REOCCURING SCIENCE CONCERNS IN THE EPA NAAQS REVIEWS

-With Suggestions for Improvements

November 9, 2009

Presentation to OMB

1. Internal EPA Reanalysis of Published Studies

- Ozone: reanalysis of selected data from Adams (2006, 2002)
  - Used to conclude clinical effects at 0.06 ppm and to support the 2008 EPA NAAQS decision
- NO_2: reanalysis of clinical studies on airway hyper-responsiveness
  - Used to conclude clinical effects at 100 ppb and to support EPA's proposed range for a NAAQS of 80-100 ppb
- SO_2: reanalysis of human clinical studies by Linn et al. (1987, 1983)
  - Used to conclude clinical effects at 200 ppb and to support EPA's proposed range for a NAAQS of 50-75 ppb

2. Transparency, Disclosure of Critical Studies/Analyses

- **Ozone: Reanalysis of selected data from Adams (2006, 2002)**
  - **Placed in the docket after the close of the Staff Paper and six days before the NPRM**

- **NO_2: Updated "meta analysis" of NO_2**
  - **Included for the first time in the final ISA thereby precluding opportunity for public comment during the ISA review**

- **SO_2: Reanalysis of clinical studies by Linn et al. (1987, 1983)**
  - **Included for the first time in final ISA thereby precluding opportunity for public comment during the ISA review**

3. Interpreting and Presenting Scientific Information

- **Studies excluded based on unclear criteria**
  - **Abrahamowicz (2003) reports thresholds for chronic PM mortality**
  - **Schildcrout (2006) reports no association: O_3, asthma symptoms**

- **Same study reported differently across NAAQS reviews**
  - **Schildcrout reports asthma symptoms for NO_2 but not O_3**
    - **described as high quality study in NO_2 ISA, poor quality study for O_3**
  - **Delfino (2002) reports asthma symptoms for peak PM_{10} exposure**
    - **results used to support PM and NO_2, alone, as causative agents**

- **Data used for purpose for which it was not intended**
  - **Human clinical studies designed to test group responses used to assess results for individuals**

- **Unproven hypothesis used to explain study inconsistencies**
  - **Mortimer (2004) only reported positive association in children on medication: hypothesis – these children have higher level of disease**
  - **Delfino (2002) only reported positive association in children not on medication: hypothesis – these children are less protected**

4. Limitations of Observational Epidemiology Studies

- **Technical concerns with observational studies**
  - **Reliance on ambient measures as a surrogate for personal exposure**
    - **HEI studies (Sarnat 2001, 2005, 2006) raise serious concerns for gases**
  - **Lack of accepted model specification criteria**
  - **Overstating the robustness of the models**
  - **Using the highest/most significant results**
  - **Reliance on small and non statistically significant results**

- **Over-reliance on observational studies to assess causality and risks**
  - **HEI 1997: not possible to identify individual pollutants as causal**
  - **HEI 2003: not possible to determine appropriate model specification**
  - **HEI 2009: unexplained regional heterogeneity; no model agreement**

- **Over-reliance on single pollutant results to assess causality and risks results in double/multiple counting of the risks of air pollution**

5. Criteria for Determining Adverse Effects

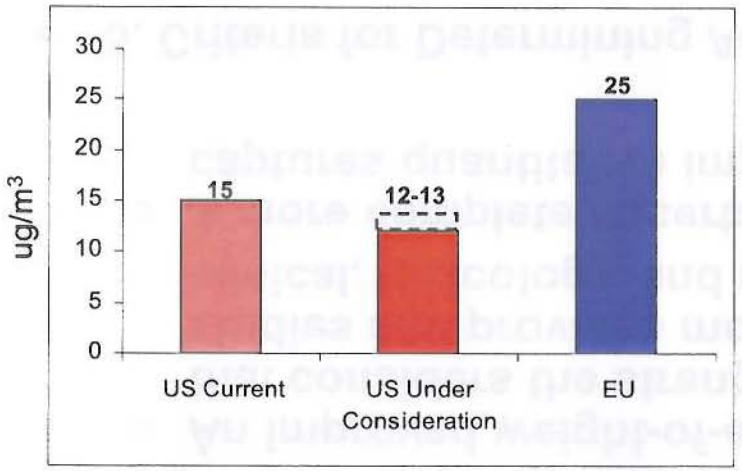
- EPA has lowered the criteria for what they consider as mild, moderate, severe reductions in lung function

- SO_2: Criteria for lung function and symptoms
  - Threshold for adversity lowered by considering a moderate change in lung function alone *or* increased symptoms alone as adverse
  - Inconsistent with medical expert advice (ATS)

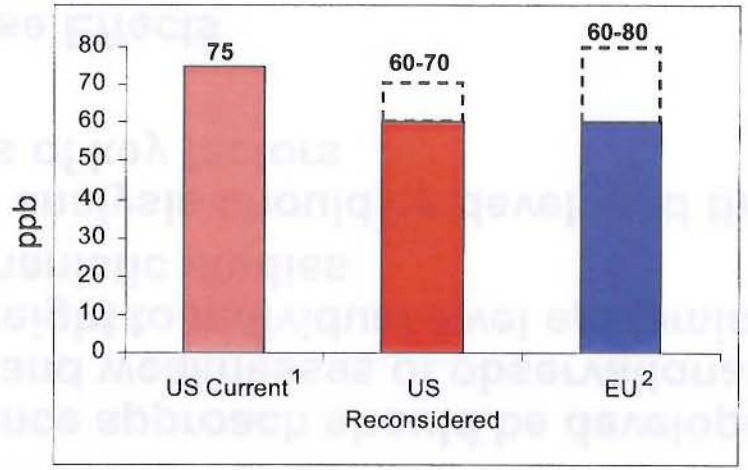
- NO_2: Criteria for Airway Hyper-responsiveness
  - Effects on AHR considered adverse even though no concentration response was observed and the actual changes in pulmonary function were very small (average FEV_1 decrease 1.5%)

Comparison of Air Quality Criteria – Based on Similar Underlying Health Data

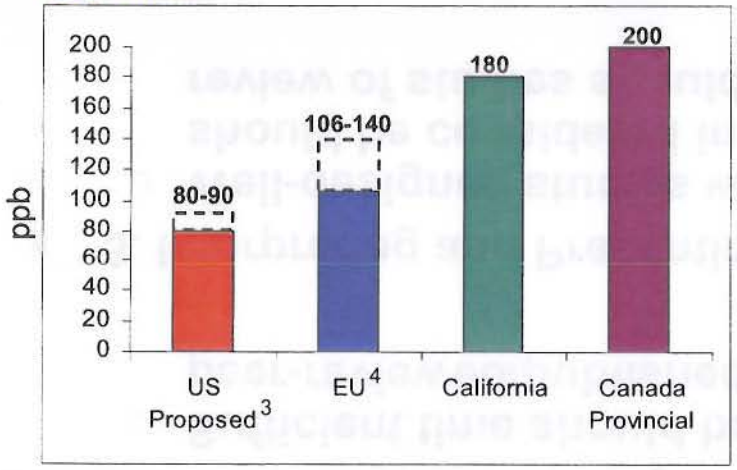
$PM_{2.5}$ Annual



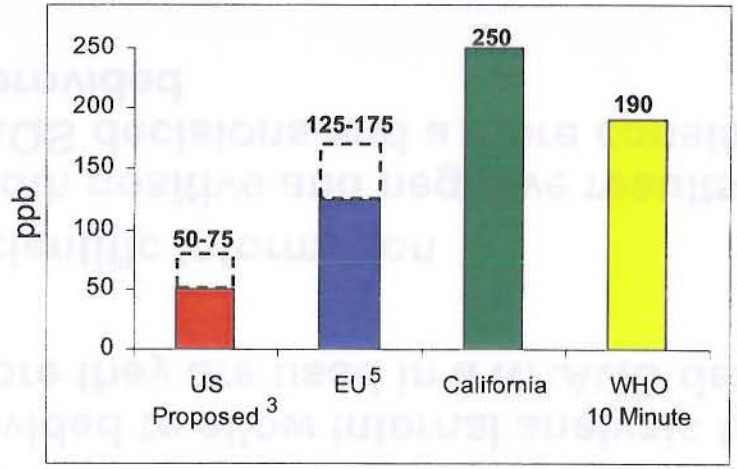
Ozone 8 Hour



NO_2 1 Hour



SO_2 1 Hour



1. 98th Percentile Form, 2. 25 Exceedences Permitted, 3. 99th Percentile Form, 4. 18 Exceedences Permitted, 5. 24 Exceedences Permitted

