

Confounding and exposure measurement error in air pollution epidemiology

Lianne Sheppard · Richard T. Burnett ·
Adam A. Szpiro · Sun-Young Kim · Michael Jerrett ·
C Arden Pope III · Bert Brunekreef

Received: 10 December 2010 / Accepted: 2 March 2011 / Published online: 23 March 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Studies in air pollution epidemiology may suffer from some specific forms of confounding and exposure measurement error. This contribution discusses these, mostly in the framework of cohort studies. Evaluation of potential confounding is critical in studies of the health effects of air pollution. The association between long-term exposure to ambient air pollution and mortality has been investigated using cohort studies in which subjects are followed over time with respect to their vital status. In such studies, control for individual-level confounders such as smoking is important, as is control for area-level confounders such as neighborhood socio-economic status. In addition, there may be spatial dependencies in the survival data that need to be addressed. These issues are illustrated

using the American Cancer Society Cancer Prevention II cohort. Exposure measurement error is a challenge in epidemiology because inference about health effects can be incorrect when the measured or predicted exposure used in the analysis is different from the underlying true exposure. Air pollution epidemiology rarely if ever uses personal measurements of exposure for reasons of cost and feasibility. Exposure measurement error in air pollution epidemiology comes in various dominant forms, which are different for time-series and cohort studies. The challenges are reviewed and a number of suggested solutions are discussed for both study domains.

Keywords Air pollution · Epidemiology · Confounding · Measurement error

L. Sheppard · A. A. Szpiro · S.-Y. Kim
University of Washington,
Seattle, WA, USA

R. T. Burnett
Health Canada,
Ottawa, Canada

M. Jerrett
University of California,
Berkeley, CA, USA

C. A. Pope III
Brigham Young University,
Provo, UT, USA

B. Brunekreef (✉)
Institute for Risk Assessment Sciences (IRAS),
Utrecht University,
PO Box 80178, 3508 TD Utrecht, The Netherlands
e-mail: B.Brunekreef@uu.nl

B. Brunekreef
Julius Center for Health Sciences and Primary Health Care,
Utrecht University,
Utrecht, The Netherlands

Introduction

Studies in air pollution epidemiology may suffer from some specific forms of confounding and exposure measurement error.

Evaluation of potential *confounding* is critical in these studies. For example, time-series studies of the daily number of adverse health events (e.g., deaths, hospital admissions, emergency room visits) and daily variation in ambient air pollution concentrations focus on how these events co-vary over time. Both the event and air pollution time series have strong temporal cycles defined by weather, day of week, seasonal, and long-term trends. Thus, time and time-varying determinants of health events are potentially important confounders to the air pollution–health event association, and these should be adequately accounted for in the analysis (Dominici et al. 2003).

Time-series studies have been principally used to understand the association between relatively short-term

temporal associations in ambient air pollution concentrations and adverse health events such as daily mortality and hospital admissions on the order of days or weeks. However, to understand how longer-term exposure to air pollution, on the order of years, plays in the development of disease and longevity, researchers have turned to cohort studies. Variation in air pollution exposures is achieved in this design by long-term follow-up of subjects who live in different communities (Dockery et al. 1993; Jerrett et al. 2009; Krewski et al. 2009; Zeger et al. 2008) or in different areas within a community (Beelen et al. 2008; Jerrett et al. 2005b). Some studies include variation in exposure generated at several spatial levels (Miller et al. 2007; Puett et al. 2008).

In population-based time-series studies, various risk factors, such as diet, smoking, or socio-demographic factors, are not likely to be confounders because they do not co-vary with pollution over relatively short time periods of interest (i. e., days) when averaged over large populations (Burnett et al. 2003). However, these risk factors clearly have spatial patterns and thus must be accounted for in the analysis of cohort studies when considering the effects of longer-term pollution exposure. In this paper, we briefly present the primary statistical model typically used to evaluate cohort data in air pollution studies, the Cox proportional hazards (CPH) model. We then present some relevant extensions to this model that allow for adequate accounting for multiple levels of clustering and for potential spatial autocorrelation. We utilize this approach to incorporating spatial aspects of cohort studies in analyses of the link between long-term ambient air pollution concentrations and mortality within the American Cancer Society's (ACS) Cancer Prevention Study II (CPS-II). Finally, we evaluate the evidence that the observed associations between air pollution and mortality are potentially due to spatial confounding.

A complete set of pertinent exposure measurements is typically not available in epidemiological studies of associations between air pollutants and a disease outcome. For this reason, the possibility of *exposure measurement error* always needs to be carefully considered because it can affect the inferences we draw from our studies. Depending on what type of error affects the exposure assignments, the effects of air pollution on some health variable may be biased and/or estimated with imprecision. Both of these may become serious enough to completely negate a study's potential to allow valid inference regarding the effect of air pollution on health. Typically, the analysis does not account for the uncertainty in the exposure assignment process, if only because the uncertainty is often not known. In this paper, we will provide a framework for exposure measurement error in air pollution epidemiology studies, with theoretical and practical examples of how measurement error can be quantified and accounted for.

This paper is divided into two sections. First, we discuss spatial confounding in cohort studies and illustrate its impact in the American Cancer Society's Cancer Prevention Study II (CPS-II). In Sect. [Measurement error in air pollution epidemiology](#), we review measurement error research in the context of air pollution epidemiology studies. We close with discussion about both confounding and measurement error.

Spatial confounding in cohort studies

Methods

Cox proportional hazards survival model

Cohort studies of outdoor air pollution have commonly used the CPH survival model to relate survival experience to exposure while simultaneously controlling for other well-known mortality risk factors. The model has the form

$$\lambda_i^{(l)}(t) = \lambda_0^{(l)}(t) \exp\left(\beta^T x_i^{(l)}(t)\right)$$

where $\lambda_i^{(l)}(t)$ is the hazard function or instantaneous probability of death for the i th subject in the l th stratum at follow-up time t . Follow-up time can either be recorded as the calendar time from the start of the study or the difference between the age of each subject when they entered the study and the age at last observation: either death, lost to follow-up, or termination of the study. We have previously shown that the air pollution association with mortality in the CPS-II cohort was similar using these two different methods of specifying follow-up time (Health Effects Institute 2000). We consider only calendar time in this paper since this time definition more easily accommodates incorporation of time-dependent definitions of air pollution exposure which are naturally defined over calendar time as opposed to age of subject.

The CPH model assumes that the baseline hazard function $\lambda_0^{(l)}(t)$ is common to all subjects within a stratum. The risk of the event is modeled by modulating the baseline hazard by the regression equation, $\left(\beta^T x_i^{(l)}(t)\right)$ which distinguishes risk among subjects within a stratum. The risk factor information such as smoking habits and air pollution concentration is contained in the vector $x_i^{(l)}(t)$ and related to the hazard function by the regression vector β , and can vary in time. Strata are often defined by age, gender, and race. For example, one would follow the event experience of all white females, aged 54 at the beginning of the study, and relate their air pollution exposure to their chances of dying at any time t . This process is repeated for all combinations of the stratification variables, and a single summary estimate of effect is determined (i.e., an estimate of β). This model also assumes that the association between

the risk factors, including air pollution, and the time to event can be represented by a single value, β , which is constant over the follow-up period. In other words, the impact of the risk factor on the hazard function is constant or proportional over the follow-up time. Since differences in values of the risk factors modulate the hazard function, this model is called the proportional hazards model.

