore the contribution from non-ambient sources. It follows that the true exposure of subject ij is given by

$$E_{ij} = \sum_{m=1}^{K} f_{ijm} t_{ijm} A_i. \tag{38}$$

We model the assessment processes of both the microenvironmental and the personal sampling methods. We assume that estimated values of f_{ijm} and t_{ijm} are normally distributed around the true values with a standard deviation proportional to the true value. Since these values are subject to physical constraints, e.g. indoor/outdoor ratios are generally required to lie between 0 and 1, we must assume that the standard deviations are small enough so that virtually all of the mass of the normal curve lies within these limits. This is the case for the standard deviation values we used in our simulations below. For the personal sampling method we specify a standard deviation ν, then generate an estimate $\hat{f}^P_{ijm} \sim N(f_{ijm}, \nu^2 f^2_{ijm})$ for each subject and microenvironment. The personal sampling estimate of exposure is then

$$\hat{E}^P_{ij} = \sum_{m=1}^{K} \hat{f}^P_{ijm} t_{ijm} A_i. \tag{39}$$

For the microenvironmental procedure, we specify a standard deviation τ_f, then generate an estimate $\hat{f}^M_{ijm} \sim N(f_{ijm}, \tau_f^2 f^2_{ijm})$ for each subject and microenvironment. To model the errors in estimating the t_{ijm}, we specify a standard deviation τ_t, then generate random errors $\varepsilon_{ijm} \sim N(0, \tau_t^2 t^2_{ijm})$ for each subject and microenvironment. We then set the estimate $\hat{t}^M_{ijm} = t_{ijm} + \varepsilon_{ijm} - \sum_{m=1}^{K} \varepsilon_{ijm}/K$. In this way $\sum_{m=1}^{K} \hat{t}^M_{ijm} = \sum_{m=1}^{K} t_{ijm}$ as required. The microenvironmental estimate of exposure is

$$\hat{E}^M_{ij} = \sum_{m=1}^{K} \hat{f}^M_{ijm} \hat{t}^M_{ijm} A_i. \tag{40}$$

Under our model, the estimates $\hat{f}^M_{ijm}$, $\hat{t}^M_{ijm}$, and $\hat{f}^P_{ijm}$ are unbiased, and for any given subject and microenvironment, the values $\hat{f}^M_{ijm}$ and $\hat{t}^M_{ijm}$ are independent. It follows that $\hat{E}^P_{ij}$ and $\hat{E}^M_{ij}$ are both unbiased for the true exposure E_{ij}, so the performances of the two estimators can be measured by comparing their variances. The variance of $\hat{E}^P_{ij}$ is

$$\mathrm{Var}(\hat{E}^P_{ij}) = \sum_{m=1}^{K} \nu^2 f^2_{ijm} t^2_{ijm} {A_i}^2. \tag{41}$$

The variance of $\hat{E}^M_{ij}$ is more complicated, since two random errors are involved, and since the errors in the $\hat{t}^M_{ijm}$ are correlated. Let $\mathbf{\Sigma}$ be the $K \times K$ covariance matrix of the $\hat{t}^M_{ijm}$. The kl element of $\mathbf{\Sigma}$ is

$$\mathrm{Cov}(t^{(M)}_{ijk}, t^{(M)}_{ijl}) = \begin{cases} \tau_t^2 \sum_{m=1}^{K} (t_{ijm}/K)^2 - t^2_{ijk}/K - t^2_{ijl}/K & \text{if } k \neq l \\ \tau_t^2 \sum_{m=1}^{K} (t_{ijm}/K)^2 + (K-2) t^2_{ijk}/K & \text{if } k = l. \end{cases} \tag{42}$$

Let $\mathbf{f_{ij}}$ be the vector $(f_{ij1}, ..., f_{ijK})$ considered as a column. The variance of $\hat{E}_{ij}^M$ is

$$\mathrm{Var}(\hat{E}_{ij}^M) = {A_i}^2(\mathbf{f_{ij}}^T \mathbf{\Sigma f_{ij}} + \tau_f^2 \sum_{m=1}^K f_{ijm}^2 t_{ijm}^2 + \tau_f^2 \tau_t^2 \sum_{m=1}^K f_{ijm}^2 [(K-2)t_{ijm}^2/K + \sum_{m=1}^K (t_{ijm}/K)^2]). \quad (43)$$

5.5 Estimating Health Effects — A Simulation Study

To assess the impact of measurement error on the estimation of health effects, we analyzed a linear model in which a health outcome (e.g. FEV_1) is modeled as a linear function of exposure. We assume we have J subjects in each of I communities. The model is

$$F_{ij} = \alpha_i + \beta E_{ij} + \varepsilon_{ij}, \quad (44)$$

where F_{ij} is FEV_1 (or other health outcome) measured on the jth subject in the ith community, expressed as a percentage of a baseline value for that subject. The baseline may be derived from other measurements of the same subject taken under relatively pollution-free conditions, or from measurements on a control group matched on variables such as height, weight, sex, race and age. The value α_i is a community-specific intercept, E_{ij} is the true exposure of subject ij, and ε_{ij} is a random error.

In order to perform the simulation, we specified values for the parameters in the model. We set the number of communities I to 12 and the number of subjects J per community to 300, to match the California Air Resources Board Children's Health Study. For the sake of simplicity, we assumed that each subject spends time in only three microenvironments: home, school, and outdoors. To compute "true" exposures E_{ij}, we specified values f_{ijm}, t_{ijm}, and A_i as follows: For the A_i we took the average daily maximum ozone concentrations during the years 1986–1990 for the 12 communities in the CARB Children's Health Study. These communities and the corresponding ambient levels are given in Table 12. To generate true values for the f_{ijm}, we specified a mean μ and a standard deviation σ for the ratios of the concentrations in each microenvironment to the ambient level A_i measured at a monitoring station. These specifications were based on the available literature on indoor/outdoor ozone ratios. Yocum (1982) concluded in a review that indoor/outdoor ratios generally fall between 0.1 and 0.7. Druzik et al. (1989) found ratios ranging from 0.24 to 0.75 for a variety of buildings. Wechsler (1989) studied three indoor sites with very different ventilation rates and found ratios ranging from 0.24 to 0.71. There appears to be little work available on residences as opposed to public buildings. To be consistent with published findings for buildings, we specified a mean ratio $\mu = 0.5$ and a standard deviation of $\sigma = 0.1$ for schools. For homes we took $\mu = 0.5$ and $\sigma = 0.15$, on the assumption that ratios would vary somewhat more from house to house than from school to school. Somewhat arbitrarily, we took $\mu = 1$ and $\sigma = 0.2$ for ratios of outdoor concentrations to the values A_i measured at monitoring stations. Then for each subject ij and each microenvironment m, we generated a true value f_{ijm} by generating an observation

from the normal distribution with mean and standard deviation appropriate to the microenvironment.

Table 12: Twelve Communities and Their Mean Peak Ozone Levels

Community	Ozone Level
Alpine	80.5
Lake Elsinore	82.7
Lake Gregory	98.9
Lancaster	70.8
Lompoc	42.7
Long Beach	45.2
Mira Loma	95.1
Riverside	95.1
San Dimas	109.2
Atascadero	58.7
Santa Maria	30.2
Upland	92.0

Measurements are average daily maximum for the period 1986–1990.
Units are ppb.

In specifying values for the times t_{ijm} spent in the various microenvironments, we were constrained by the requirement that $\sum_{m=1}^{K} t_{ijm} = 24$, when t_{ijm} is measured in hours per day. We therefore generated preliminary values t^*_{ijm} for each m and took $t_{ijm} = 24t^*_{ijm}/\sum_{m=1}^{K} t^*_{ijm}$. To generate the values t^*_{ijm} we used the normal distribution, specifying a mean μ and standard deviation σ for each microenvironment. To specify the mean value for percentage of time spent outdoors, we relied on findings of Wiley (1991) that indicated that California schoolchildren spend about 10% of their time outdoors on average. That percentage stays approximately constant over different times of the year, and on weekends as well as weekdays. We therefore took $\mu = 3$ and $\sigma = 1$ for the number of hours per day spent outdoors. We took $\mu = 6$ and $\sigma = 2$ for the number of hours in school, and $\mu = 15$ and $\sigma = 2$ for the number of hours spent at home.

