Biostatistics (2003), **4**, 4, *pp.* 539–553
Printed in Great Britain

Control for confounding in the presence of measurement error in hierarchical models

JOEL SCHWARTZ†

Department of Environmental Health, Harvard School of Public Health, Suite 415-L, 401 Park Drive West, Boston, Massachusetts 02115, USA
jschwrtz@hsph.harvard.edu

BRENT A. COULL

Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, Massachusetts 02115, USA

SUMMARY

Hierarchical modeling is becoming increasingly popular in epidemiology, particularly in air pollution studies. When potential confounding exists, a multilevel model yields better power to assess the independent effects of each predictor by gathering evidence across many sub-studies. If the predictors are measured with unknown error, bias can be expected in the individual substudies, and in the combined estimates of the second-stage model. We consider two alternative methods for estimating the independent effects of two predictors in a hierarchical model. We show both analytically and via simulation that one of these gives essentially unbiased estimates even in the presence of measurement error, at the price of a moderate reduction in power. The second avoids the potential for upward bias, at the price of a smaller reduction in power. Since measurement error is endemic in epidemiology, these approaches hold considerable potential. We illustrate the two methods by applying them to two air pollution studies. In the first, we re-analyze published data to show that the estimated effect of fine particles on daily deaths, independent of coarse particles, was downwardly biased by measurement error in the original analysis. The estimated effect of coarse particles becomes more protective using the new estimates. In the second example, we use published data on the association between airborne particles and daily deaths in 10 US cities to estimate the effect of gaseous air pollutants on daily deaths. The resulting effect size estimates were very small and the confidence intervals included zero.

Keywords: Air pollution; Confounding; Hierarchical models; Measurement error.

1. INTRODUCTION

A major concern of epidemiologists is whether an observed association between one exposure and an outcome is due, all or in part, to the correlation between that exposure and a second exposure. If the most likely candidates for the confounding exposures are known, the best approach is to measure and control for them. In many cases, there are two correlated exposures of interest, or one exposure of interest and one principal confounder. We focus on this scenario as a simplification of the general case of multiple confounders, but also report the sensitivity of the conclusions to cases with multiple confounders.

†To whom correspondence should be addressed

Biostatistics 4(4) © Oxford University Press; all rights reserved.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

Under the standard modeling assumptions, the results of the multiple regression will give unbiased estimates of the effects of both exposures. Three events tend to spoil this happy scenario. First, it may not be possible to measure the confounding exposure. Second, the correlation between the two exposures may be high enough to yield relatively imprecise estimates of each. For example, in air pollution studies, the correlation between concentrations of airborne particles and carbon monoxide often exceeds 0.80. Finally, both of the exposures are almost certainly measured with error. While measurement error generally biases effect estimates toward the null, the effect of one variable may be partially transferred to the better measured one, resulting in an upward bias if there are substantial differences between the measurement errors in the two exposures (Zeger *et al.* 2000). In many cases, interest focuses on combining evidence across multiple studies or samples in a hierarchical model. In the presence of measurement error, bias will remain in this hierarchical setting, including the risk of upward bias. The latter is of special concern in a regulatory setting, or for risk assessment. Regulatory decisions may be based on estimated effect sizes, and while governments are often willing to make decisions in the presence of uncertainty, bias is another matter. Some assurance that effect sizes are not overstated may be critical. In the hierarchical model setting, it is possible to obtain two alternative estimates of the unconfounded effect of each exposure using hierarchical modeling. This paper describes those two alternatives and explores the performance of the standard method and the alternative methods in the presence of measurement error both analytically and via simulation. Finally, we apply the proposed approaches to two examples of current interest. One involves the relative importance of particles of different size in the association between airborne particles and daily deaths. The second examines the association between daily deaths and gaseous air pollution.

