HEALTH EFFECTS OF AIR POLLUTION: A STATISTICAL REVIEW

Francesca Dominici[1], Lianne Sheppard[2] and Merlise Clyde[3]

[1]*Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205-3179, USA. E-mail:fdominic@jhsph.edu*

[2]*Departments of Biostatistics and Environmental Health, School of Public Health and Community Medicine, University of Washington, Seattle, WA, USA. E-mail:sheppard@u.washington.edu*

[3]*Merlise Clyde, Institute of Statistics and Decisions Sciences, Duke University, Durham, NC 27708-0251, USA. E-mail:clyde@stat.duke.edu*

Summary

We critically review and compare epidemiological designs and statistical approaches to estimate associations between air pollution and health. More specifically, we aim to address the following questions:

1. *Which epidemiological designs and statistical methods are available to estimate associations between air pollution and health?*
2. *What are the recent methodological advances in the estimation of the health effects of air pollution in time series studies?*
3. *What are the the main methodological challenges and future research opportunities relevant to regulatory policy?*

In question 1, we identify strengths and limitations of time series, cohort, case-crossover and panel sampling designs. In question 2, we focus on time series studies and we review statistical methods for: 1) combining information across multiple locations to estimate overall air pollution effects; 2) estimating the health effects of air pollution taking into account of model uncertainties; 3) investigating the consequences of exposure measurement error in the estimation of the health effects of air pollution; and 4) estimating air pollution-health exposure-response curves. Here, we also discuss the extent to which these statistical contributions have addressed key substantive questions. In question 3, within a set of policy-relevant-questions, we identify research opportunities and point out current data limitations.

Key words: Time series, Air Pollution, Epidemiological Designs, Hierarchical Models, Exposure Measurement error, Model Choice

1 Introduction

The potential for air pollution at high concentrations to cause excess deaths was established in the mid-twentieth century by a series of air pollution "disasters" in the US and Europe (Logan and Glasg, 1953; Ciocco and Thompson, 1961) which followed striking increases in mortality. By the early 1990's, time series studies, each conducted at a single location (Lipfert and Wyzga, 1993; Pope et al., 1995a; Schwartz, 1995; American Thoracic Society, 1996a,b), showed that air pollution levels, even at much lower concentrations, were associated with increased rates of mortality and morbidity in several cities in the United States, Europe and other developed countries. At present, although these relative rates are small (a few percentage points increase in mortality or morbidity over a realistic exposure range), the burden of disease attributable to air pollution may be substantial considering the large population exposed to air pollution and to whom the relative rates of mortality/morbidity apply (Murray and Lopez, 1996).

The statistical analysis of epidemiological data on air pollution and health presents several methodological issues (National Research Council, 1998, 1999, 2001). The health effects (acute or chronic) and exposure (short-term, lifetime) of interest motivate the choice of the epidemiological design for data collection, and the characteristics of the design influence the statistical methods to be used. Identification of the unique effects of a specific pollutant requires careful adjustment for simultaneous exposure to a complex mixture of co-pollutants. Extensive covariate adjustment is required to control for confounding factors (as for example age effects, and weather variables); key confounding variables will differ depending on the choice of design. In addition, exposure measurement error can potentially lead to bias in estimates of the health effects of air pollution. As much of modern environmental epidemiology is oriented towards the analysis and evaluation of complex data on the health effects of air pollution, advanced statistical methods, such as generalized regression models for count and binary time series data (Liang and Zeger, 1986; McCullagh and Nelder, 1989), generalized additive models (Hastie and Tibshirani, 1990), Cox proportional hazard models (Cox and Oakes, 1984), and hierarchical models (Lindley and Smith, 1972; Morris and Normand, 1992), have figured prominently in epidemiological applications.

In this article, we review some of the most relevant sampling designs and statistical methods for investigating associations between exposure to air pollution and health outcomes, and discuss how these statistical methods have provided a clearer understanding of the extent to which air pollution affects human health. Although there has been progress in the investigation of the association between air pollution and health, several important questions still need to be addressed, and there exists considerable need for further data collection and methodological development. Future research opportunities, directly relevant to regulatory policy, are identified. We also highlight the main limitations of the current epidemiological investigations in terms of the data collection, design and statistical methods, and identify current gaps in research.

More specifically, we address the following questions related to the estimation and interpretation of the health effects of air pollution:

1. Which epidemiological designs and statistical methods are available to estimate associations between air pollution and health? What are their strengths and limitations? (Section 2)

2. What are the recent methodological advances in the estimation of the health effects of air pollution in time series studies? To what extent have recent statistical contributions addressed key substantive issues? (Section 3)

3. What are the future research opportunities relevant to regulatory policy? What novel statistical methods need to be developed to address these questions? (Section 4)

In summary, in this paper we bring together methodological and substantive contributions in the estimation of the health effects of air pollution. These two research focus have been reviewed and discussed, separately: for example, see Patil et al. (1998); Thomas (2000); Guttorp (2000) for environmental statistical reviews and Lipfert (1994); Schwartz (1994a); Dockery and Pope (1994); Pope et al. (1995a); Thurston (1996) for reviews of epidemiological evidence of health effects of particulate air pollution.

2 Epidemiological Designs and Statistical Methods

Most of the air pollution studies on humans have been observational studies.[1] Furthermore, the great majority of air pollution studies are also "opportunistic" studies that arise out of the opportunity to combine data from sources compiled for other purposes, and are often limited by the type, structure, and amount of data available. One of the main data opportunities in air pollution research is the wealth of available ambient monitoring data (see for example the Aerometric Retrieval Information System `http://www.epa.gov/airs/airs.html`). Monitors are located at fixed sites, often chosen for regulatory purposes, and thus are best at capturing time-varying ambient and local source concentrations of selected pollutants.

Here, we assume that assessing the health effects of ambient air pollution on individuals in a population is the primary goal of air pollution epidemiology studies. Health effects can be either acute or chronic. Acute effects are transient and due to time-varying exposures. Chronic effects are more likely due to the cumulative effects of exposure, but could be associated with more complex functions of lifetime exposure. Outcomes can be major or minor life events (e.g. death or onset of symptoms) or changes in function (e.g. vital capacity, lung growth, symptom severity). The nature of the outcome (e.g. binary or continuous) and the structure of the data lead to the selection of the model and the types of effects to be estimated. Regression models are generally the method of choice.

A key design consideration for statistical analysis is whether there is sufficient variation in the exposure. For exposure to ambient air pollution, the sources of variation come from the temporal

[1] The few exceptions are chamber studies of gaseous pollutants and microarray gene expression analysis of PM constituents and other pollutants; these will not be discussed in this review.

and/or spatial distribution of pollution. When we consider personal exposure to pollution, the spatial distribution is broadened to include personal behaviors and individual micro-environments. Because accurate estimation of relevant exposures for an individual participant can be daunting, most air pollution studies have used ambient air pollution measurements, often taken from central site monitors. Misclassification of exposure is a well-recognized limitation of these studies, and it will be discussed in the next section.

Biases due to confounding and correlation among covariates can never be completely ruled out in observational data. As a result bias may often dominate efficiency considerations when choosing among designs. Confounding is present when a covariate that is not part of the exposure casual pathway, is associated with both the exposure and outcome.

A related issue is the high intercorrelation of different pollutants in the atmosphere. One of the ways to address the separate effects of highly intercorrelated pollutants has been to conduct studies in locations where one or more pollutants are absent or nearly so. For multi-location studies, this can be achieved by designing the study to have large exposure contrasts.

In practice, most air pollution epidemiology study designs have fallen into four types: ecological time series, case-crossover, panel, and cohort studies. Cross-sectional studies have also been done but are less common. Conceptually, panel studies collect individual time and space varying outcomes, exposures, and confounders and therefore they encompass all other epidemiological designs which are based on spatially and/or temporally aggregated data. In practice, panel studies also rely on group-level data. The time series, case-crossover, and panel studies are best suited for estimating the acute effects of air pollution, while cohort studies estimate acute and chronic effects combined. This section is devoted to a brief description of each study design, identification of the corresponding approaches to statistical analysis, and presentation of examples. It concludes with a comparison of these designs.

2.1 Time Series Studies

Time series studies associate time-varying pollution exposures to time-varying event counts. These are a type of ecologic study because they analyze daily population-averaged health outcomes and exposure levels. If the health effects are small and the disease outcomes are rare, the bias from ignoring the data-aggregation across individuals should be small (Wakefield and Salway, 2001).

Generalized Linear Models (GLM) with parametric splines (e.g. natural cubic splines) (McCullagh and Nelder, 1989) or Generalized Additive Models (GAM) with non parametric splines (e.g. smoothing splines or lowess smoothers) (Hastie and Tibshirani, 1990), are used to estimate effects associated with exposure to air pollution while accounting for smooth fluctuations in the mortality that confound estimates of the pollution effect. GAM is the most widely applied method because it allows for non-parametric adjustments for non-linear confounding effects of seasonality, trends, and weather variables, and it is a more flexible approach than fully-parametric alternatives like the GLM with natural cubic spline. However, recently the GAM software implementation in S-Plus has been called into question and more systematic comparisons between GAM and GLM are underway

(Dominici et al., 2002b; Ramsay et al., 2002). Further details are in section 4.

GAM or GLM for the mortality counts, Y_t, and daily levels of exposures, X_t, take the following additive form:

$$E(Y_t) = \exp\{\beta_0 + \beta X_{t-\ell} + S(\text{time}, \lambda_1) + S(\text{temp}, \lambda_2) + \gamma \times \text{DOW}\} \quad (1)$$

where temp is a measure of temperature (average daily and/or dew point), DOW are indicator variables for day of the week and ℓ is the lag of the pollution exposure which is generally restricted from 0 to 7 days. The smooth function $S(\cdot, \lambda)$ denotes a smooth function of a covariate (calendar time, temperature, humidity) often constructed using smoothing splines, loess smoothers, or natural cubic splines with a smoothing parameter λ. The smoothing parameter λ represents the number of degrees of freedom in the smoothing spline, the span in the loess smoother and $\lambda - 2$ interior knots in the natural cubic splines. The parameter of interest β describes the change in the logarithm of the population average mortality count per unit change in $X_{t-\ell}$, and it is generally interpreted as the % increase in mortality per 10 unit increase in ambient air pollution levels at lag ℓ. Typically residual serial dependence is negligible in mortality time series with adequate control for temporal trends and can be ignored in the analysis. Residual serial dependence is more common in morbidity time series (Sheppard et al., 1999) and adjustment is possible (Lumley and Heagerty, 1999).

Rather than focus on the effect associated with a single lag of a pollution variable, distributed lag models and time-scale models can be used to estimate cumulative and longer time-scale health effects of air pollution. Distributed lag models (Almon, 1965; Zanobetti et al., 2000b, 2002) are used to estimate associations between health outcome on a given day, and air pollution several days prior by replacing βX_t in (1) with $\theta \sum_{\ell=1}^{L} \eta_\ell X_{t-\ell}$, $\sum_{\ell=1}^{L} \eta_\ell = 1$ where θ measures the cumulative effect, and η_ℓ measures the contribution of the lagged exposure $X_{t-\ell}$ to the estimation of θ. Time-scale models (Zeger et al., 1999; Schwartz, 2000b, 2001; Dominici et al., 2002c) are used to estimate associations between smooth variations of air pollution and daily health outcomes by replacing βX_t in (1) with $\sum_{k=1}^{K} \beta_k W_{tk}$, $\sum_{k=1}^{K} W_{kt} = X_t$ where $W_{1t}, \ldots, W_{kt}, \ldots, W_{Kt}$ is a set of orthogonal predictors obtained by applying a Fourier decomposition to X_t. The parameters β_k denote the log relative rate of the health outcome for increases in air pollution at time scale k. Time scales of interest are short-term air pollution variations (1 to 4 days) and longer-term variations (1 to 2 months) capturing acute and less acute health effects, respectively. Beyond two months it is likely that any effects are dominated by seasonal confounding.

The smooth function $S(\text{time}, \lambda_1)$ in the model is used to adjust for smooth fluctuations in mortality over time, so that only shorter-term variation in mortality and air pollution (less than 2 months) is used to estimate β (Samet et al., 1995; Kelsall et al., 1997). The smoothing parameter λ_1 can be pre-specified based upon prior epidemiologic knowledge of the time scale of the major possible confounders. For non-accidental mortality studies, a reasonable choice is to set λ_1 equal to seven degrees of freedom per year of data so that little information from time scales longer than approximately two months is included when estimating β. This choice largely eliminates expected confounding from seasonal changes in mortality (e.g. influenza epidemics), and from

longer-term trends (e.g. changes population size, population composition, medical practice, or health behaviors), while retaining as much information as possible.

Similarly, to control for weather, smooth functions of temperature variables, $S(\text{temp}, \lambda_2)$, such as same day temperature, average temperature for the three previous days, maximum temperature, and the analogous functions for dew point and/or humidity are included in the model. In many United States cities, mortality decreases smoothly with increased temperature until reaching a relative minimum and then increases quite sharply at higher temperatures. Here a smoothness parameter greater than three is necessary to capture the highly non-linear bend in mortality as a function of temperature. See Kelsall et al. (1997); Cakmak et al. (1998) for further discussion of adjustments for confounding factors in time series studies.

The National Morbidity, Mortality and Air Pollution Study (NMMAPS) (Samet et al., 2000c,b,a) is the largest multi-site time series study yet conducted. In each of 90 cities the daily total mortality count was regressed on PM_{10} air pollution (particulate matter less than 10 microns in aerodynamic diameter) using model (1) with lags $\ell = 0, 1,$ and 2 to estimate the relative increase in mortality rates associated with an incremental increase of 10 $\mu g/m^3$ of PM_{10}.

The NMMAPS modeling approach uses the same set of confounding variables in the different locations, with the same number of degrees of freedom. Alternative approaches would have been to select these confounding variables differently within each city, for example by inspection of the autocorrelation function of the residuals of the mortality time series or by minimizing the Akaike Information Criterion (AIC) (Akaike, 1973). Although there are several reasonable alternatives for obtaining adequate control for confounding, an optimal method has not been identified.

Unlike most other air pollution time series studies that concentrate on a single city, the goal of NMMAPS is to estimate city-specific, regional, and national effects of PM_{10} on mortality. Hierarchical models are particularly suitable for combining relative rates across locations. These methods and further NMMAPS results are discussed in the following sections.

2.2 Case-Crossover Studies

The case-crossover design is used to estimate the risk of a rare event associated with a short-term exposure. It was originally proposed by Maclure (1991) to study acute transient effects of intermittent exposures. The case-crossover design can be viewed heuristically as a modification of the matched case-control design (Breslow and Day, 1980; Schlesselman, 1994) where each case acts as his/her own control, and the distribution of exposure is compared between cases and controls. The distinction from a case-control study is that exposures are sampled from an individual's time-varying distribution of exposure. More specifically, the exposure at the time just prior to the event (the *case* or *index time*) is compared to a set of *control* or *referent times* that represent the expected distribution of exposure for non-event follow-up times. In this way, the measured and unmeasured time-invariant characteristics of the subject (such as gender, age, smoking status) are matched, minimizing the possibility of confounding.

In the last decade of application, it has been shown that the case-crossover design is best suited

to study intermittent exposures inducing immediate and transient risk, and abrupt rare outcomes (Maclure and Mittleman, 2000). This design has been found to be topical for estimating the risk of a rare event associated with a short-term exposure because the widespread availability of ambient monitoring data presents opportunities to further analyze existing cases series from case-control studies.

The data for the case-crossover design consist of exposure measures X_{it} for subjects $i = 1, \ldots, n$ at times $t = 1, \ldots, T$, the *index times* t_i, $i = 1, \ldots, n$, and individual characteristics that can be used to evaluate effect modification (e.g. smoking status). The referent sampling scheme determines the referent sets W_i of exposure that are used in the analysis. Estimates of the relative risks are obtained by solving the following estimating function:

$$U_i(\beta) = X_{it_i} - \sum_{t \in W_i} \frac{X_{it} e^{\beta X_{it}}}{\sum_{s \in W_i} e^{\beta X_{is}}}. \tag{2}$$

For some referent sampling schemes, this estimating function is identical to the score equation derived from a conditional likelihood. For others there is no corresponding conditional likelihood (Lumley and Levy, 2000). Standard conditional logistic regression software can be used for this analysis.