Our simulation was based on 1000 artificial data sets. For each data set we first generated for each subject a "true" exposure E_{ij} computed from (38). We then generated measured exposures $\hat{E}^M_{ij}$ and $\hat{E}^P_{ij}$ as described in section 5.4, using one of several values for the standard deviations ν, τ_f, and τ_t. To fit model (44), we took ε_{ij} to be normally distributed with mean 0 and standard deviation 13. This standard deviation was estimated from between-individuals data on approximately 3600 Southern California school children tested in the spring of 1993 in the CARB Children's Health Study.

Our model assumes that measured values are normally distributed around the true values. Therefore it should provide reasonably accurate results whenever measured values are

distributed reasonably symmetrically around the true values. If measurements are severely biased, or skewed as a result of outliers, the results may be different that those described here.

We took the exposure-response slope β equal to -0.10, indicating a 1% decrease in FEV_1 per 10 ppb increase in exposure. All the α_i were taken to have the common value $100 - \beta\bar{E}$, where $\bar{E}$ is the average of the true exposures E_{ij} over all subjects in all communities. In this way the expected value of F_{ij} is 100, as required.

The variances of $\hat{E}_{ij}^M$ and $\hat{E}_{ij}^P$ were well estimated. Table 13 shows true and estimated variances of $\hat{E}_{ij}^M$ and $\hat{E}_{ij}^P$ averaged over i and j for several values of τ_f^2, τ_m^2, and ν^2. True and estimated variances were computed as described in section 5.4.

Table 13: True and Estimated Variances of Microenvironmental and Personal Sampler Estimates of Exposure

Exposure Estimate	True Variance	Estimated Variance
Microenvironmental ($\hat{E}_{ij}^M$)		
$\tau_f = .1$, $\tau_t = .1$	10.25	10.31
$\tau_f = .1$, $\tau_t = .2$	14.58	14.78
$\tau_f = .1$, $\tau_t = .3$	21.78	22.15
$\tau_f = .2$, $\tau_t = .1$	36.95	37.18
$\tau_f = .2$, $\tau_t = .2$	42.00	42.88
$\tau_f = .2$, $\tau_t = .3$	50.42	52.26
$\tau_f = .3$, $\tau_t = .1$	81.43	81.91
$\tau_f = .3$, $\tau_t = .2$	87.68	89.63
$\tau_f = .3$, $\tau_t = .3$	98.18	102.27
Personal Sampler ($\hat{E}_{ij}^P$)		
$\nu = .05$	2.20	2.20
$\nu = .10$	8.81	8.82
$\nu = .15$	19.84	19.84
$\nu = .20$	35.27	35.26
$\nu = .25$	55.11	55.10
$\nu = .30$	79.39	80.38

Variances are the averages over all subjects ij of $\text{Var}(\hat{E}_{ij}^M)$ and $\text{Var}(\hat{E}_{ij}^P)$.

We fit model (44) using $\hat{E}_{ij}^M$ and $\hat{E}_{ij}^P$ in turn in place of the unknown E_{ij}. We also fit an ecological version of the model in which all the α_i were assumed equal, and in which the ambient level A_i was used to estimate the true exposure E_{ij} for all the subjects in community i. Table 14 gives results based on 1000 simulated data sets. For each data set the least squares estimate $\hat{\beta}$ and the conventional estimate of its standard error was

Table 14: Performance of Dose-response Slope Estimator When FEV_1 is Regressed Against Various Estimates of Exposure: No Bias Correction Applied

Exposure Estimate	Mean	SD	Nominal SE	Bias	Root Mean Square Error
True Exposures	-0.099	0.027	0.027	0.001	0.027
Outdoor Ambient Level	-0.057	0.009	0.009	0.043	0.044
Microenvironmental ($\hat{E}_{ij}^M$)					
$\tau_f = .1$, $\tau_t = .1$	-0.086	0.025	0.025	0.014	0.029
$\tau_f = .1$, $\tau_t = .2$	-0.083	0.024	0.024	0.017	0.030
$\tau_f = .1$, $\tau_t = .3$	-0.076	0.024	0.023	0.024	0.034
$\tau_f = .2$, $\tau_t = .1$	-0.064	0.021	0.021	0.036	0.042
$\tau_f = .2$, $\tau_t = .2$	-0.061	0.020	0.021	0.039	0.044
$\tau_f = .2$, $\tau_t = .3$	-0.057	0.020	0.020	0.043	0.047
$\tau_f = .3$, $\tau_t = .1$	-0.045	0.018	0.018	0.055	0.058
$\tau_f = .3$, $\tau_t = .2$	-0.042	0.018	0.018	0.058	0.060
$\tau_f = .3$, $\tau_t = .3$	-0.040	0.017	0.017	0.060	0.063
Personal Sampler ($\hat{E}_{ij}^P$)					
$\nu = .05$	-0.097	0.026	0.026	0.003	0.026
$\nu = .10$	-0.087	0.025	0.025	0.013	0.028
$\nu = .15$	-0.077	0.023	0.023	0.022	0.032
$\nu = .20$	-0.065	0.022	0.022	0.035	0.041
$\nu = .25$	-0.053	0.021	0.020	0.046	0.051
$\nu = .30$	-0.045	0.019	0.018	0.055	0.058

True value of parameter is -0.10. Sample size is 3600.

calculated. The column labeled "Mean" is the average of the 1000 values of $\hat{\beta}$. The column labeled "SD" is the sample standard deviation of these 1000 estimates. This is a consistent estimate of the true standard error when the true exposure is estimated. The column labeled "Nominal SE" is the square root of the average of the 1000 conventional estimates of $\text{Var}(\hat{\beta})$. The column labeled "Bias" gives the difference between the mean and the true value of -0.10. The column labeled "Root Mean Square Error" gives the quantity $\sqrt{\text{SE}^2 + \text{Bias}^2}$, which is the usual measure of the accuracy of an estimator. The first row of the table gives the results when the true exposures E_{ij} are used as the dependent variable. The estimate is unbiased. In comparison, estimates based on the ambient level, $\hat{E}_{ij}^M$, or $\hat{E}_{ij}^P$, are biased severely toward 0. In the case of $\hat{E}_{ij}^M$ and $\hat{E}_{ij}^P$, the bias is due to nondifferential measurement error. In the case of the ambient level, it is due to the fact that the ambient level strongly overestimates the exposure, since it is equivalent to assuming that subjects are outside all the time. In principle, the conventional estimates of the standard deviation of $\hat{\beta}$ should tend to underestimate the true standard error slightly. This should be the case

Table 15: Performance of Exposure-response Slope Estimator When FEV_1 is Regressed Against Various Estimates of Exposure: Bias Correction Applied

Exposure Estimate	Mean	SD	Nominal SE	Bias	Root Mean Square Error
True Exposures	-0.099	0.027	0.027	0.001	0.027
Outdoor Ambient Level	-0.057*	0.009	0.009	0.043	0.044
Microenvironmental ($\hat{E}_{ij}^M$)					
$\tau_f = .1$, $\tau_t = .1$	-0.099	0.029	0.029	0.001	0.029
$\tau_f = .1$, $\tau_t = .2$	-0.101	0.029	0.030	-0.001	0.029
$\tau_f = .1$, $\tau_t = .3$	-0.101	0.032	0.031	-0.001	0.032
$\tau_f = .2$, $\tau_t = .1$	-0.100	0.034	0.034	-0.000	0.034
$\tau_f = .2$, $\tau_t = .2$	-0.101	0.033	0.034	-0.001	0.033
$\tau_f = .2$, $\tau_t = .3$	-0.102	0.035	0.036	-0.002	0.035
$\tau_f = .3$, $\tau_t = .1$	-0.102	0.040	0.040	-0.002	0.040
$\tau_f = .3$, $\tau_t = .2$	-0.100	0.042	0.041	0.000	0.042
$\tau_f = .3$, $\tau_t = .3$	-0.100	0.043	0.043	-0.000	0.043
Personal Sampler ($\hat{E}_{ij}^P$)					
$\nu = .05$	-0.100	0.029	0.030	-0.000	0.027
$\nu = .10$	-0.099	0.029	0.029	0.001	0.029
$\nu = .15$	-0.101	0.030	0.031	-0.001	0.030
$\nu = .20$	-0.100	0.034	0.033	-0.000	0.034
$\nu = .25$	-0.099	0.038	0.036	0.001	0.038
$\nu = .30$	-0.099	0.041	0.040	0.001	0.041

*No bias correction is applied to the ambient level estimate.