2. A MOTIVATING EXAMPLE

Particulate air pollution has been associated with daily deaths in many studies over the past decade (Schwartz and Dockery, 1992; Fairley, 1992; Schwartz ,1993; Pope *et al.*, 1995). These studies have used Poisson regression to model the dependence of daily death counts on air pollution and other factors, typically using several thousand days of data. Recently, attention has focused on hierarchical models of multiple cities [e.g. six cities (Schwartz *et al.*, 1996), 12 cities (Katsouyanni *et al.*, 1997), 10 cities (Schwartz, 2000), 20 cities (Daniels *et al.*, 2000; Samet *et al.*, 2000)] which allow examination of sources of heterogeneity. For example, Samet and co-workers (2000) examined the effect of daily variations in PM_{10} (particles less than 10 μm in diameter) concentrations and daily deaths in 20 cities. Those associations were shown to be independent of SO_2, CO, NO_2, and O_3 concentrations in multipollutant models. In contrast, no significant associations were seen with any of the gases. While such studies strongly support the conclusion that a causal independent effect of PM_{10} exists, Lipfert and Wyzga (1997) have argued that if one pollutant is measured with substantially more error than another, the better-measured pollutant could capture the effect of the more poorly measured one. The goal of this paper is to develop an estimate of the independent effect of PM_{10} and gaseous air pollutants that is free of upward bias, or ideally unbiased, in the presence of measurement error.

3. MODELS

Hierarchical modeling describes analyses for data that have several levels. In the air pollution setting, consider a set of 10 studies examining the association between an outcome of interest and two exposures of interest. In the following, we use a two-stage approach to estimate overall associations. Dominici *et al.* (2000) noted that such an approach works well given that the sample size for each study is large, which is typically the case in multi-city epidemiological mortality studies. We first consider the simple setting in which associations of interest are homogeneous across study before moving to the heterogeneous case.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

3.1 *Homogeneous effects model*

For day t, $t = 1, \ldots, T_i$, in city i, $i = 1, \ldots, N$, consider the true model

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \epsilon_{it} \tag{3.1}$$

where

$$\mathbf{X}_{it} = (X_{1it}, X_{2it})^{\mathsf{T}} \sim N\left(\boldsymbol{\mu}_{x,i}, \begin{bmatrix} \sigma_{x1}^2 & \sigma_{x1}\sigma_{x2}\rho_{x,i} \\ \sigma_{x1}\sigma_{x2}\rho_{x,i} & \sigma_{x2}^2 \end{bmatrix}\right), \tag{3.2}$$

and $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ and is independent of $\mathbf{X}_{it}$. If we measure $\mathbf{X}_{it}$ exactly, then fitting the model

$$X_{2it} = \gamma_{0i} + \gamma_{1i} X_{1it} + \eta_{it}, \tag{3.3}$$

yields

$$E(\widehat{\gamma}_{1i}) = \frac{\sigma_{x2}\rho_{x,i}}{\sigma_{x1}} = \gamma_{1i}.$$

Now, suppose X_{hit} is measured with error. Denote $Z_{hit} = X_{hit} + U_{hit}$, $h = 1, 2$, with

$$\mathbf{U}_{it} = (U_{1it}, U_{2it})^{\mathsf{T}} \sim N\left(\mathbf{0}, \begin{bmatrix} \sigma_{u1}^2 & \sigma_{u1}\sigma_{u2}\rho_{u,i} \\ \sigma_{u1}\sigma_{u2}\rho_{u,i} & \sigma_{u2}^2 \end{bmatrix}\right).$$

Consider first the setting in which the measurement error for the two exposures is independent; that is, $\rho_{u,i} = 0$ for all i. Suppose we fit the model

$$Z_{2it} = \gamma_{0i}^* + \gamma_{1i}^* Z_{1it} + \eta_{it}^*$$

separately for each city. Then, from the distributional assumptions on $\mathbf{X}_{it}$, $\mathbf{Z}_{it}$,and $\mathbf{U}_{it}$, we have

$$E(\widehat{\gamma}_{1i}^*) = \frac{\sigma_{x1}\sigma_{x2}\rho_{x,i}}{\sigma_{x1}^2 + \sigma_{u1}^2} = \frac{\sigma_{x1}^2}{\sigma_{x1}^2 + \sigma_{u1}^2}\gamma_{1i} = \alpha\gamma_{1i},$$

where α is typically termed the attenuation coefficient. Now consider the second model

$$Y_{it} = \delta_{0i}^* + \delta_{1i}^* Z_{1it} + \zeta_{it}^*.$$

Due to the measurement error and the exclusion of Z_2, we have

$$E(\widehat{\delta}_{1i}^*) = \frac{\sigma_{x1}^2}{\sigma_{x1}^2 + \sigma_{u1}^2}(\beta_1 + \beta_2\gamma_{1i}) = \alpha(\beta_1 + \beta_2\gamma_{1i}).$$