Potential Process Improvements

- 1. and 2. Internal Reanalysis of Data and Transparency
  - Sufficient time should be provided to allow internal analysis to be peer-reviewed/published before they are used in a NAAQS decision
- 3. Interpreting and Presenting Scientific Information
  - Well-designed studies with both positive and negative results should be considered in NAAQS decisions and a more consistent review of studies should be provided
- 4. Limitations of Observational Studies
  - An improved weight-of-evidence approach should be developed that considers the strengths and weaknesses of observational studies and provides more weight to individual level epidemiology, clinical, toxicology, and mechanistic studies
  - A more complete uncertainty analysis should be developed that captures quantitative impacts of key factors
- 5. Criteria for Determining Adverse Effects
  - Deference given to accepted clinical standards of adversity

NO_2 and Respiratory Effects

Julie E. Goodman, Ph.D., DABT

Principal Scientist
Gradient

Adjunct Faculty Member
Department of Epidemiology
Harvard School of Public Health

November 9, 2009

NO_2 and Airway Hyper-responsiveness

- US EPA conducted an analysis of controlled exposure studies in asthmatics
  - Concluded for the first time that short-term exposure to 0.1 ppm NO_2 can cause increases in response to non-specific airway challenges
  - **Used to support US EPA's proposed NAAQS of 80-100 ppb**
- Goodman *et al.* (2009) conducted more rigorous analysis
  - Found no evidence to suggest that NO_2 leads to significant adverse effects at any of the exposures tested, up to 0.6 ppm
  - Does not support US EPA's proposed NAAQS

Airway Hyper-responsiveness (AHR)

- Exaggerated airway-narrowing response to many environmental triggers, such as allergens and exercise; characteristic of asthma
- Normally measured by histamine or methacholine challenge
- Studies reviewed compared AHR after exposure to NO_2 *vs.* air
- What denotes an "adverse" clinically relevant effect is subjective

US EPA Analysis of AHR in Controlled Human Exposure Studies

- Evaluated the fraction of individuals in each study with increased AHR after NO_2 exposure *vs.* air
- Tested for statistical significance using a "sign test"
  - Essentially looked at how many people had increased *vs.* decreased AHR after NO_2 *vs.* air
- Concluded there was a statistically significant increase in AHR

Limitations of US EPA Analysis

- Has not undergone peer-review
- Because first appeared in final ISA, not informed by public comment
- Methods are not fully transparent
  - Inclusion/exclusion criteria for studies not stated
  - Data included/excluded from individual studies not stated
  - No sensitivity analyses
- Use of sign test to evaluate statistical significance is inappropriate
- No evaluation of the magnitude of AHR as a function of NO_2 exposure
  - Many subjects may have experienced a small, but not clinically relevant, change
  - Could not evaluate exposure-response

Goodman *et al.* (2009) Meta-Analysis

- Comprehensive literature search conducted by professional Information Research Specialist
  - Final dataset included data from 26 studies; 38 exposure scenarios
- Inclusion/exclusion criteria and analysis methods established *a priori*
- Evaluated three measures of AHR:
  - Fraction of subjects with increased AHR
    - $PD_{NO2} < PD_{Air}$ or $\Delta FEV_{1,NO2} < \Delta FEV_{1,Air}$
  - Change in provocative dose (ΔPD)
  - Change in ΔFEV_1 ($\Delta\Delta FEV_1$)

Goodman *et al.* (2009) Meta-Analysis (cont.)

- Conducted meta-analysis for whole dataset, and stratified by
  - Airway challenge (specific/non-specific)
  - Exposure method (mouthpiece/whole chamber)
  - Activity level during exposure (rest/exercise)
- Conducted meta-regressions to assess exposure-response
- Conducted influence and sensitivity analyses
- Peer-reviewed and published in *Critical Reviews in Toxicology*

Goodman *et al.* (2009) Results

- Small, statistically significant associations for overall meta-analysis
- Magnitude of effects is similar between 0.1 and 0.6 ppm NO_2
- Individual studies with multiple exposure concentrations have not observed exposure-response association
- Meta-effects on two measures of AHR, ΔPD and $\Delta\Delta FEV_1$, are not adverse
  - $\Delta\Delta FEV_1$ of -1.75% *vs.* adverse effect of -10%
  - ΔPD_g of -27% *vs.* adverse effect of -50%

Meta-Analysis Conclusions

- NO_2 is *not* associated with clinically relevant effects on AHR at exposures up to 0.6 ppm
  - Small magnitude of effects
  - Lack of exposure-response
- Conclusions consistent with WHO (2005):
  - *"[T]he small size of the decrements and concerns expressed over the level of statistical significance of some of these results suggest that great caution should be exercised in accepting these findings as demonstrating acute effects."*

Epidemiology Studies Do Not Support a Causal Association Between NO_2 and Respiratory Effects

- Statistically significant findings are not large, robust, or consistent
- Most studies report statistically significant findings in single- but not multi-pollutant models
- NO_2 could be surrogate for other traffic pollutants, particles, fungi, pollen, relative humidity, and maximum temperature
- Use of measurements from central monitors likely leads to high degree of exposure misclassification

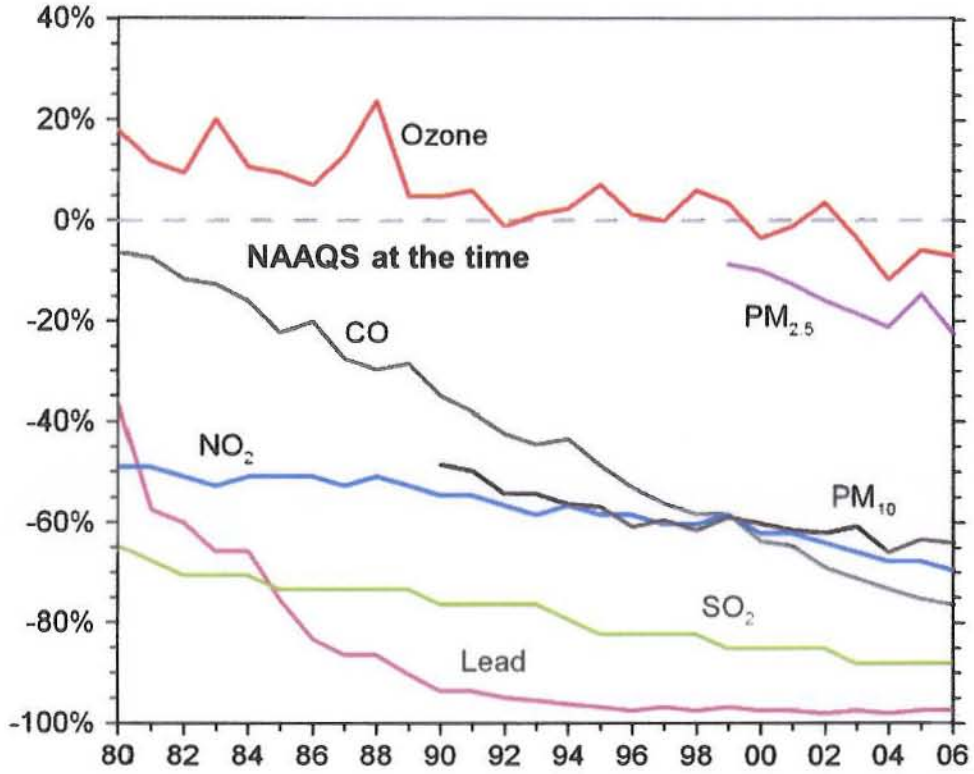
Overall Conclusions

- Clinical studies do not support adverse effects at exposures up to 0.6 ppm
- Epidemiology studies do not support adverse effects at ambient exposures