True value of parameter is -0.10. Sample size is 3600.

because the conventional estimator $\hat{\sigma}^2 / \sum_{i=1}^{I} \sum_{j=1}^{J} (\hat{E}_{ij} - \hat{E}_{i.})^2$ is a consistent estimator of the expectation of the conditional variance $\mathrm{Var}(\hat{\beta}|\hat{E}_{ij})$. The desired variance is the unconditional variance $\mathrm{Var}(\hat{\beta})$, which is equal to $\mathrm{E}[\mathrm{Var}(\hat{\beta}|\hat{E}_{ij})] + \mathrm{Var}[\mathrm{E}(\hat{\beta}|\hat{E}_{ij})]$, where $\mathrm{E}(\hat{\beta}|\hat{E}_{ij}) = \beta \sum_{i=1}^{I} \sum_{j=1}^{J} (\hat{E}_{ij} - \hat{E}_{i.})(E_{ij} - E_{i.}) / \sum_{i=1}^{I} \sum_{j=1}^{J} (\hat{E}_{ij} - \hat{E}_{i.})^2$. Since this quantity is not constant, its variance is strictly positive. However, the variance tends to 0 as the product $IJ \to \infty$, and appears to be negligible in our simulation.

If the parameters τ_f^2, τ_t^2, and ν^2 are known, the variances of $\hat{E}_{ij}^P$ and $\hat{E}_{ij}^M$ can be estimated, and the bias in $\hat{\beta}$ can be corrected. A consistent estimator of $\hat{\beta}$ is obtained by multiplying the conventional estimator by the correction factor

$$F = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} (\hat{E}_{ij} - \hat{E}_{i.})^2}{\sum_{i=1}^{I} \sum_{j=1}^{J} (\hat{E}_{ij} - \hat{E}_{i.})^2 - \frac{J}{J-1} \sum_{i=1}^{I} \sum_{j=1}^{J} \mathrm{Var}(\hat{E}_{ij})} \tag{45}$$

The denominator of the expression (45) is an unbiased and consistent estimator of $\sum_{i=1}^{I}\sum_{j=1}^{J}(E_{ij}-E_{i.})^2$. Since by Jensen's inequality $\mathrm{E}(1/X) > 1/\mathrm{E}(X)$ for positive random variables X, we might in principle expect a slight bias away from 0 when this correction factor is applied. A consistent estimate of the standard error can be obtained by multiplying the conventional estimate by the correction factor F. Table 15 gives results when the bias correction is applied. No remaining bias is noticeable.

6 Conclusions

The bidirectional case-crossover design is a valuable method for estimating acute effects of environmental exposures that takes advantage of two features of studies involving environmental exposures: (1) accurate information about past exposure is often available, so model-based estimates of exposure are not needed, and (2) levels of exposure are unaffected by the occurrence of failure of a subject. In contrast to other case-crossover designs, control information is assessed both before and after failure, which eliminates confounding due to time trends in exposure.

The proposed two-stage analysis of the multilevel analytic design provides asymptotically unbiased and efficient estimation of effects in a complex model involving unmeasured between-groups differences, measurement error, and a complex measurement model combining individual and aggregate exposure data. In particular, in cases where the within-groups exposure variance is less than the between-groups variance, estimates obtained through pooling can be more efficient than estimates based on either individual level or aggregate level analyses alone. Simulation techniques can be used to optimize the various trade-offs between the design parameters if reasonable estimates of the model parameters are available. Since variables are measured on the individual level, the model should be relatively free from nonadditivity bias, but could be subject to bias in situations where its assumptions are seriously violated. We believe this design and its associated analysis offers considerable promise for resolving some of the difficulties of between-group confounding, measurement error, and restricted variability that have historically plagued environmental epidemiology.

Our work on measurement error in exposure assessment represents an initial attempt to address some issues in the accuracy of estimated health effects of air pollution under various forms of exposure assessment. Refinement of the techniques used here will lead to more accurate results. It is quite clear from our work so far that neither the microenvironment nor the personal sampler method produces reliable estimates of exposure-response slope when measurement error is uncorrected. In our examples, the bias was toward 0 under the assumptions of our model, because of non-differential measurement error. If the measurement error were to involve bias that was correlated with the response, a bias away from 0 could result as well.

Our simulation studies show that the standard error of the estimate of a health effect increases as the errors in exposure assessment increase. When the fraction of the ambient level in a microenvironment is estimated with a standard error of 30%, the standard error of the estimate is 50% higher than it would be if the true exposures were known. Interestingly, it appears that errors in estimating indoor/outdoor ratios have much more influence on the accuracy of the microenvironmental approach than do errors in estimating the time spent in the microenvironments.

In order to adjust for measurement error, something about its distribution must be known. To correct the bias in $\hat{\beta}$, the variance (and bias, if any) in the measurement error must be known. Because of the great impact of measurement error, it is clear that careful validation studies are necessary before exposure assessment methods can be relied on to provide accurate estimates of health effects of pollutants.

7 Acknowledgments

We would like to thank Dr. Aaron Cohen of HEI for his unflagging encouragement and support of this project. Dr. Cohen led us to find many fruitful avenues of research that otherwise would have gone unexplored. This project owes a lot to his inspiration. We also appreciate the efforts of Dr. Geoffrey Sunshine of HEI, who assisted us in writing this report and made valuable suggestions regarding its organization. Ms. Virgi Hepner of HEI developed creative solutions to the many problems involved with getting the manuscript into print, which enabled us to present our findings in our own style.

Several of our colleagues on the USC Children's Health Study research team were instrumental in the success of our work. Dr. John Peters helped us put the data in a realistic context, and his knowledge of the epidemiology of respiratory disease enabled us to formulate more realistic statistical models. Fred Lurmann provided expertise on exposure assessment that lies behind much of the work in section 5. Ed Avol provided us with much background knowledge of air pollution and pulmonary function assessment that enabled us to formulate our results in ways more relevant to actual practice.

8 References

Avol, E., K. Anderson, J. Whynot, B. Daube, W. Linn, and J. Hackney (1989). Progress in personal and area ozone sampling using solid reagents. Paper presented at the Conference of Total Exposure Assessment Methodology: New Horizons, U.S. EPA and the A&WMA, Las Vegas, Nevada, November 27–30, 1989.

Armstrong, B.G. (1990). The effects of measurement errors on relative risk regressions. Am. J. Epidemiol. 132, 1176–1184.

Brenner H., D.A. Savitz, K.H. Jöckel, and S. Greenland (1992). Effects of nondifferential exposure misclassification in ecologic studies. Am. J Epidemiol 135, 85–95.

Druzik, J.R., M.S. Adams, C. Tiller, and G.R. Cass (1989). The measurement and model predictions of indoor ozone concentrations in museums. Atmos. Environ. 7, 1813–1823.

Duan, N. (1982). Models for human exposure to air pollution. Environ. Int. 8, pp 305–309.

Feldman, U. (1993a). Epidemiologic assessment of risks of adverse reactions associated with intermittent exposure. Biometrics 49, 419–428.

Feldman, U. (1993b). Design and analysis of drug safety studies. J Clin Epidemiol 46, 237–244.

Fugas, M. (1975) Assessment of the total exposure to an air pollutant. Proceedings of the International Symposium on Environmental Monitoring, Las Vegas, Nevada. Institute of Electrical and Electronic Engineers, Inc., NY, 2, pp 38–45.

Geisling, K.L., M.K. Tashima, J.R. Girman, R.P. Miksch, and S.M. Rappaport (1982). A Passive sampling device for determining formaldehyde in indoor air. Environ. Int. 8, pp 153–158.

Geyh, A., J.M. Wolfson, P. Koutrakis, and J. Mulik (1994). The development of an active personal ozone sampler using a hollow tube diffusion denuder. Paper submitted for the Proceedings form; the Air & Waste Management Association Toxics and Ozone Precursor Measurement Conference, Durham, North Carolina, May 3–6, A&WMA.

Greenland, S. (1980). The effect of misclassification in the presence of covariates. Am. J. Epidemiol. 112, 564–569.

Greenland, S. (1988). Statistical uncertainty due to misclassification: implications for validation substudies. J. Clin. Epidemiol. 41, 1167–1174.

Greenland, S. (1996). Confounding and exposure trends in case-crossover and case-time-control designs. Epidemiology 7, p231.

Greenland, S. and J. Robins (1994). Ecologic studies: biases, misconceptions, and counterexamples. Am J Epidemiol, 139, 747–760.