Statistical tests of the proportional hazards assumption are available. Flexible modeling of the interaction between a predictor variable of interest and follow-up time has been proposed as a means of assessing the proportional hazards assumption of no such interaction (Abrahamowicz et al. 2003). We suggest an alternative approach to formal statistical testing by examining the sensitivity of the regression parameter estimate and its estimate of uncertainty to the proportional hazards assumption by first dividing the follow-up time into distinct periods, performing the survival regression analysis on each period, and then summarizing the effect of a predictor variable on survival among the time periods by a weighted average of the time-specific parameter estimates, with weights given by the inverse of the square of the parameter estimate error. In this approach, effects of all mortality predictors are allowed to vary by the separate follow-up periods. Thus, for example, the effect of air pollution is estimated in a model which simultaneously allows the effects of other risk factors, such as smoking, to also vary in time.

Spatial survival model

The Cox proportional hazards model assumes that the survival experience among all subjects is statistically independent. However, subjects living in the same community and/or neighborhood within a community intrinsically have some risk factors in common that are not included in the model. These unmeasured factors tend to correlate the experience of subjects within geographic areas. To accommodate this potential correlation or spatial clustering of events, we extend the Cox proportional hazards model to include multiple levels of clustering (Ma et al. 2003). For illustration purposes, we only consider two cluster levels. The model is presented by

$$\lambda^{(l)}_{isr}(t) = \lambda^{(l)}_0(t)U_{sr}(i) \exp\left(\beta'x^{(l)}_{isr}(t)\right)$$

where $U_{sr}(i)$ represents the random effect or unexplained variation for the r th sub-cluster within the s th cluster which contains subject i . Let the set of random effects for the clusters be denoted by $U_* = (U_1, \dots, U_S)$ for the S clusters units. Then we have

$$E(U_s) = 1 \text{ Cov}(U_s, U_{s'}) = \sigma^2 \rho^{d_{ss'}}$$

for cluster units s and s' . The cluster-level random effect

variance is σ^2 , common to all clusters and the correlation among the cluster units $d_{ss'}$ apart is $\rho^{d_{ss'}}$, for $-1 < \rho < 1$. The sub-cluster units, conditional on the cluster-level random effects, are stochastically characterized by

$$E(U_{sr}|U_*) = U_s$$

$$\text{Cov}(U_{sr}, U_{s'r'}|U_*) = 0$$

$$\text{Cov}(U_{sr}, U_{sr'}|U_*) = \tau^2 \pi^{d_{sr,sr'}} U_s$$

where τ^2 and π are the sub-cluster random effect variance and autocorrelation parameters, respectively. Given the random effects, responses between subjects are independent, and given the cluster random effects, sub-cluster units in different clusters are also uncorrelated.

Sub-clusters in practice are selected based on factors likely to influence the correlation between individuals. For example, cities may vary in the provision of programs that protect health and have therefore been extensively used as cluster units. Likewise, evidence exists suggesting that the neighborhood in which a person lives may influence their health beyond individual risk factors, so sub-clusters are often selected to represent these neighborhood effects.

We have specified this correlation structure on the random effects since it is also possible that subjects living in clusters or different sub-clusters within the same cluster unit close together will share some lifestyle and environmental risk factors which are not as strongly shared between subjects living in cluster or sub-cluster units farther apart. Distance between cluster units can be defined in a number of ways, including nearest neighbors or Euclidian distance. In the example below, we will define distance $d=1$ if two cluster (sub-cluster) units are adjacent and $d = \infty$ if not. Adjacency is defined by constructing Thiessen polygons for each cluster (sub-cluster) unit. Any two connected polygons are assumed to be adjacent.

The error specification is completed by noting that

$$\text{Cov}(U_{sr}, U_{s'r'}) = \sigma^2 \rho^{d_{ss'}}$$

$$\text{Cov}(U_{sr}, U_{sr'}) = \sigma^2 + \tau^2 \pi^{d_{sr,sr'}}$$

Estimates of the regression and dispersion parameters are given by methods described by Krewski et al. (2009).

Illustration: Confounding in the American Cancer Society’s Cancer Prevention Study II

We will illustrate our approach to incorporating spatial aspects of cohort studies in analysis of the link between long-term ambient air pollution concentrations and mortality with the American Cancer Society’s CPS-II, an ongoing prospective mortality study of approximately 1.2 million

adults. Cohort participants were enrolled by CPS-II volunteers in the fall of 1982 to the winter of 1983. They resided in all 50 states, the District of Columbia, and Puerto Rico. Most were friends, neighbors, or acquaintances of the CPS-II volunteers. Enrollment was restricted to persons who were at least 30 years of age and who were members of households with at least one individual 45 years of age or more. Participants completed a questionnaire which included questions about age, sex, weight, height, demographic characteristics, smoking history, alcohol use, occupational exposures, and other characteristics. For this analysis, the analytic cohort has been restricted to include those who resided in US metropolitan areas within the 48 contiguous states (including the District of Columbia) and within metropolitan areas that had available pollution data. Mortality of the study participants was ascertained by volunteers in 1984, 1986, and 1988, and subsequently with automated linkage using the National Death Index. For the purposes of this illustration, we considered vital status follow-up until 2000. We also obtained information on the concentrations of fine particulate matter (PM_{2.5}) for 1999 and 2000 in 116 metropolitan statistical areas (MSAs). A total of almost 500,000 subjects resided in these 116 MSAs. The average PM_{2.5} concentration was 14.02 $\mu\text{g}/\text{m}^3$, with variance 9.12 $(\mu\text{g}/\text{m}^3)^2$. Concentrations ranged from 5.8 to 22.2 $\mu\text{g}/\text{m}^3$. See Krewski et al. (2009) for more details on the health and exposure data.

We initially examined the appropriateness of the proportional hazards assumption on inferences of the regression parameters by subdividing the entire follow-up time from 1982 to 2000 into six intervals: 1982–1985, 1986–1988, 1989–1991, 1992–1994, 1995–1997, and 1998–2000. We then ran the Cox proportional hazards model for each interval separately and estimated the PM_{2.5} regression coefficient in addition to all the coefficients of the corresponding mortality risk factors. Thus, all the coefficients were allowed to vary by time period. We then summarized the effects of each predictor on mortality by taking a weighted average of the time-period-specific coefficients with the weights defined by the inverse of their respective squared standard errors. This is not a formal test for proportional hazards but a sensitivity analysis on this assumption with respect to the estimate of the regression coefficients and their uncertainty. Thus, we are not testing whether the coefficients vary in time per se but whether a summary of these time-dependent coefficients and standard error is similar to that obtained from the CPH model based on a single analysis of the entire follow-up period.

Evidence of confounding by individual risk factors To control for age, sex, and race, cohort subjects were stratified by 1-year age categories, sex, and race (white versus other), which allowed each age–sex–race stratum to have its own

baseline hazard. To control for potential confounding by individual risk factors, a host of known mortality risk factors were also included in the survival model that were measured for each individual, including: smoking history, education and marital status, body mass index, alcohol consumption, occupation and dust exposure, and diet. As expected, these covariates were often significant predictors of mortality risk. However, in previous analyses and as reported elsewhere (Pope et al. 2002), we evaluated evidence of confounding by these individual risk factors by sequentially adding the smoking, education and marital status, body mass index, alcohol, occupational dust exposures, and diet factors into the model in a controlled step-wise fashion. After controlling for smoking variables, the inclusion of the additional individual risk factor variables had little influences on the air pollution mortality associations—suggesting minimal residual confounding by these individual risk factors.