Overall Conclusions

- Clinical studies do not support adverse effects at exposures up to 0.6 ppm
- Epidemiology studies do not support adverse effects at ambient exposures

EPA Air Quality Trends Report - 2008



- **Ozone - 275 sites (4th maximum 8-hour average)**
- **$PM_{2.5}$ - 752 sites (annual average)**
- **PM_{12} - 391 sites (2nd maximum 24-hour average)**
- **NO_2 - 87 sites (annual average)**
- **CO - 144 sites (2nd maximum 8-hour average)**
- **SO_2 - 154 sites (annual average)**
- **Lead - 15 sites (maximum quarterly average)**

Comparison of national levels of the six principal pollutants to National Ambient Air Quality Standards (NAAQS). National levels are average across all sites.

Risk Analysis, Vol. 22, No. 6, 2002

Exposure Misclassification and Threshold Concentrations in Time Series Analyses of Air Pollution Health Effects

M. Brauer,[1]* J. Brumm,[2] S. Vedal,[3] and A. J. Petkau[2]

Linear, no-threshold relationships are typically reported for time series studies of air pollution and mortality. Since regulatory standards and economic valuations typically assume some threshold level, we evaluated the fundamental question of the impact of exposure misclassification on the persistence of underlying personal-level thresholds when personal data are aggregated to the population level in the assessment of exposure-response relationships. As an example, we measured personal exposures to two particle metrics, $PM_{2.5}$ and sulfate (SO_4^{2-}), for a sample of lung disease patients and compared these with exposures estimated from ambient measurements. Previous work has shown that ambient:personal correlations for $PM_{2.5}$ are much lower than for SO_4^{2-}, suggesting that ambient $PM_{2.5}$ measurements misclassify exposures to $PM_{2.5}$. We then developed a method by which the measured:estimated exposure relationships for these patients were used to simulate personal exposures for a larger population and then to estimate individual-level mortality risks under different threshold assumptions. These individual risks were combined to obtain the population risk of death, thereby exhibiting the prominence (and the value) of the threshold in the relationship between risk and estimated exposure. Our results indicated that for poorly classified exposures ($PM_{2.5}$ in this example) population-level thresholds were apparent at lower ambient concentrations than specified common personal thresholds, while for well-classified exposures (e.g., SO_4^{2-}), the apparent thresholds were similar to these underlying personal thresholds. These results demonstrate that surrogate metrics that are not highly correlated with personal exposures obscure the presence of thresholds in epidemiological studies of larger populations, while exposure indicators that are highly correlated with personal exposures can accurately reflect underlying personal thresholds.

KEY WORDS: Exposure assessment; exposure misclassification; environmental exposure; time series; air pollution; threshold; exposure-response

1. INTRODUCTION

Numerous time series studies have indicated that current levels of air pollution are associated with adverse health outcomes, including daily mortality.[1–3] These studies have been conducted in a variety of locations, using a variety of data analytic approaches, and have been performed by different investigators. In nearly all cases, the studies suggest a linear association between air pollution and increased risk of death, with no apparent threshold.[4]

1 University of British Columbia, School of Occupational and Environmental Hygiene and School of Medicine.
2 University of British Columbia, Department of Statistics.
3 University of British Columbia, School of Medicine. Current address: National Jewish Medical and Research Center, Division of Environmental and Occupational Health Sciences, Denver, CO.
*Address correspondence to Dr. Michael Brauer, Sc.D., University of British Columbia, School of Occupational and Environmental Hygiene, 2206 East Mall, Vancouver BC V6T 1Z3 Canada; tel.: (604) 822-9585; fax: (604) 822-9588; brauer@interchange.ubc.ca.

0272-4332/02/1200-1183$22.00/1

Although these epidemiological studies appear to withstand criticism focused on statistical methodology, coherence of results, and limitations of administrative health outcome data, they have generally relied on estimates of exposure in which the ambient air pollution concentration on each day is used to represent the exposure of the entire study population. Estimating individual exposure to air pollutants from central outdoor pollution monitors may result in considerable error.[5,6] The accuracy of using central outdoor air pollution monitors as indicators of personal exposure is one component of measurement error in such epidemiological studies.

The impacts of measurement error on the exposure-response relationship have been addressed recently by Zeger and colleagues,[7] who present a framework for evaluating the impact of measurement error and describe how the inadequate characterization of personal exposure can bias the magnitude of the effect estimates in time series epidemiological studies of ambient air pollution. Carrothers and Evans described how differential measurement error of multiple pollutants can lead to biased regression coefficients.[8] In addition to the impact on effect estimates, measurement error may also affect the ability to observe a threshold level, should one exist.

Cakmak and colleagues performed simulations to evaluate whether nonparametric smoothing is capable of detecting population-level thresholds in the presence of exposure measurement error.[9] Specific functional forms for the relationship between population-level risk and ambient concentrations were assumed and then simulated Poisson-distributed death counts corresponding to simulated log-normal ambient concentration levels were analyzed. These simulations examined the ability of different data analytic approaches (nonparametric smoothing and weighted nonlinear regression) to detect and estimate threshold concentrations in the presence of exposure measurement error.

In contrast, we focus specifically on the more fundamental question of the impact of exposure misclassification on the persistence of underlying personal-level thresholds when personal data are aggregated to the population level. The ability to identify a threshold level is critical to economic valuations[10] and to regulatory standards.[11] The issue of thresholds for time series studies of particulate air pollution has recently been examined in combined analyses of time series data from multiple cities.[12,13] Consistent with time series analyses from individual cities, these analyses have indicated that no population-level threshold is apparent when flexible modeling approaches are applied to the data. In this article we use measured personal exposures in a simulation approach to examine the extent to which exposure misclassification may obscure the presence of a threshold concentration in ecologic exposure-response relationships.

The inability to observe a threshold may be due, in part, to the fact that there is a distribution of individual exposures in a population.[14,15] If one assumes that all individuals in a population have the same function relating risk of an effect to the ambient pollutant concentrations, then in a population as a whole there will always be some observed effect, even for very low concentrations. This is because some individuals in the population will have greater exposures than others for any given ambient concentration and this will therefore result in a distribution of risks. Therefore, even if a common underlying threshold does in fact exist at an individual level, it may not be possible to observe it in a study that uses ambient concentrations to estimate individual exposures. The simulations reported here are intended to quantify the extent to which this may occur.