A key difficulty in case-crossover studies is how to properly define the referent sets W_i. Control for bias in the estimation of the relative risk β is the dominant concern in the choice of the referent sampling strategy, although the size of the referent set also affects efficiency. Two main sources of bias in case-crossover studies have been identified. The first arises from the trend and seasonality in the air pollution time series. Since case-crossover comparisons are made within subjects at different points in time, the case-crossover analysis implicitly depends on the assumption that the exposure distribution is stationary. The long-term time trends and seasonal variation inherent in air pollution time series violate this assumption (Navidi, 1998; Bateson and Schwartz, 1999; Lumley and Levy, 2000; Bateson and Schwartz, 2001; Levy et al., 2001a).

The second source of bias is called *overlap bias*. If the referent windows W_i, $i = 1, \ldots, n$, are exactly determined by the case period and are not disjoint, then the independent sampling inherent in the conditional likelihood approach is invalidated (Austin et al., 1989; Lumley and Levy, 2000). Lumley and Levy (2000) quantified the overlap bias analytically, and through simulations studies, showed that the direction is unpredictable and the magnitude is a function of the size of the coefficient β. However, for the small effects seen in air pollution epidemiology, current experience suggests that overlap bias is similar to the small-sample bias, e.g. the bias obtained by estimating β in (2) with a small number of referents.

Referent sampling approaches that have been proposed to control for bias in cases-crossover studies include: *Ambi-directional*; *Symmetric bidirectional*; *Semi-symmetric bidirectional*; and *Time-stratified* sampling. The *Ambi-directional* sampling, first proposed by Navidi (1998), selects referents from both before and after the event. While the approach reduces bias due to to lack of stationarity[2], a limitation of Navidi's implementation is that the entire exposure time period is used as a

[2]Note that sampling exposures after the event is only valid when the subjects are still at risk and the event has

referent set for each case time. When only a single ambient exposure series is available, Navidi's approach reduces to a Poisson regression model without adjustment for time-varying confounders (Lumley and Levy, 2000), although inclusion of time varying confounders is possible.

Symmetric bidirectional sampling selects referent sets based upon a pre-specified structure exactly determined by the index time. For examples, Schwartz and Lee (1999) and Neas et al. (1999) used referents at 7, 14, 21 days before and 7, 14, 21 days after the index time. Bateson and Schwartz (1999) showed that using this approach to referent sampling can adjust for temporal confounding by design. However, since the symmetric bidirectional referent sampling scheme produces a referent set that is completely defined by the index time, it is subject to overlap bias.

Semi-symmetric bidirectional sampling (Navidi and Weinhandl, 2002) almost completely eliminates overlap bias by randomly selecting a single referent time among a prespecified set of times before and after the case time. This scheme can be viewed as randomly selecting referents from one of two time stratified samples.

Time-stratified sampling (Lumley and Levy, 2000) divides *a priori* the time period into fixed strata and uses the remaining days in a stratum as referents for a case that falls in that stratum. For example days of the week within calendar months could define strata. Thus, a case on the second Sunday in December would be compared with all the other Sundays in December that year. Because the pre-specified strata are fixed and disjoint, the time-stratified referent sampling scheme is not subject to overlap bias, and it therefore preserves the validity of the conditional logistic regression. In addition, by using strata from short time periods that are not restricted to be prior to the event, the time-stratified case-crossover design also avoids bias due to lack of stationarity in the exposure series (specifically due to season and day of week). Note that *a priori* time stratification also avoids the selection bias discussed by Bateson and Schwartz (2001) that arises from sampling from larger set of referent exposures than those giving rise to the cases.

Recently the case-crossover design has been applied to many studies of air pollution and health (Schwartz and Lee, 1999; Neas et al., 1999; Peters et al., 2001; Levy et al., 2001b). For example, Levy et al. (2001b) analyzed the effect of short-term changes in PM exposure on the risk of sudden cardiac arrest. The sample consisted of cases of paramedic-attended out-of-hospital primary cardiac arrest who we free of other life-threatening conditions and did not have a history of clinically detected cardiovascular disease. The cases were obtained from a previously conducted population-based case-control study and were combined with ambient air monitoring data. The results did not show any evidence of a short-term effect of PM air pollution on the risk of sudden cardiac arrest in people without previously recognized heart disease.

2.3 Panel Studies

Panel studies enroll a cohort or panel of individuals and follow them over time to investigate changes in repeated outcome measures. They are most effective for studying short-term health

not affected subsequent exposure. However, under a rare disease assumption the bias due to sampling after a fatal event is very small (Lumley and Levy, 2000).

effects of air pollutants, particularly in a susceptible subpopulation. Traditionally, a panel study design involves collecting repeat health outcome data for all N subjects over the entire time period of length T although this can be relaxed with proper accomodation in the analyses. The most suitable health outcomes vary within a person over the time period of observation. The pollution exposure measurement could be from a fixed-site ambient monitor, as well as personal monitors.

A variety of models have been used to estimate the effects of pollutants on health in a panel study setting. A basic model for the mean $E(Y_{it}|X_{it})$ (ignoring confounders) is the marginal mean model, which takes the form

$$g(E(Y_{it}|X_{it})) = \beta_0 + X_{it}\beta^M \tag{3}$$

where Y_{it} represents the response to individual $i = 1, \ldots, N$ on day $t = 1, \ldots, T$, X_{it} is the exposure, and $g(\cdot)$ is a link function (McCullagh and Nelder, 1989). In practice, some of the predictors are likely to be common to all individuals at time t. Covariates that are time invariant can be incorporated into an extended model for the intercept. It is often reasonable to assume that individuals are independent, and the repeat measures over time within individual are not. To accommodate correlation between repeated measures on the same individual, a related conditional mean model can be written

$$g(E(Y_{it}|X_{it}, b_i)) = \beta_0 + b_i + X_{it}\beta^C \tag{4}$$

where b_i is a random effect with a hypothesized distribution. The random effect allows for individual variation in the intercept, or when integrated out, induces exchangeable correlation between measurements on the same individual.

The transition model incorporates previous outcomes into the mean model:

$$g(E(Y_{it}|X_{it}, Y_{it-1}, X_{it-1})) = \beta_0 + X_{it}\beta^T + (Y_{it-1}\gamma - X_{i(t-1)}\beta^T). \tag{5}$$

The parameters β, β^C and β^T are defined as marginal, conditional, and transitional effects, respectively (Diggle et al., 1994). Linear models can be formulated so β^M, β^C, and β^T have the same interpretation, whereas in generalized linear models discrepancies between these parameters can occur.

Modern approaches to longitudinal data analysis (Diggle et al., 1994), including mixed, marginal, and transition models and Bayesian hierarchical models can accommodate the complex variance structure induced by these data. Analysis can be performed using software for linear or generalized linear mixed models (Breslow and Clayton, 1993) such as SAS GENMOD, MIXED or the GLIMMIX macro (Littell et al., 1996) or Bayesian software such as the BUGS or WinBugs package (Thomas et al., 1992).

Although estimation of the health effect of the air pollution exposure is often the primary goal of a panel study, the interpretation of the estimated health effects is often less clear. This is particularly true when the traditional panel design of observing all N individuals on the same T days is not realized. Whenever panel members do not share the same observation period, parameterization and estimation of exposure effects should be considered carefully. In many applications it is safest

to separate the components of exposure into average personal exposure ($\overline{X}_i$) and the day-to-day within-person deviations from average personal exposure ($X_{it} - \overline{X}_i$). A linear regression model for a continuous outcome would then become

$$E(Y_{it} \mid X_{it}, \overline{X}_i) = \beta_0 + (X_{it} - \overline{X}_i)\beta_W + \overline{X}_i\beta_B$$

where β_W, β_B represent the within and between-person exposure effect parameters, respectively. Failing to separate these effects results in an overall exposure effect that combines effects due to temporal variation in air pollution within subjects and variation of the average exposure between subjects. An exception to this recommendation is the traditional panel study design which observes all the subjects at every time point and focuses on daily ambient air pollution measurements. This is because in traditional panel studies there is no variation in $\overline{X}_i$, and therefore the parameter β_B is not identifiable. However, studies that observe individuals over different times and only use ambient measurements have variation in $\overline{X}_i$ that is due to seasonal differences in pollution levels. The separation of the within and between effects can at least partially take into account uncontrolled seasonal confounding.

When interpretation of panel study effects includes the within-person effect (β_W), then it is important to adjust for time-varying confounders such as indicators for day of week, functions of season and weather, and time-varying personal behaviors. Additional variables that do not vary within individual (e.g. gender, baseline lung function), can be incorporated into the model to improve efficiency of estimation. They also are useful variables for investigating effect modification, e.g. for testing for differential effects of air pollutants across different subpopulations.

Beyond the structure induced by covariates, variation in panel study data can be due to clustering of outcomes within individuals, similarities of observations close in time, and/or measurement error. Conditional and transition models induce some dependence structure through specification of the mean model; although additional dependence may be added through specification of the covariance structure. In contrast, all dependence in marginal models is specified separately through the covariance model. In generalized estimating equations (GEE) this is called the working covariance matrix (Zeger et al., 1988). One choice is to use an independence working model, where the covariance matrix is proportional to an identity matrix and ignores any serial correlation that may arise because of the longitudinal nature of panel data. Although working independence may result in loss of efficiency, it protects against bias that may result when the full covariate conditional mean does not equal the partly conditional mean specified in the analysis (Pepe and Anderson, 1994).

Although not recommended, a historically popular approach to analyzing panel data is to aggregate over all measurements on day t to estimate the panel average, also called the panel attack rate when the outcome is binary (Korn and Whittemore, 1979). Most typically, air pollution studies regress $Y_t = \sum_i Y_{it}$ on a vector of time-varying exposure and confounding variables, such as ambient pollutant and meteorological variables. This approach is problematic for several reasons, many of them related to problems with ecological studies (Sheppard et al., 1996; Sheppard, 2002a,b). The linear model makes a constant variance assumption over time, and assumes the outcomes on

successive days are statistically independent. Estimates could be biased by the pattern of drop-outs because subject-specific missing data patterns are not taken into account. For binary outcomes there are additional difficulties due to: 1) model specification, 2) the lack of range restriction for the outcome, and 3) the dependence of the variance on the mean. Biases due to using ecologic analysis may be even more problematic when the outcome is not rare (Wakefield and Salway, 2001). Adjustment for some of these problems is possible but rarely undertaken. Although this approach has been discouraged since the late seventies (Korn and Whittemore, 1979), until recently it was the most common approach to analysis in asthma panel studies. Even by the late nineties, with alternative software widely available, the practice had not disappeared completely.

The panel study design is commonly used to study chronic disease exacerbations such as daily asthma symptoms or lung function. For example, Yu et al. (2000) studied asthma exacerbation by obtaining daily self-reports of asthma symptoms. These were regressed on air pollution concentrations from ambient monitors to assess the effects of air pollutant excursions from typical levels for each child's observation period. This study found the odds of asthma symptoms to be increased by 11% (95% CI 3-20%) for a 10 μg/m^3 increase in PM_{10}. Further analysis of this dataset (Slaughter et al., 2002) using ordinal linear regression (McCullagh, 1980; Heagerty and Zeger, 1996) indicated there were effects on the severity of attacks as well as on presence of any asthma symptoms.

Studies that follow cohorts of individuals over longer time periods, say multiple years, are typically referred to as cohort or longitudinal studies rather than panel studies. While the exposure and outcome characterization will be different, in general the recommended methods of analysis for longitudinal studies are similar to those outlined either for panel studies or cohort studies, depending upon the goals of the analyses.

The Southern California Children's Health Study (Peters et al., 1999b,a; Gauderman et al., 2000) is one example of a longitudinal study of air pollution health effects. Children from grades 4, 7, and 10 residing in twelve communities near Los Angeles were followed annually. Communities were selected based on diversity in their historical air pollution levels. Longitudinal analyses of lung function growth using linear mixed models indicated associations of exposure to ambient particles, NO_2, and inorganic acid vapor with reduced lung function growth (Gauderman et al., 2000).

2.4 Cohort Studies

Air pollution cohort studies associate long-term exposure with health outcomes. Either a prospective or retrospective design is possible. In a prospective design, participants complete a questionnaire at entry into the study to elicit information about age, sex, weight, education, smoking history, and other subject-specific characteristics. They are followed over time for mortality or other health events. A measure of cumulative air pollution is often used as the exposure variable. A key design consideration for air pollution cohort studies is identifying a cohort with sufficient exposure variation. Individuals from multiple geographic locations must be studied in order to assure sufficient variation in cumulative exposure, particularly when ambient air pollution measurements are used. However, by maximizing the geographical variability of exposure, the relative risk estimates from

cohort studies are likely to be confounded by area-specific characteristics.

Survival analysis tools can evaluate the association between air pollution and mortality. Typically the Cox proportional-hazards model (Cox and Oakes, 1984; Clayton and Hills, 1993) is used to estimate mortality rate ratios for airborne pollutants while adjusting for potential confounding variables. Relative risk is estimated as the ratio of hazards for an exposed relative to an unexposed or reference group. The hazard for individual i, $\lambda_i(t)$, is modeled as

$$\lambda_i(t) = \lambda_0(t) \exp(X_i\beta + \text{confounders}),$$

where $\lambda_0(t)$ represents the baseline hazard at time t, X_i is an average long-term exposure, and β is the relative risk parameter. The time axis may be defined as calendar time or age. Stratification of the baseline hazard function into disjoint groups can relax the proportional hazards assumption by allowing for separate and not necessarily proportional hazards across strata.

The Harvard Six Cities study and the American Cancer Society (ACS) study (Dockery et al., 1993; Pope et al., 1995b) are among the largest air pollution prospective cohort studies. In the Harvard Six Cities study (Dockery et al., 1993) a random sample of 8,111 adults who resided in one of the six U.S. communities at the time of the enrollment were followed for 14 to 16 years. An analysis of all-cause mortality revealed an increased risk of death associated with increases in particulate matter and sulfate air pollution after adjusting for individual-level confounders. Because of the small number of locations, findings of this study cannot generalized easily.

The ACS study (Pope et al., 1995b) evaluated effects of pollution on mortality using data from a large cohort drawn from 151 metropolitan areas. Ambient air pollution from these areas was linked with individual risk factors for 552,138 adult residents. While the ACS study covered a larger number of areas, the subjects were not randomly sampled as in the Six Cities study. However, both studies reported similar results – the relative risk of all-cause mortality was 1.26 (95% CI 1.08-1.47) for an 18.6 μg/m^3 change in fine particulate matter in the Six Cities study and 1.17 (95% CI 1.09-1.26) for a 24.5 μg/m^3 change in fine particulate matter in the ACS study. A detailed reanalysis of these two studies (Krewski et al., 2000) and a new ACS study including data for a longer period of time (Pope et al., 2002) replicated and extended these results by incorporating a number of new ecologic covariates and applying several models for spatial autocorrelation.

2.5 Design Comparisons

Ultimately, the choice of an optimal design depends upon the research question and the availability of data. No single design is best for all applications. Each design targets specific types of effects, outcomes, and exposure sources. An optimal design should have sufficient power to detect the effect of exposure; this depends on the variability of exposure and the size of the study.

Table 1 compares and contrasts the four main designs discussed in this section. Time series, case-crossover, and panel studies are best suited to the study of short-term exposures. Cohort studies are best suited to the study of medium to long-term exposures. An important distinction

between the cohort (and longitudinal) studies and the other three designs (ecologic time series, case-crossover, and panel studies) is their inclusion of long-term (cumulative) exposure variables.

The panel and cohort studies can study events or continuous outcomes. Time series and case-crossover studies focus on events, and these events should be rare. One key difference between the time series and the case-crossover designs is the approach to control for seasonality and long-term time trends. The case-crossover study controls seasonality and trends by design through restriction of eligible referent samples. In contrast, time series studies use statistical adjustment in the regression model by including smooth functions of calendar time.

Acute effects can be estimated from panel, time series, and case-crossover studies. These studies rely exclusively on estimating associations between variation over time in exposure and variation over time in the outcome. Time scale analyses of time series studies (Zeger et al., 1999; Schwartz, 2000b, 2001; Dominici et al., 2002c) allow estimation of such associations at different time scales: 1 month to 2 months, 1 to 2 weeks, 1 week to 3 days, and less than 3 days.