Now suppose we know the true $\gamma_{1i}^* = \alpha\gamma_{1i}$, instead of the estimate $\widehat{\gamma}_{1i}^*$. Then we have a random variable $\widehat{\delta}_{1i}^*$ with mean $\alpha\beta_1 + \beta_2(\alpha\gamma_{1i}) = c_0 + \beta_2(\alpha\gamma_{1i})$, a linear function of variable $\alpha\gamma_{1i}$. Thus, we can estimate c_0 and β_2 using ordinary least squares, yielding an attentuated estimate of β_1 but an unbiased estimate of β_2 in the presence of independent measurement error $\mathbf{U}$. This development suggests several additional considerations:

1. Although we use the multivariate normal formulation to make our estimation strategy clear, normality is not required as long as linear models (3.1) and (3.3) hold. For instance, Lyles and Kupper (1997) noted that, when a linear model contains a lognormal covariate measured with multiplicative error, ordinary least squares estimation yields similarly attenuated slope estimates.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

2. In practice, we have also estimated γ_{1i}^* for each city. When the sample sizes become large within each study, which is often the case in the air pollution context, this error will be neglible and the estimate of β_2 is unbiased. Note that this is in contrast to fitting the full model with Z_{1it} and Z_{2it}, as the estimator of β_2 is biased regardless of the sample sizes for each city. Simulation results presented in Section 4 suggest that even though each γ_{1i}^* is estimated, this two-stage approach outperforms the full model estimator.
3. In order for this two-step approach to work, variation in the γ_{1i}^* must be present. Otherwise, the approach is analogous to a regression with no variation in the explanatory variable. This of course precludes the artificial use of this approach that divides a single sample into multiple 'locations'.
4. The proposed approach requires the attentuation factor, α, to be constant across locations. The above development satisfies this assumption since the model assumes σ_{x1}^2, σ_{x2}^2, σ_{u1}^2, and σ_{u2}^2 are constant across locations. An alternative development that also satisfies this assumption is the case in which the measurement error variances for a given city are a fixed proportion of the true exposure variance. We use this alternative assumption in the simulations of Section 4.
5. As long as the individual study sample sizes are sufficiently large and there is heterogeneity in the linear relationship between X_{1it} and X_{2it}, we have shown that this meta-regression approach works when the effects of interest β_1 and β_2 do not vary across studies. We relax this assumption in the next section.
6. The above development substitutes the expression of X_2 that is a function of X_1. We can also reverse this substitution, obtaining an estimate of β_1 by expressing X_1 as a function of X_2 in (3.1). We refer to these as slope estimates of β_1 and β_2, respectively. In addition, the estimated intercepts $\widehat{c}_0$ yield attentuated estimates of β_1 and β_2, with these estimates attentuated by α. Thus, this use of hierarchical modeling yields three possible estimators of the independent effects of each of the two exposures: the standard (two-exposure model) estimator, the intercept estimator, and the slope estimator. When measurement error does not exist, the standard estimator will be more efficient. However, this estimator incurs bias in the presence of measurement error.

We also note that Marcus and Kegler (2001) criticized the proposed meta-regression approach on the basis that it yields biased estimates when one omits an important confounder from the model. However, this is true of most modeling strategies; an important assumption required for the validity of the resulting estimates is that the model is correctly specified. For instance, the approach based on the full model also produces biased estimates when an important confounder is omitted from the model. If interest focuses on additional confounders, the above development extends naturally beyond the two exposure setting. Suppose the full model consists of exposures X_1, X_2, and X_3. The first stage regresses Y, X_2, and X_3 on X_1 for each city. The second stage consists of a multiple regression using the resulting Y coefficients as the response and the X_2 and X_3 coefficients as covariates. In this setting, the slope estimates are not unique. For instance, one could also obtain an unbiased slope estimate of β_2 using X_3 as the covariate in the first stage of this approach.

3.2 *Correlated measurement error*

Now consider model (3.1) with correlated measurement errors, so that $\rho_{u,i} \neq 0$. The expected value of the slope in the regression model

$$Z_{2it} = \gamma_{0i}^* + \gamma_{1i}^* Z_{1it} + \eta_{it}^*$$

is now

$$E(\gamma_{1i}^*) = \alpha\gamma_{1i} + \frac{\rho_{u,i}\sigma_{u1}\sigma_{u2}}{\sigma_{x1}^2 + \sigma_{u1}^2}. \tag{3.4}$$

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

Table 1. *Bias for the second coefficient (β_2) from the two-stage estimator based on the full model and the proposed slope estimator under the homogeneous effects model (3.1) in the presence of measurement error correlation. In all cases, $\beta_2 = 1.0$*