Grosjean, D. and M.W.M. Hishani (1992). A passive sampler for atmospheric ozone. J. Air Waste Manage. Assoc. 42, pp 169–173.

Hayes, S.R. (1989). Estimating the effect of being indoors on total personal exposure to outdoor air pollution. JAPCA 39, pp 1453–1461.

Johnson, T., and R. Paul (1983). The NAAQS model (NEM) applied to carbon monoxide. U.S. Environmental Protection Agency, EPA 450/5-83-003.

Johnson, T., R. Paul, J. Capel, and T. M^{c}Curdy (1990). Estimation of ozone exposure in Houston using a probabilistic version of NEM. Paper No. 90–150.1, 83rd annual meeting of A&WMA.

Kanno, S. and Y. Yanagisawa (1992). Passive ozone/oxidant sampler with coulometric determination using I_2/Nylon-6 charge-transfer complex. Environ./ Sci. Technol. 26, pp 744–749.

Koutrakis, P., J.M. Wolfson, H.H. Suh, and J.D. Mulik (1991). Use of personal measurements for ozone exposure assessment. Paper presented at the International Society of Exposure Analysis Conference: Measuring, Understanding, and Predicting Exposures in the 21st Century, November 20, 1991.

Koutrakis, P., J. Wolfson, A. Bunyaviroch, S.E. Forchlich, K. Hirano, and J.D. Mulik (1993). Measurements of ozone using a nitrite coated filter. Analytical Chemistry 65, 209–214.

Langholz, B. M. and Goldstein, L. (1997). Fitting logistic models using the conditional logistic likelihood when there are large strata. Technical Report #113, Department of Preventive Medicine, University of Southern California.

Liu, L.-J. S., P. Koutrakis, H.H. Suh, J.D. Mulik, and R.B. Burton (1993). Use of personal measurements for ozone exposure assessment: a pilot study. Environmental Health Perspectives 101, 318–324.

Liu, L.-J. S., M.P. Olson III, G.A. Allen, and P. Koutrakis (1994). Evaluation of the Harvard ozone passive sampler on human subjects. Environmental Science and Technology 28, 504–514.

Liu, L.-J. S., P. Koutrakis, J. Leech, and I. Broder (1995). Assessment of ozone exposures in the greater metropolitan Toronto area. J. Air & Waste Manag. Assoc. 45, 223–234.

Lurmann, F.W., A.M. Winer, and S.D. Colome (1989). Development and application of a new regional human exposure (REHEX) model. Paper presented at the Conference on Total Exposure Assessment Methodology: New Horizons, U.S. EPA and the A&WMA, Las Vegas, Nevada November 27–30, 1989.

Lurmann, F.W. and S.D. Colome (1991). 1991 AQMP revision, Technical Report V-H: An assessment of current and future human exposure to ozone in the South Coast air basin. Report by Sonoma Technology Inc., Santa Rosa, CA, STI-90070-1022-FR.

Maclure, M. (1991). The case-crossover design: A method for studying transient effects on the risk of acute events. The American Journal of Epidemiology 133, 144–153.

Marshall, R. and Jackson, R. (1993). Analysis of case-crossover designs. Statistics in Medicine 12, 2333–2341.

Mittleman M.A., Maclure, M., and Robins, J.M. (1995). Control sampling strategies for case-crossover studies: An assessment of relative efficiency. American Journal of Epidemiology, 142, 91–98.

M^{c}Curdy, T., and T. Johnson (1989). Development of a human exposure model for ozone. Paper presented at the Conference on Total Exposure Assessment Methodology: New Horizons, U.S. EPA and the A&WMA, Las Vegas, Nevada November 27–30, 1989.

M^{c}Curdy, T. (1994). Human exposure to ambient ozone. In Trops. Ozone, D. McKee, ed., Lewis Publishers, Ann Arbor.

Ott, W.R. (1984). Exposure estimates based on computer generated activity patterns. J. Toxic. Clin. Toxicol. 21, pp 97–128.

Ott, W., J. Thomas, D. Mage, and L. Wallace (1988). Validation of the simulation of human activity and pollutant exposure (SHAPE) model using paired days from the Denver, CO, carbon monoxide field study. Atmos. Environ. 22, pp 2101–2113.

Palmes, E.D., A.F. Gunnison, J. Dimattio, and C. Tomczyk (1976). Personal Sampler for NO_2. Amer. Ind. Hyg. Assoc. J. 37, pp 570–577.

Penrose, W.R., L. Pan, and W. Ollison (1994). Real time electrochemical measurement of ozone in the presence of nitrogen dioxide. In: Measurement of Toxic and Related Air Pollutants, Air and Waste Management Association, Report VIP-39.

Peters, J. (1996). Epidemiologic investigation to identify chronic health effects of ambient air pollutants in Southern California: Phase II final report, prepared for the California Air Resources Board under Research Contract #A033–186.

Prentice, R.L., and Sheppard, L. (1990). Dietary fat and cancer: consistency of the epidemiologic data, and disease prevention that may follow from a practical reduction in fat consumption. Cancer Causes Control 1, 81–97.

Richardson, S., I. Stücker, and D. Hémon (1987). Comparison of relative risks obtained in ecological ind individual studies. Int J Epidemiol, 16, 111–120.

Rosner, B., and Willett, W.C. (1988). Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. Am. J. Epidemiol. 127, 377–386.

Rosner, B., Willett, W.S., and Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat. Med. 8, 1031–1040.

Ryan, P.B., J.D. Spengler, and R. Letz (1986). Estimating personal exposures to NO_2. Environ. Int. 12, pp 394–400.

Ryan, P.B. (1991). An overview of human exposure modeling. J. Exposure Analysis and Environ. Epidem. 1, pp 453–474.

Schwartz, J. (1994). What are people dying of on high air pollution days? Environmental Research, 64, 26–35.

Schwartz, J. and Dockery, D. (1992). Increased mortality in Philadelphia associated with daily air pollution concentrations. Am Rev Respir Dis, 145, 600–604.

Selvin, H.C. (1958). Durkheim's suicide and problems of empirical research. Am. J. Sociol. 63, 607–619.

Sexton, K., and P.B. Ryan (1988). Assessment of human exposure to air pollution: methods, measurements and models. In: Air Pollution, the Automobile and Public Health, A.Y. Watson, R.R. Bates, and D. Kennedy (eds.), National Academy Press, New York, pp 207–238.

Spiegelman, D., and Gray, R. (1991). Cost efficient study designs for binary response data with Gaussian covariate measurement error. Biometrics 47, 851–869.

Strock, T.H., D.J. Kotchman, C.F. Contant, et al. (1985). The estimation of personal exposures to air pollutants for a community-based study of health effects in asthmatics — design and results of air monitoring. JAPCA 35, 1266–1276.

Thomas, D., Stram, D., and Dwyer, J. (1993). Exposure measurement error: influence on exposure-disease relationships and methods of correction. Ann. Rev. Publ. Health 14, 69–93.

Topham, L.A., G.I. Mackay, and H.I. Schiff (1992). Performance assessment of the portable and lightweight LOZ-3 chemiluminescent type ozone monitor. In: Measurement of Toxic and Related Air Pollutants, Air and Waste Management Association, Report VIP-25.

Von Korff, M., T. Koepsell, S. Curry, and P. Diehr (1992). Multilevel analysis in epidemiologic research on health behaviors and outcomes. Am J Epidemiol 135, 1077–1082.

Wechsler, C.J., H.C. Shields, and D.V. Naik (1989). Indoor ozone exposures. JAPCA 39, pp 1562–1568.

Wiley, J.A. (1991). Study of children's activity patterns. Final report to the California Air Resources Board by the Survey Research Center, University of California at Berkeley, CA on Contract No. A7-33-149.

Winer, A.M., F.W. Lurmann, L.A. Coyner, S.D. Colome, and M. Po (1989). Characterization of air pollutant exposures in the South Coast air basin: Application of a new regional human exposure (REHEX) model. Final report to South Coast Air Quality Management District/California State University Fullerton Foundation, Report No. TSA 106-01-88.

Yocum, J.E. (1982). Indoor-outdoor air quality relationships: a critical review. JAPCA 32, pp 500–520.

9 About the Authors

William Navidi received his B.A. in mathematics from New College, and his Ph.D. in statistics from the University of California at Berkeley. He was Associate Professor of Preventive Medicine at the University of Southern California, and is currently Associate Professor of Mathematical and Computer Sciences at the Colorado School of Mines.