Cohort studies estimate a combination of acute and chronic effects because the outcomes accumulate over long time periods and could be triggered by either cumulative or short-term peak exposures. Thus, although estimation of chronic effects is one goal of cohort studies, these may not be separable from the acute effects of exposure (Vedal, 1996; Dockery et al., 1993; Pope et al., 1995b; Krewski et al., 2000). Longitudinal studies use repeat measures over time, so they are better able to model medium- to long-term effects. Future cohort studies may be able to separate chronic and acute effects by including both spatially varying chronic exposure metrics and time-dependent short-term excursions in the regression models.

The effect of exposure may vary across susceptible subpopulations. The case-crossover, panel, and cohort study designs are better suited to directly assessing effect modification across population groups than the time series design. The time series design aggregates events over a large population. Typically individual risk factors or other information about the underlying population at risk are not available. In contrast, because each case is included individually in the analysis, the remaining three designs have the advantage of being able to target well-defined subgroups and to more directly evaluate personal characteristics as exposure effect modifiers.

Restrictions in the type of exposure variable used in a design limit the study interpretation and the exposure effects that can be estimated. For examples, most air pollution epidemiology studies rely on ambient monitor measurements of exposure, thus the interpretation of health effects from these studies is restricted to ambient-source exposure.

3 Methodological Advances in Time Series Studies of Air Pollution and Health

The estimation and interpretation of an association between air pollution and health presents several methodological challenges in addition to the selection of the sampling design and the choice of the statistical model. In this section we review statistical methods for: 1) combining information

across several data sources to estimate overall air pollution effects; 2) estimating the health effects of air pollution taking into account model uncertainties; 3) investigating the consequences of exposure measurement error in the estimation of the health effects of air pollution; and 4) estimating air pollution-health dose-response curves. These methods are presented by first introducing the substantive questions, then summarizing the recent statistical developments to address these questions; and finally discussing the extent that these substantive issues have been addressed. Although we discuss methods for time series studies, most of these methodological solutions apply to the other sampling designs as well.

3.1 Combining Information in Multi-Site Time Series Studies

In the past, single-site time series have been criticized because of the un-representativeness of the study locations and the heterogeneity of the statistical approaches used to estimate the associations between air pollution and health (Lipfert and Wyzga, 1993; Li and Roth, 1995). These criticisms have been addressed by using multi-site studies (Katsouyanni et al., 1997; Samet et al., 2000c) where site-specific data on air pollution and health are assembled under a common framework and analyzed with an uniform analytic approach. Hierarchical models provide an appropriate approach for summarizing and integrating the findings of research studies in a particular area (Lindley and Smith, 1972; Morris and Normand, 1992; Gelman et al., 1995; Carlin and Louis, 1996). Hierarchical models have been familiar to statisticians for the last four decades, and recently, because of the development of computational tools that facilitate their implementation (Thomas et al., 1992; Gilks et al., 1996), they have been applied widely in many disciplines. Recently they have been applied to analysis of multi-site time series data (Burnett and Krewski, 1994; Katsouyanni et al., 1997; Roemer et al., 1998; Dominici et al., 2000a; Zanobetti et al., 2000a; Schwartz, 2000a).

Bayesian hierarchical modeling is one appropriate and unified approach for combining evidence across studies, quantifying the sources of variability, and identifying effect modification (see Dominici (2002) for a more detailed discussion about the use of hierarchical models in multi-site time series studies). For example, we can assume a two-stage hierarchical model with the following structure: I) Given multiple sites with time series of air pollution and daily mortality/morbidity counts, the association between air pollution and health is described using a site-specific regression model (1) which takes into account potential confounding factors such as trend, season, and climate; II) The information from multiple sites is combined in a linear regression model where the outcome variable (β^s) is the true relative mortality rate associated with site-specific air pollution indexes within each site, and the explanatory variables (X_j^s) are site-specific characteristics (population density, yearly averages of the pollutants and temperature). Formally:

$$\beta^s = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_j^s + \text{error}. \tag{6}$$

If the predictors X_j^s are centered about their means, the intercept (α_0) can be interpreted as the pooled effect for a site with mean predictors. The regression parameters (α_j) measure the change

in true relative rate of mortality associated with a unit change in the corresponding site-specific variable.

The sources of variation in the estimation of health effects of air pollution are specified by the levels of the hierarchical model. The variation of $\hat{\beta}^s$ about β^s is described by the within-site variance (v^s), which depends on the number of days with available air pollution data, and on the predictive power of the site-specific regression model. The variation of β^s about α_0 is described by the between-site variance (τ^2) which measures the heterogeneity of the true air pollution effects across cities. The specification of a Bayesian hierarchical model is completed with the selection of the prior distributions for the parameters at the top level of the hierarchy. If there is no desire to incorporate prior information into the analysis, then conjugate priors with large variances are a default choice. However, it is important to complete the Bayesian analysis by investigating the sensitivity of the substantive findings to the prior distributions.

Posterior distributions of the pooled estimate (α_0), of the between-site variance (τ^2), and of the second-stage regression parameters (α_j) provide an overall summary of the site-specific relative rates of mortality, a characterization of the heterogeneity of the air pollution effects across the several locations, and the identification of site-specific characteristics that modify the association between air pollution and health. The two-stage hierarchical approach described above can be extended to include additional levels of the hierarchical models (for example, sites within geographical regions, geographical regions within nations, etc) which lead to the estimation of additional sources of variability (within-site, between-site within region, and between regions), and potential effect modifiers at the site or regional level (see, for example, Dominici et al. (2002a)).

Complex hierarchical models can be fitted using simulation-based methods (Tierney, 1994; Gilks et al., 1996) which provide samples from the posterior distributions of all parameters. Alternatively, a point estimate of the pooled effect can be obtained by assuming a random effects model and by taking a weighted average of the site-specific estimates as, for example, suggested by DerSimonian and Laird (1986). Under the weighted average approach for a random effects model, the weights of the site-specific estimates are modified to take into account the variability between locations, e.g. by including a point estimate of τ^2. The heterogeneity of the effects across locations is more completely assessed using a Bayesian approach because inspection of the posterior distribution of τ^2 provides a better characterization of the degree of heterogeneity of the effects across site than a point estimate of τ^2 and/or the classical χ^2 test of $\tau^2 = 0$.

Dominici et al. (2002a) estimated city-specific, regional and national air pollution effects by applying a three-stage hierarchical model to the NMMAPS data base. Results are reported for 88 of the largest metropolitan areas in the United States from 1987 to 1994. Figure 1 shows maximum likelihood estimates and 95% confidence intervals of the log-relative rates of mortality per 10 $\mu g/m^3$ increase in PM_{10} for each location. These estimates were obtained by fitting the log-linear model (1) with natural cubic splines to the data for each location independently. The solid squares with the bold segments denote the posterior means and 95% posterior intervals of the pooled regional effects. The pooled regional estimates of the PM_{10} effects varied somewhat across the regions,

and were estimated to be greatest in the Northeast, with a relative rate of 0.41 % per $10\mu g/m^3$ (95% posterior interval $0.04, 0.78$). At the extreme right, marked with a triangle, is displayed the national averaged relative rate which is equal to 0.21% increase in mortality per $10\mu g/m^3$ increase in PM_{10} (95% posterior interval $0.09, 0.33$) (Dominici et al., 2002d).

3.2 Model Uncertainties

There are many uncertainties in the choice of the statistical model used to assess associations between air pollution and mortality. For example, in reanalyses of data from the same city, the choice of: 1) the meteorological variables to include into the model; 2) the lag structure for the air pollution variables; and 3) the degree of adjustment for trend and seasonality (e.g. choice of degrees of freedom and knot locations) has been shown to impact estimates of relative risks (Smith et al., 2000a; Clyde, 2000; Schwartz, 1994a; Sullivan, 2000; Clyde et al., 2000). As a result, many authors use data to identify the best single lag or combination of lags for the meteorological and pollution variables, and to identify the best degree of smoothing to adjust for trend and seasonality. Recent work, based on a meta-analysis of peer-reviewed literature and simulation studies, shows that to select the most significant lag among a set of possible lags overstates the chance of finding a significant association (Smith et al., 2001).

While it is standard practice to report confidence intervals for summarizing uncertainty in the relative risk estimates once a model has been selected, it is far from routine to acknowledge uncertainty of model choice. Althought, it is generally accepted that different statistical models might lead to different conclusions, typically only a "best" final model is presented, giving a false impression that this is the "only" model suitable for estimating parameters of interest. Ignoring uncertainty due to model choice can lead to over-confident inferences and predictions (Draper, 1995; Hodges, 1987).

Bayesian model averaging (BMA) (Draper, 1995; Raftery et al., 1995; Hoeting et al., 1999; Volinsky et al., 1997; Clyde, 1999, 2000; Chipman et al., 2001) has emerged as an elegant approach for making inferences that account for uncertainties in model choice and for providing a coherent method for combining inferences under different models given the same data. BMA can be implemented by using Bayesian hierarchical models. At the first stage, we specify the distribution of the data given a model and model specific parameters. At the second stage, we elicit the prior distributions for the model specific parameters. At the third stage, we elicit prior probabilities on each model to take into account of model uncertainity. Posterior distributions for quantities of interest such as relative risks (Δ), can be obtained as a weighted average of posterior distributions of the model specific relative risks $p(\Delta|M,\ \text{data})$ weighted by the support that each model M received from the data $p(M|\text{data})$. More specifically:

$$P(\Delta|\ \text{data}) = \sum_{M\ \text{in}\ \mathcal{M}} p(\Delta|M,\ \text{data})p(M|\text{data}) \tag{7}$$

where the sum is over all models under consideration in the space of models $\mathcal{M}$. Because closed-form expressions for BMA do not exist for generalized linear or additive models (except for normal

models), methods for implementing BMA for health effect studies have ranged from reversible jump Markov chain Monte Carlo sampling over parameters and models (Clyde, 1999) to asymptotic approximations (Clyde, 2000; Clyde et al., 2000; Sullivan, 2000). The latter approach can be implemented using standard output from routines commonly used to fit generalized linear or additive models.

The Bayes Information Criterion (BIC) (Schwarz, 1978) can be used to calculate posterior model probabilities (Raftery, 1999; Clyde, 2000). BIC is defined as

$$BIC(M) = \text{deviance}(M) + \dim(M)\log(n), \tag{8}$$

where the deviance(M) is -2 times the maximized log likelihood, and $\dim(M)$ are the number of parameters under model M. The second term acts as a penalty for adding variables and can be viewed as a penalized likelihood procedure. Assuming that all models are equally likely *a priori*, the posterior probability of model M can be approximated using BIC as

$$p(M|\text{data}) = \frac{\exp\left(-\frac{1}{2}BIC(M)\right)}{\sum_{m \text{ in } \mathcal{M}} \exp\left(-\frac{1}{2}BIC(m)\right)} \tag{9}$$

where the sum in the denominator is over all models.

To illustrate model uncertainty and BMA, we did an analysis of daily non-accidental mortality for individuals over 65 from Phoenix, AZ. Variables under consideration included orthogonal polynomials in temperature and specific humidity, coarse PM (PM_{10} - $PM_{2.5}$), fine PM ($PM_{2.5}$) and 0-3 day lags of each. The temporal trend was modeled using B-splines with unknown degrees of freedom and knot locations. The model search started by estimating a temporal trend with approximately 7 knots per year. The maximum likelihood estimate of the linear predictor for the temporal trend was then treated as an offset in the GLM, and leaps and bounds (Volinsky et al., 1997) was used to identify the best subsets of the remaining confounding and exposure variables with different sizes. Each of these models was then re-fit varying the number of degrees of freedom in the B-spline from 1 to 15 per year to simultaneously estimate all model parameters (additive and linear terms). The posterior distribution for the degrees of freedom was highly concentrated on 4 per year, indicating little uncertainty about the level of smoothness for the temporal trend.

Figure 2 shows the top 25 models ranked by posterior model probabilities and associated posterior intervals for relative risks, with the highest posterior probability model at the top. Uncertainty regarding the variables other than the time trend is illustrated with an image map of the model space. Each column corresponds to a variable (labeled at the bottom) and each row corresponds to a model. The uncertainty is represented by differential shading of variables that are included. This provides a visual representation of model uncertainty regarding the choice of pollution and meteorological variables, transformations and lags. These 25 models represent 87 % of the total posterior probabilities across models. Less than half of this is attributed to the highest probability model, which has a posterior probability of 0.46, indicating substantial model uncertainty.

Model M_1 can be easily compared to model M_2 by using the Bayes Factor (BF) (e.g. the ratio between the posterior and the prior odds for a specific hypothesis). The log BF is simply the

difference in BIC values, -.5 (BIC(M_1) - BIC(M_2)), and if we assume that all models are equally likely a priori, then the posterior odds for comparing one model to a second depend only on the BF and not the prior odds.

Log BF for comparing each model to the worst model identified in the search are reported in the Y-axis for the left hand plot. Differences in the log BF between these models vary from 1 to 4, and can be used to compare models within the top 25. Based on Jeffreys' scale of interpretation of BF (Kass and Raftery, 1995), a log BF of 4 is relatively small and indicates that even the largest difference between these models is not important. However, the scale also indicates that these models are much better than the model receiving the lowest posterior probability. While there is some uncertainty about which lag of PM coarse to include (lag 1 or lag 2), none of the top twenty-five models include a fine PM effect.

The plot on the right shows the corresponding model-specific 95% posterior intervals for relative risks associated with a one interquartile range increase for any of the particulate matter variables included in the model (a 17.9 $\mu g/m^3$ and 9.1 $\mu g/m^3$ increase for coarse and fine PM, respectively) obtained using a normal approximation for the log relative risk centered at the maximum likelihood estimate and with variance based on the inverse Fisher information. For models in the top 25 that did not include any particulate matter variables, the posterior distribution of the relative risk is degenerate at one. Using BMA, the posterior mean of the relative risk associated with a one interquartile change in fine and coarse particles is 1.028 (95% posterior interval = 1,1.045). Note that BMA reports higher uncertainty in the estimated air pollution effect than the one reported under the highest probability model (95% posterior interval 1.014 to 1.047).

BMA is a suitable approach for taking into account model uncertainity and has great potential in air pollution studies where model choice is critical. Unlike p-values, posterior model probabilities can be used to quantify the evidence in favor of an air pollution effect, taking into account model selection. Using all models from the search, the probability that there is no PM effect (the relative risk equals one) can be obtained by summing the posterior model probabilities over all models that do not include any PM variables. For this example, the posterior probability of no effect is 0.0522, which corresponds to posterior odds that the relative risk is not 1 of roughly 18 to 1.

However, standard epidemiological modelling often involves first selecting confounding variables, and then including exposure variables after adjustment for confounders. This is a critical step in air pollution studies where the air pollution effect on health is much smaller than the effects of the other time varying covariates (such as weather, and trend). In this situation, because BMA treats exposure and confounders symmetrically, it is likely to ascribe the potential air pollution effects to other correlated confounders.

To date model averaging has only been explored in single city analyses. In multi-site time series studies, BMA provides more accurate estimates of within city-pollutant effects. The distributions of quantities of interest can then be averaged over models within cities, and the pooled across cities within region, and over regions.

3.3 Effects of Misclassification of Exposure

One barrier in interpreting the observational evidence concerning the adverse health effects of air pollution is the measurement error inherent in estimates of exposure based on ambient pollutant monitors. Exposure assessment studies have shown that data from monitors at central sites do not adequately represent personal exposure (Lioy et al., 1990; Mage and Buckley, 1995; Janssen et al., 1997, 1998; Ozkaynak et al., 1996; Haran et al., 2002). Thus, the exposure error resulting from using centrally measured data as a surrogate for personal exposure can potentially lead to bias in estimates of the health effects of air pollution (Thomas et al., 1993). However, because regulations are based on ambient air and most epidemiologic studies rely on exposure measurements from central site monitors, health effects from ambient exposures are the target of many epidemiologic studies, regardless of their intent.