$\sigma_{u,1}$	$\sigma_{u,2}$	Range for $\rho_{x,i}$	Range for $\rho_{u,i}$	FULL	SLOPE
0.5	0.5	[0.0,0.6]	0.0	0.84	1.00
0.5	0.5	[0.0,0.6]	[0.0,0.1]	0.83	1.00
0.5	0.5	[0.0,0.6]	[0.0,0.2]	0.82	0.99
0.5	0.5	[0.0,0.6]	[0.0,0.5]	0.80	0.96
1.0	0.5	[0.0,0.6]	0.0	0.92	1.00
1.0	0.5	[0.0,0.6]	[0.0,0.1]	0.91	0.99
1.0	0.5	[0.0,0.6]	[0.0,0.2]	0.91	0.97
1.0	0.5	[0.0,0.6]	[0.0,0.5]	0.89	0.87
1.0	1.0	[0.0,0.6]	0.0	0.57	1.00
1.0	1.0	[0.0,0.6]	[0.0,0.1]	0.55	0.98
1.0	1.0	[0.0,0.6]	[0.0,0.2]	0.54	0.92
1.0	1.0	[0.0,0.6]	[0.0,0.5]	0.51	0.59
0.5	0.5	[0.0,0.3]	0.0	0.82	1.00
0.5	0.5	[0.0,0.3]	[0.0,0.1]	0.81	0.99
0.5	0.5	[0.0,0.3]	[0.0,0.2]	0.81	0.97
0.5	0.5	[0.0,0.3]	[0.0,0.5]	0.79	0.86
0.5	1.0	[0.0,0.3]	0.0	0.51	1.00
0.5	1.0	[0.0,0.3]	[0.0,0.1]	0.50	0.98
0.5	1.0	[0.0,0.3]	[0.0,0.2]	0.49	0.89
0.5	1.0	[0.0,0.3]	[0.0,0.5]	0.46	0.57
1.0	1.0	[0.0,0.3]	0.0	0.53	1.00
1.0	1.0	[0.0,0.3]	[0.0,0.1]	0.52	0.91
1.0	1.0	[0.0,0.3]	[0.0,0.2]	0.51	0.72
1.0	1.0	[0.0,0.3]	[0.0,0.5]	0.48	0.27

$\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = 1.0$

Equation (3.4) shows that if both the measurement error correlation $\rho_{u,i}$ and the variances σ_{u1}^2, σ_{u2}^2 and σ_{x1}^2 are constant across studies, then the resulting two-stage estimate for β_2 will also be unbiased. However, if the second term on the right-hand side of (3.4) varies significantly across studies, the resulting two-stage estimator that regresses $\widehat{\delta}_{1i}$ on $\widehat{\gamma}_{1i}^*$ will be biased. Because the second stage simply relies on least squares, this bias has a simple analytic form. Again assuming that city-specific sample sizes are large enough so that the sampling error in $\widehat{\gamma}_{1i}^*$ is neglible, the expected value of $\widehat{\mathbf{c}} = (\widehat{c}_0, \widehat{c}_1)$ is

$$\left(\mathbf{W}^{\mathsf{T}}\mathbf{W}\right)^{-1}\mathbf{W}^{\mathsf{T}}\mathbf{v} \tag{3.5}$$

where the ith row of $\mathbf{W}$ is $(1, w_i)$, where w_i is the expectation given in (3.4), and the ith element of $\mathbf{v}$ is $\alpha\,(\beta_1 + \beta_2\gamma_{1i})$.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

Contrast this expectation to the traditional two-stage approach that first fits the full model to each city separately. If $\mathbf{\Sigma}_{x,i}$ and $\mathbf{\Sigma}_{u,i}$ are the variance–covariance matrices for $\mathbf{X}_i$ and $\mathbf{U}_i$, respectively, then the expected value of the city-specific estimator is

$$(\beta_1, \beta_2)\mathbf{C}_i, \tag{3.6}$$

where $\mathbf{C}_i = \mathbf{\Sigma}_{x,i}(\mathbf{\Sigma}_{x,i} + \mathbf{\Sigma}_{u,i})^{-1}$ (Carroll *et al.*, 1995). When the city-specific sample sizes are large, we can think of the full multilevel estimator as an average of these biased city-specific estimators.