Accordingly, we repeatedly measured individual personal exposures of a panel of chronic obstructive pulmonary disease (COPD) patients and evaluated the impact of different exposure estimates on the population exposure-response relationship. As an example, we measured personal exposure to particle mass and compared this with sulfate, which is a better marker of exposure to ambient particles than fine particle mass ($PM_{2.5}$).[16,17] Using the exposure data, we then performed simulations to evaluate whether enhanced assessment of individual exposure improves correspondence between an underlying common individual threshold and the population-level threshold. In the process of illustrating this example we provide a general empirical methodology for addressing this general issue.

2. METHODS

2.1. Exposure Monitoring

Sixteen subjects, ages >60 years, currently nonsmokers and currently not living with a smoker, with physician-diagnosed moderate COPD were recruited for the exposure monitoring study. Personal (24-hour) particulate ($PM_{2.5}$ and SO_4^{2-}) exposures were then monitored during 5–7 measurement sessions, randomly spaced approximately 1.5 weeks apart. Details are reported elsewhere.[16] Ambient $PM_{2.5}$ and SO_4^{2-} concentrations were measured during periods corresponding to the personal monitoring sessions

at five fixed-location monitoring stations within the Vancouver, Canada study region.

2.2. Modeling and Simulation

Our approach involves the estimation of the relationship between personal exposure and ambient ($PM_{2.5}$ and sulfate) concentration, followed by simulation of the relationship between personal exposure and risk, here defined as the probability of death [Pr(death)]. The sequential approach includes analysis of the measured exposure data, the application of these estimated relationships to a larger population, modeling the individual-level relationship between exposure and risk of death, and two final components focused on the simulation of the population average risk of death.

2.2.1. Modeling Assumptions

2.2.1.1. Analysis of the Relationship Between Ambient Concentration and Personal Exposure. For the purposes of this illustration, we assume that the relationship between personal exposure and ambient concentration for each individual is linear. Different individuals can have different slopes and intercepts describing this relationship and the variability of the residuals can also be different from individual to individual. These modeling assumptions are motivated by the form of the measured $PM_{2.5}$ and sulfate data for the 16 subjects (Figs. 1 and 2). However, the

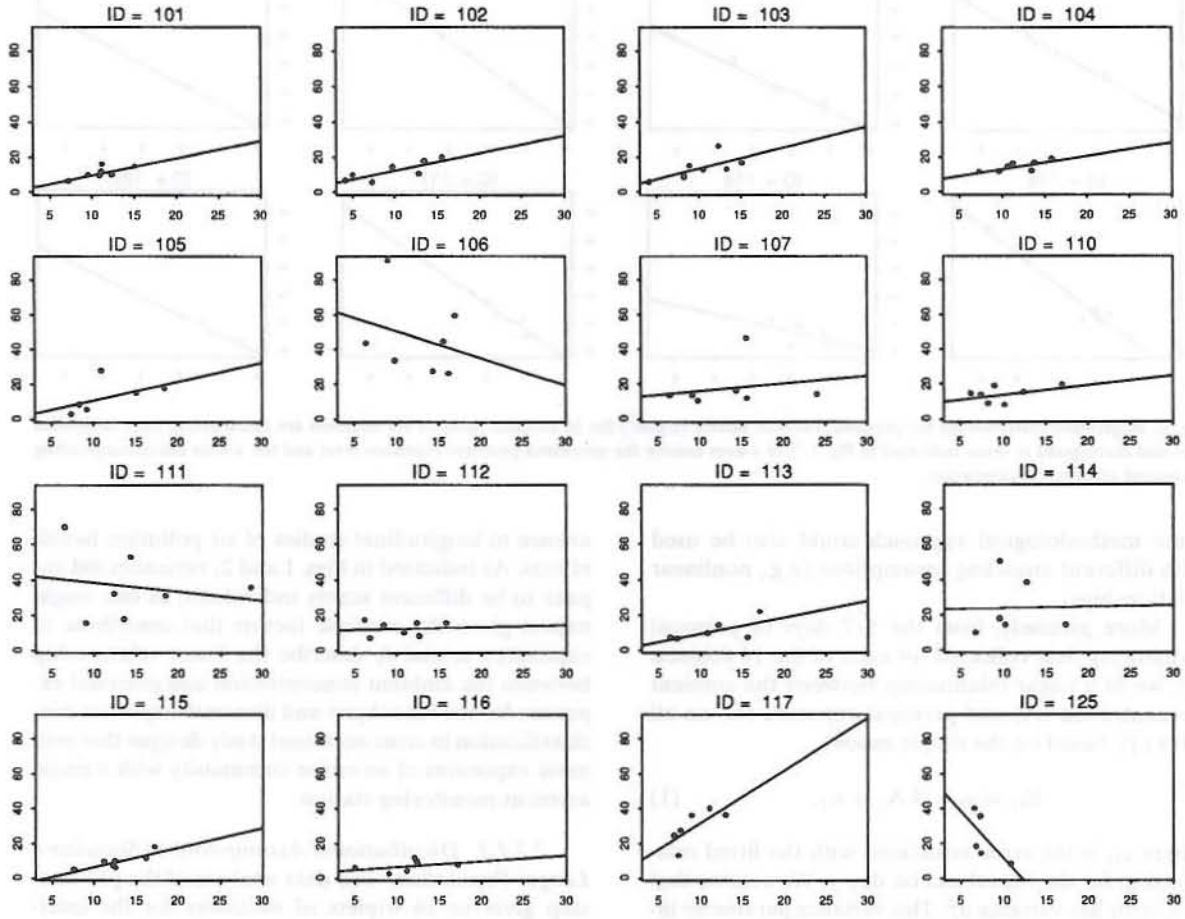


Fig. 1. Regression relationships for personal:ambient $PM_{2.5}$ ($\mu g/m^3$) for 16 subjects. Subject ID numbers are listed above each individual plot and correspond to those indicated in Fig. 2. The y-axes denote the measured personal exposure level and the x-axes the corresponding measured ambient concentration.