Both ambient and personal air pollutant exposures vary over time. In addition personal exposures vary significantly over individuals. The micro-environments where individuals spend their time have different dominant sources and all contribute additively to total personal exposure. Micro-environmental modeling is a popular approach for exposure modeling, particularly when direct measurement of total personal exposure is not possible (Duan, 1991; Rodes et al., 2001).

Individual personal exposure can be partitioned into to ambient versus non-ambient sources, of which personal exposure to ambient sources is of interest from a regulatory perspective. A simple model for total personal exposure for individual i at time t, X_{it}^P, has the form

$$X_{it}^P = X_{it}^N + \alpha_{it} X_{it}^A$$

where X_{it}^N is personal exposure from non-ambient sources, X_{it}^A is the ambient concentration at individual i's spatial location, and α_{it} is an attenuation parameter defined as the fraction of ambient concentration experienced as exposure (Ott et al., 2000; Wilson et al., 2000; Sheppard and Damian, 2000). The attenuation parameter depends upon the penetration, deposition and decay rates for a specific pollutant and micro-environment, as well as on individual behavior (particularly how much time is spent outdoors).

Current research suggests that for large populations it may be reasonable to assume that $\alpha_{it} = \alpha$, $X_{it}^A \propto X_t^A$, and that the ambient and non-ambient sources are independent. This has important consequences for study design, particularly since most studies rely on a measurement $\widehat{X}_t^A$ of X_t^A from a fixed-site ambient monitor.

In considering the consequences for estimating health effects of air pollution by using surrogate measures of exposures, one approach (Zeger et al., 2000) begins by decomposing the pollution measurement difference between X_{it}^P and $\hat{X}_t^A$ into three components:

$$X_{it}^P = \hat{X}_t^A + (X_{it}^P - \bar{X}_t^P) + (\bar{X}_t^P - X_t^A) + (X_t^A - \widehat{X}_t^A) \tag{10}$$

where, $(X_{it}^P - \bar{X}_t^P)$ is the error due to having aggregated rather than individual exposure data; $(\bar{X}_t^P - X_t^A)$ is the difference between the average personal exposure and the true ambient pollutant level; and $(X_t^A - \widehat{X}_t^A)$ represents the difference between the true and the measured ambient concentration.

Regardless of the epidemiological design, each of these differences may introduce some degree of exposure measurement error.

For example, in an ecological time series study of air pollution and health, suppose that we are interested in drawing inferences about the relationship between Y_t and $\bar{X}_t^P$ given substantial information about the relationship between Y_t and $\widehat{X}_t^A$ at a particular site, and separate studies of the association between $\widehat{X}_t^A$ and $\bar{X}_t^P$. Here we assume $\widehat{X}_t^A = X_t^A$. In this scenario, a reasonable approach is to build a two-stage model where at the first stage we assume:

$$E[Y_t] = \exp(\bar{X}_t^P \beta_P + \text{confounders}) \tag{11}$$

where β_P is the log relative rate of death associated with a unit change in average personal exposure, $\bar{X}_t^P$ is a missing covariate, and the confounders are the same as in model (1). At the second stage we assume a measurement error model for the relationship of average personal exposures and ambient concentrations, taking account of variations within and across locations. This modelling approach combines a log-linear model for Y_t given $\bar{X}_t^P$ with measurement error model for $\bar{X}_t^P$ given X_t^A to make inference about β_P. As discussed below, the direct regression of Y_t on X_t^A giving β is also of interest from a regulatory perspective as only ambient concentrations are currently regulated. This modelling approach (Dominici et al., 2000b) can be described as a combination of Bayesian hierarchical modeling (Lindley and Smith, 1972; Morris and Normand, 1992) and data augmentation (Tanner, 1991), and is an example of regression calibration which is widely used for handling measurement error in non-linear models (Carroll et al., 1995). Recently, alternative statistical models have been developed (Clayton, 1991; Thomas et al., 1993; Richardson and Gilks, 1993; Zidek et al., 1996; Cakmak et al., 1999; Fung and Krewski, 1999; Dominici et al., 2000b; Zeger et al., 2000; Sheppard and Damian, 2000) that can be used to describe the relationship of a health outcome with air pollution exposure using surrogate measurements of exposure.

Figure 3 shows the posterior distributions of log relative rates (% increase in mortality for 10 $\mu g/m^3$ increase of PM_{10} exposure) of mortality from ambient exposures β, and from total personal exposure β_P. Note that measurement error tends to bias the results towards zero and that the IQR of β_P is larger than the IQR of β. This is because we have taken into account the variability due to having $\hat{X}_t^A$, not $\bar{X}_t^P$.

Relative rate of mortality corresponding to short-time scales of PM_{10} exposure might be attenuated more by misclassification of exposure than the relative rate of mortality corresponding to longer-time scales. To further investigate this issue, we reanalyzed data from Particle Total Exposure Assessment Methodology Study (PTEAM) Ozkaynak et al. (1996) where personal and ambient PM_{10} exposure measurements were recorded for 48 consecutive days (September 22-November 9, 1990) in Riverside, California.

To account for the scale-varying association between ambient and personal concentrations, we can extend the measurement error modeling approach to incorporate a time-scale dependent

measurement error model. More specifically, we consider the following model:

$$\log \mu_t = \sum_k X^P_{kt}\beta^P_k + \text{confounders} \quad (12)$$

$$X^P_{kt} = \alpha_{0k} + \alpha_{1k}X^A_{kt} + N(0,\sigma^2_k),\ k = 1,\ldots,K \quad (13)$$

where: X^P_{kt} is the average personal exposure at day t at time scale k, and β^P_k measures the percentage increase in mortality per 10 $\mu g/m^3$ increase in the average personal pollution level at time scale k. The slopes α_ks measure the changes in the personal exposure for unit change in the shorter, medium, and longer time scale variations of the ambient concentrations. The intercept α_{0k} represents the remaining portion of personal exposure to particles unexplained by the ambient concentrations, that arise from personal activities or unmeasured micro-environments not well-represented by fixed indoor or outdoor monitors, at time scale k. Under this modeling approach, the estimate of β^P_k has the simple form $\hat{\beta}^P_k = \hat{\beta}_k/\alpha_{1k}$, and its standard deviation approximately equals the standard deviation of $\hat{\beta}_k$ times $1/\alpha_{1k}$. This is the well-known regression calibration Carroll et al. (1995) approach to adjust for exposure measurement error.

We applied this strategy to estimate the slopes (α_{1k}s) by using the PTEAM data. Results are showed in the Table 2. As expected, the association between average personal exposures and ambient concentrations is stronger at the longer time scales than at the shorter time scales. Therefore, the time-scale coefficient at the long time scales is affected less by the measurement error.

In summary, regardless of the sampling design, studies of air pollution and health are subject to the limitations posed by the available air pollution measurements. Various aspects of the air pollution exposure measurements, such as the three exposure components described in equation (10), and time scales of variation in ambient and personal exposure described in equation (12), produce different types of bias in the health effect estimates, and such biases are more or less important in the various sampling designs. Furthermore, rexamination of simplified assumptions will be needed as better understanding of exposure measurement distributions develops. A systematic treatment of exposure measurement error and modelling issues for different sampling designs, including recommended data collection and novel statistical methods, is still lacking in the literature.

3.4 Shape of the Exposure Response Curve

One approach for estimation of a dose-response curve in time series studies is to model the logarithm of the expected value of daily mortality as a smooth function of air pollution after adjusting for other confounders:

$$E(Y_t) = \exp\{S(X_t, \text{knots} = \boldsymbol{\nu}) + \text{confounders}\} \quad (14)$$

where $S(\cdot,\cdot)$ is modeled as a natural cubic spline with k knots at locations $\boldsymbol{\nu} = (\nu_1,\ldots,\nu_k)$, and the term confounders includes smooth functions of time, and temperature, as described in model (1). The number and locations of knots are generally fixed in advance as in Daniels et al. (2000) in

the analysis of the 20 largest US cities, or can be estimated from the data using Reversible Jump Monte Carlo Markov Chain (RJMCM) (Green, 1995), as in Dominici et al. (2002a) in an analysis of the 88 largest US cities. Other methods can be used as an alternative to the natural cubic spline, such as smoothing splines, loess or B-splines (Smith et al., 2000b; Schwartz, 2000a; Schwartz and Zanobetti, 2000). Shapes of dose-response curves are generally not very sensitive to the smoothing method (except at extremes), but can be sensitive to the number of knots or smoothness penalty.

For a fixed number of knots, the dose-response model (14) can be expressed as a parametric model

$$E(Y_t) = \exp\{\boldsymbol{\beta} B_t + \text{confounders}\} \tag{15}$$

where B_t is the t-row of the design matrix for the natural cubic spline with knots at locations $\boldsymbol{\nu}$ of X_t and $\boldsymbol{\beta}$ is the corresponding vector of coefficients.

To examine the question of whether the health effects of air pollution are negligible below some level, a linear threshold model can be used:

$$E(Y_t) = \exp\{\theta(X_{t-\ell} - h)^+ + \text{confounders}\} \tag{16}$$

where the $X_{t-\ell}$ term in model (1) is replaced with a term of the form $(X_{t-\ell} - h)^+$, where ($x^+ = x$ if $x \geq 0$ and $x^+ = 0$ if $x < 0$) and h is an unknown change-point that is estimated from the data. Bayesian methods can be used to estimate marginal posterior distribution of the location of the threshold h (Cakmak et al., 1999; Smith et al., 2000b; Daniels et al., 2000; Schwartz, 2000a). The parameters β in the generalized additive model (1) and θ in model (16) have different interpretations: β measures the percentage increase in mortality per unit increase in air pollution at any level of the pollutant; θ measures the percentage increase in mortality per unit increase in air pollution *when the pollutant level is higher than* h. For pollutants value smaller than h, the log relative rate β is set to zero.

In multi-site time series studies, it is of interest to estimate a pooled dose-response relationship by combining dose-response curves across locations. The hierarchical approach described earlier can be used here, where equation (6) is generalized to the multivariate case as

$$\boldsymbol{\beta}^s = \boldsymbol{\alpha}_0 + N(0, \Sigma) \tag{17}$$

where $\boldsymbol{\beta}^s$ and $\boldsymbol{\alpha}_0$ are the city-specific and the overall vectors of the spline coefficients, respectively. Effect modification of the shape of the dose-response curve can be also investigated by including city-specific covariates in equation (17).

Dominici et al. (2002a) recently used the hierarchical spline model defined in equations (15) and (17) with an unknown number and locations of knots to estimate regional and national exposure-response relationships in the NMMAPS data base (Samet et al., 2000c). Figure 4 shows regional PM_{10}-mortality dose-response curves for each region in bold solid lines. The curves with empty dots were estimated using a simplified version of the Reversible Jump Monte Carlo Markov Chain (RJMCMC) algorithm with one fixed knot at $40\mu g/m^3$. At the bottom right are the national dose-response curves. The bold line represents the estimated national dose-response curve, allowing for

unknown numbers and locations of knots, obtained by pooling information across the 88 cities. This national curve, represented by the empty dots, overlaps with the national curve estimated by using RJMCMC. The hierarchical spline and linear models can be compared easily, because the linear model is a special case of the spline model with 0 knots.

The national dose-response curve, obtained by combining information across all the cities, is clearly linear, and the posterior probability of zero knots from the RJMCMC is almost 1. At the regional level, the data from cities in several regions (Northwest, Upper Midwest, and Southeast) indicate some modest departure from a linear model. In particular, the Northwest and Upper Midwest regions show a leveling off (saturation effect) at higher levels of PM. However, the uncertainty boundaries for these regions indicate compatibility of the data with a linear relationship, and we currently have no specific hypotheses as to why these regions might deviate from a linear dose-response curve.

These findings are consistent with previous analyses of time series data in a number of locations including London, Cincinnati, Birmingham, Utah Valley, and Shenyang (Pope, 2000), and with findings of the two major long-term studies, the Harvard Six-Cities and the American Cancer Society cohort studies (Dockery et al., 1993; Pope et al., 1995b). The consistent finding of a linear relationship, now confirmed at the national level, places a difficult burden on policy-makers, charged by the Clean Air Act with setting protective standards for public health that include a margin of safety. However, such linearity might also reflect averaging out nonlinearities at the region, and at the city levels, and at susceptible subgroups of the populations (e.g. elderly or those with pre-existing diseases). At the other end, substantial statistical power would be needed to provide strong evidence of nonlinearity in concentration-response relationship for very specific strata of geographical and populations subgroups.

These methods provide evidence in a form that would be directly applicable to policy-development, as shown by our finding that the concentration-response relationship for air pollution and mortality is linear with a high degree of certainty. With repeated application, our methods would offer an approach for tracking the health effects of air pollution over time as control measures are implemented; they should have application to other environmental health problems as well.

4 Discussion and Research Opportunities

Risk assessment in environmental epidemiology is challenged by the complexity of the human exposure to environmental agents and by the difficulty of accurately measuring exposure. Large populations are exposed at very low air pollution levels, and in order to detect small relative risks, high statistical power is necessary. Environmental agents vary in characteristics and concentrations over time and across space, they tend to be highly autocorrelated, and they are usually measured imprecisely. Both environmental agents and health outcomes are typically measured at different levels of temporal and spatial resolution.

These statistical issues are central to the goal of determining the health effects of chronic and

acute exposures to particulate matter and other air pollutants. Although several state-of-the-art statistical approaches for the analyses of environmental data have been proposed, considerable need remains for further methodological development that would support more effective risk analysis. Research opportunities that relate to development or enhancement of methods are discussed below.

Generalized Additive Models for analyses of time series studies in air pollution and health. In time series studies of air pollution and mortality, GAM has been the most widely applied method because it allows for non-parametric adjustments for non-linear confounding effects of seasonality, trends, and weather variables (Schwartz, 1994b; Kelsall et al., 1997; Samet et al., 2000a; Dominici et al., 2000a; Katsouyanni et al., 1997; Moolgavkar, 2002).

Recent work by Dominici et al. (2002b) and by Ramsay et al. (2002) called for caution in the use of GAM for estimating relative rates of mortality/ morbidity in time series studies of air pollution and health. Dominici et al. (2002b) reported that when the data to which the GAM are applied have the following two characteristics: 1) the estimated regression coefficients are small; and 2) there exist confounding factors that are modeled using at least two non-parametric smooth functions, the defaults in the S-PLUS software (Version 3.4) package `gam` do not assure convergence of its iterative estimation procedure, and can provide biased estimates of the regression coefficients and standard errors.

This phenomenon is likely to occurr in time series analyses of contemporary data on air pollution and mortality. To evaluate the impact of default implementation of the `gam` software and understimation of standard errors on published NMMAPS analyses, the NMMAPS data has been reanalyzed with three different methods: 1) Generalized Linear Models with natural cubic splines; 2) GAM with default convergence parameters; and 3) GAM with more stringent convergence parameters than the defaults. In the re-analyses, the pooled estimate across 90 cities at lag 1 moves from 0.41% (posterior standard error 0.05%) percent increase in total mortality for 10 unit increase in PM_{10} to 0.27% (posterior standard error 0.05) when more stringent parameters in the GAM are used. When GLM with natural cubic splines are used, then the pooled estimate is 0.21% (posterior standard error 0.06). In every analysis, however, there is a strong evidence for a positive association of acute exposure to PM_{10} and death, even with very conservative adjustments for trend, seasonality and weather (Dominici et al., 2002d).

In parallel, Ramsay et al. (2002) reported that inability of the GAM to detect concurvity can lead to under-estimation of the standard error of relative rate estimates (see also Chambers and Hastie (1992) pages 303-304). The NMMAPS re-analyses Dominici et al. (2002d) empirically confirm theoretical results of Ramsay et al. (2002), and show that the degree of bias in the standard errors is proportional to the size of the standard errors, and that this underestimation remains even when more stringent convergence parameters are used.

In summary, although simulation studies for NMMAPS showed that Generalized Linear Models with natural cubic splines provide a less biased estimate of the pooled relative rate, than GAM with smoothing splines (Dominici et al., 2002b) additional statistical investigations are required be-

fore making recomandation on which of the two methods should be used. These recent discoveries have generated substantial costernation among environmental scientist, EPA members and policy makers. Sources of model uncertainty call for a systematic assessment of model choice and for development of new methods. Importantly, the weight given by this scientific evidence in setting policy requires a level of confidence in findings that is difficult to attain in the small effects/many potential confounders context, regardless of the sophistication of the statistical approach.