Expression (3.5) demonstrates that when a regression coefficient is zero, the slope estimator for that coefficient is unbiased even in the presence of correlated measurement error. This is a major advantage of this estimator in the regulatory context. To investigate this bias for non-zero values of the coefficients, we followed Zeger *et al.* (2000) and compared (3.5) and the average (over city) of (3.6) in the presence of correlated measurement error. We assumed that (1) we have data from 10 cities (2) $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = 1.0$, and (3) both the correlation $\rho_{x,i}$ among the two exposure variables X_1 and X_2 and the correlation $\rho_{u,i}$ among the two measurement errors vary uniformly within specified intervals. Table 1 presents results. We see that, as the theory suggests, when the amount of heterogeneity in the correlation between U_{1i} and U_{2i} is small, the proposed slope estimator is biased only slightly. In particular, the proposed estimator outperforms the approach based on the full model except in cases in which the amount of heterogeneity in the correlations $\rho_{u,i}$ is large, the range of slopes γ_{1i} relating the two exposures is small, and the measurement error variances are large. In all cases examined, the proposed estimator is biased downward, and we observed the same general patterns for negative $\rho_{u,i}$ and other ranges of $\rho_{x,i}$. Furthermore, these calculations agree with those of Zeger *et al.* (2000), who noted that, even for the estimators based on the full model, estimates are not likely to incur upward bias unless substantial correlation exists either between the measurement errors or the exposures themselves. For instance, when $\rho_{u,i} = -0.95$ $\sigma_{u1}^2 = 2.0$ and $\sigma_{u2}^2 = 0.2$ and $\rho_{x,i}$ varies uniformly between 0.0 and 0.3, the expected value of the estimator for β_2 based on the full model is 1.04.

The above considerations suggest that in the presence of small amounts of measurement error correlation, our proposed two-stage approach remains essentially unbiased. In the specific case of air pollution studies, there are now limited data on the magnitude of these correlations. We examined data from Sarnat *et al.* (2001), who measured personal exposure to multiple pollutants simultaneously in 56 subjects in Baltimore, MD. Defining the difference between personal and ambient pollution measurements as the exposure error, we found the correlation between measurement error in $PM_{2.5}$ and other pollutants to be quite low, with a maximum correlation of 0.18 between the errors in $PM_{2.5}$ and NO_2. In order to investigate how these measurement errors vary across study, we compared these correlations to ones observed in recently collected data from Boston, MA. These unpublished correlations were similarly low, with the maximum correlation of 0.17 again corresponding to the pair of measurement errors associated with $PM_{2.5}$ and NO_2. Thus, in the air pollution context, early data from two cities suggests that the homogeneous measurement error correlation assumption may be reasonable, although more studies are necessary to confirm this observation.

3.3 *Heterogeneous effects model*

Often the mean number of events per day differs substantially across communities, as do the seasonal and weather patterns, making an overall pooled analysis inappropriate. For such situations, consider the heterogeneous effects model

$$Y_{it} = \beta_{0i} + \beta_{1i} X_{1it} + \beta_{2i} X_{2it} + \epsilon_{1it} \tag{3.7}$$

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

where

$$\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i}, \beta_{2i})^{\mathsf{T}} \sim N\left(\boldsymbol{\beta}, \begin{bmatrix} \sigma_{b0} & \sigma_{b0}\sigma_{b1}\rho_{b,01} & \sigma_{b0}\sigma_{b2}\rho_{b,02} \\ \sigma_{b0}\sigma_{b1}\rho_{b,01} & \sigma_{b1}^2 & \sigma_{b1}\sigma_{b2}\rho_{b,12} \\ \sigma_{b0}\sigma_{b2}\rho_{b,02} & \sigma_{b1}\sigma_{b2}\rho_{b,12} & \sigma_{b2}^2 \end{bmatrix}\right),$$

with $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^{\mathsf{T}}$. In the standard two-stage approach, the first stage estimates a separate regression for each location. In the second stage, these results are pooled to obtain an overall estimate:

$$\widehat{\boldsymbol{\beta}}_i = \boldsymbol{\beta} + \mathbf{b}_i + \boldsymbol{\nu}_i,$$

where $i = 1, \ldots, N$, $\widehat{\boldsymbol{\beta}}_i$ is the relative risk in city i, $\boldsymbol{\beta}$ is the mean underlying risk, $\mathbf{b}_i$ is the random between cities effect, and $\boldsymbol{\nu}_i$ is the known within-study error. It is common in this second stage of analysis to consider additional city-specific covariates that are potential effect modifiers that explain some of the random city-to-city variation in $\boldsymbol{\beta}_i$. These predictors could be characteristics of the populations in the individual studies or differences in the quality of exposure data. For simplicity of presentation we suppress these second stage covariates. Methods for fitting such models have been well-described (DerSimmonian and Laird, 1986; Berkey *et al.*, 1995). Tosteson *et al.* (1998) noted that, in the presence of measurement error, the 'naive' estimator that ignores the measurement error incurs bias for this model as well.