Evidence of confounding by spatial contextual factors We have also conducted analyses that included several known mortality predictors defined on various spatial levels (Jerrett 2010; Krewski et al. 2009). These “contextual” effects occur when individual differences in health outcome associate with the grouped variables that represent the social, economic, and environmental settings where the individuals live, work, or spend time (e.g., poverty in a neighborhood).

For this analysis, we obtained information on neighborhood socio-demographic variables that could potentially contribute to spatial confounding. These ecological variables were collected and compiled for the 11,334 zip code areas (ZCAs) within the 116 MSAs from the 1979 US Census and included: median household income, 125% of poverty line, percentage of unemployed persons over the age of 16 years, percentage of adults with less than grade 12 education, percentage of homes with air conditioning, the GINI coefficient of income inequality, and percentage of population that are not white. We used boundary averaging methods to overlay census information at the census sub-division level and the ZCA level for which we have location information from the ACS subjects. We only used those ZCAs which contained ACS subjects to more accurately represent the social environment of the ACS participants for metropolitan areas. We included information on potential ecological risk factors recorded near the commencement of the follow-up period (1982) in order to be temporally consistent with the information on individual risk factors, such as smoking habits. No additional information on the individual risk factors was available during follow-up. However, information may be obtained on the ecological risk factors from subsequent censuses and could be included in the survival model as time-dependent covariates.

Because we were concerned that comparing zip code characteristics between cities does not fully capture

potential confounding, we also created two other variables for inclusion in the survival models. The first involves aggregating all ZCAs with ACS subjects within an MSA to obtain an average estimate of the ecologic confounder. For the second variable, we deviated the zip code specific values from their metropolitan area means (DIFF). This deviation ensures that all comparisons are made within communities where the social variables are most likely to have interpretable results because cost of living and other factors affecting the comparisons are controlled within cities.

Results

For the proportional hazards assessment, the $PM_{2.5}$ coefficient and standard error were virtually identical between the two models (separate and single follow-up time periods). This suggests that the proportional hazards assumption on all the predictor variables was not critical in estimating the effect of $PM_{2.5}$ on mortality throughout the follow-up period, and we will assume a proportional hazards model in any further analysis reported here. Because this proportionality assumption appeared not to influence our inferences concerning $PM_{2.5}$, we proceeded to examine the influence of adjusting for ecological covariates and spatial autocorrelation in the survival data.

The ischemic heart disease (IHD) mortality hazard ratio for a $10\text{-}\mu\text{g}/\text{m}^3$ change in $PM_{2.5}$ is presented in Table 1 for various combinations of adjustment for the ecological covariates (none, defined at the ZCA level and defined at both the MSA and DIFF level) and survival model error specification. We considered clustering only at the MSA level, at both the MSA and ZCA within MSA level in addition to clustering at the state and MSA within state levels.

Adjusting for all seven ecological covariates simultaneously tended to increase the hazard ratio but also increased the width of the 95% confidence interval due to confounding between the ecological covariates and $PM_{2.5}$. The percentage of homes with air conditioning and income were negatively associated with mortality due to ischemic heart disease, while in a proportion of MSA population that did not achieve a high school graduation, unemployment and poverty were positively associated with mortality. Both income disparity and percent non-white were not related to IHD mortality (Table 7 of Krewski et al. 2009). Air conditioning was positively correlated with $PM_{2.5}$ exposure while education, unemployment, and income disparity were negatively associated. Percent non-white, household income, and percent in poverty were not clearly linked to $PM_{2.5}$ concentrations (Table 4 of Krewski et al. 2009). Adjustment of the $PM_{2.5}$ association with IHD mortality for air

conditioning, education, and/or income increased the hazard ratio while adjusting for the other four ecological covariates had little influence on the $PM_{2.5}$ effect (Table 7 of Krewski et al. 2009).

Including random effects at the MSA and ZCA levels also increased the $PM_{2.5}$ hazard ratios (Table 1) and increased the width of the confidence intervals, suggesting that there is a spatial pattern of unexplained IHD mortality in the cohort which was not accounted for in the standard Cox survival model which assumes independence of observations. The addition of the ecological covariates explained only a small amount of the residual variation. This variation was reduced from 10.51×10^{-3} without the ecological covariates, to 10.24×10^{-3} after adjusting for them at the ZCA level, and to 9.47×10^{-3} after adjusting for them at both the MSA and DIFF levels (Table 1).

The assumption that the random effects were spatially uncorrelated was examined by including a correlation structure on the random effects at both the MSA and ZCA within MSA levels using the nearest neighbor specification. We assumed that the random effects of two MSAs that were neighbors are correlated and those MSAs that were not neighbors are uncorrelated. A similar assumption was made on the ZCAs within each MSA. We observed positive spatial autocorrelation of the random effects at both MSA and ZCA cluster levels, and including this error specification reduced that $PM_{2.5}$ hazard ratios and increased the width of the confidence intervals (Table 1) as expected.

The MSA random effect variance was similar to the ZCA within MSA random effect variance, suggesting that there was as much unexplained variation in mortality between ZCAs within a MSA as there was between MSAs. The 95% coverage interval of the random effect estimates ranged from approximately 0.80 to 1.20 for both cluster levels, indicating that the mortality rate in the upper end of this distribution is 50% larger than at the lower end. The $PM_{2.5}$ hazard ratio comparing the maximum concentration of 22.2 to $5.8 \mu\text{g}/\text{m}^3$ is 1.49, assuming a hazard ratio of 1.276 per $10 \mu\text{g}/\text{m}^3$, a value in the middle of the range of observed hazard ratios (Table 1). These results suggest that the amount of unexplained spatial variation in IHD mortality across the USA in this cohort is about the same as that explained by $PM_{2.5}$.

A second spatial definition of the clusters was considered which was geographically broader than MSA–state. States are responsible for the implementation of health care and social assistance and thus health in general and mortality in particular could be clustered at this geographic level. However, similar results were observed when either state or state and MSA within state cluster levels were included in the survival model as random effects compared to the model with MSA and ZCA within MSA.

Table 1 PM_{2.5}-Ischemic heart disease mortality association sensitivity analysis to adjustment for ecological covariates and the error specification of random effects spatial survival model

Cluster level	Sub-cluster level	Ecological covariate adjustment	Hazard ratio ^a (95% confidence limit)	Cluster variance ^b and correlation	Sub-cluster variance ^b and correlation
None	None	None	1.153 (1.111–1.197)	0, 0	0, 0
None	None	ZCA	1.210 (1.163–1.258)	0, 0	0, 0
None	None	MSA and DIFF ^c	1.240 (1.189–1.293)	0, 0	0, 0
MSA	ZCA	None	1.181 (1.092–1.278)	10.51, 0	10.36, 0
MSA	ZCA	ZCA	1.243 (1.147–1.346)	10.24, 0	8.40, 0
MSA	ZCA	MSA and DIFF	1.287 (1.170–1.404)	9.47, 0	8.05, 0
MSA	ZCA	None	1.168 (1.065–1.280)	10.49, 0.36	15.61, 0.30
MSA	ZCA	ZCA	1.229 (1.120–1.347)	10.15, 0.36	10.64, 0.28
MSA	ZCA	MSA and DIFF	1.276 (1.156–1.409)	9.79, 0.36	10.06, 0.27
State	None	MSA and DIFF	1.284 (1.189–1.387)	13.14, 0	0, 0
State	None	MSA and DIFF	1.232 (1.141–1.330)	14.04, 0.39	0, 0
State	MSA	MSA and DIFF	1.320 (1.192–1.460)	9.74, 0	3.05, 0
State	MSA	MSA and DIFF	1.241 (1.112–1.382)	8.83, 0.39	4.07, 0.15