Misaligned data. Data on environmental pollutants and health outcomes are typically measured at different scales of temporal and spatial resolution. For example in time series studies hourly, daily, 3-day, and 6-day air pollution data are recorded from a limited number of monitoring stations, whereas daily health outcomes are typically aggregated over large spatial areas. To analyze misaligned data, a pragmatic modeling strategy is to convert both health outcomes and environmental data to common spatial and temporal resolutions. For example, the environmental pollutants can be aligned with the outcome data by calculating 24-hour averages and aggregating these measurements across monitoring stations.

Recently novel statistical approaches have been proposed to estimate associations between environmental exposures and disease from epidemiological data collected at different levels of spatial aggregation (Carlin and Mugglin, 1997; Mugglin and Carlin, 1998; Mugglin et al., 1998; Gelfand et al., 2001). For example, Gelfand et al. (2001) proposed a hierarchical modelling approach for estimating associations between point-level ozone measurements and zip-code level pediatric asthma ER visits. Best et al. (2000) developed hierarchical Poisson-Gamma models to handle misaligned data by relating spatial variables to an underlying continuous random field. When such data are available, these models can accommodate individual characteristics, such as age and gender.

Because of the opportunities arising from the increasing availability of large, complex data bases on exposure and disease at different temporal and spatial scales, new statistical methods are needed to estimate health effects of exposure to environmental agents which appropriately take into account: 1) spatially varying individual exposures and risk factors within areas; 2) spatially varying outcomes and population risk factors between geographical areas; and 3) exposure measurement error. Neglecting any of these may induce bias in the health effect estimates.

Opportunities for aligning data can also come from combining data from different sampling designs. For instance, time series studies rely upon day-to-day variations in pollution within a population, and they estimate the health risk associated with shorter-term exposures, on the order of days, weeks or at most months. In contrast, cohort studies rely upon site-to-site variations of cumulative pollution exposures and estimate the health risk associated with long-term exposures.

Time series studies and cohort studies provide different numerical estimates of the relative risks associated with exposure to air pollution. More specifically, relative rate estimates from cohort studies are an order of magnitude larger than relative rates estimates from time series studies, and a direct comparison between these two relative rates has not been possible because time series and cohort estimates have been obtained from different populations, and from different exposures.

Recently, there are been attempts to reconcile differences in time series and cohort studies by estimating "sub-chronic effects" in time series studies from distributed lag (Zanobetti et al., 2000b; Braga et al., 2001) or time-scale models (Schwartz, 2001; Dominici et al., 2002c). For example, Zeger et al. (2002) estimated sub-chronic effects in 4 NMMAPS locations by using averages of past exposure to PM_{10} as predictors. However, when several days in the past are used to calculate the averages (say mor than two weeks), then the health effects associated with such slowly varying air pollution exposures are highly confounded by the other slowly varying covariates.

A promising research opportunity would be to merge cohort and time series national data bases, and to develop statistical methods to jointly estimate acute and chronic effects of air pollution on the same population.

Exposure assessment and study design. Environmental exposures typically affect large populations at very low levels. Such low exposures induce very small health risks which, in turn, require huge sample sizes to be detected. Exposure characteristics, both underlying and due to measurement, can have a large impact on the health effect estimates. Limited or inaccurate exposure data is the norm in environmental epidemiology, and it induces large measurement error making even more difficult to properly estimate such small risks. Integrating exposure assessment with study design can enable scientists to detect small effects that would otherwise be undetectable; strengthen the interpretation of the effects to be estimated, and facilitate recommendations for more targeted exposure assessment studies.

The time series design is an excellent example of integration between exposure assessment and study design. The success of time series studies in detecting small effects may be due to the fortuitous opportunity to combine reasonably representative time series of population exposure with time series of rare events derived from huge populations. In the case of particulate matter, the assumption of reasonably accurate population exposure assessment follows from both the characteristics of population exposure, and the current ambient monitoring program. Ambient particulate matter concentrations are much more variable in time than space (Goswami et al., 2002). Although time-varying ambient concentration of particulate matter is only a small fraction of the variation of total personal exposure, ambient measurements may be a good surrogate for average population exposure. In time series studies, because ambient and non-ambient sources for particulate matter are likely to be independent (Ott et al., 2000), ignoring non-ambient source may not induce bias.

An alternate perspective on time series studies would argue the estimates are inherently flawed. For instance, several of the exposure assumptions could be wrong. Further, failure to recognize the importance of spatial variation of ambient particulate matter and the effects of aggregation in these studies may lead to the *ecological fallacy* where characteristics of individuals are wrongly inferred from characteristics of groups (Greenland and Morgenstern, 1989; Richardson, 2000; Greenland, 1992; Sheppard et al., 1996; Wakefield and Salway, 2001; Morgenstern and Thomas, 1993; Sheppard, 2002a).

A better temporal and spatial resolution can be achieved with alternative designs, such as per-

sonal exposure panel studies on individuals. However, most alternative studies require unrealistic exposure-assessment effort and unaffordable sample sizes. This highlights the inter-relationship between feasibility, study design, and exposure assessment.

Measurement Error in Monitoring Data. In addition to biases from using central site monitors to represent personal exposure, measurement error in central site monitoring data has received little attention. Recent work by McBride et al. (2002) has examined biases in measured PM concentrations based on monitor type: Federal Reference Method (FRM) monitors, FRM-equivalent monitors, and monitor-inlet type for co-located monitors in Phoenix, AZ. Although the FRM-monitor is more accurate and precise, measurements are taken every third day, while the FRM-equivalent monitor provides daily observations. Daily concentrations from the FRM-equivalent-monitor exceeded those of the FRM-monitor by roughly 40% (95% posterior probability interval: 1.20,1.61) for the period prior to the inlet change in the FRM-equivalent monitor. After the inlet change the direction of bias reversed with daily concentrations of the reference monitor exceeding those of the FRM-equivalent monitor by roughly 50% (95% posterior probability interval: 0.56,0.78). Moreover, the probability of 1:1 agreement between the two monitors during either the pre- or post-inlet periods is very close to zero. Even though the FRM-equivalent-monitor has better temporal resolution with fewer missing observations, the use of unadjusted measurements could lead to substantial biases (in either direction) or lead to non-significant effects in health effect studies.

Multiple pollutant models. Because EPA regulates each criteria pollutant independently, most of the current research on air pollution and health outcomes focuses on estimating health effects of a specific pollutant. Identification of the unique effect of a specific pollutant requires careful adjustment for simultaneous exposure to a complex mixture of co-pollutants.

Given the observational nature of studies on health effects of air pollution, disentangling the effects of one pollutant from others is difficult because of correlation among the pollution variables. In addition, interpretation of individual coefficients separately from multi-pollutant models can be problematic, since multi-collinearity can alter parameter estimates, e.g. some coefficients may be negative because of positive correlation between pollutants. In addition very limited data is currently available on particulate matter composition to better characterize the risk.

Multi site analyses provide a more sound approach for exploring confounding and effect modification to other pollutants and weather. Pooling of city-specific relative rates from multiple pollutant models can be performed in an univariate or multivariate fashion. In an univariate fashion, city-specific coefficients for each pollutant are pooled separately using fixed effects, random effects, or Bayesian models. However univariate pooling ignores within-city statistical correlation between relative rates estimates for the pollutants jointly included in the model. Because of this limitation, multivariate pooling – which instead takes into account of such correlation – is more suitable in multipollutant analyses. Multivariate pooling can be easily performed by using Bayesian two-stage multivariate normal models (Lindley and Smith, 1972) and implemented by the software TLNISE

(Everson and Morris, 2000) which allow specification of non-informative priors on the heterogeneity covariance matrix.

Alternatively, it might be more reasonable to assume that there exists a mixture of pollutants that can be considered harmful. While individual effects may not be estimated reliably, combined effects can be. This is one of the priority research topics identified by the National Research Council (1998). However, in calculating the health effect for simultaneous increases in several pollutants, it is not clear how to determine these simultaneous changes. While a one interquartile range change is often used (Sheppard et al., 1999; Yu et al., 2000; Clyde et al., 2000), this is an arbitrary choice. In addition, an interquartile range change it is not necessarily consistent with what would happen to one pollutant if a second pollutant changed or if a latent variable that affects both pollutants were modified. In general one would need a joint probability model for the multiple pollutants to address the issue of how much to change each pollutant when estimating the air pollution effect. This requires a better understanding of sources, the dynamics of secondary aerosol formation, weather, and interventions.

Inferences in multipollutant models is further complicated by the presence of exposure measurement error in the pollutant variables (Zidek et al., 1996; Zeger et al., 2000). One important limitation of most published measurement error formulations is that only a single pollutant, generally PM_{10}, is considered. Work is needed to extend these models to multiple pollutants – for example, PM_{10}, $PM_{2.5}$, NO_2, SO_2, and CO, combustion by products, and PM mass speciated data.

Harmful PM constituents. Although many epidemiological studies have found consistent evidence of effects of particulate matter on health, a research priority is identifying which component(s) of particulate matter are actually responsible for these adverse health effects (National Research Council, 1998).

One approach to this problem is to use multi-site time series studies to investigate whether between-location differences in particle composition explain the heterogeneity of the relative rates of mortality across cities (Samet et al., 2000c,b; Katsouyanni et al., 1997; Dominici et al., 2002a).

An alternate approach is to use speciated PM time series. Speciated data are starting to become available for a few locations, but they are not measured in a systematic fashion yet. Source apportionment models apportion time series of multiple pollutants and/or chemical species into multple sources. Common approaches to source apportionment are principal components anlysis (Henry, 1987), mass balance equations (Miller et al., 1972; Watson et al., 1984), and positive matrix factorization (Paatero, 1997; Ramadan et al., 2000). Regardless of the approach, the source estimates are uncertain and subject to measurement error. Work is needed to directly link the source apportionment models with health effect estimates in the time series regression models.

Mortality Displacement. Even if the associations between particulate air concentrations and mortality are causal, it is still important to know who is affected: if only those very near death are

susceptible then the public health importance of particulate air pollution might still be fairly small. Short-term mortality displacement implies a model where a pool of frail, near-death individuals is created and maintained by the effects of other diseases such as pneumonia or congestive heart failure. These individuals have a very short life expectancy, and the effect of high air pollution is to hasten their deaths by a matter of days. This depletes the pool of frail individuals, so the higher mortality immediately following a day of high air pollution is followed by lower mortality for the next few days. Fundamental to this model is the assumption that particulate air pollution does not affect the rate at which people enter the frail pool.

Several statistical approaches have been developed to investigate short-term mortality displacement. Poisson regression methods in the frequency domain (Kelsall et al., 1999; Zeger et al., 1999; Schwartz, 2001) and in the time domain (Schwartz, 2001; Dominici et al., 2002c) estimate associations between air pollution mortality at different time-scales of variation in the exposure and outcome. These approaches can be used to test the mortality displacement hypothesis: if the association between air pollution and mortality is evident at "only" the shortest time-scales, then pollution-related deaths are advanced by only few days and therefore support mortality displacement.

Findings from recent time-scale analyses are inconsistent with the hypotheses of short-term mortality displacement (Zeger et al., 1999; Schwartz, 2001). In fact, they suggest that smoother variations in the air pollution time series (14 days to 3 days) tend to be more strongly associated with mortality than less smooth variations (less than 2 days).

Although innovative, frequency domain and time scale approaches have limitations and their findings need to be interpreted with caution. Relative rates of mortality associated with long-time scales variations in air pollution (say larger than 2 months) tend to be heavily confounded by seasonality, and interpretation of the time-scale relative rates is not straighforward. Distributed lag models (Zanobetti et al., 2000b) are reasonable alternatives and allow a better interpretation of the coefficients, but are still affected by the long-term confounding.

Alternative state-space models that attempt to capture the rates of transition into near-death and death (Nelson and Murray, 2000) have suggested that most of the mortality does occur in those very near death. Findings from different studies and statistical approaches need to be reconcilied.

Spatial confounding in chronic studies. As detailed in the subsection 2.4, cohort studies (Dockery et al., 1993; Pope et al., 1995b, 2002) compare air pollution-mortality associations across geographical sites rather than day-to-day variations in air pollution and mortality within one site, as done in time series studies. Chronic studies have reported that persons living in more polluted cities have higher mortality risks than "otherwise similar" persons in less polluted cities. Here, "otherwise similar" refers to statistical control for age, gender, socio-economic factors, smoking and additional person-level confounding variables.

When spatial dependence between health outcomes and exposures is due to unmeasured spatially varying confounders, then bias in the health effect estimates can occurr. In addition, when

spatial variation in the health outcomes and exposures are on the same scale and it is not due to confounding, then lack of adjustment for the spatial dependence reduces the efficiency of the health effect estimate (Guthrie et al., 2002). Re-analyses of the ACS data includes modeling of spatial dependence and indicates the possiblity of residual spatial confounding (Krewski et al., 2000). However, later analysis of these data with longer follow-up indicated no residual spatial structure (Pope et al., 2002).

Air pollution cohort studies generally use ambient community-level measure of cumulative air pollution exposure. This approach has merit when most of the exposure variation is between communities but is limited when most of the exposure variation is within communities (Sheppard and Prentice, 1995; Greenland, 1992; Wakefield and Salway, 2001). However, innovative work on modeling within-community variation in long-term ambient exposure as a function of local traffic sources and residence location is increasing the capability of these studies to capture additional exposure information (Briggs et al., 1997; Lebret et al., 2000; Hoek et al., 2001).

	Sampling Designs			
	Time-series	**Case-Crossover**	**Panel and Longitudinal**	**Cohort**
General Data Structure	Daily time series of air pollution, health event counts and weather variables	Case series with exposures at event and referent times	Longitudinal exposure and outcome data on a group of subjects	Survival times for a cohort of people linked with long-term average air pollutants and subject-specific covariates
Outcome	Daily event counts in a defined area	Event times for a rare event	Repeated measures of symptoms, function, or growth	Event times
Typical exposure measurement	Time-varying ambient	Time-varying ambient	Time-varying total personal or ambient	Cumulative city-specific ambient
Exposure duration	short-term (days)	short-term (days or hours)	short-term (days or hours) to longer-term (years)	long-term (years)
Typical confounders	day of week, season and time trends, weather	day of week, season, time (note adjustment is by referent sampling)	day of week, season, weather	personal characteristics (age, race, gender, smoking history, occupational exposures, SES, body mass)
Other possible confounders	epidemics, other pollutants	weather variables, other pollutants, time-varying personal behaviors	personal characteristics, other pollutants	other pollutants, area-level covariates (e.g. poverty index, altitide)
Regression model	Generalized additive model (GAM) or Generalized Linear Model (GLM)	Conditional logistic regression	Marginal, conditional, or transitional model for repeated measures	Cox proportional hazards model
Parameter of interest	relative rate expressed as % increase in mortality/ morbidity per unit increase in daily air pollution	relative risk of the health event per unit increase in air pollution	increase in the odds of symptoms or change in function/ growth per unit increase in air pollution	relative risk of death or disease incidence for areas with high versus low air pollution
Advantages	inexpensive; applicable to large publicly available data bases; targets large populations; design doesn't depend on exposures that are independent of ambient source exposures	key confounders controlled by design; targets a specific population; population-based sampling readily incorporated; ability to assess effect modification	most feasibly incorporates total personal exposure; ability to separately estimate within and between-person effects; ability to control for subject-specific covariates; can easily target sensitive subpopulations	evaluates effects of long term exposure; control for subject-specific covariates
Disadvantages	difficult to asses adequate control for season and trend; results sensitive to lag choices; aggregation over the population may cause bias or blur differential effects; doesn't allow for direct evaluation of effect modification	difficult to choose the referent period; time-varying personal exposure and covariate data often not available; results sensitive to lag choices	tend to rely on small samples; often sample of convenience is used; difficult to determine adequate control for season and trend; results sensitive to lag choices; collection of high quality data requires intensive effort	expensive and time-consuming; does not differentiate acute from chronic effects; difficult to differentiate historical from recent exposure; difficult to assess adequate control for spatial confounding
Examples	(NMMAPS) Samet et al. (2000)	Levy et al. (2001)	(CAMP) Yu et al. (2000)	(Six Cities) Dockery et al. (1993), (ACS) Pope at al. (1995)

All designs may have additional bias due to ecological exposures and exposure measurement error

Table 2: Results of the measurement error analysis. $\hat{\alpha}_k$ is the estimated change in the personal exposure for unit change in the k-th time scale variation of the ambient concentrations, $\hat{\beta}_k$ is the percentage increase in total mortality per 10 $\mu g/m^3$ increase in the ambient pollution level at time scale k, and $\hat{\beta}_k^P$ is the percentage increase in total mortality per 10 $\mu g/m^3$ increase in the average personal pollution level at time scale k. Numbers between () are the 95% confidence intervals.

time-scale	$\hat{\alpha}_k$	$\hat{\beta}_k$	$\hat{\beta}_k^P$
< 21	$0.83(0.73, 0.93)$	$0.55(0.20, 0.91)$	$0.66(-0.19, 1.51)$
10	$0.53(0.43, 0.53)$	$0.54(0.24, 0.84)$	$1.01(0.41, 1.61)$
5	$0.52(0.34, 0.70)$	$0.27(0.01, 0.56)$	$0.52(-0.03, 1.07)$
< 3.5	$0.42(0.07, 0.75)$	$-0.01(-0.33, 0.32)$	$-0.02(-0.67, 0.63)$

Acknowledgments

The authors wish to express their appreciation to Sandra McBride, Thomas Lumley, and Pamela Shaw for helpful input. Research described in this article was partially supported by a contract and grant from Health Effects Institute (HEI), an organization jointly funded by the Environmental Protection Agency (EPA R824835) and automotive manufacturers. The contents of this article do not necessarily reflect the views and policies of HEI, nor do they necessarily reflect the views and policies of EPA, or motor vehicles or engine manufacturers. Funding for Francesca Dominici was provided by a grant from the Health Effects Institute (Walter A. Rosenblith New Investigator Award). Funding for Lianne Sheppard was provided by grants from EPA and NIEHS (EPA R827355 and NIHES R29ES08062). Funding for Merlise Clyde was provided by EPA grant CR 827868-01-0.