Duncan Thomas received his B.A. in mathematics and astronomy from Haverford College, and his Ph.D. in epidemiology from McGill University. He is currently Professor of Preventive Medicine at the University of Southern California.

Bryan Langholz received his B.A. in mathematics from Humboldt State University and his Ph.D. in Biostatistics from the University of Washington. He is currently Associate Professor of Preventive Medicine at the University of Southern California.

Daniel Stram received his B.A. in mathematics from Tufts University and his Ph.D. in statistics from Temple University. He is currently Associate Professor of Preventive Medicine at the University of Southern California.

10 Publications Resulting from this Research

Navidi, W., Thomas, D., Stram, D., and Peters, J. (1994). Design and analysis of multilevel analytic studies with applications to a study of air pollution. Environmental Health Perspectives 102, (Suppl 8), pp 25–32.

Navidi, W. and Lurmann, F. (1995). Measurement error in air pollution exposure assessment. Journal of Exposure Analysis and Environmental Epidemiology, 5, pp 111–124.

Navidi, W. (1997). Bidirectional case-crossover designs for exposures with time trends. Biometrics, 54, 596–605.

11 Abbreviations

α_g	baseline outcome for group g
β	increase in log odds of failure due to a one-unit increase in exposure
$\hat{\beta}$	an estimate (e.g., maximum likelihood estimate) of β
$\hat{\beta}_{1-1}$	bidirectional case-crossover estimator with a single control day for each subject randomly chosen from among all control days
$\hat{\beta}_{BI}$	bidirectional case-crossover estimator using the full stratum of control days for each subject
$\hat{\beta}_{lag14}$	conventional case-crossover estimator with control times 14 days before failure
$\hat{\beta}_{lag28}$	same as above, with control times 28 days before failure
$\hat{\beta}_{TH}$	total history version of case-crossover estimator in which all days prior to failure are used as controls
λ_i	baseline log odds for subject i
A_i	set of days on which subject i is absent
D_{n_i}	the collection of all sets of n_i days
E	expectation
$I(\beta)$	Fisher information, used to estimate the variance of the maximum likelihood estimator $\hat{\beta}$
$L(\beta)$	log likelihood
N_{ij}	number of subjects at risk in community i on day j
n_i	number of days that subject i is absent
p_{ij}	probability that subject i is absent on day j
PM_{10}	particulate matter smaller than 10 microns in diameter
RMSE	root mean square error
S	arbitrary set containing n_i days
SD	standard deviation
SE	standard error
$t_1, ..., t_M$	list of days a subject is at risk
v_{gi}	confounder variable for subject i in group g
X_g	mean exposure for group g
X_{ij}	exposure of subject i on day j
x_{gi}	exposure variable for subject i in group g
y_{gi}	outcome variable for subject i in group g

INTRODUCTION

The Health Effects Institute (HEI)* has had a long-standing interest in the long-term effects of air pollutants on human health. In 1991, HEI set up the Environmental Epidemiology Planning Project to identify key methodological issues that needed to be addressed in the design and interpretation of future environmental epidemiologic studies of the health effects of air pollutants. Leading scientists identified several major concerns, including the relatively low levels of air pollutants to which populations are generally exposed and the consequent difficulty in establishing health risk; the difficulties inherent in measuring exposure to different pollutants; and the frequently high correlation among different pollutants, which makes it difficult to identify the effect of a particular agent. At the outcome of the Environmental Epidemiology Planning Project, further research on the design, measurement, data analysis, and interpretation of epidemiologic studies was recommended (Morgenstern and Thomas 1994; Hatch and Thomas 1994; Prentice and Thomas 1994). To meet these research needs, HEI issued Request for Applications (RFA) 91-1, *Epidemiological Studies of the Health Effects of Long-Term Ozone Exposure*, in 1991. The major goals of the RFA were to support (1) the development of epidemiologic methods for studying the effects of long-term ozone exposure, and (2) the development and testing in actual data sets of innovative approaches to the control of confounding of ozone effect measurements by other air pollutants.

In response to the RFA, Dr. Navidi and colleagues at the University of Southern California (USC) submitted a proposal entitled "Statistical Methods for Epidemiological Studies of the Health Effects of Air Pollution." The goal of the proposed study was to develop statistical designs to provide efficient estimates of the health effects of exposure to air pollutants in epidemiologic studies. As part of their research plan, Navidi and colleagues also intended to evaluate the effects of measurement error in exposure assessment (that is, the difference between true and estimated exposures) on the accuracy of estimated health effects. This is an area of critical interest, because problems in accurately measuring an individual's exposure may result in uncertainty in assessing the health risk of an air pollutant.

The HEI Research Committee, commending the caliber of the study team and the quality of the proposed methods, approved funding. In addition, they noted that a key strength of Navidi's application was that the statistical designs he developed would be applied in an ongoing cohort study conducted by a team of investigators at USC. This ten-year study, *Epidemiologic Investigation to Identify Chronic Health Effects of Ambient Air Pollutants in Southern California*, referred to as the "USC Children's Health Study" and sponsored by the California Air Resources Board (CARB), seeks to determine the health effects of long-term exposure to ozone and other air pollutants. Approximately 3,600 children from twelve communities in the southern California air basin exposed to differing components and levels of air pollutants are involved. The USC investigators, led by Dr. John Peters, are examining a variety of health endpoints, including changes in pulmonary function, days of school absence, and the appearance or augmentation of clinical symptoms, in relation to long-term exposure to ozone, particulate matter, nitrogen oxides, and acid (H^+). As described in the accompanying Investigators' Report and in subsequent sections of this Commentary, Dr. Navidi tested several of his epidemiologic models with data from the USC Children's Health Study.

Navidi's was one of six studies funded under RFA 91-1; four of the other five have already been published by HEI (Loomis et al. 1996; Avol et al. 1998; Kinney et al. 1998; Tager et al. 1998a,b). During the review of the study, the HEI Review Committee and the investigators exchanged comments and clarified issues in the Investigators' Report and in the HEI Review Committee's Commentary. This Commentary is intended to aid HEI sponsors and the public by highlighting the strengths of the study, pointing out alternative interpretations, and placing the report into scientific perspective.

BACKGROUND AND AIMS

The investigators begin their report by noting that there are three fundamental challenges common to many environmental epidemiologic studies: measuring exposure, dis-

* A list of abbreviations appears at the end of the Investigators' Report.

Dr. William Navidi's 3-year study, *Statistical Methods for Epidemiologic Studies of the Health Effects of Air Pollution*, began in March 1993 and had total expenditures of $285,915. The Investigators' Report from Dr. Navidi and colleagues was received for review in December 1996. A revised report, received in July 1997, was accepted for publication in October 1997. During the review process, the HEI Review Committee and the investigators had the opportunity to exchange comments and to clarify issues in the Investigators' Report and in the Review Committee's Commentary.

ease, and related variables; finding populations with different degrees of exposure that are comparable with respect to potentially confounding factors; and distinguishing effects at the individual level.

Dr. Navidi and his colleagues undertook a study with three specific aims to address the methodological implications of these challenges. The aims were

1. to develop efficient statistical designs for panel studies,
2. to develop methods that allowed for the effects of exposure measurement error, and
3. to develop methods for combining individual and aggregate level comparisons.

The investigators developed a bidirectional case-crossover design to address the first aim. They addressed the second aim by developing mathematical measurement error models to compare community measures of exposure (termed "microenvironmental" or "indirect" by Navidi) with the direct, "personal sampler" approach. They also used this model to assess the impact of measurement errors. Finally, they developed multilevel analytic designs to meet the third aim. For each statistical model and design the investigators used the USC Children's Health Study data for illustration. These three approaches are discussed in the subsequent sections.

BIDIRECTIONAL CASE-CROSSOVER APPROACH

The case-crossover design (Maclure 1991) is a statistical design in which only cases are studied, and their exposure at the time of the event of interest is compared with some estimate of their typical level of exposure. This design is an adaptation of a case-control design and a crossover design. In a study utilizing a case-control design, cases (that is, subjects with a certain condition, for example, absence from school) are compared with controls (for example, subjects not absent). In a crossover study, subjects are compared under two different conditions (for example, when absent and when not absent from school). The case-crossover design blends the two approaches. The same subject is both a case and a control, and "crosses over" from being a case to being a control. That is, the health status of the subject is observed both when the patient is absent from school (case status) and when the subject is not absent (control status).