^a Per 10 µg/m³

^b Random effect variance multiplied by 10⁻³

^c DIFF is the difference between ZCA and MSA average

Measurement error in air pollution epidemiology

Background

Exposure framework

Scientific understanding of exposure and its sources of variation is crucial for epidemiologic study design and inference. Often, the primary exposure of interest is total personal exposure for a specific time period. In the air pollution application, personal exposure can be partitioned into the ambient plus non-ambient sources, i.e., $E^P = E^A + E^N$, where ambient source is the product of ambient concentration and ambient attenuation ($E^A = C^A \times \alpha$) (Mage et al. 1999; Wilson et al. 2000; Allen et al. 2004). (Note that indices for individual (or location) and time can be incorporated into this notation.) Ambient concentration contributes to exposure both outdoors and indoors due to the infiltration of ambient pollution into indoor environments. The ambient exposure attenuation factor is

$$\alpha = [f_o + (1 - f_o)F_{inf}], \quad (1)$$

a weighted average of infiltration through the building filter (F_{inf}), weighted by time spent outdoors (f_o) (Wilson et al. 2000; Allen et al. 2004). In many air pollution epidemiology studies, the exposure of interest is ambient source (E^A) or total personal (E^P).

Air pollution exposure scales encompass temporal and spatial domains; the temporal and spatial scales of interest

vary by study design. Examples of temporal scales include lifetime, weekly, or daily averages. Spatial scales include regional and local where local can be defined at the level of a subject's residence. Individuals move around in space over time, and even without considering individual time-activity, spatio-temporal variation can be present across locations. Rich exposure data at all the spatial and temporal scales of interest are almost never available, and indeed, in most environmental epidemiology applications, exposure data are limited. Notwithstanding, relative to most environmental exposures, air pollution exposure data are extensive, thanks to the large amount of existing regulatory monitoring data. However, these data may not represent the locations or temporal and spatial scales of interest in a specific epidemiological study. Exposure models and simplified conceptualizations of exposure to individuals are often necessary in air pollution epidemiology applications.

Exposure models Even with rich data from regulatory monitoring networks, models are needed to predict individual exposures. Special data collection and modeling efforts are required for some components of individual exposure, specifically non-ambient source exposures, individual time-activity, and building- and season-specific infiltration. Even for ambient concentration, models are needed to predict concentrations at locations without monitors. Land use regression models are popular for predicting spatially varying concentrations measured over a fixed time period (e.g., Hoek et al. 2008). Spatio-temporal models are being developed (Szpiro et al. 2010; Lindström

et al. 2010; Yanosky et al. 2008; Paciorek et al. 2009) that explicitly acknowledge spatially varying trends in concentration data, can use both existing monitoring data and special sampling campaigns, and avoid oversimplification of data into spatial averages. Existing data from regulatory monitoring networks have inherent design features that can affect the model results because data availability is driven by regulation. For instance, ambient concentration data from regulatory monitors are rich in time (often with daily or hourly measurements), but they are collected at a very limited number of fixed locations. In addition, the monitor siting criteria are pollutant dependent—monitors are preferentially sited close to or far from sources depending upon the pollutant. Diggle et al. (2010) have shown biased geostatistical predictions from a class of preferentially sampled designs that oversample realizations of high (or low) exposure relative to the predictable surface. The implications of preferential sampling on health effect estimates has not received much attention in the literature although Szpiro and Sheppard (2010), in their comment on Diggle et al's paper, demonstrated by simulation that this class of preferential sampling does induce bias and uncertainty in the health effect estimate. Regulatory monitoring networks are an example of a different class of preferential sampling designs; it is possible this class may have similar consequences for health effect estimates.

Conceptual framework for measurement error in epidemiology

Exposure measurement error is a challenge in epidemiology because inference about health effects can be incorrect when the measured or predicted exposure used in the analysis is different from the underlying true exposure. Epidemiologic inference is based on estimating a regression parameter in a disease model that relates the exposure to the health outcome. While regression models naturally handle error in the outcome variable, they typically assume all the covariates, including the exposure, are fixed and known, i.e., measured without error. Measurement error can be differential when exposure measurements are related to the outcome or non-differential when they are not. Much measurement error research focuses on the impact of non-differential measurement error since differential errors can be minimized through the design and implementation of the study.

A general framework for non-differential measurement error in epidemiological studies proposed by Clayton (1992) is a useful foundation for conceptualizing specific applications. This framework has three sub-models: the exposure model to describe the distribution of exposure over space, time, and individuals; the measurement model to link exposure measurements to the underlying true exposure; and the disease model that specifies the association between exposure and the health outcome. Thomas (2009, p 223) presents this framework as a directed acyclic graph that highlights that all three

models rely on the unknown true exposure. Given that we do not observe the true exposure, all three sub-models are needed to obtain a health effect estimate with correct coverage.

As one example of this framework applied to air pollution cohort studies, Szpiro et al. (2011) describe a linear disease model for a continuous outcome Y as

$$Y = \beta_0 + X\beta_X + \mathbf{Z}\beta_Z + \varepsilon \quad (2)$$

where X is the unknown true exposure, \mathbf{Z} is vector of confounding and other adjustment variables, ε is the error, and β_X is the parameter of interest. The exposure and measurement models are described jointly using the geostatistical model

$$\begin{pmatrix} X \\ X^* \end{pmatrix} = \begin{pmatrix} S \\ S^* \end{pmatrix} \gamma + \begin{pmatrix} \eta \\ \eta^* \end{pmatrix} \quad (3)$$

where “*” represents locations with exposure measurements, (S^T, S^{*T}) are known predictors of exposure, γ are their unknown coefficients, and (η^T, η^{*T}) is assumed to have a normal distribution with mean 0, and variance that captures the residual spatial structure with parameter θ_η . Under a geostatistical model parameterization, θ_η has two variance parameters (partial sill and nugget) and a range parameter to capture the degree of spatial dependence. Given this joint model, the true exposures can be predicted given the measured exposures and the estimated parameters:

$$W = E\left(X \mid X^*; \hat{\gamma}, \hat{\theta}_\eta\right).$$

This example assumes that the true “exposure” is determined exclusively by spatially varying features and it does not incorporate any adjustments due to individual characteristics or behaviors.

Study designs In addition to the conceptual framework, the designs for the health and exposure studies must be considered, specifically the amount and details of the data available. Study design affects how exposure can be related to the health outcome in the analysis, how the data quantify the exposure distribution, and what information is available to assess these relationships. For instance, two contrasting study designs commonly used in air pollution epidemiology, time-series, and cohort studies rely on different sources of variation in exposure and address different scientific questions. Time-series studies use aggregated outcomes (population counts) and focus on temporal variability of exposure, typically relying on exposure data representing entire metropolitan areas (by, e.g., spatial aggregation of daily observations at multiple monitors). Cohort studies use individual-level data and focus on spatial variability of exposure. As we discuss in the next section, exposure is often predicted from a model or based on a proxy covariate such as distance from nearest

major road. There must be enough inherent variability in the aspect of the underlying exposure field being linked to the health data to make a study worthwhile. The dominant measurement error challenges depend upon the study design.