References

Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle." In *Second International Symposium on Information Theory*, eds. B. Petrox and F. Caski, 267.

Almon, S. (1965). "The Distributed Lag Between Capital Appropriations and Expenditures." *Econometrica*, 33, 178–196.

American Thoracic Society, Bascom, R. (1996a). "Health effects of outdoor air pollution, Part 1." *American Journal of Respiratory and Critical Care Medicine*, 153, 3–50.

— (1996b). "Health effects of outdoor air pollution, Part 2." *American Journal of Respiratory and Critical Care Medicine*, 153, 477–498.

Austin, H., Flanders, W. D., and Rothman, K. J. (1989). "Bias arising in case–control studies from selection of controls from overlapping groups." *International Journal of Epidemiology*, 18, 713–716.

Bateson, T. F. and Schwartz, J. (1999). "Control for seasonal variation and time trend in case–crossover studies of acute effects of environmental exposures." *Epidemiology*, 10, 539–544.

— (2001). "Selection Bias and Confounding in Case-Crossover Analyses of Environmental Time-Series Data." *Epidemiology*, 12, 654–661.

Best, N. G., Ickstadt, K., and Wolpert, R. L. (2000). "Spatial Poisson regression for health and exposure data measured at disparate resolutions." *Journal of the American Statistical Association*, 95, 1076–1088.

Braga, A., Zanobetti, A., and Schwartz, J. (2001). "The lag structure between particulate air pollution and respiratory and cardiovascular deaths in 10 US cities." *joem*, 43, 11, 927–933.

Breslow, N. and Clayton, D. (1993). "Approximation Inference in generalized linear mixed models." *Journal of the American Statistical Association*, 88, 9–25.

Breslow, N. E. and Day, N. E. (1980). "Statistical Methods in Cancer Research. Volume I - The analysis of case-control studies." In *IARC Scientific Publication Series, No 32.*. Internatina; Agency for Research on Cancer: Lyon.

Briggs, D. J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., Pryl, K., van Reeuwijk, H., Smallbone, K., and van der Veen, A. (1997). "Mapping Urban Air Pollution Using GIS: A Regression-Based Approach." *International Journal of Greographical Information Science*, 11, 7, 699–718.

Burnett, R. and Krewski, D. (1994). "Air Pollution effects of hospital admission rates: A random effects modelling approach." *The Canadian Journal of Statistics*, 22, 441–458.

Cakmak, S., Burnett, R., and D., K. (1999). "Methods for detecting and estimating population threshold concentrations for air pollution-related mortality with exposure measurement error." *Risk Analysis*, 19, 487–496.

Cakmak, S., Burnett, R., and Krewski, D. (1998). "Adjusting for temporal variation in the analysis of parallel time series of health and environmental variables." *Journal of Exposure Analysis and Environmental Epidemiology*, 8, 129–144.

Carlin, B. P. and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, New York.

Carlin, B. P. and Mugglin, A. S. (1997). "Hierarchical Bayes Population Interpolation Models for Geographic Information Systems." In *ASA Proceedings of the Biometrics Section*, 9–18. American Statistical Association (Alexandria, VA).

Carroll, R. J., Ruppert, D., and Stefanski, L. (1995). *Measurement Error in Nonlinear Models*. Chapman & Hall.

Chambers, J. and Hastie, T. (1992). *Statistical Models in S*. Chapman and Hall, London.

Chipman, H., George, E., and McCulloch, R. (2001). "The Practical Implementation of Bayesian Model Selection." In *IMS Lecture Notes - Mongraph Series, Volume 38*, 65–134.

Ciocco, A. and Thompson, D. (1961). "A follow-up of Donora ten years after: methodology and findings." *American Journal of Public Health*, 51, 155–164.

Clayton, D. (1991). "Models for the Longitudinal Analysis of Cohort and Case-control Studies With Inaccurately Measured Exposures." In *Statistical Models for Longitudinal Studies of Health*, 301–331.

Clayton, D. and Hills, M. (1993). *Statistical Models in Epidemiology*. Oxford: Claredon Press.

Clyde, M. (1999). "Bayesian Model Averaging and Model Search Strategies (with discussion)." In *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 157–185. Oxford: Oxford University Press.

— (2000). "Model Uncertainty and Health Effect Studies for Particulate Matter." *Environmetrics*, 745–763.

Clyde, M., Guttorp, P., and Sullivan, E. (2000). "Effects of Ambient Fine and Coarse Particles on Mortality in Phoenix, Arizona." Discussion Paper 00-05, Duke University ISDS, USA.

Cox, D. R. and Oakes, D. O. (1984). *Analysis of Survival Data*. New York: Chapman & Hall.

Daniels, M., Dominici, F., Samet, J. M., and Zeger, S. L. (2000). "Estimating PM10-Mortality Dose-Response Curves and Threshold Levels: An Analysis of Daily Time-Series for the 20 Largest US Cities." *American Journal of Epidemiology*, 152, 397–412.

DerSimonian, R. and Laird, N. (1986). "Meta-analysis in Clinical Trials." *Controlled Clinical Trials*, 7, 177–188.

Diggle, P. J., Liang, K.-Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Clarendon Press.

Dockery, D. and Pope, C. A. (1994). "Acute respiratory effects of particulate air pollution." *Annual Review of Public Health*, 15, 107—132.

Dockery, D., Pope, C. A., Xu, X., Spengler, J., Ware, J., Fay, M., Ferris, B., and Speizer, F. (1993). "An association between air pollution and mortality in six U.S. cities." *New England Journal of Medicine*, 329, 1753–1759.

Dominici, F. (2002). "Air Pollution and Health: what can we learn from an hierarchical approach? Invited commentary." *American Journal of Epidemiology*, 1, 11–15.

Dominici, F., Daniels, M., Zeger, S. L., and Samet, J. M. (2002a). "Air Pollution and Mortality: Estimating Regional and National Dose-Response Relationships." *Journal of the American Statistical Association*, 97, 100–111.

Dominici, F., McDermott, A., Zeger, S. L., and Samet, J. M. (2002b). "0n Generalized Additive Models in Time Series Studies of Air Pollution and Health." *American Journal of Epidemiology*, 156, 3, 1–11.

— (2002c). "Airborne Particulate Matter and Mortality: Time-Scale Effects in Four US Cities." *American Journal of Epidemiology (under revision).*

— (2002d). *A Report to the Health Effects Institute on Reanalyses of the NMMAPS Database*. The Health Effects Institute, Cambridge, MA.

Dominici, F., Samet, J. M., and Zeger, S. L. (2000a). "Combining Evidence on Air pollution and Daily Mortality from the Twenty Largest US cities: A Hierarchical Modeling Strategy (with discussion)." *Royal Statistical Society, Series A, with discussion*, 163, 263–302.

Dominici, F., Zeger, S. L., and Samet, J. M. (2000b). "A measurement error correction model for time-series studies of air pollution and mortality." *Biostatistics*, 2, 157–175.

Draper, D. (1995). "Assessment and Propagation of Model Uncertainty (Disc: P71-97)." *Journal of the Royal Statistical Society, Series B, Methodological*, 57, 45–70.

Duan, N. (1991). "Stocahstic Microenvironmental Models for Air pollution Exposure." *Journal of Exposure Analysis and Envronmental Epidemiology*, 1, 2, 235–257.

Everson, P. and Morris, C. (2000). "Inference for multivariate Normal hierarchical models." *Journal of the Royal Statistical Society, series B*, 62, 399–412.

Fung, Y. and Krewski, D. (1999). "On measurement error adjustement methods in Poisson regression." *Environmetrics*, 10, 213–224.

Gauderman, W. J., McConnell, R., Gilliland, F., London, S., Thomas, D., Avol, E., Vora, H., Berhane, K., Rappaport, E. B., Lurmann, F., Margolis, H. G., and Peters, J. (2000). "Association between Air Pollution and Lung Function Growth in Southern California Children." *American Journal of Respiratory and Critical Care Medicine*, 162, 1383–1390.

Gelfand, A. E., Zhu, L., and Carlin, B. (2001). "On the change of support problem for spatio-temporal data." *Biostatistics*, 95, 31–45.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., eds. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Goswami, E., Larson, T., and Liu, L. S. (2002). "Spatial Characteristics of Fine Particulate Matter: Identifying Representative Monitoring Locations in Seattle." *Journal of the Air and Waste Management Association*, 52, 3, 324–333.

Green, P. J. (1995). "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination." *Biometrika*, 82, 711–732.

Greenland, S. (1992). "Divergeng Biases in Ecologic and Individual-Level Studies." *Statistics in Medicine*, 11, 1209–1223.

Greenland, S. and Morgenstern, H. (1989). "Ecological Bias, Confounding, and Effect Modification." *International Journal of Epidemiology*, 18, 269–274.

Guthrie, K., Sheppard, L., and Wakefield, J. (2002). "A Hierarchical Aggregate Data Model with Spatially Correlated Disease Rates." *Biometrics (to appear)*.

Guttorp, P. (2000). "Environmental Statistics." *Journal of the American Statistical Association*, 95, 289–292.

Haran, M., Carlin, B. P., Adgate, J. L., Ramachandran, G., Waller, L. A., and Gelfand, A. E. (2002). "Hierarchical Bayes Models for Relating Particulate Matter Exposure Measures." In *Cases Studies in Bayesian Statistics VI*. New York: Springer-Verlag.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*. Chapman and Hall, New York.

Heagerty, P. and Zeger, S. (1996). "Marginal Regression Models for Clustered Ordinal Measurements." *Journal of the American Statistical Association*, 91, 1024–1036.

Henry, R. C. (1987). "Curren Factor Analysis Models Are Ill-Posed." *Atmospheric Environment*, 21, 1815–1820.

Hodges, J. S. (1987). "Uncertainty, Policy Analysis and Statistics." *Statistical Science*, 2, 259–275.

Hoek, G., Fischer, P., van den Brant, P., Goldbohm, S., and Brunekreef, B. (2001). "Estimation of Long-term Average Exposure to Outdoor Air Pollution for a Cohort Study on Mortality." *Journal of Exposure Analysis and Environmental Epidemiology*, 11, 459–469.

Hoeting, H. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). "Bayesian model averaging: A tutorial (with discussion)." *Statistical Science*, 14, 382–417.

Janssen, N., Hoek, G., Brunekreef, B., Harssema, H., Mensink, I., and Zuidhof, A. (1998). "Personal Sampling of Particles in Adults: Relation among Personal, Indoor, and Outdoor Air Concentrations." *American Journal of Epidemiology*, 147, 537—544.

Janssen, N., Hoek, G., Harssema, H., and Brunekreef, B. (1997). "Childhood exposure to PM10: relation between personal, classroom, and outdoor concentrations." *Occupational and Environmental Medicine*, 54, 1—7.

Kass, R. E. and Raftery, A. E. (1995). "Bayes Factors." *Journal of the American Statistical Association*, 90, 773–795.

Katsouyanni, K., Toulomi, G., Spix, C., Balducci, F., Medina, S., Rossi, G., Wojtyniak, B., Sunyer, J., Bacharova, L., Schouten, J., Ponka, A., and Anderson, H. R. (1997). "Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project." *British Medical Journal*, 314, 1658–1663.

Kelsall, J., Samet, J. M., and Zeger, S. L. (1997). "Air Pollution, and Mortality in Philadelphia, 1974-1988." *American Journal of Epidemiology*, 146, 750–762.

Kelsall, J., Zeger, S. L., and Samet, J. M. (1999). "Frequency Domain Log-Linear Models; Air pollution and Mortality." *The Royal Statistical Society, Series C*, 48, 331–344.

Korn, E. L. and Whittemore, A. S. (1979). "Methods for Analyzing Panel Studies of Acute Health Effects of Air Pollution." *Biometrics*, 35, 795–802.

Krewski, D., Burnett, R. T., Goldberg, M. S., Hoover, K., Siemiatycki, J., Jerrett, M., Abrahamowicz, M., and White, W. H. (2000). *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality*. Health Effects Institute, Cambridge, MA.

Lebret, E., Briggs, D., van Reeuwijk, H., Fischer, P., Smallbone, K., Harssema, H., Kriz, B., Gorynski, P., and Elliot, P. (2000). "Small Area Variations in Ambient NO_2 Concentrations in Four European Areas." *Atmospheric Environment*, 34, 177–185.

Levy, D., Lumley, T., Sheppard, L., Kaufman, J., and Checkoway, H. (2001a). "Referent Selection in Case–Crossover Analyses for Acute Health Effects of Air Pollution." *Epidemiology*, 12, 186–192.

Levy, D., Sheppard, L., Checkoway, H., Kaufman, J., Lumley, T., Koenig, J., and Siscovick, D. (2001b). "A Case-Crossover Analysis of Particulate Matter Air Pollution and Out-of-Hospital Primary Cardiac Arrest." *Epidemiology*, 12, 193–199.

Li, Y. and Roth, H. (1995). "Daily Mortality analysis by using different regression models in PHiladelphia county, 1973-1990." *Inhalation Toxicology*, 7, 45–58.

Liang, K.-Y. and Zeger, S. L. (1986). "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*, 73, 13–22.

Lindley, D. V. and Smith, A. F. M. (1972). "Bayes Estimates for the Linear Model (with Discussion)." *Journal of the Royal Statistical Society, Series B, Methodological*, 34, 1–41.

Lioy, P., Waldman, J., Buckley, T., Butler, J., and Pietarinen, T. (1990). "The Personal, Indoor, and Outdoor Concentrations of PM10 measured in an industrial community during the winter." *Atmospheric Environment*, 24, 57—66.

Lipfert, F. (1994). *Air Pollution and Community Health: A Critical Review and Data Sourcebook*. New York: Van Nostrand Reinold.

Lipfert, F. and Wyzga, R. (1993). "Air Pollution and Mortality: Issues and Uncertainty." *Journal Air Waste Manage. Assoc*, 45, 949—966.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996). *SAS System for Mixed Models*. Cary, North Carolina: SAS Instistute.

Logan, W. and Glasg, M. (1953). "Mortality in London fog incident, 1953." *Lancet*, 1, 336–338.

Lumley, T. and Heagerty, P. (1999). "Weighted Empirical Adaptive Variance Estimators for Correlated Data Regression." *Journal of the Royal Statistical Society, Series B*, 61, Part 2, 459–477.