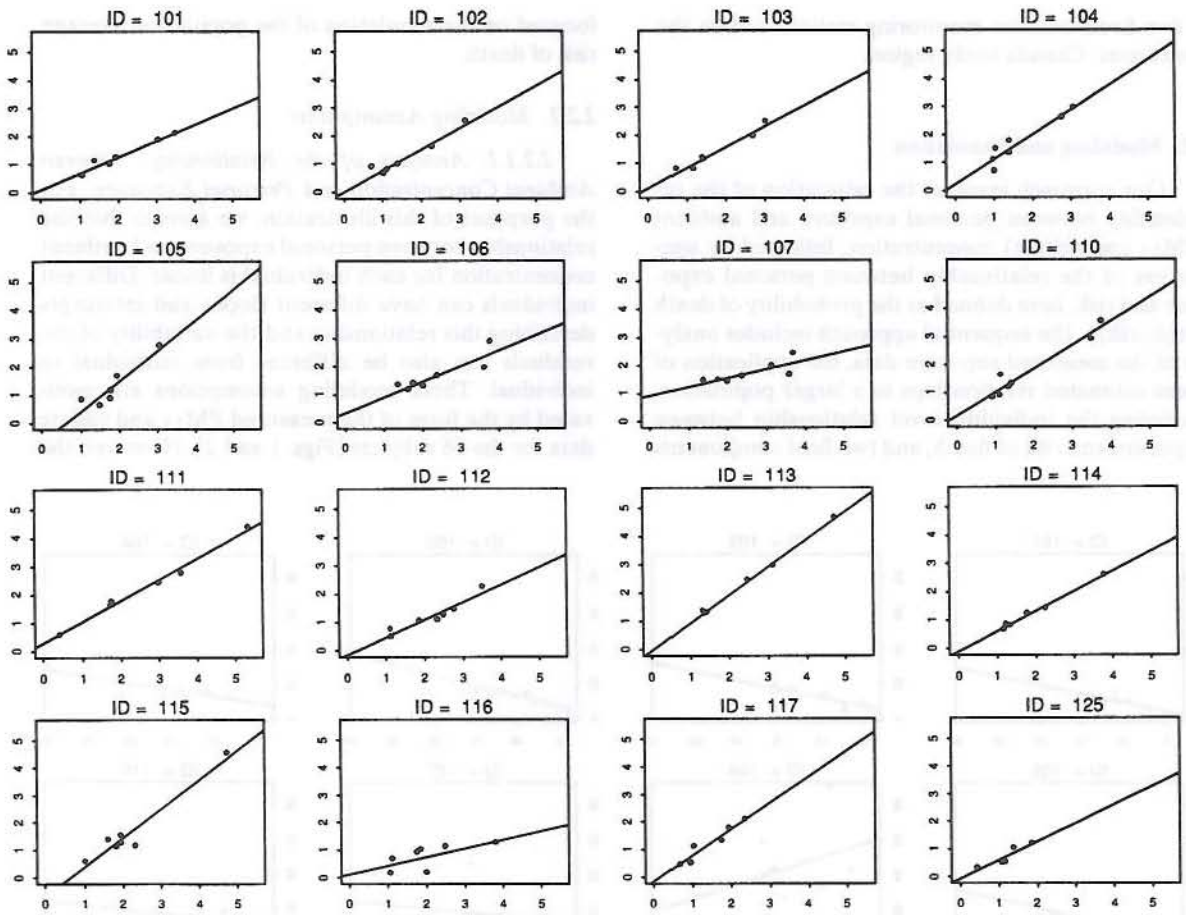


Fig. 2. Regression relationships for personal:ambient sulfate ($\mu g/m^3$) for 16 subjects. Subject ID numbers are listed above each individual plot and correspond to those indicated in Fig. 1. The y-axes denote the measured personal exposure level and the x-axes the corresponding measured ambient concentration.

same methodological approach could also be used with different modeling assumptions (e.g., nonlinear relationships).

More precisely, from the 5–7 days of personal monitoring data collected for each of the 16 subjects (i), we fit a linear relationship between the ambient concentration (A) and personal exposure (E) on all days (j), based on the simple model

$$E_{ij} = \alpha_i + \beta_i A_j + \varepsilon_{ij}, \qquad (1)$$

where ε_{ij} is the error associated with the fitted relationship for the ith subject on day j. We assume that this error has variance σ_i^2. This variance parameter indicates the extent to which exposures of the ith subject track ambient concentrations and has particular relevance to longitudinal studies of air pollution health effects. As indicated in Figs. 1 and 2, variances did appear to be different across individuals, as one might expect given the multiple factors that contribute to exposures. α_i and β_i describe the linear relationship between the ambient concentration and personal exposure for the ith subject and illustrate exposure misclassification in cross-sectional study designs that estimate exposures of an entire community with a single ambient monitoring station.

2.2.1.2. Distributional Assumptions to Simulate a Larger Population. The data analysis of the previous step gives us 16 triplets of estimates for the intercepts α_i, the slopes β_i, and the error variances σ_i^2. We now create a simulated personal exposure profile

corresponding to any particular ambient concentration. We assume that the pairs of intercepts and slopes are a sample from a bivariate normal distribution and the standard deviations from an independent log-normal distribution. To simulate a population of triplets that represent individuals, we draw a pair of values from the bivariate distribution of intercepts and slopes and then independently draw a value from the distribution of error variances. We had insufficient data to support a particular form of dependence between the variances and slopes/intercepts and therefore assumed that these distributions were independent. This assumption enabled us to describe the distribution of the parameters over the population of individuals in a relatively simpler form. Assuming a trivariate distribution of slopes, intercepts, and variances, which also incorporated such dependence, would provide a more comprehensive approach but also would require a substantially larger data set to support the necessary modeling. The means, variances, and co-variances of these distributions are set equal to the sample means, variances, and covariances based on the 16 COPD subjects (Table I, Figs. 1 and 2). At any particular ambient concentration A_j, the personal exposure E_{ij} for the ith individual in this population is then simulated according to Equation (1), where the error E_{ij} is simulated as Gaussian with a mean of 0 and variance σ_i^2.

2.2.1.3. Assumed Form of the Relationship Between Individual Risk and Personal Exposure with a Given Threshold, δ. We model the relationship between the probability of death for the ith individual (P_i) and personal exposure (E_i) as a deterministic relationship $P_i = t(E_i)$, where $t(\cdot)$ is a piecewise linear threshold function (Fig. 3) of the form:

$$t(E) = \gamma_1 + \gamma_2(E - \delta) \quad \text{if} \quad E > \delta$$
and
$$t(E) = \gamma_1 \quad \text{if} \quad E \leq \delta \tag{2}$$

Here we assume that all individuals have a common threshold δ that is determined in advance; below this threshold the probability of dying is constant at the baseline level γ_1. If the personal exposure exceeds this threshold δ, the risk increases linearly with slope γ_2. In a real population one might expect a distribution of individual-level thresholds (δ), and of exposure-response relationships (γ_2). Together, these components determine an individual's susceptibility to a given exposure. For the purposes of our simulations, however, we assumed a common threshold and exposure-response relationship for all individuals so as to isolate the impact of variable exposures on the relationship between population risk and the ambient concentration.

Table I. Parameters of $PM_{2.5}$ and Sulfate Distributions of Personal Exposure:Ambient Concentration Relationships (as Depicted in Figs. 1 and 2) Used in Simulations

Parameter	Mean	Standard Deviation
$PM_{2.5}$		
Slope	0.27	1.78
Intercept	14.75	22.85
Log SD	1.86	0.79
*Slope-intercept r	−0.84	
Sulfate		
Slope	0.74	0.23
Intercept	0.03	0.37
Log SD	−1.61	0.48
Slope-intercept r	−0.72	

*The slope-intercept r refers to the correlation coefficient between the estimated slopes and intercepts.

2.2.2. Simulation

2.2.2.1. Obtain the Population Risk for the Simulated Population at a Given Ambient Concentration for a Particular Threshold Function (Fixed Values of δ, γ_1, and γ_2). We now simulate the individual risks (P_i) at a given ambient level A. The simulated value of $P_i(\mathrm{A}) = \Pr(\text{death of individual } i \mid \text{ambient concentration} = \mathrm{A})$ is denoted by $\hat{P}_i(A)$. We use the sampled values of the regression parameters, from Section 2.2.1.2, to obtain the personal exposures for the simulated population, which in turn are translated into individual risks with the threshold function shown in Equation (2).

We denote the population probability of death (the probability of death at an ambient concentration A for a randomly chosen individual from the population) by $\mathbf{P}(A) = \Pr(\text{death} \mid \text{ambient concentration} = \mathrm{A})$. A given ambient level leads to a distribution of personal exposures in the simulated population, which in turn determines the distribution of the individual probabilities of death $P_i = P_i(\mathrm{A})$. We estimate the population probability of death at each ambient concentration A as the average of the simulated individual probabilities at that ambient concentration:

$$\hat{\mathbf{P}}(A) = \frac{1}{n}\sum_{i=1}^{n} \hat{P}_i(A) \tag{3}$$

The size of the simulated population ($n = 10{,}000$) was chosen to be large enough to control the estimation error involved.