Measurement error impacts

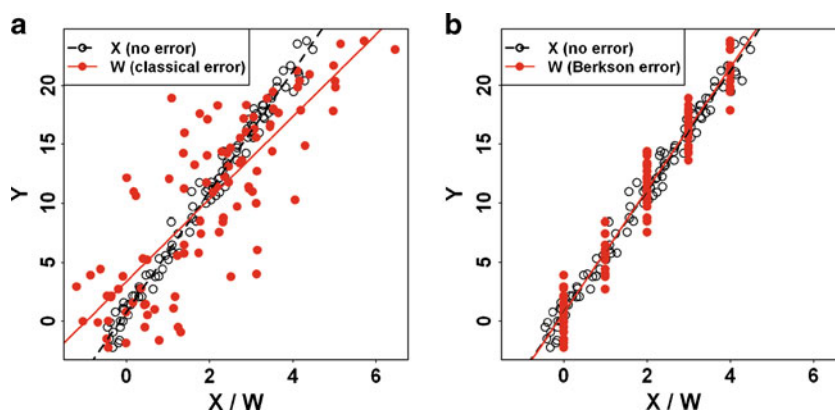
Impact of pure Berkson or classical measurement error For expository purposes, we frame the disease model as a linear regression model as in Eq. 2. Measurement error effects in generalized linear disease models commonly used in epidemiology are broadly similar to effects in the linear model, with more divergence as the generalized models become more nonlinear (Carroll et al. 2006; Buzas et al. 2005). In disease models, the error in the outcome is subsumed in the probability model; for the outcome in Eq. 2, the error is represented explicitly by ε . In routine regression, the disease model is conditional on the covariates, and measurement error in the exposure is not incorporated. There are two general classes of exposure measurement error models: Berkson and classical. Conceptually, we have Berkson measurement error when we measure part of the true exposure, while we have classical measurement error when the exposure measurement includes the true exposure plus noise. For additive and unbiased error models, Berkson error is defined as $X = W + U$ while classical error is defined as $W = X + U$ where X is the true exposure, W is the exposure measurement, and U is a mean zero random variable that is independent of W in the Berkson case and X in the classical case. The technical definitions of Berkson and classical error rely on the exposure measurement W being a surrogate. Specifically, W is a surrogate (or equivalently there is non-differential measurement error) when the distribution of the health outcome is the same when we condition on the true exposure X and other covariates Z as when we also condition on the measured exposure W . In the linear model setting with normally distributed random variables, this definition simplifies— W is a surrogate when it is not correlated with the disease model error ε . Figure 1 gives an example of a linear disease model relationship and the impact on the regression results when the exposure is measured with pure classical or Berkson measurement error. In this example, the linear relationship given the true exposure is shown by open black circles while the relationship given the measured exposure is given by solid red circles. The true effect β_X is 5; given the true exposure X , it is estimated to be 5.11 with a standard error of 0.066. When W is measured with classical measurement error, the effect estimate is attenuated at 3.5 with standard error 0.256. When W follows a Berkson error model, the effect estimate is unbiased with an estimate of 5.21, but with larger standard error of 0.122 than with the true exposure X . This example shows that both types of measurement error have an impact on health effect estimates,

where typically Berkson error leads to unbiased but more variable health effect estimates while classical error leads to biased and incorrect standard error estimates (these could be more or less variable leading to incorrect coverage of the 95% confidence intervals). Often the exposure measurement error structure will have features of both types; it is not uncommon for methods that address exposure measurement error to assume that only one type is present.

Air pollution health effect estimates from “plug-in” exposures Most air pollution epidemiology studies report estimates of health effects conditional on measured or predicted exposures without regard to how these exposure estimates were obtained. In other words, the reported health effect estimates do not account for uncertainties in the estimated exposures, i.e., exposure measurement error. Thus, many reported health effect estimates may have poor coverage properties due to bias and/or incorrect standard errors. Coverage is based on the sampling distribution of the health effect estimate and is defined to be the proportion of 95% confidence intervals that indeed cover the true value. Since inference about health effects is based on confidence intervals, poor coverage properties for effect estimates will lead to poor or misleading inference. We are just beginning to understand how various features of the underlying exposure distribution, exposure assignment/prediction approach, and study designs (for both exposure and health data, and including sample sizes) affect coverage.

Measurement error impact on time-series studies The time-series model estimates the association between daily event counts and daily average ambient concentration of a pollutant. The seminal paper by Zeger et al. (2000) presents a framework for measurement error in time-series studies. They presume that the ideal conceptual design is an individual-level study to estimate the effects of total personal exposure on a given day and develop the time-series design and measurement error framework from this underlying model. They aggregate an individual-level log-linear model for risk and partition the difference between the target individual exposure and the measured ambient concentration in the aggregated model into three components using a linear expansion of the exponential risk model. (The second-order terms, a possible source of specification bias, are dropped to simplify the presentation on the grounds that they are an order of magnitude smaller and inconsequential for studies of mortality.) The three exposure components are (1) the difference between risk-weighted and unweighted average personal exposure, (2) the difference between average personal exposure and true ambient concentration, and (3) the difference between true and measured ambient concentration. Zeger et al. (2000) argue that the first and third differences are likely to behave like Berkson measurement error and are thus unlikely

Fig. 1 Examples of a linear disease model relationship with the true exposure X overlaid with the empirical relationship given the mis-measured exposure W , measured with **a** classical or **b** Berkson measurement error



to induce bias in the model while the second could be a substantial source of bias. They suggest that a time-series study estimate will be biased toward the null because of correlation between this second term and measured ambient concentration.

In a later note, Zeger and Diggle (2001) present the exposure formulation given above and argue a time-series study estimate based on ambient concentration will differ from the estimate that would have been obtained from using average personal exposure by a factor that is approximately the temporal average of ambient attenuation (α , see Eq. 1). Similarly, Sheppard et al. (2005) conclude that the parameter in the time-series model differs from the one induced from the aggregated individual-level model conditional on personal exposure. In simulation studies, they show that the time-series study health effect parameter estimate is scaled by the average ambient attenuation parameter (α) with sensitivity to its seasonal distribution and temporal association with ambient concentration. They also show that non-ambient source exposures do not affect time-series study results when non-ambient exposures are not correlated with ambient concentration. However, substituting a measure of average personal exposure from a sample of the population does induce bias because of the additional non-ambient exposure variability in the sample; unless the entire population is sampled, this behaves like classical measurement error. Given that concentration is the exposure surrogate in time-series studies, these results suggest that one source of differences between health effect estimates in different cities is variations in population exposures. This is supported by Jansen et al.'s (2002) results showing city-specific estimates of PM_{10} effects on CVD, COPD, and pneumonia hospital admissions vary by the use of air conditioning as well as whether PM_{10} was highest in winter or summer (see Fig. 1 of that paper).

More recent papers (Sarnat et al. 2010; Peng and Bell 2010) focus on the third term in Zeger et al.'s decomposition and consider spatial misalignment in time-series studies with respect to monitor locations and spatial

heterogeneity of pollutants. Sarnat et al. (2010) compared time-series study effect estimates by monitor location; they found similar estimates for spatially homogenous pollutants and discrepancy for spatially heterogeneous pollutants in urban versus rural monitors. Peng and Bell (2010) evaluated regression calibration and two-stage Bayesian approaches to correcting for the misalignment error for application to counties with limited monitoring data. Their approach relies on conditioning on a fitted spatio-temporal concentration model and assuming a classical measurement error model for the spatial misalignment. Given the available data, these papers only focus on one component of the error decomposition.