In this health effects study utilizing a case-crossover design, pollutant exposure when a subject is a case (is absent from school) was compared with the "typical" exposure when the subject is a control (is not absent). Further, in the usual case-crossover setting, the case data (exposure during time interval of case) is compared to some estimate of a previous control setting (exposure during prior time periods) when the individual is a noncase. For this example, the pollutant exposure when a child is absent from school is compared to some usual estimate of prior exposure when the child is not absent from school. For environmental studies, use of this previous time period control exposure data can introduce a number of serious problems and biases. In particular, the control period exposure data may be systematically lower or higher than the case period exposure data (possibly related to time trends) (see Greenland 1996).

As the investigators noted, the case-crossover design has been introduced and investigated by a number of others in the social sciences (Maclure 1991; Feldman 1993a,b; Mittleman et al. 1995). One useful contribution of Navidi and colleagues comes from understanding and applying the method in the air pollution exposure setting.

The case-crossover method is often used to investigate if some recent behavior has triggered or is related to the cause of an event (for example, in Marshall and Jackson [1993], the objective was to investigate whether the onset of a myocardial infarction [MI] was related to recent alcohol consumption). The method has at least two problems, which may lead to biased results: (1) it is very sensitive to the mathematical model employed, and (2) the control period data often used in the analysis may not be "typical" and so may not be good control data. In addition, as with the application of any epidemiologic method to observational data, obtaining statistically significant associations does not establish causality. Establishing causality usually requires synthesizing a wide range of evidence obtained using a number of approaches.

THE INVESTIGATORS' APPROACH

The investigators addressed the problem of biases in case-crossover studies by developing a bidirectional case-crossover where control exposure data were obtained both before and after the subject was a case (or, in the vocabulary of the investigators, before and after the subject was a failure). In addition, the investigators advocated using observed (recorded) exposure data for the control data rather than some estimated typical exposure.

The investigators conducted two simulation studies to evaluate the standard case-crossover design and their bidirectional case-crossover. For this particular model, they demonstrated bias in the standard case-crossover design and the lack of bias in their bidirectional method.

Lastly, they tested the bidirectional case-crossover approach by relating the acute effects of exposure to air pollution to absences from school for the USC Children's Health Study. The pollutants considered were ozone, particulate matter with a diameter less than 10 micrometers,

and nitrogen dioxide (NO_2). The analysis controlled for day of the week, temperature, and humidity. Because the subject acted as his or her own control, the analysis automatically controlled for subject-specific variables such as sex, age, race, community of residence, and prior health conditions. The analysis suggested that a relation between school absences and exposure to NO_2 lagged by two days. It should be noted that this effect was the only statistically significant one of many comparisons.

CRITIQUE OF THE BIDIRECTIONAL CASE-CROSSOVER METHOD

Have Navidi and colleagues eliminated or diminished the problems of the case-crossover by introducing the bidirectional aspect (that is, by using control data from both before and after the case period)? In the example they cite (absence from school in the USC Children's Health Study), the bias appears to be reduced. This is due to more appropriate control data. The issue of sensitivity to the mathematical model has not been resolved, however, because they did not investigate this issue.

In addition, the application of the bidirectional method may be in question for endpoints more extreme than absence from school. For example, in trying to employ the method for investigating the relation of pollutant exposure to MI, the subject may be extremely different before the event than after it. Thus, the properties of the subject in the control periods before and after the MI may be different. In the standard crossover design this would be related to what is called a "carryover effect": the MI changes the person. The current version of the bidirectional case-crossover design does not appear to accommodate this possibility.

In summary, the bidirectional case-crossover is an advance over the case-crossover design and shows the potential bias in the latter. The bidirectional case-crossover can incorporate better control data into the analysis. It needs investigation to study the sensitivity to the mathematical model, however, including carryover effects.

MEASUREMENT ERRORS IN AIR POLLUTION EXPOSURE ASSESSMENT

In this phase of the project, the investigators studied two ways of measuring cumulative exposure, the indirect or "microenvironmental" approach and the direct or "personal sampler" approach. The microenvironmental approach estimates the time spent in each microenvironment and also estimates an exposure time to each pollutant in each microenvironment. From these estimates a time-weighted average exposure for an individual is determined. The personal sampler approach requires that an individual wear a sampling device. The individual's average exposure is then estimated from the integrated concentration measured by the sampler and the amount of time the sampler was worn.

The two approaches differ in significant ways. With the microenvironmental approach, the exposure metric is flexible—exposure could be estimated as peak exposure above a certain level or as a cumulative exposure. In the personal sampling approach, only the cumulative exposure value is available. A second important difference involves the number of parameters to be estimated. The microenvironmental method requires estimates of both concentration of the pollutant and length of time spent in each environment; the only uncertainty in the personal sampler method is in the concentration, since the amount of time spent in each microenvironment is always correct.

The investigators used a simple model for the true exposure and compared the results from the two approaches under a variety of assumptions about the magnitude of the measurement errors involved. They performed a simulation study varying the measurement errors. Their conclusion was that neither the personal sampler nor the microenvironmental model gave correct answers when uncorrected measurement errors were present.

CRITIQUE OF MEASUREMENT ERRORS IN AIR POLLUTION EXPOSURE ASSESSMENT

Again, this component of the study was a careful, well-presented analysis. The investigators' conclusion that reliable results will not be achieved when there are measurement errors was based on a simple model. The investigators believe that more elaborate models can be constructed that will be self-correcting, and this will result in reliable estimates of cumulative exposures, but it is not apparent how this will be done. In addition, given the lack of literature concerning the validation of microenvironmentally based methods of exposure assessment and the limitations of the personal sampling methods to estimate cumulative exposure that Navidi and colleagues document so well, it is not clear how the comparison results should be viewed. More basic research is needed to determine whether the two approaches (microenvironmental and personal samplers) will be capable of estimating cumulative exposures accurately, and, if so, how well. Further, more extensive modeling and simulations are still needed to make any definite conclusions.

MULTILEVEL ANALYTIC DESIGN

To address the third aim of the project, to develop methods for combining individual and aggregate level comparisons, the investigators propose multilevel analytic designs.

In epidemiology there are two general types of studies for obtaining data to quantify the association between an exposure and a disease: ecologic studies and analytic studies. In the former, disease rates in groups of individuals are related to the average exposure rates in the groups. In the latter, the individual's disease outcomes are related to his or her own exposure values. Cohort and case-control studies are examples of analytic studies.

Often the results of ecologic and analytic studies are different. The resolution of the differences usually rests on three issues:

1. between-group confounding: a characteristic of groups that is not accounted for in the model but is a real risk factor. The outcome of interest may occur not as a result of the exposure, but instead as a result of some other group characteristics, including differences in medical diagnosis and reporting.
2. measurement error: the inability to measure without error the exposure variable(s). In analytic studies, for example, effects (the quantification of the impact of an exposure on an outcome) may be diluted due to measurement errors.
3. restricted variability: this results when, for example, a study is conducted on only one set of subjects and there may not be a wide enough variability in an exposure variable to reveal a true association.

Ecologic studies are often less expensive than analytic studies, are less prone to measurement error, and usually have good variation across groups. They may, however, not be well suited for the control of confounders. Analytic studies are usually better able to deal with confounders, and usually have better quality data, but may suffer more from measurement errors. Both of these types of studies are important, and they may often complement each other.

THE INVESTIGATORS' MODEL

The investigators suggest a multilevel analytic design where a number of groups are selected and within each group a sample of subjects is drawn. The analysis can incorporate variables at the subject level and also at the group level. This design incorporates the positive features of the ecological and analytic studies while addressing the deficiencies.

Their model, given as Equation 12 in the Investigators' Report, is as follows:

$$y_{gi} = \alpha_g + \beta x_{gi} + \gamma \upsilon_{gi} + \varepsilon_{gi}$$

where y_{gi} is the outcome variable for individual i in group g (for example, a lung function test value for a child in community g); and α_g is the average of the outcome variable for group g, which may be fixed or random effect. In this context, a fixed effect would mean that the groups in the study are the only groups of interest, while a random effect would mean that the communities in the study represent a random sample of some bigger collection of groups. The variable x is the exposure or exposures of interest, υ represents the set of confounders, and ε are the error terms. The confounders may be fixed or random. Often the ε are assumed to be normally distributed. The main interest lies in estimating β, the regression coefficient for the exposures x.