Lumley, T. and Levy, D. (2000). "Bias in the case–crossover design: implications for studies of air pollution." *Environmetrics*, 11, 689–704.

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

Maclure, M. and Mittleman, M. A. (2000). "Should we use a case-crossover design?" *Annual Review Public Health*, 21, 193–221.

Mage, T. and Buckley, T. J. (1995). "The Relationship between Personal Exposures and Ambient Concentrations of Particulate Matter." In *Air and Waste Management, 88-th Annual Meeting, Pittsburgh, PA*, 1–16.

McBride, S. J., Clyde, M. A., and Marcus, A. (2002). "Modeling Airborne Particulate Matter Concentrations by Combining Data from Multiple Monitors." Tech. rep., Institute of Statistics and Decision Sciences.

McCullagh, P. (1980). "Regression Models for Ordinal Data." *Journal of the Royal Statistical Society, Series B*, 42, 109–142.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models (Second Edition)*. New York: Chapman & Hall.

Miller, M. S., Friedlander, S. K., and Hidy, G. M. (1972). "A chemical element balance for the Pasadena aerosol." *J. Colloid Interface Sci.*

Moolgavkar, S. (2002). "Air pollution and daily mortality in three U.S. counties." *Environmental Health Perspectives*, 108, 777–784.

Morgenstern, H. and Thomas, D. (1993). "Principles of Study Design in Environmental Epidemiology." *Environmental Health Perspectives Supplement*, 101, Suppl 4, 23–38.

Morris, C. N. and Normand, S.-L. (1992). "Hierarchical models for combining information and for meta-analysis." In *Bayesian Statistics 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 321–344. Oxford: Oxford University Press.

Mugglin, A. S. and Carlin, B. P. (1998). "Hierarchical Modeling in Geographic Information Systems: Population Interpolation Over Incompatible Zones." *Journal of Agricultural, Biological, and Environmental Statistics*, 3, 111–130.

Mugglin, A. S., Carlin, B. P., and Gelfand, A. E. (1998). "Fully model based approaches for spatially misaligned data." *Journal of American Statistical Association*, 95, 877–887.

Murray, C. J. and Lopez, A. D. (1996). *The Global Burden of Disease*. Harvard School of Public Health and World Health Organization, MA USA.

National Research Council (1998). "Research Priorities for Airborne Particulate Matter." *National Academy Press. Washington, DC.*

— (1999). "Research Priorities for Airborne Particulate Matter. Part II. Evaluating Research Progress and Updating the Portfolio." *National Academy Press. Washington, DC.*

— (2001). "Research Priorities for Airborne Particulate Matter. Part III. Early Research Progress." *National Academy Press. Washington, DC.*

Navidi, W. (1998). "Bidirectional Case–Crossover Designs for Exposures with Time Trends." *Biometrics*, 54, 596–605.

Navidi, W. and Weinhandl, E. (2002). "Risk Set Sampling for Case-Crossover Designs." *Epidemiology*, 13, 100–105.

Neas, L. M., Schwartz, J., and Dockery, D. (1999). "A Case-Crossover Analysis of Air Pollution and Mortality in Philadelphia." *Environmental Health Perspectives*, 107, 629–631.

Nelson, C. R. and Murray, C. J. (2000). "State-Space Modeling of the Relationship Between Air Quality and Mortality." *Journal of the Air and Waste Management Association*, 50, 7, 1075–1080.

Ott, W., Wallace, L., and Mage, D. (2000). "Predicting Particulate (PM_{10}) Personal Exposure Distributions Using a Random Component Superposition Model." *Journal of the Air and Waste Management Association*, 50, 1390–1406.

Ozkaynak, H., Xue, J., Spengler, J., Wallace, L., Pellizzari, E., and Jenkins, P. (1996). "Personal Exposure to Airborne Particles and Metals: Results from the Particle Team Study in Riverside, California." *Journal of Exposure Analysis and Environmental Epidemiology*, 6, 57—78.

Paatero, P. (1997). "Least squares formulation of robust, non-negative factor analysis." *Chemom. Intell. Lab. System.*, 37, 23–35.

Patil, G. P., Rao, C. R. E., and Bucher Bartelson, B. R. (1998). "Review of Environmental Statistics." *Journal of the American Statistical Association*, 93, 1530–1530.

Pepe, M. S. and Anderson, G. L. (1994). "A Cautionary Note on Inference for Marginal Regression Models with Longitudinal Data and General Correlated Response Data." *Communications in Statistics – Simulation*, 23, 4, 939–951.

Peters, A., Dockery, D. W., Muller, J. E., and Mittleman, M. A. (2001). "Increased particulate air pollution and the triggering of myocardical infarction." *Circulations*, 103, 2810–2810.

Peters, J. M., Avol, E., Gauderman, W. J., Navidi, W., London, S. J., Margolis, H., Rappaport, E., Vora, H., Gong, Jr., H., and Thomas, D. C. (1999a). "A Study of Twelve Southern California Communities with Differing Levels and Types of Air Pollution: II. Effects on Pulmonary Function." *American Journal of Respiratory and Critical Care Medicine*, 159, 768–775.

Peters, J. M., Avol, E., Navidi, W., London, S. J., Gauderman, W. J., Lurmann, F., Linn, W. S., Margolis, H., Rappaport, E., Gong, Jr., H., and Thomas, D. C. (1999b). "A Study of Twelve Southern California Communities with Differing Levels and Types of Air Pollution: I. Prevalence of Respiratory Morbidity." *American Journal of Respiratory and Critical Care Medicine*, 159, 760–767.

Pope, C. A. (2000). "Invited Commentary: Particulate Matter-Mortality Exposure-Response Relations and Threshold." *American Journal of Epidemiology*, 152, 407–412.

Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., and Thruston, G. D. (2002). "Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution." *Journal of the American Medical Association*, 287, 9, 1132–1141.

Pope, C. A., Dockery, D., and Schwartz, J. (1995a). "Review of Epidemiological evidence of health effects of particulate air pollution." *Inhal.Toxicol.*, 47, 1—-18.

Pope, C. A., Thun, M., Namboodiri, M., Dockery, D., Evans, J., Speizer, F., and Heath, C. (1995b). "Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults." *American Journal Respiratory Critical Care Medicine*, 151, 669–674.

Raftery, A. E. (1999). "Comment on "A Critique of the Bayesian Information Criterion for Model Selection"." *Sociological Methods & Research*, 27, 411–427.

Raftery, A. E., Madigan, D., and Volinsky, C. T. (1995). "Accounting for Model Uncertainty in Survival Analysis Improves Predictive Performance (with Discussion)." In *Bayesian Statistics 5*, 323–349. Oxford University Press.

Ramadan, R., Song, X., and Hopke, P. K. (2000). "Identification of sources of Phoenix aerosol by positive matrix factorization." *Journal of the Air and Waste Management Association*, 50, 1308–1320.

Ramsay, T., Burnett, R., and Krewski, D. (2002). "The effect of concurvity in generalized additive models linking mortality and ambient air pollution." *Epidemiology (to appear)*.

Richardson, S. (2000). "Statistical methods for geographical correlation studies." In *Geographical and Environmental Epidemiology: Methods for Small area studies*. Oxford: Oxford University Press.

Richardson, S. and Gilks, W. R. (1993). "A Bayesian Approach to Measurement Error Problems in Epidemiology Using Conditional Independence Models." *American Journal of Epidemiology*, 138, 430–442.

Rodes, C., Lawless, P., Evans, G., Sheldon, L., Williams, R., Vette, A., Creason, J., and Walsh, D. (2001). "The relationships between personal PM exposures for elderly populations and indoor and outdoor concentrations for three retirement center scenarios." *J. Expos. Assess. Environ. Epid.*, 11, 2, 103–115.

Roemer, W., Hoek, G., Brunekreef, B., Kalandidi, A., and Pekkanen, J. (1998). "Daily variations in air pollution and respiratory health in a multicenter study: the PEACE project." *European Respiratory Journal*, 12, 1354–1361.

Samet, J. M., Dominici, F., Curriero, F., Coursac, I., and Zeger, S. L. (2000a). "Fine Particulate air pollution and Mortality in 20 U.S. Cities: 1987-1994." *New England Journal of Medicine (with discussion)*, 343, 24, 1742–1757.

Samet, J. M., Zeger, S. L., and Berhane, K. (1995). *The Association of Mortality and Particulate Air Pollution*. Health Effects Institute, Cambridge, MA.

Samet, J. M., Zeger, S. L., Dominici, F., Curriero, F., Coursac, I., Dockery, D., Schwartz, J., and Zanobetti, A. (2000b). *The National Morbidity, Mortality, and Air Pollution Study Part II: Morbidity and Mortality from Air Pollution in the United States*. Health Effects Institute, Cambridge, MA.

Samet, J. M., Zeger, S. L., Dominici, F., Dockery, D., and Schwartz, J. (2000c). *The National Morbidity, Mortality, and Air Pollution Study Part I: Methods and Methodological Issues*. Health Effects Institute, Cambridge, MA.

Schlesselman, J. J. (1994). *Case Control Studies: Design, Conduct, Analysis.*. Oxford University Press, New York.

Schwartz, J. (1994a). "Air Pollution and Daily Mortality: A Review and Meta Analysis." *Environmental Research*, 64, 36–52.

— (1994b). "Nonparametric Smoothing in the Analysis of Air Pollution and Respiratory Illness." *The Canadian Journal of Statistics*, 22, 471–487.

— (1995). "Air Pollution and Daily Mortality in Birmingham, Alabama." *American Journal of Epidemiology*, 137, 1136–1147.

— (2000a). "Assessing Confounding, Effect Modification, and Thresholds in the Associations between Ambient Particles and Daily Deaths." *Environmental Health Perspective*, 108, 563–568.

— (2000b). "Harvesting and Long Term Exposure Effects in the Relation between Air Pollution and Mortality." *American Journal of Epidemiology*, 151, 440–448.

— (2001). "Is there harvesting in the association of airborne particles with daily deaths and hospital admission?" *Epidemiology*, 12, 55–61.

Schwartz, J. and Lee, J. T. (1999). "Reanalysis of the effects if air pollution on daily mortality in Seoul, Korea: A case-crossover design." *Environmental Health Perspective*, 107, 633–636.

Schwartz, J. and Zanobetti, A. (2000). "Is there a threshold in the relation between particulate air pollution and daily deaths: a meta-smoothing approach." *Epidemiology*, 6, 666– 672.

Schwarz, G. (1978). "Estimating the Dimension of a Model." *The Annals of Statistics*, 6, 461–464.

Sheppard, L. (2002a). "Ecologic Study Design." In *Encyclopedia of Environmentrics*, 673–705. John Wiley and Sons, New York.

— (2002b). "Insights on bias and information in group-level studies." *Biostatistics*, to appear.

Sheppard, L. and Damian, D. (2000). "Estimating Short-Term PM Effects Accounting for Surrogate Exposure Measurements from Ambient Monitors." *Environmetrics*, 11, 675–687.

Sheppard, L., Levy, D., Norris, G., Larson, T. V., and Koenig, J. Q. (1999). "Effects of Ambient Air Pollution on Nonelderly Asthma Hospital Admissions in Seattle, Washington, 1987–1994." *Epidemiology*, 10, 23–30.

Sheppard, L. and Prentice, R. L. (1995). "On the Reliability and Precision of Within and Between Population Estimates of Relative Risk Parameters." *Biometrics*, 51, 853–863.

Sheppard, L., Prentice, R. L., and Rossing, M. A. (1996). "Design Considerations for Estimation of Exposure Effects on Disease Risk, Using Aggregate Data Studies." *Statistics in Medicine*, 15, 1849–1858.

Slaughter, J. C., Lumley, T., Sheppard, L., Koenig, J. Q., and Shapiro, G. S. (2002). "Effects of ambient air pollution on symptoms severity and medication usage in asthmatic children." *To Be Submitted*.

Smith, R., Guttorp, P., Sheppard, L., Lumley, T., and Ishikama, N. (2001). "Comments on the PM Criteria Document." Technical report (nrcse-trs no.66) http://www.nrcse.washington.edu/research/reports.html, University of Washington, Seattle.

Smith, R., J.M., D., Sacks, J., and Speckman, P. (2000a). "Regression models for air pollution and daily mortality: analysis of data from Birmingham, Alabama." *Environmetrics*, 11, 6, 719–743.

Smith, R., Kim, Y., and Fuentes, M. (2000b). "Threshold dependence of mortality effects for fine and coarse particles in Phoenix, Arizona." *Journal of Air Waste Management Association*, 50, 1367– 1379.

Sullivan, E. M. (2000). "Estimating the Association Between Ambient Particulate Matter and Elderly Mortality in Phoenix and Seattle Using Bayesian Model Averaging." Tech. Rep. 54, National Research Center for Statistics and the Environment, University of Washington.

Tanner, M. A. (1991). *Tools for Statistical Inference – Observed Data and Data Augmentation Methods*, vol. 67 of *Lecture Notes in Statistics*. New York: Springer-Verlag.

Thomas, A., Spiegelhalter, D. J., and Gilks, W. R. (1992). "BUGS: A Program to Perform Bayesian Inference Using Gibbs Sampling." In *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting*, 837–842. Clarendon Press: Oxford.

Thomas, D. (2000). "Some Contributions of Statistics to Environmental Epidemiology." *Journal of the American Statistical Association*, 95, 315–318.

Thomas, D., Stram, D., and Dwyer, J. (1993). "Exposure Measurement error: Influence on Exposure-Disease Relationships and Methods of Corrections." *Annual Review Public Health*, 14, 69–93.

Thurston, G. (1996). "A critical review of PM_{10}-Mortality Time-Series studies." *Journal of Exposure Analysis and Environnmental Epidemiology*, 6, 3–21.

Tierney, L. (1994). "Markov chains for exploring posterior distributions (with discussion)." *Annals of Statistics*, 22, 4, 1701–1762.

Vedal, S. (1996). "Ambient Particles and Health: Lines that Divide." *Journal of the Air Waste Management Association*, 47, 551—-581.

Volinsky, C. T., Madigan, D., Raftery, A. E., and Kronmal, R. A. (1997). "Bayesian Model Averaging in Proportional Hazard Models: Assessing the Risk of a Stroke." *Applied Statistics*, 46, 433–448.

Wakefield, J. C. and Salway, R. E. (2001). "A Statistical Framework for Ecological and Aggregate Studies." *Journal of the Royal Statistical Society, Series A*, 164, 119–137.

Watson, J. G., Cooper, J. A., and Huntzicker, J. J. (1984). "The Effective Variance Weighting for Least Squares Calculations Applied to the Mass Balance Receptor Model." *Atmospheric Environment*, 18, 1347–1355.

Wilson, W. E., Mage, D. T., and Grant, L. D. (2000). "Estimating Separately Personal Exposure to Ambient and Nonambient Particulate Matter for Epidemi9ology and Risk Assessment: Why and How." *Journal of the Air and Waste Management Association*, 50, 1167–1183.

Yu, O. C., Sheppard, L., Lumley, T., Koenig, J. Q., and Shapiro, G. G. (2000). "Effects of ambient air pollution on symptoms of asthma in Seattle-area children enrolled in the CAMP study." *Environmental Health Perspectives*, 108, 12, 1209–1214.

Zanobetti, A., Schwartz, J., and Dockery, D. (2000a). "Airborne particles are a risk factor for hospital admissions for heart and lung disease." *Environmental Health Perspective*, 108, 1071–1077.

Zanobetti, A., Wand, M., Schwartz, J., and Ryan, L. (2000b). "Generalized Additive Distributed Lag Models." *Biostatistics*, 1, 279–292.

Zanobetti, A. et al. (2002). "The temporal pattern of mortality responses to air pollution: a multicity assessment of mortality displacement." *Epidemiology*, 13, 87–93.

Zeger, S., Dominici, F., McDermott, A., and Samet, J. (2002). "Modeling the Health Effects of Environmental Exposures: Particulate Air Pollution and Mortality in Urban America." In *Monitoring the Health of the Populations: Statistical Methods for Public Health Surveillance*, ?–? Oxford: Oxford University Press.