Measurement error impact on cohort studies The cohort study disease model relates individual exposure to individual disease outcomes. Typically unknown exposures are predicted using a model, and predicted exposures are “plugged in” and treated as known in the analysis. Approaches to spatial prediction have varied across cohort studies with the earliest studies relying on city-wide averages (Dockery et al. 1993; Pope et al. 2002). More recently studies have used nearest monitor interpolation (e.g., Miller et al. 2007), land use regression (e.g., Brauer et al. 2003), geostatistical methods such as universal kriging (e.g., Jerrett et al. 2005a), and semi-parametric smoothing (e.g., Puett et al. 2008). Exposure data are “spatially misaligned” in the cohort study setting when exposure data are only available at monitoring sites and not at subject locations (see, e.g., Banerjee et al. 2004). Measurement error resulting from spatially misaligned data has only recently begun to receive attention in the literature. We discuss several recent results below.

A recent set of simulation studies showed that health effect estimates from cohort studies that ignore exposure uncertainty give less than the nominal coverage (Kim et al. 2009). This paper used relative risk estimates and other model details from the Women's Health Initiative observational cohort (Miller et al. 2007) as a foundation for

simulating health outcome and exposure data in a single city. The authors assumed the exposures followed geostatistical models with varying degrees of spatial dependence captured in the mean and covariance models. Realizations from Eq. 3 were assumed to be the “true” exposures. For the health analysis, the “true” exposures were observed only at the 22 monitor locations and exposures at subject residences were predicted conditional on the monitored data. Two exposure prediction approaches were considered: nearest monitor and kriging. Kim et al. (2009) found that estimated subject exposures were more predictable when the underlying exposure distribution had large-scale spatial structure (as parameterized by a larger range parameter) and these predictions gave better health effect estimates (i.e., closer to nominal coverage), even without explicit acknowledgment of the exposure measurement error in the health analysis (see Table 2). In contrast, predictions of exposures with shorter range explained less exposure variability and produced poorer health effect estimates with more bias and incorrect standard errors. It is tempting to conclude from this work that good exposure predictions will give good health effect estimates even without measurement error correction, but these simulations only assessed a limited number of conditions and it is premature to draw such a general conclusion.

Szpiro et al. (2011) develop a conceptual framework for measurement error for spatially misaligned data. They show measurement error in this scenario can be decomposed into two components: Berkson-like and classical-like. The Berkson-like component results from smoothing the exposure surface in the prediction model. Berkson-like measurement error can be viewed as the part of the true exposure not captured by the model and its behavior is similar to standard Berkson error (Carroll et al. 2006). The term “like” refers to the fact that the errors in this model are correlated in space and that the standard Berkson model is based on treating the predicted exposures as fixed; in this context, that is essentially equivalent to treating the monitoring data as fixed. The classical-like measurement error is due to the uncertainty of estimating the parameters in the exposure prediction model. Similar to classical measurement error, this uncertainty adds variability to the predicted exposures and can induce bias in the health effect estimates. The term “like” indicates this error is not purely classical because the additional variability is shared across all locations and is not independent.

Both Gryparis et al. (2009) and Szpiro et al. (2011) discuss methods for measurement error correction in the presence of spatially misaligned data. Gryparis et al. (2009) demonstrate the pitfalls with the intuitively simplest approach—exposure simulation. Here, the exposure is simulated multiple times in order to account for exposure variability not captured by the predicted exposures. One

generates multiple realizations from the estimated exposure distribution, plugs each set of these in turn into the disease model and estimates parameters, and then averages the resulting estimates and fixes the variance using the formula developed for multiple imputation. Exposure simulation differs from multiple imputation because the disease outcome is left out of the imputation. Because exposure simulation includes only the exposure and not the disease outcome in the multiple imputation procedure, the resulting health effect estimates are biased (Gryparis et al. 2009; Little 1992).

Gryparis et al. (2009) compare four correction approaches to the plug-in approach of conditioning on the predicted exposures: exposure simulation, out of sample regression calibration, a joint model using a Bayesian estimation approach, and a two-stage Bayesian approach to separately estimate parameters in the exposure and health models. Use of predicted exposures from a model is already a form of regression calibration; the regression calibration approach used by Gryparis et al. attempts to correct for the potential classical measurement error bias in the predicted exposures. Similar to the findings of Kim et al. (2009), for the linear health model, the performance of the plug-in estimate of exposure worsened as the underlying exposure surface became less predictable (described by Gryparis et al. as more rough). Exposure simulation performed worse than the plug-in exposure, with greater discrepancy between these results with the rougher underlying surfaces. Regression calibration gave larger estimated standard errors than any of the procedures but did remove the attenuation bias in the plug-in estimate. The fully Bayesian and two-stage Bayesian approaches had the least bias and the smallest standard errors of any of the correction approaches. However, they did not perform as consistently well as one would expect, particularly given results in one scenario that were far from the nominal 95% coverage.

Szpiro et al. (2011) develop two-stage correction approaches based on the bootstrap. These involve using the predicted exposures in the health model but with bootstrap resampling to correct for the bias and uncertainty resulting from using the predicted exposures. The general parametric bootstrap approach is to initially estimate the parameters from the exposure and health models given the data, simulate new exposure and health data given these parameters and the data-generating models, re-estimate new predicted exposures and parameters given the bootstrap sample, and use these to obtain the bootstrap estimate and its standard error. Szpiro et al.’s implementation of the parametric bootstrap (Davison and Hinkley 1997) is computationally intensive because it involves complete estimation of all the exposure and health model parameters in each bootstrap iteration. They also present the parameter bootstrap, a less computation-

Table 2 Example simulation results based on Kim et al. (2009) showing that in models conditioning on the exposure estimate, health effect estimate properties vary by the predictability of the underlying exposure surface as well as the approach to exposure prediction

True Exposure	Fitted Exposure (R^2)	Bias ²	Variance	Mean Square Error	Coverage probability of 95% confidence intervals	
Least predictable (shortest range)	True	0	9	9	0.95	
	Nearest Kriging (0)	327	23	350	0.03	
		342	778	1,120	0.58	
	Nearest Kriging (0.20)	True	0	31	31	0.95
		33	58	91	0.76	
	Nearest Kriging (0.40)	1	734	735	0.74	
True		0	69	69	0.95	
Most predictable (longest range)	Nearest Kriging (0.47)	30	125	155	0.87	
		1	426	427	0.89	
	True	0	56	56	0.96	
		34	105	139	0.85	
	Nearest Kriging (0.47)	0	153	153	0.92	

ally intensive alternative that avoids the computationally costly step of re-estimating the exposure model parameters in each bootstrap sample by relying on the original estimated distribution for the exposure parameters. In simulations, the bootstrap corrections give nominal coverage, even in scenarios where the plug-in approach gives confidence intervals with poor coverage.

The asymptotically optimal approach to correcting for exposure measurement error is to use a joint model to estimate the exposure and disease model parameters. Very few published examples exist, but Molitor et al. (2007) fit a joint model for the effect of NO₂ exposure on lung function (FVC) based on a conditional autoregressive (CAR) model. They focused on the changes in health effect estimates from incorporating spatial structure at both the inter- and intra-community levels in both the exposure and the disease models as compared to models without this spatial structure. They found that models with spatial structure gave effect estimates that were 5–30% smaller and had narrower confidence intervals. Their spatial model estimates revealed that most of the variability in the exposure models was spatial while only a fraction of the overall variability in the disease models was spatial. Since they did not show results for models that handled the disease and exposure models separately or for models with spatial structure in only one of these models, it is difficult to identify the primary factors causing the change in estimates.