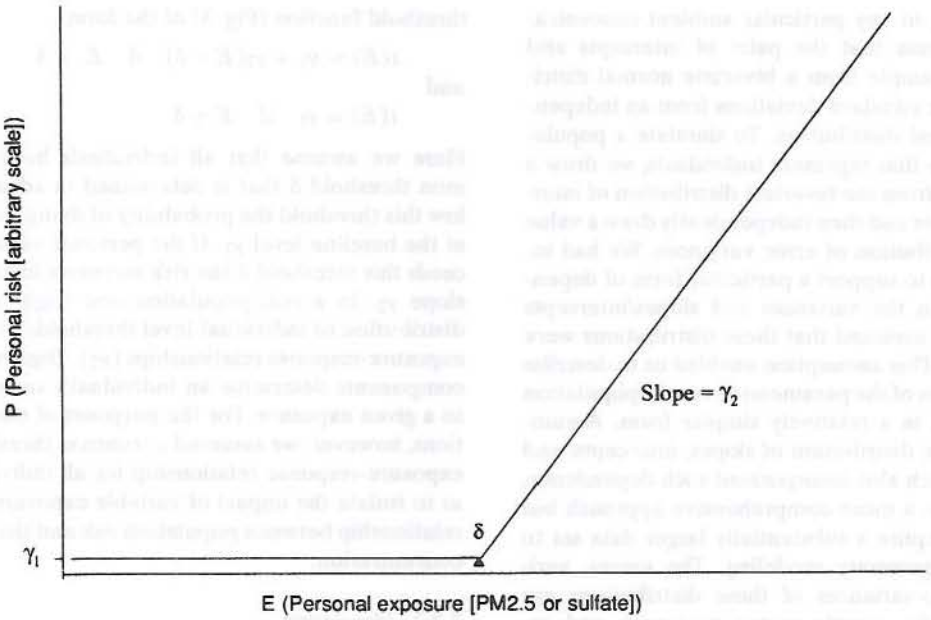


Fig. 3. Illustration of an underlying (individual-level) threshold function.

2.2.2.2. Calibrate the Relationship Between Ambient Exposure and Population Risk. By simulating population risks as described above for a sequence of ambient concentrations, we obtain a relationship between ambient concentration and population risk. To get realistic values for the parameters γ_1 and γ_2 of the threshold function of Equation (2), we varied these parameters until we obtained a relationship between population risk and ambient concentrations that was compatible with mortality rates in the Vancouver area. For the choice of $\delta = 0$, the values of γ_1 and γ_2 are varied and the simulation in Section 2.2.2.1 above is repeated until the simulated population probability of death for an ambient concentration of 0 $\mu g/m^3$ matches the baseline daily mortality (17 per 1,000,000) of the Vancouver area (15) and the percent increase per unit of ambient concentration in the simulation study matches a target risk function, the mortality risk estimates from the WHO Air Quality Guidelines:[11]

$$\% \text{ increase in daily mortality} = 0.151^{*} \text{ PM}_{2.5}$$

$$\% \text{ increase in daily mortality} = 0.60^{*} \text{ sulfate}$$

We then used these parameter values ($\gamma_1 = 1.70 \times 10^{-6}$ and 1.70×10^{-6}, $\gamma_2 = 7.99 \times 10^{-9}$ and 3.20×10^{-8} for $PM_{2.5}$ and sulfate, respectively) in all the simulations.

Initially, we specify only the functional form and we fix the threshold value. We then perform simulations for different values of δ to explore the impact of this parameter on the population risk-ambient concentration relationship.

With this approach, we illustrate the extent to which a common individual-level threshold is obscured by the error in using ambient concentrations as surrogates for personal exposure. In this way, we can determine, for example, that if a common underlying individual-level threshold exists, it would have to be above a certain concentration to manifest itself at the population level.

3. RESULTS

Figs. 1 and 2 display the personal exposure versus ambient concentration relationships measured for the 16 subjects for $PM_{2.5}$ and sulfate, respectively. The distributions of the key features of these regression fits are summarized in Table I. Note that the intercept term, which describes the personal exposure at an ambient concentration of 0, is much higher for $PM_{2.5}$ than for sulfate, due to the impact of indoor sources and personal activities on personal exposure to $PM_{2.5}$. Also note that the standard deviation of the slopes is much larger for the $PM_{2.5}$ relationships, reflecting the

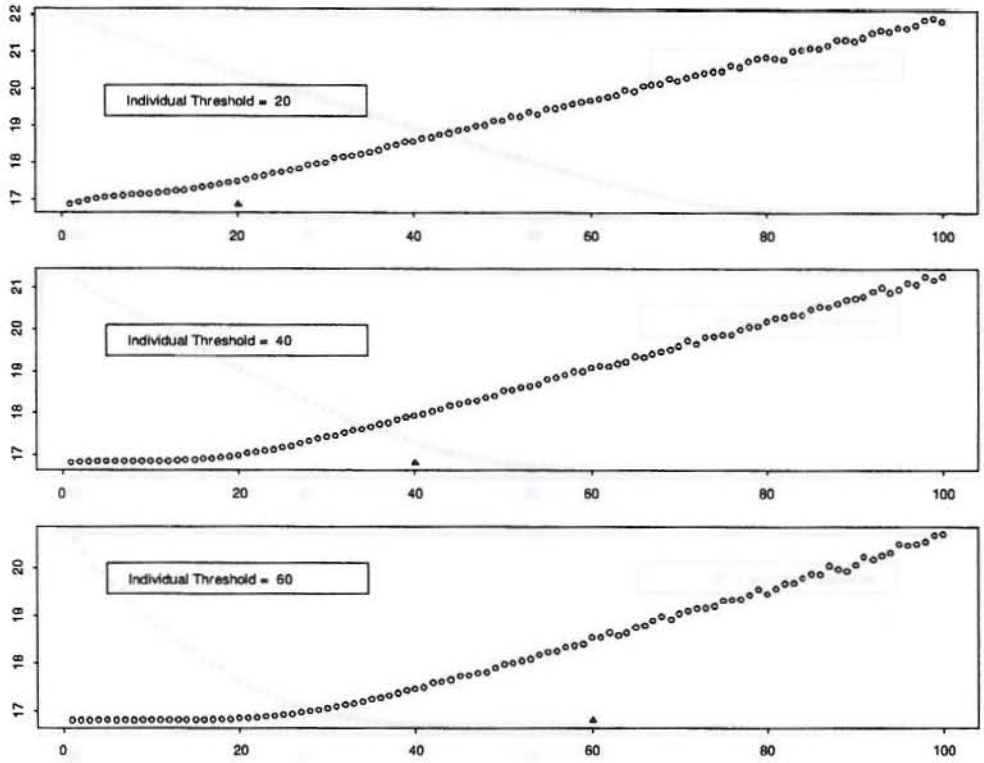


Fig. 4. Expected number of deaths per 1,000,000 (y-axis) versus ambient concentrations ($\mu g/m^3$) of $PM_{2.5}$ (x-axis). Individual-level thresholds, δ, are denoted by ▲.

much greater consistency across subjects in the sulfate relationships apparent in Figs. 1 and 2.

Fig. 4(a–c) depicts some of the results of the simulation for $PM_{2.5}$. Here, three individual-level thresholds (20, 40, and 60 $\mu g/m^3$) were specified and the resulting population risk is plotted against the ambient particle levels. No threshold is apparent on a population basis if an individual threshold of 20 $\mu g/m^3$ is specified (Fig. 4(a)). Even at higher individual-level thresholds, the apparent threshold at the population level is 20–30 $\mu g/m^3$ lower than the specified personal threshold (Fig. 4(b–c)). In contrast, for sulfate (Fig. 5(a–c)), the personal-level thresholds closely match the population-level thresholds.