The investigators discussed weighted least squares and two-stage least squares as methods for estimating the model regression coefficients. In addition to the random effect models, the investigators incorporated two exposure measurement error models into their multilevel model. The Berkson model assigns individuals the group exposure (Armstrong 1990); the classical model assigns a random value whose expected value is the true exposure (Thomas et al. 1993). In this current work, they attempted to estimate the true effects of exposure (variables ε of the above model) because of the impossibility of measuring ε without error.

The investigators examined some "optimal" properties of their models as they related to the number of groups and the sample sizes within groups. These were performed under simplifying assumptions. An example was given from the USC Children's Health Study, which involved approximately 3,600 children from 12 southern California communities who underwent spirometric assessment of pulmonary function. The investigated relation was between a measurement of lung function, FEV_1, and lifetime exposure to ozone.

CRITIQUE OF THE MULTILEVEL ANALYTIC DESIGN

The designs that the investigators propose are not new, but have been used for several years in education, economics, and other social sciences. Currently they are employed extensively in health-related fields. These methods have undergone, and continue to undergo, intense development and extensions and are known under headings such as "multilevel models" (Goldstein 1995), "random effects models" (Chapter 2 of Goldstein 1995; Searle et al. 1992), and "hierarchical models" (Bryk and Ravdenbush 1992). Von Korff and colleagues (1992) recommended their use in epidemiology, and they have been used in some air pollution studies. In addition, the estimation procedures (weighted least squares and two-stage estimates of a [â] and measurement error models) are standard for these problems.

As described in the critique of the case-crossover design, the main contribution of Navidi and colleagues has been to apply these methods appropriately to data on the health

effects of air pollution. The investigators present a useful advance in this field by identifying in detail the components needed to model the phenomena and by supplying reasonable mathematical models. Further, they provide appropriate ways of obtaining efficient estimates of the relevant model parameters that quantify the effect of the exposure variables in the models.

There are concerns, however. The multilevel models and measurement errors models are sensitive to model assumptions. The investigators present neither discussions about the sensitivity of the models to assumptions nor diagnostics for assessing the fit (or appropriateness) of their models. For example, the assumption of random uncorrelated errors is an important component of the models. This may be unrealistic in practice and may limit the applicability of these models to air pollution studies. Further, if the model is not valid, its application may result in serious bias. The mathematical sophistication of the models does not ensure applicability. The investigators have presented the models very well. Correct applications of these models will have to address the above concerns.

Last, the discussion of bias (in Section 4.3) is incomplete and does not relate directly, as the investigators state, to the bias discussed by Greenland and Robins (1994). The problem the investigators call bias relates to how interactions at the individual level (Equation 16 in the Investigators' Report) could bias main effects estimates for the ecologic analysis. Greenland and Morgenstern (1989) deal with this, as do Greenland and Robins (1994). They go on, however, to describe how nonlinearities in confounder dose-response at the individual level could bias exposure main-effect estimates in the ecologic analysis, even if there are no interactions at all, and even if the exposure effects are only linear. At a minimum, major simulation studies are needed to evaluate the effects of nonlinearities on the results of analyses from the investigators' model, referred to as Equation 12 in the Investigators' Report and in the previous section of the Commentary.

FINAL COMMENTS

The investigators' main contribution has been not so much in the development of new methods as in transporting and extending existing methods to the air pollution exposure field. Their simulations demonstrate the potential for bias in the usual case-crossover method and sound a cautionary warning to any who desire to use it. Their extension to the bidirectional case-crossover study design offers promise for reducing bias. In addition, the investigators' discussion of measurement error in air pollution exposure, although simple, does document the impact of measurement error of estimates of cumulative exposures.

Finally, the application of the multilevel model to exposure problems also offers potential for advancement. There are major concerns, however, related to model appropriateness and validation that need to be addressed before it can be recommended for general use. The use of these models will always require a careful consideration of the appropriateness of the assumptions made and an understanding of potential biases inherent in the models.

In general, Navidi and his collaborators have made advances in statistical methodology, but have also, as is to be expected, left some issues unresolved. Applications of their methods to exposure problems, which include careful model development and validation, are needed to resolve these questions.

ACKNOWLEDGMENTS

The Health Review Committee would like to thank the ad hoc reviewers for their help in evaluating the Investigators' Report; Dr. Geoffrey Sunshine for his help in preparing the Commentary; and Thomas Atwood, Julia Campeti, Victoria Casana, Sean Donahue, Virgi Hepner, and Hope Steele for their roles in publishing this report.

REFERENCES

Armstrong BG. 1990. The effects of measurement errors on relative risk regression. Am J Epidemiol 132:1176–1184.

Avol EL, Navidi WC, Rappaport EB, Peters JM. 1998. Acute Effects of Ambient Ozone on Asthmatic, Wheezy and Healthy Children. Research Report Number 82, Health Effects Institute, Cambridge, MA.

Bryk, AS, Ravdenbush SW. 1992. Hierarchical Linear Models: Applications and Data Analysis Methods. Sage, Newbury Park, CA.

Feldman U. 1993a. Epidemiologic assessment of risks of adverse reactions associated with intermittent exposure. Biometrics 49:419–428.

Feldman U. 1993b. Design and analysis of drug safety studies. J Clin Epidemiol 46:237–244.

Goldstein H. 1995. Multilevel Statistical Models. Edward Arnold, New York, NY.

Greenland S. 1996. Confounding and exposure trends in case-crossover and case-time-control designs. Epidemiology 7:231.

Greenland S, Morgenstern H. 1989. Ecological bias, confounding and effect modification. Int J Epidemiol 18:269–274.

Greenland S, Robins JM. 1994. Ecologic studies: Biases, misconceptions, counterexamples. Am J Epidemiol 139: 747–760.

Hatch M, Thomas D. 1994. Measurement issues in environmental epidemiology. In: Environmental Epidemiology Planning Project. HEI Communication Number 3, pp. 49–58. Health Effects Institute, Cambridge, MA.

Kinney PL, Aggarwal M, Nikiforov SV, Nadas A. 1998. Methods Development for Epidemiologic Investigations of the Health Effects of Prolonged Ozone Exposure, Part III: An Approach to Retrospective Estimation of Lifetime Ozone Exposure Using a Questionnaire and Ambient Monitoring Data (U.S. Sites). Research Report Number 81. Health Effects Institute, Cambridge, MA.

Loomis DP, Borja-Aburto VH, Bangdiwala SI, Shy CM. 1996. Ozone Exposure and Daily Mortality in Mexico City: A Time-Series Analysis. Research Report Number 75. Health Effects Institute, Cambridge, MA.

Maclure M. 1991. The case-crossover design: A method for studying transient effects on the risk of acute events. Am J Epidemiol 133:144–153.

Marshall RJ, Jackson RT. 1993. Analysis of case-crossover designs. Stat Med 12:2333–2341.

Mittleman MA, Maclure M, Robins JM. 1995. Control sampling strategies for case-crossover studies: An assessment of relative efficiency. Am J Epidemiol 142:91–98.

Morgenstern H, Thomas D. 1994. Principles of study design in environmental epidemiology. In: Environmental Epidemiology Planning Project, HEI Communication Number 3, pp. 23–38. Health Effects Institute, Cambridge, MA.

Prentice RL, Thomas D. 1994. Methodological research needs in environmental epidemiology: Data analysis. In: Environmental Epidemiology Planning Project, HEI Communication Number 3, pp. 39–48. Health Effects Institute, Cambridge, MA.

Searle S, Casella G, McCulloch CE. 1992. Variance Components. Wiley and Sons, New York, NY.

Tager IB, Künzli N, Ngo L, Balmes J. 1998a. Methods Development for Epidemiologic Investigations of the Health Effects of Prolonged Ozone Exposure, Part I: Variability of Function Measures. Research Report Number 81. Health Effects Institute, Cambridge, MA.

Tager IB, Künzli N, Lurmann F, Ngo L, Segal M, Balmes J. 1998b. Methods Development for Epidemiologic Investigations of the Health Effects of Prolonged Ozone Exposure, Part II: An Approach to Retrospective Estimation of Lifetime Ozone Exposure Using a Questionnaire and Ambient Monitoring Data (California Sites). Research Report Number 81. Health Effects Institute, Cambridge, MA.