Zeger, S., Thomas, D., Dominici, F., Cohen, A., Schwartz, J., Dockery, D., and Samet, J. (2000). "Measurement Error in Time-Series Studies of Air Pollution: Concepts and Consequences." *Environmental Health Perspectives*, 108, 5.

Zeger, S. L., Dominici, F., and Samet, J. M. (1999). "Harvesting-resistant estimates of pollution effects on mortality." *Epidemiology*, 89, 171–175.

Zeger, S. L., Liang, K., and Albert, P. (1988). "Models for longitudinal data: A generalised estimating equation approach." *Biometrics*, 44, 1049–1060.

Zidek, J., Wong, H., Le, N., and Burnett, R. (1996). "Causality, Measurement Error and Multicollinearity in Epidemiology." *Environmetrics*, 7, 441–451.

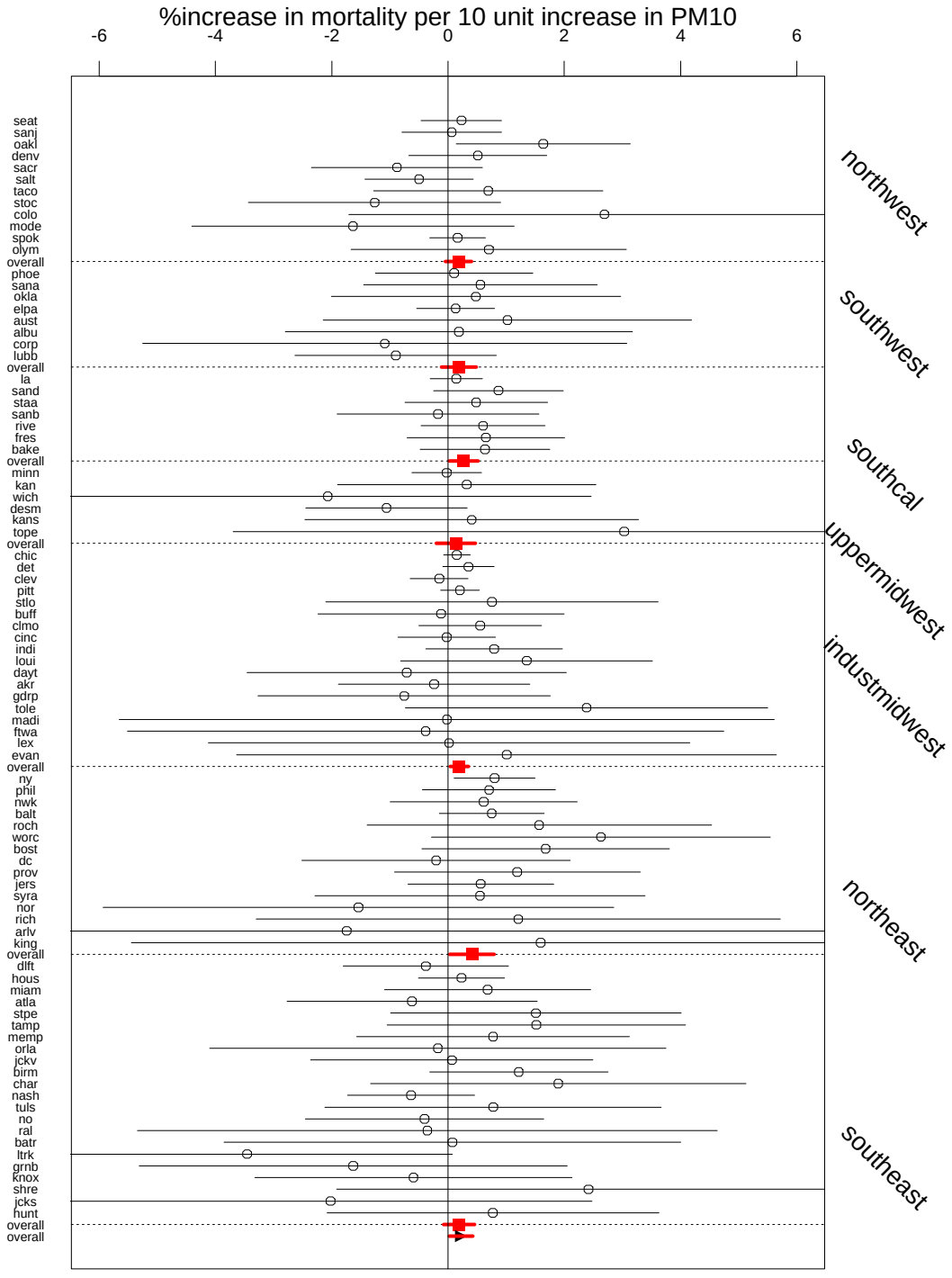

Figure 1: *Maximum likelihood estimates and 95% confidence intervals of the log-relative rates of mortality per* $10\mu g/m^3$ *increase in* PM_{10} *for each location. The solid circles with the bold segments denote the posterior means and 95% posterior intervals of the pooled regional effects. At the bottom, marked with a triangle are the overall effects for* PM_{10} *for all the cities. These results are obtained by using generalized linear models with natural cubic splines, and also reported in the HEI Re-analyses Report (Dominici et al. 2000b).*

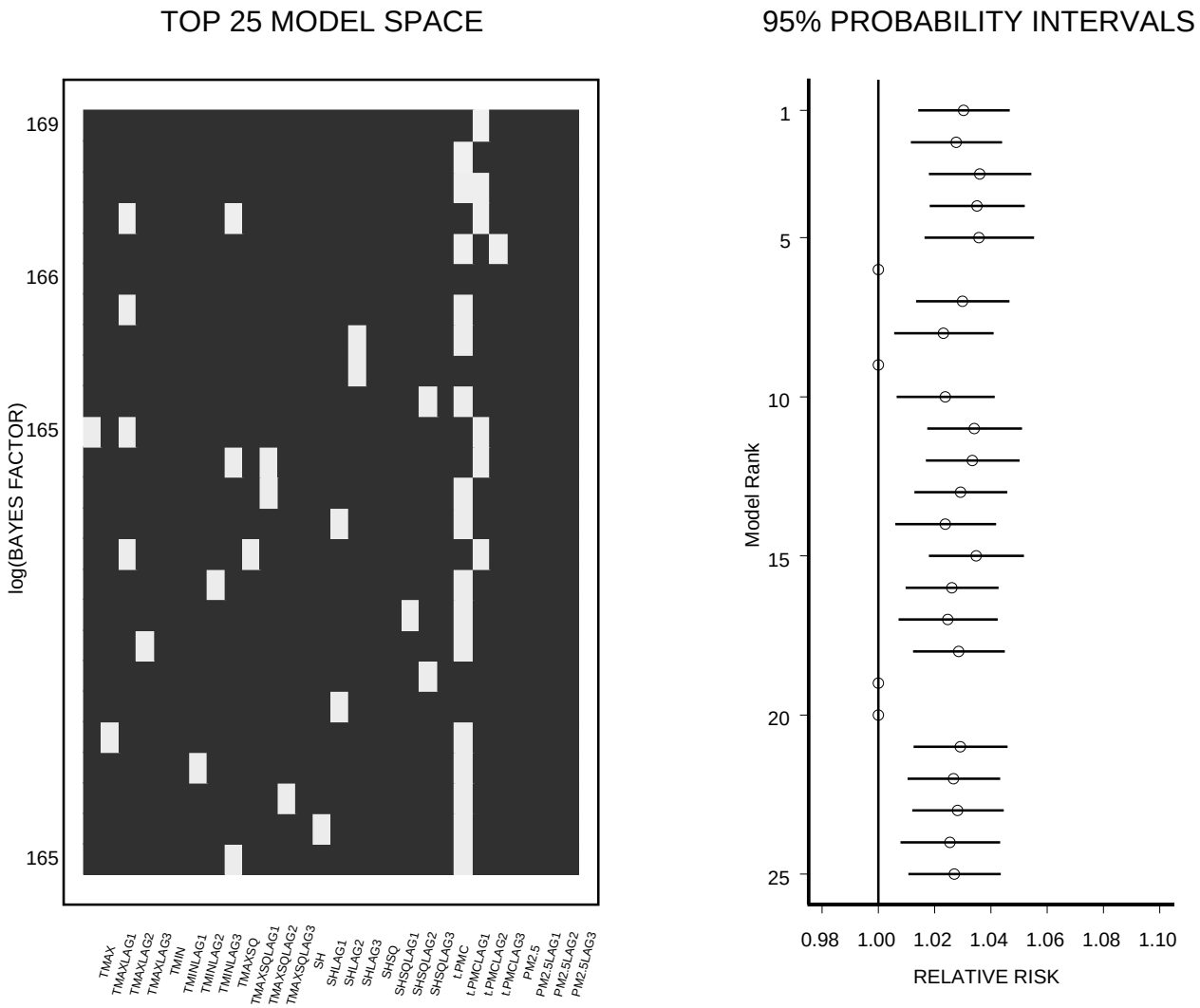

Figure 2: Top 25 models and associated 95 % probability intervals for the relative risk associated with a one interquartile range increase in particulate matter for Phoenix, AZ. In the left plot of the model space, rows correspond to models and columns to variables, with white rectangles indicating that the variable for that column is included in the model corresponding to that row. The particulate matter variables (coarse, fine, and lags of each) are in the last 8 columns.

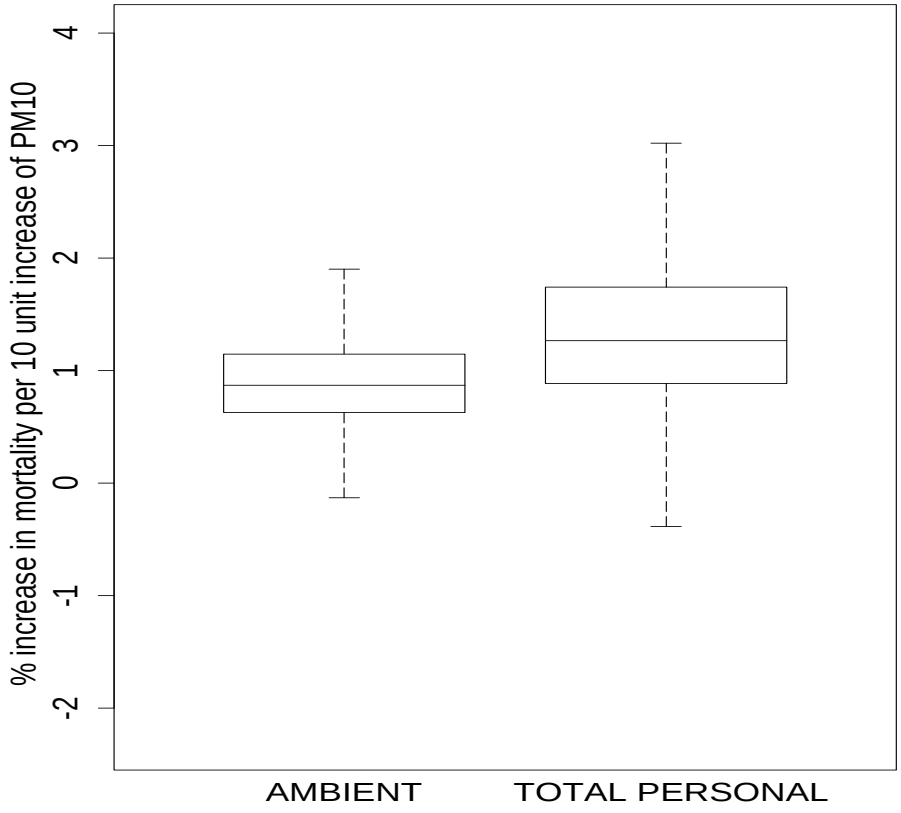

Figure 3: *Comparison between the log relative rates (% increase in mortality for 10 μ_g/m^3increase of PM_{10} exposure), obtained by fitting the generalized additive model having as predictors the ambient concentrations and the average personal exposures, respectively. The boxplots represents the posterior distributions of the two log relative rates (β_z, β_x).*

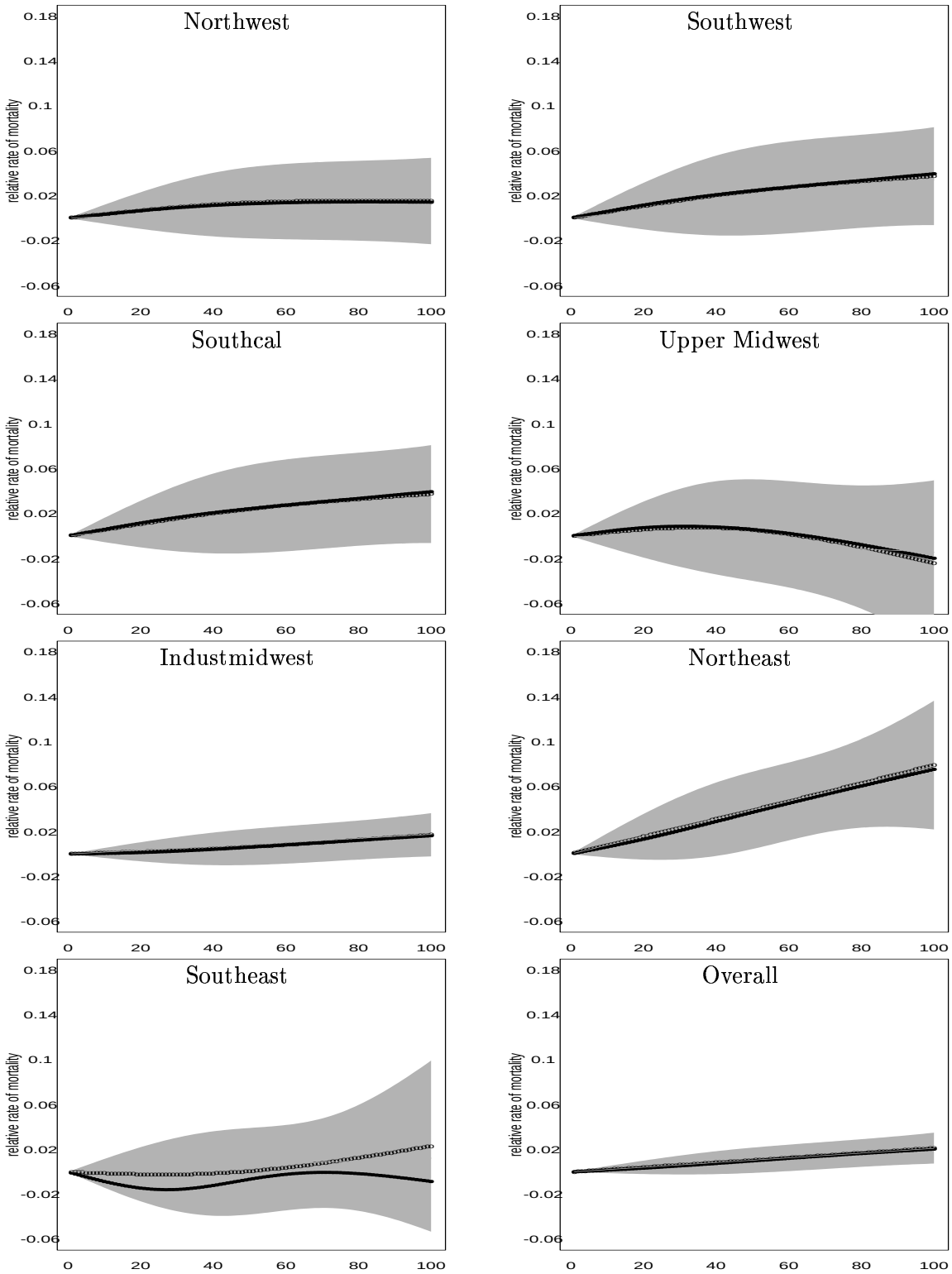

Figure 4: *Regional and national PM_{10}-mortality dose response curves. Solid black curves are obtained by fitting the spline model with the RJMCMC and allowing for an unknown number and location of knots. The curves with the empty dots are obtained by setting one knot at $40\mu_g/m^3$ and fitting the spline model with a Gibbs sampler. The linear curves are obtained by fitting the hierarchical linear model with a Gibbs sampler without borrowing strength across regions. The shaded area denotes the 95% confidence bands for the curve with a fixed knot. These results are obtained by using generalized linear models with natural cubic splines.*