4. DISCUSSION

In these simulations we have demonstrated that the use of surrogate measures that are not highly correlated with personal exposures can obscure a threshold at the population level, even if a common threshold exists for individuals within the population. However, if exposure misclassification is reduced by the use of appropriate exposure metrics (in this example, measured exposures that are highly correlated with ambient concentrations), then common underlying individual thresholds result in similar population-level thresholds. Although we have conducted these simulations for data regarding ambient particulate air pollution, the same principles apply to any situation where exposure is misclassified by the use of surrogate measures to estimate individual exposures. In the Appendix we describe an analytic derivation of the relationship between personal exposure and population-level risk.

In our simulations we have assumed that all individuals in a population have the same threshold concentration and the same slope of their concentration-response relationship. This simple situation was examined as it was our intention to isolate the potential impact of exposure misclassification on threshold detection. The simulation and the analytic solution could also be generalized to incorporate a more realistic scenario with a distribution of individual

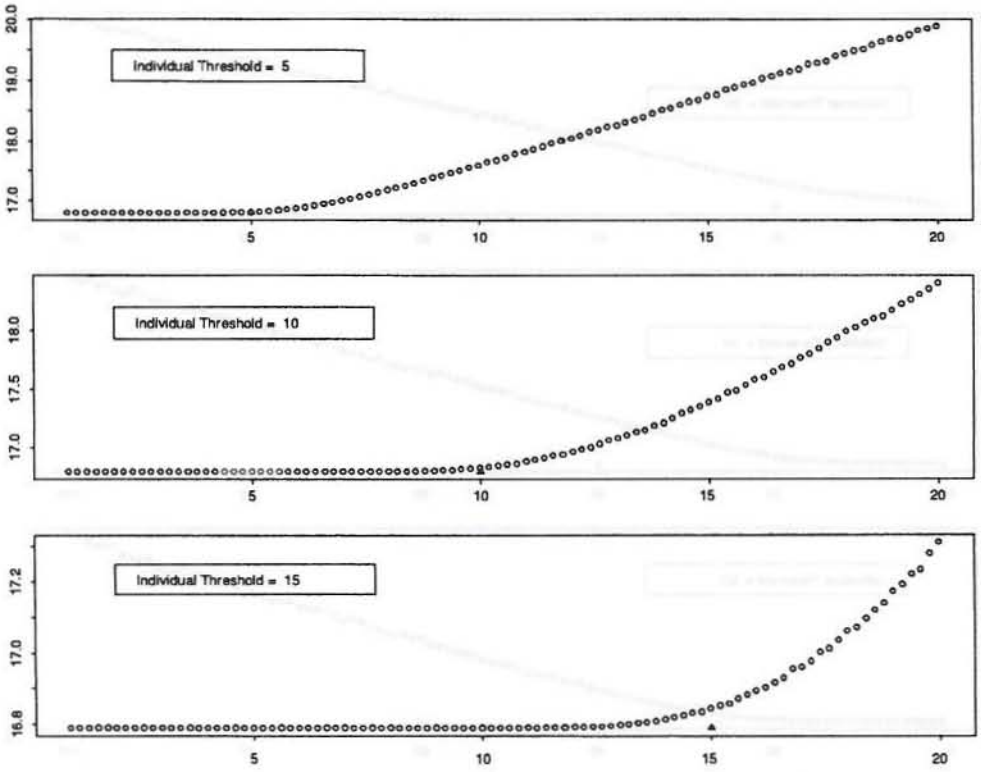


Fig. 5. Expected number of deaths per 1,000,000 (y-axis) versus ambient concentrations ($\mu g/m^3$) of sulfate (x-axis). Individual-level thresholds, δ, are denoted by ▲.

susceptibilities (both slopes and thresholds), although that was not the purpose of this exercise. The obscuring of thresholds, as observed in our simulations, would clearly be even greater if the simulations incorporated thresholds that vary across individuals. Our simulations also focused on a single dose-response function to illustrate the methodology and general findings in one specific, simple, yet realistic, scenario. Our goal was to present a general empirical methodology that could also incorporate alternative assumptions, such as other dose-response functions.

Although it is generally understood that measurement error and, more specifically, exposure misclassification, can lead to biased estimates of effect, the impact on thresholds has received less attention. Watt and colleagues measured PM_{10} exposures of traffic officers and used these data to assess the effect of exposure estimates based on ambient concentrations on the shape of the air pollution exposure-response curve.[15] Personal exposures were 6–10 times higher than ambient measurements and, in a limited simulation, this difference had the effect of almost completely obscuring the assumed threshold for health effects. Here we have built on that preliminary study by generalizing the model used in the simulations and by comparing ambient concentrations and personal exposures for the same particle measurements, over the same averaging period, in a group that is representative of individuals affected by particulate air pollution.

Although the comparison between $PM_{2.5}$ and sulfate is presented here as an example, readers may infer that personal $PM_{2.5}$ exposure is the "gold standard" against which $PM_{2.5}$ measured at central monitoring sites is to be compared. In truth, the gold standard for ambient $PM_{2.5}$ is that component of personal $PM_{2.5}$ that is due to exposure to ambient $PM_{2.5}$.[18] Sources of personal $PM_{2.5}$ that do not derive from ambient $PM_{2.5}$, including all indoor sources, should not be considered when attempting to assess measurement error resulting from the use of centrally monitored concentrations. Based on the relatively strong correlation between centrally monitored and personal sulfate, the

correlation between centrally monitored $PM_{2.5}$ and that component of personal $PM_{2.5}$ due to ambient $PM_{2.5}$ is also strong. Therefore, the specific findings of this simulation do not apply to $PM_{2.5}$. The findings do apply to pollutant measures for which the correlation between centrally monitored concentrations and personal exposures to the ambient pollutant are weak. Carbon monoxide and ozone may be examples of such pollutants.

5. LIMITATIONS

As discussed by Zeger and colleagues,[7] there are many causes of exposure misclassification. Here we have shown an example in which misclassification occurs due to measurement error and the use of a nonspecific exposure metric. Using the general methodology we have provided would show, for example, that a reduction in the amount of measurement error in our data would lead to improved agreement between an underlying individual threshold and that based on aggregated population-level data. Additional limitations in the data that we used for the simulation are the low concentrations that were measured in relation to the thresholds that were assumed for the simulations and the small number of repeated measurements (5–7) for each subject. It is possible that for examples with higher ranges of exposures the correlations between personal and ambient measurements would be higher. As with the reduction in measurement error, this would lead to a smaller difference between apparent and underlying thresholds. Further, while it was our intention to recruit more subjects and to collect more repeat measurements, this was not feasible logistically. The simulations we have described could be repeated if larger data sets become available. Finally, the results of the simulation exercises depend on the specific form of risk function that is used. In our example we used a function from the WHO Air Quality Guidelines, although the methodology allows any function to be used. These limitations indicate that our quantitative results are sensitive to the input data used. However, the methodology and analytical solution that we present are general and can be applied to other data sets. Further, our general illustration of the ability of exposure misclassification to obscure thresholds remains despite these limitations.

6. CONCLUSION

The identification of threshold levels is important for regulatory standards, risk assessments, and economic valuations, which are often incorporated in cost-benefit analyses. Specifically, for environmental exposures such as air pollution, which tend to be episodic, concentrations tend to be low for the majority of the time. In performing economic valuations or risk assessments, a decision must be made as to whether the given exposure-response relationship is applied to low levels, or if some threshold is set under which it is assumed that no effects occur.[11] Exactly what level is chosen for this threshold can have a dramatic influence on the results of the assessment, since concentrations in many locations are below these levels most of the time. For standard setting, regulators often are faced with a dilemma of incorporating epidemiological results that do not indicate a threshold, with regulatory requirements that stipulate that a specific level should be indicated. The results described in this simulation suggest that the inability to detect a threshold in many epidemiological studies does not, in fact, mean that no threshold exists. Further, the results of this simulation imply that improved characterization of exposure will improve the ability of epidemiological studies to identify threshold levels that are consistent with those actually experienced by the individuals in the study population.