In general, there are several practical problems with the joint modeling approach. One important consideration is computational feasibility. Some state of the art exposure prediction models are computationally very intensive, such as the spatio-temporal model being developed for the MESA Air study (Lindström et al. 2010; Szpiro et al. 2010), and thus it is impractical to jointly fit exposure and

disease models. Interestingly, while some published simulation examples based on a joint model exhibit good properties (Szpiro et al. 2011), others have not behaved as well as one would expect from an asymptotically optimal approach (Gryparis et al. 2009; Madsen et al. 2008), suggesting there may be additional computational challenges that can affect results. Another consideration is the potential for feedback between the exposure and health models that may impact the health effect estimates and possibly give misleading inference. Health models are descriptive and most likely misspecified by their very nature. In most applications, the exposure data are much sparser than the health data; this data imbalance along with a joint model that allows feedback from the health data to the concentration data may influence the exposure estimation. This in turn may result in a poorly estimated exposure surface and distortion of the health effects (Bennett and Wakefield 2001; Wakefield and Shaddick 2006). However, the contribution of air pollution to overall disease risk is typically small; this is likely to limit the impact of disease model misspecification on the exposure predictions. The importance of and implications for feedback in epidemiological studies is an important research topic.

Discussion

Confounding

Long-term exposure to ambient air pollution is just one of several known risk factors for cardiovascular disease and mortality (Brook et al. 2010). The quantitative assessment of this association requires that information on other known mortality risk factors, such as smoking and diet, be

collected and related simultaneously to mortality as they may be potential confounders. In addition to risk factors measured the individual level, an assessment of the environment that the subject lives, works, and plays is required to more fully understand how mortality varies in space. Ecological or grouped information may be used to predict mortality effects of the contextual risk factors that represent the living environment. This type of risk factor can play an important role in explaining spatial variation in mortality because both ambient air pollution and contextual risk factors intrinsically vary in space.

Spatial patterns in mortality can persist after adjusting for individual risk factors, contextual risk factors, and air pollution. This residual spatial pattern induces dependencies in the data that will not be accounted for by the standard Cox survival model. There is no clear means of modeling possibly complex spatial patterns and several recent papers have shown that the spatial scale of the residual dependencies and their relationship with predictors of interest can affect the bias and precision of the air pollution risk estimates (Paciorek 2010; Hodges and Reich 2010). We have suggested one possible model in which the spatial mortality pattern is characterized by multiple levels of location of each subject, such as the community and neighborhood that subjects live in and spatial dependencies of these locations. Our extension of the Cox proportional hazards survival model to include location-based random effects does allow for an assessment of the nature of spatial dependency in mortality. However, this model is dependent on the definition of the random effects and their hypothesized spatial interdependence. In this example, we hypothesized that adjacent ZCAs within an MSA were correlated in addition to adjacent MSAs within the entire USA. More complex spatial dependencies could be considered such as higher-order nearest neighbors and distance-decay structures.

Although analyses of the ACS cohort have supported the hypothesis that long-term exposure to fine particulate matter is a positive risk factor for cardiovascular mortality, quantitative estimates of the association are somewhat variable depending on the statistical model employed and the risk factors that are included for adjustment. For example, the $PM_{2.5}$ -IHD mortality association was sensitive to the adjustment for ecological covariates and the error specification of the spatial random effects survival model with hazard ratios ranging from 1.153 to 1.320 (Table 1). The standard error of the $PM_{2.5}$ association with IHD deaths also was sensitive to the model specification with variation in the standard error ranging over a threefold span.

We recommend that alternatives to the standard Cox proportional hazards be investigated both in terms of the proportional hazards assumption and the stochastic structure of the survival data. Since air pollution exposure is intrinsi-

cally linked to space, a proper assessment of its effects on health should not be made without consideration of spatial dependencies.

Exposure measurement error

The implications of exposure measurement error for health effect estimates will differ by application. In general, for predictions from spatially misaligned exposure data, the measurement error structure is complex and not purely classical or Berkson. Exposure predictions are estimates from a model; the uncertainty in estimating the parameters for this model induces classical-like measurement error. The importance of the classical-like contribution to measurement error does not diminish with the size of the health study. In contrast, Berkson-like measurement error results from predictions being smoother than the underlying surface. While Berkson-like error limits the information available about exposure (by reducing exposure variability) and thus results in wider confidence intervals for the health effect estimates, its impact decreases with the size of the health study. Furthermore, in linear models, Berkson-like error does not bias the exposure–response relationship. Research is needed to understand how generally this property extends to nonlinear disease models such as logistic and survival models.

Exposure assessment for epidemiology should be evaluated in the context of the health effect estimation goal. It is important to design the exposure assessment to capture the underlying exposure variability for the pollutants of interest, obtain exposure data that are directly relevant to the study population (e.g., representative of residence locations), and ensure there are sufficient exposure data to support good predictions. Research is needed to identify optimal sampling designs for exposure assessment in epidemiology studies.

Many factors contribute to the quality of epidemiological study inference. In addition to the quality of the exposure estimates, it is important to consider study design, ability to control for confounding, data structure, and distribution of the underlying exposure as it relates to both the study design and data structure. It is not well understood how these factors combine to produce reliable health effect estimates. It is tempting to believe that inference in the presence of exposure measurement error will improve with better exposure predictions, even without correction for the measurement error. However, more research is needed to understand the generality of this belief.

Acknowledgments The research was supported by a grant from the California Air Resources Board (grant no. 55245A) and the Health Effects Institute. Support for this research has also been provided by US EPA through assistance agreement CR-834077101-0 and grant RD831697, Health Effects Institute through agreement 4749-RFA05-1A/06-10, and NIEHS through R01-ES009441 and P50-ES015915.