Thomas D, Stram D, Dwyer J. 1993. Exposure measurement error: Influences on exposure-disease relationships and methods of correction. Annu Rev Public Health 14:69–93.

Von Korff M, Koepsell T, Curry S, Diehr P. 1992. Multilevel analysis in epidemiologic research on health behaviors and outcomes. Am J Epidemiol 135:1077–1082.

RELATED HEI PUBLICATIONS

Report No.	Title	Principal Investigator	Publication Date*
Human Studies			
Research Reports			
63	Development of Samplers for Measuring Human Exposure to Ozone, *Part I:* Active and Passive Ozone Samplers Based on a Reaction with a Binary Reagent	J. D. Hackney	1994
70	Oxident and Acid Aerosol Exposure in Healthy Subjects and Subjects with Asthma, *Part II:* Effects of Sequential Sulfuric Acid and Ozone Exposures on the Pulmonary Function of Healthy Subjects and Subjects with Asthma	M. J. Utell	1995
75	Ozone Exposure and Daily Mortality in Mexico City: A Time-Series Analysis	D. P. Loomis	1996
78	Effect of Ozone on Normal and Potentially Sensitive Human Subjects, *Part I:* Airway Inflammation and Responsiveness to Ozone in Normal and Asthmatic Subjects	J. R. Balmes	1997
	Part II: Airway Inflammation and Responsiveness to Ozone in Nonsmokers and Smokers	M. W. Frampton	1997
	Part III: Mediators of Inflammation in Bronchoalveolar Llavage Fluid from Nonsmokers, Smokers, and Asthmatic Subjects Exposed to Ozone: A Collaborative Study	M. W. Frampton	1997
83	Daily Changes in Oxygen Saturation and Pulse Rate Associated with Particulate Air Pollution and Barometric Pressure	D. W. Dockery	1999
Methods Development			
Research Reports			
81	Methods Development for Epidemiologic Investigations of the Health Effects of Prolonged Ozone Exposure		
	Part I: Variability of Pulmonary Function Measures	I. B. Tager	1998
	Part II: An Approach to Retrospective Estimation of Lifetime Ozone Exposure Using a Questionnaire and Ambient Monitoring Data (California Sites)	I. B. Tager	1998
	Part III: An Approach to Retrospective Estimation of Lifetime Ozone Exposure Using a Questionnaire and Ambient Monitoring Data (U.S. Sites)	P. L. Kinney	1998
HEI Communications			
1	New Methods in Ozone Toxicology: Abstracts of Six Pilot Studies		1992
3	Environmental Epidemiology Planning Project		1993
	Introduction	A. J. Cohen and L. Gordiis	
	Methodologic Issues:		
	Methodologic Frontiers in Environmental Epidemiology	K. J. Rothman	
	Principles of Study Design in Environmental Epidemiology	H. Morgenstern and D. Thomas	
	Methodologic Research Needs in Environmental Epidemiology: Data Analysis	R. L. Prentice and D. Thomas	
	Measurement Issues in Environmental Epidemiology	M. Hatch and D. Thomas	

Basic Problems in Interaction Assessment	S. Greenland
The Place of Epidemiology in Environmental Decisions: Needed Support for the Development of Risk Assessment Policy	R. Neutra and D. Trichopolous
Tropospheric Ozone:	
Introduction to Working Group on Tropospheric Ozone, Health Effects Institute Environmental Epidemiology Planning Project	I. B. Tager
Examining Acute Health Outcomes Due to Ozone Exposure and Their Subsequent Relationship to Chronic Disease Outcomes	B. D. Ostro
Design and Analysis of Studies of the Health Effects of Ozone	A. Munoz
Summary of Papers and Research Recommendations of Working Group on Tropospheric Ozone, Health Effects Institute Environmental Epidemiology Planning Project	I. B. Tager

**Reports published since 1990. Copies of these reports or a list of reports published prior to 1990 can be obtained by writing or calling the Health Effects Institute, 955 Massachusetts Avenue, Cambridge, MA 02139.*
Phone (617) 876-6700 FAX (617) 876-6709 E-mail pubs@healtheffects.org www.healtheffects.org

HEI

The Board of Directors

Archibald Cox *Chairman*
Carl M. Loeb University Professor (Emeritus), Harvard Law School

Douglas Costle
Chairman of the Board and Distinguished Senior Fellow, Institute for Sustainable Communities

Alice Huang
Senior Councilor for External Relations, California Institute of Technology

Donald Kennedy
President (Emeritus) and Bing Professor of Biological Sciences, Stanford University

Susan B. King
Fellow, Sanford Institute of Public Policy, Duke University

Richard B. Stewart
Professor, New York University School of Law

Robert M. White
President (Emeritus), National Academy of Engineering, and Senior Fellow, University Corporation for Atmospheric Research

Health Research Committee

Bernard D. Goldstein *Chairman*
Director, Environmental and Occupational Health Sciences Institute

Glen R. Cass
Professor of Environmental Engineering and Mechanical Engineering, California Institute of Technology

Seymour J. Garte
Professor of Environmental and Community Medicine, Environmental and Occupational Health Science Institute

Rogene Henderson
Senior Scientist, Lovelace Respiratory Research Institute

Meryl H. Karol
Professor of Environmental and Occupational Health, University of Pittsburgh Graduate School of Public Health

Stephen I. Rennard
Larson Professor, Pulmonary and Critical Care Medicine Section, Department of Internal Medicine, University of Nebraska Medical Center

Jonathan M. Samet
Professor and Chairman, Department of Epidemiology, School of Hygiene and Public Health, Johns Hopkins University

Robert F. Sawyer
Class of 1935 Professor of Energy (Emeritus), Professor of the Graduate School, University of California at Berkeley

Frank E. Speizer
Edward H. Kass Professor of Medicine, Channing Laboratory, Harvard Medical School, Department of Medicine, Brigham and Women's Hospital

Gerald van Belle
Professor, Departments of Environmental Health and Statistics, University of Washington

Health Review Committee

Daniel C. Tosteson *Chairman*
Professor of Cell Biology, Dean Emeritus, Harvard Medical School

John C. Bailar III
Professor, Department of Health Studies, Biological Sciences Division, The University of Chicago

A. Sonia Buist
Professor of Medicine and Physiology, Oregon Health Sciences University

Ralph D'Agostino
Professor of Mathematics/Statistics and Public Health, Boston University

Thomas W. Kensler
Professor, Division of Toxicological Sciences, Department of Environmental Sciences, Johns Hopkins University

Edo D. Pellizzari
Vice President for Analytical and Chemical Sciences, Research Triangle Institute, NC

Donald J. Reed
Distinguished Professor of Biochemistry, Department of Biochemistry and Biophysics, and Environmental Health Sciences Center, Oregon State University

David J. Riley
Professor of Medicine, University of Medicine and Dentistry of New Jersey–Robert Wood Johnson Medical School

Sverre Vedal
Professor of Medicine, University of British Columbia

Officers and Staff

Daniel S. Greenbaum *President*
Richard M. Cooper *Corporate Secretary*
Howard E. Garsh *Director of Finance and Administration*
Kathleen M. Nauss *Director for Scientific Review and Evaluation*
Robert M. O'Keefe *Director of Program Strategy*
Jane Warren *Director of Research*
Aaron J. Cohen *Senior Scientist*
Maria G. Costantini *Senior Scientist*
Debra A. Kaden *Senior Scientist*
Alison Geyh *Staff Scientist*
Bernard Jacobson *Staff Scientist*
Diane J. Mundt *Staff Scientist*
Martha E. Richmond *Staff Scientist*
Geoffrey H. Sunshine *Staff Scientist*
Gail V. Allosso *Office and Contracts Manager*
Thomas Atwood *Manager of Publications and Information*
Julia F. Campeti *Publications Assistant*
Terésa Fasulo *Senior Administrative Assistant*
Darlene Jones *Senior Administrative Assistant*
L. Virgi Hepner *Senior Scientific Editor*
Judith Lopez *Receptionist*
Francine Marmenout *Senior Executive Assistant*
Teresina McGuire *Accounting Assistant*
Beverly Morse *Administrative Assistant*
Jacqueline C. Rutledge *Controller*
Hope Steele *Acting Managing Editor*
Mary L. Stilwell *Senior Administrative Assistant*

HEI HEALTH EFFECTS INSTITUTE 955 Massachusetts Avenue, Cambridge, MA 02139 (617) 876-6700

Research Report Number 86 May 1999