APPENDIX

Evaluation of the Population Risk

The body of the article described how we simulated a population of personal exposures that leads to a distribution of individual risks and hence to an estimate of the population risk (probability of death) at a given ambient exposure level. Here we present the corresponding analytical expressions.

In what follows, relationships are considered for a fixed day, so the argument corresponding to the day (j) is suppressed in all the expressions. We modeled the personal exposure E_i for the ith individual on a day with ambient exposure A as

$$E_i = \alpha_i + \beta_i A + \varepsilon_i$$

where ε_i was assumed to be normally distributed with mean 0 and variance σ_i^2. This assumption specifies the conditional distribution of the personal exposure E_i given the ambient exposure A and the individual's vector of parameters $\theta_i = (\alpha_i, \beta_i, \sigma_i)$.

Individuals are characterized by their vector of parameters θ_i, which are distributed across the population according to a trivariate density $\pi(\theta_i)$. In the simulations, we assumed that the slope and intercept

parameters (α_i, β_i) were bivariate normal and the variances (σ_i^2) were independently log-normal, in which case $\pi(\theta_i)$ factors into a product of bivariate normal and log-normal densities.

The probability of death for the ith individual, P_i, was modeled as a deterministic function $t(\cdot)$ of the personal exposure: $P_i = t(E_i) = t(\alpha_i + \beta_i A + \varepsilon_i)$. The error term ε_i in this expression implies that, given the individual's vector of parameters θ_i, the probability of death for the ith individual is a random quantity. Thus, interest focuses on the expected probability of death for the ith individual, which is given by

$$\mathbf{P}(A \mid \theta_i) = \int t(\alpha_i + \beta_i A + \sigma_i z)\phi(z)dz,$$

the quantity estimated by the simulated individual risk $\hat{P}_i(A)$; see Equation (3). The corresponding expected population risk (probability of death) is given by

$$\mathbf{P}(A) = \int \mathbf{P}(A \mid \theta_i)\pi_i(\theta_i)d\theta_i$$

the quantity estimated by $\hat{\mathbf{P}}(A)$ of Equation (3).

For the special case (2) of the threshold function $t(\cdot)$ used in the simulations, we have

$$\mathbf{P}(A \mid \theta_i) = \gamma_1 + \gamma_2\sigma_i\Psi\left(\frac{\delta - (\alpha_i + \beta_i A)}{\sigma_i}\right)$$

where $\Psi(u) = \phi(u) - u[1 - \Phi(u)]$, with $\phi(\cdot)$ and $\Phi(\cdot)$ the standard normal density and distribution function, respectively. Evaluation of the expected population risk when the ambient exposure is A then requires three-dimensional integration of this function with respect to $\pi(\theta_i)$, the joint trivariate distribution of the parameters. Some simplification results from the assumption made in the simulations that σ_i is distributed independently of (α_i, β_i), but this evaluation (and the simulations) would be no more difficult for other choices of distributions for these parameters.

We modeled the relationship between the personal exposure and the individual risk as a deterministic threshold function $t(\cdot)$ that is the same for all individuals. This could easily be generalized in a variety of ways, but we do not pursue this here.

This Appendix has focused on evaluation of the expected population risk (probability of death) as that is the function estimated in the simulations reported in the article. Expressions could also be obtained for other functions of the distribution of personal exposures. For example, an expression for the distribution of individual risks when the ambient exposure is A could be obtained. In general, explicit evaluation of such expressions would still require multidimensional integration.

ACKNOWLEDGMENTS

This research was supported in part by grants from the British Columbia Lung Association, a British Columbia Lung Association/Medical Research Council of Canada Scientist Award (M. Brauer), and an American Lung Association Career Investigator Award (M. Brauer).

REFERENCES

1. Pope III, C. A., Dockery, D. W., & Schwartz, J. (1995). Review of epidemiological evidence of health effects of particulate air pollution. *Inhalation Toxicology, 7*, 1–18.
2. Dockery, D. W., & Pope III, C. A. (1994). Acute respiratory effects of particulate air pollution. *Annual Review of Public Health, 15*, 107–132.
3. Vedal, S. (1997). Ambient particles and health: Lines that divide. *Journal of the Air and Waste Management Association, 47*, 551–581.
4. Brunekreef, B., Dockery, D. W., & Krzyzanowski, M. (1995). Epidemiologic studies on short-term effects of low levels of major ambient air pollution components. *Environmental Health Perspectives, 103*(2), 3–13.
5. Gamble, J. F., & Lewis, R. J. (1996). Health and respirable particulate (PM10) air pollution: A causal or statistical association? *Environmental Health Perspectives, 104*, 838–850.
6. Gamble, J. F. (1998). PM2.5 and mortality in long-term prospective cohort studies: Cause-effect or statistical associations? *Environmental Health Perspectives, 106*, 535–549.
7. Zeger, S. L., Thomas, D., Dominici, F., Samet, J. M., Schwartz, J., Dockery, D., & Cohen, A. (2000). Exposure measurement error in time-series studies of air pollution: Concepts and consequences. *Environmental Health Perspectives, 108*, 419–426.
8. Carrothers, T. J., & Evans, J. S. (2000). Assessing the impact of differential measurement error on estimates of fine particle mortality. *Journal of the Air and Waste Management Association, 50*, 65–74.
9. Cakmak, S., Burnett, R. T., & Krewski, D. (1999). Methods for detecting and estimating population threshold concentrations for air pollution-related mortality with exposure measurement error. *Risk Analysis, 19*, 487–496.
10. Hall, J. V., Winer, A. M., Kleinman, M. T., Lurmann, F. W., Brajer, V., & Colome, S. D. (1992). Valuing the health benefits of clean air. *Science, 255*, 812–817.
11. WHO. (1999). *Air quality guidelines*. World Health Organization. Available at www.who.int/peh/air/Airqualitygd.htm.
12. Schwartz, J. (2000). Assessing confounding, effect modification, and thresholds in the association between ambient particles and daily deaths. *Environmental Health Perspectives, 108*, 563–568.
13. Daniels, M. J., Dominici, F., Samet, J. M., & Zeger, S. L. (2000). Estimating particulate matter-mortality dose-response curves and threshold levels: An analysis of daily time-series for the 20 largest US cities [see comments]. *American Journal of Epidemiology, 152*, 397–406.
14. Armstrong, B. (1995). Study design for exposure assessment in epidemiological studies. *Science of the Total Environment, 168*, 187–194.
15. Watt, M., Godden, D., Cherrie, J., & Seaton, A. (1995). Individual exposure to particulate air pollution and its relevance

to thresholds for health effects: A study of traffic wardens. *Occupational and Environmental Medicine, 52*, 790–792.
16. Ebelt, S. T., Petkau, A. J., Vedal, S., Fisher, T. V., & Brauer, M. (2000). Exposure of chronic obstructive pulmonary disease (COPD) patients to particles: Relationship between personal exposure and ambient air concentrations. *Journal of the Air and Waste Management Association, 50*, 1081–1094.
17. Sarnat, J. A., Koutrakis, P., & Suh, H. H. (2000). Assessing the relationship between personal particulate and gaseous exposures of senior citizens living in Baltimore, MD. *Journal of the Air and Waste Management Association, 50*, 1184–1198.
18. Mage, D., Wilson, W., Hasselblad, V., & Grant, L. (1999). Assessment of human exposure to ambient particulate matter. *Journal of the Air and Waste Management Association, 49*, 1280–1291.