Although research described in this article has been funded in part by the United States Environmental Protection Agency, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Abrahamowicz M, Schopflocher T, Leffondré K et al (2003) Flexible modeling of exposure-response relationship between long-term average levels of particulate air pollution and mortality in the American Cancer Society Study. *J Toxicol Environ Health A* 66:1625–1654
- Allen R, Wallace L, Larson T, Sheppard T, Liu LJS (2004) Estimated hourly personal exposures to ambient and non-ambient particulate matter among sensitive populations in Seattle, Washington. *J Air Waste Manage Assoc* 54:1197–1211
- Banerjee S, Carlin BP, Gelfand AE (2004) Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, Boca Raton
- Beelen R, Hoek G, van den Brandt PA et al (2008) Long-term effects of traffic-related air pollution on mortality in a Dutch cohort (NLCS-AIR study). *Environ Health Perspect* 116:196–202
- Bennett J, Wakefield J (2001) Errors-in-variables in joint population pharmacokinetic/pharmacodynamic modeling. *Biometrics* 57:803–812
- Brauer M, Hoek G, van Vliet P, Meliefste K, Fischer P, Gehring U, Heinrich J, Cyrys J, Bellander T, Lewne M, Brunekreef B (2003) Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. *Epidemiology* 14:228–239
- Burnett RT, Dewanji A, Dominic F et al (2003) On the relationship between time series studies, dynamic population studies, and estimating life lost due to short-term exposure to environmental risks. *Environ Health Perspect* 111:1170–1174
- Buzas JS, Stefanski LA, Tosteson TD (2005) Measurement error. In: Ahrens W, Pigeot I (eds) *Handbook of epidemiology*. Springer, New York, Also available online http://www4.stat.ncsu.edu/~stefansk/mimeo_series_mem_version_master_file.pdf
- Carroll RJ, Ruppert D, Stefanski LS, Crainiceanu CM (2006) *Measurement error in nonlinear models: a modern perspective*, 2nd edn. Chapman and Hall/CRC, Boca Raton
- Clayton DG (1992) Models for the longitudinal analysis of cohort and case-control studies with inaccurately measured exposures. In: Dwyer JH, Lippert P, Feinleib M, Hoffmeister H (eds) *Statistical models for longitudinal studies of health*. Oxford University Press, Oxford, pp 301–331
- Davison AC, Hinkley DV (1997) *Bootstrap methods and their applications*. University of Cambridge Press, London
- Diggle PJ, Menezes R, Su T (2010) Geostatistical inference under preferential sampling. *Appl Stat* 59:191–232
- Dockery DW, Pope CA III, Xu X et al (1993) An association between air pollution and mortality in six U.S. cities. *N Engl J Med* 329:1753–1759
- Dominici F, Daniels M, McDermott A, Zeger SL, Samet J (2003) Mortality among residents of 90 cities. In: *Revised analyses of time-series of air pollution and health*. Special Report. Health Effects Institute, Boston, pp 9–24
- Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull BA (2009) Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics* 10:258–274
- Health Effects Institute (2000) *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality: a special report of the Institute's Particle Epidemiology Reanalysis Project*. Health Effects Institute, Cambridge
- Hodges JS, Reich BJ (2010) Adding spatially-correlated errors can mess up the fixed effect you love. *Research Report 2010-002*, Division of Biostatistics, University of Minnesota. Submitted to *The American Statistician*
- Hoek G, Beelen R, de Hoogh K, Gulliver J, Fischer P, Briggs DJ (2008) A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos Environ* 42:7561–7578
- Janssen NAH, Schwartz J, Zanobetti A, Suh HH (2002) Air conditioning and source-specific particles as modifiers of the effect of PM10 on hospital admissions for heart and lung disease. *Environ Health Perspect* 110:43–49
- Jerrett M (2010) Spatial modeling in environmental and public health research. *Int J Environ Res Publ Health* 7:1302–1329
- Jerrett M, Burnett RT, Ma R et al (2005a) Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 16:727–736
- Jerrett M, Buzzelli M, Burnett RT, DeLuca PF (2005b) Particulate air pollution, social confounders, and mortality in small areas of an industrial city. *Soc Sci Med* 60:2845–2863
- Jerrett M, Burnett RT, Pope CA III et al (2009) Long-term ozone exposure and mortality. *N Engl J Med* 360:1085–1095
- Kim SY, Sheppard L, Kim H (2009) Health effects of long-term air pollution: influence of exposure prediction methods. *Epidemiology* 20:442–450
- Krewski D, Jerrett M, Burnett RT et al (2009) *Extended analysis of the American Cancer Society Study of Particulate Air Pollution and Mortality*. Research report 140. Health Effects Institute, Boston
- Lindström J, Szpiro AA, Sampson PD, Sheppard L, Oron A, Richards M, Larson T (2010) A flexible spatio-temporal model for air pollution: Allowing for spatio-temporal covariates (January 19, 2011). UW Biostatistics Working Paper Series. Working Paper 370. <http://www.bepress.com/uwbiostat/paper370>
- Little RJA (1992) Regression with missing X's: a review. *J Am Stat Assoc* 87:1227–1237
- Ma R, Krewski D, Burnett R (2003) Random effects Cox models: a Poisson modeling approach. *Biometrika* 90:157–169
- Madsen L, Ruppert D, Altman NS (2008) Regression with spatially misaligned data. *Environmetrics* 19:453–467
- Mage D, Wilson W, Hasselblad V, Grant L (1999) Assessment of human exposure to ambient particulate matter. *J Air Waste Manage Assoc* 49:1280–1291
- Miller KA, Siscovick DS, Sheppard L et al (2007) Long-term exposure to air pollution and incidence of cardiovascular events in women. *N Engl J Med* 356:447–458
- Molitor J, Jerrett M, Chang CC, Molitor NT, Gauderman J, Berhane K, McConnell R, Lummann F, Wu J, Winer A, Thomas D (2007) Assessing uncertainty in spatial exposure models for air pollution health effects assessment. *Environ Health Perspect* 115:1147–1153
- Paciorek CJ (2010) The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Stat Sci* 25:107–125
- Paciorek CJ, Yanosky JD, Puett RC, Laden F, Suh HH (2009) Practical large-scale spatio-temporal modeling of particulate matter concentrations. *Ann Appl Stat* 3:370–397
- Peng RD, Bell ML (2010) Spatial misalignment in time series studies of air pollution and health data. *Biostatistics* 11:720–740
- Pope CA, Burnett RT, Thun MJ, Calle EE, Ito K, Krewski D, Thurston GD (2002) Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* 287:1132–1141

- Puett RC, Schwartz J, Hart JE, Yanosky JD, Speizer FE, Suh H, Paciorek CJ, Neas LM, Laden F (2008) Chronic particulate exposure, mortality, and coronary heart disease in the Nurses' Health Study. *Am J Epidemiol* 168:1161–1168
- Sarnat SE, Klein M, Sarnat JA, Flanders WD, Waller LA, Mulholland JA, Russell AG, Tolbert PE (2010) An examination of exposure measurement error from air pollutant spatial variability in time-series studies. *J Expo Sci Environ Epidemiol* 20:135–146
- Sheppard L, Slaughter JC, Schildcrout JS, Liu L-JS, Lumley T (2005) Exposure and measurement contributions to estimates of acute air pollution effects. *J Expo Sci Environ Epidemiol* 15:366–376
- Szpiro AA, Sheppard L (2010) Comment on "Diggle PJ, Menezes R, Su T (2010). Geostatistical inference under preferential sampling.". *Applied Statistics* 59(Part 2):224–226
- Szpiro AA, Sampson PD, Sheppard L, Lumley T, Adar SD, Kaufman JD (2010) Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. *Environmetrics* 21:606–631
- Szpiro AA, Sheppard L, Lumley T (2011) Efficient measurement error correction with spatially misaligned data. *Biostatistics* (in press) (published online 20 Jan, 2011)
- Thomas DC (2009) *Statistical methods in environmental epidemiology*. Oxford University Press, Oxford
- Wakefield J, Shaddick G (2006) Health-exposure modeling and the ecological fallacy. *Biostatistics* 7:438–455
- Wilson W, Mage DT, Grant LD (2000) Estimating separately personal exposure to ambient and non-ambient particulate matter for epidemiology and risk assessment: why and how. *J Air Waste Manage Assoc* 50:1167–1183
- Yanosky JD, Paciorek CJ, Schwartz J, Laden F, Puett R, Suh HH (2008) Spatio-temporal modeling of chronic PM10 exposure for the Nurses' Health Study. *Atmos Environ* 42:4047–4062
- Zeger SL, Diggle PJ (2001) Correction: exposure measurement error in time-series air pollution studies. *Environ Health Perspect* 109:A517
- Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, Cohen A (2000) Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ Health Perspect* 108:419–426
- Zeger SL, Dominici F, McDermott A, Samet JM (2008) Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000–2005). *Environ Health Perspect* 116:1614–1619