Consider the meta-regression approach where we regress Y on X_1 and X_2 on X_1 individually for each location. Following the arguments provided in the previous section, we have

$$E(\widehat{\gamma}_{1i}^*) = \alpha\gamma_{1i}$$

and

$$E(\widehat{\delta}_{1i}^* | \boldsymbol{\beta}_i) = \alpha(\beta_{1i} + \beta_{2i}\gamma_{1i}).$$

This last equation yields

$$E(\widehat{\delta}_{1i}^*) = E\left[E(\widehat{\delta}_{1i}^* | \boldsymbol{\beta}_i)\right] = \alpha(\beta_1 + \beta_2\gamma_{1i}).$$

Thus, regressing $\widehat{\delta}_{1i}^*$ on γ_{1i}^* once again yields a consistent estimator of β_2.

4. A SIMULATION STUDY

The above analytical considerations are still restrictive in that we only consider Gaussian outcomes with two predictors. However, it demonstrates that the three estimates of the effects of the correlated exposures behave differently in the presence of measurement error, and that the meta-regression-based estimates may be less prone to bias.

In the mortality examples in Section 5 the response is often a count (number of deaths each day in a community) and so Poisson regression with a log link is appropriate. However, measurement error models are more complex in the Poisson setting (Zeger *et al.*, 2000). To examine the properties of these estimators in this setting, we conducted a simulation study. Specifically, we focused on the scenario in which there is more measurement error in one variable than another, as this case is common in exposure assessment scenarios. In addition, in view of the arguments presented in Section 3.2, we also consider the case with correlated measurement error.

Environmental epidemiology typically involves small relative risks of common outcomes. Thus, we used small relative risks in simulations to ensure that conclusions would be valid even in this more challenging situation. A baseline simulation consisted of the following:

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

Table 2. *Settings for simulation scenarios of Poisson health data with two covariates. Parameters for each city were generated uniformly from a given interval*

Scenario	Number of cities	Baseline count	Range for β_1	Range for β_2	Range for $\rho_{x,i}$	Range for γ_{1i}	$\rho_{u,i}$
a	10	22	0.05	0.05	0.8	8-fold	0.0
b	10	22	0.05	0.05	0.6	8-fold	0.0
c	10	22	0.05	0.05	0.8	3-fold	0.0
d	10	22	0.05	0.05	0.6	3-fold	0.0
e	20	22	0.05	0.05	0.8	8-fold	0.0
f	10	22	0.05	0.05	-0.8	8-fold	0.0
g	10	5	0.05	0.05	0.8	8-fold	0.0
h	10	22	0.05	0.05	[0.7,0.9]	8-fold	0.0
i	10	22	[0.035,0.065]	[0.035,0.065]	0.8	8-fold	0.0
j	10	22	[0.035,0.065]	[0.035,0.065]	[0.7,0.9]	8-fold	0.0
k	10	22	[0.035,0.065]	[0.035,0.065]	[0.7,0.9]	8-fold	[0.0,0.2]
l	10	22	0.05	0.0	0.8	8-fold	0.0
m	10	22	[0.035,0.065]	0.0	[0.035,0.065]	8-fold	0.0

1. Two exposure variables were simulated from a multivariate normal distribution with 2000 observations, and mean zero. The correlation between the two variables was set to either 0.8 or 0.6. The slope (γ_1) of the relationship between the two exposures was randomly varied over a range of either 3-fold or 8-fold by choosing the standard deviation of one of the exposures from a uniform distribution, while fixing the standard deviation of the other exposure at one.
2. A random Poisson count was simulated using the two exposures, with a true log relative risk of 0.05 per one standard deviation change in exposure for each variable and a baseline of 22, which is typical of the daily counts seen in air pollution studies.
3. Measurement error was then added to each variable, with a mean of 0 and a standard deviation of 0.3 of the standard deviation of the first exposure, and 0.6 of the standard deviation of the second exposure. This yields a dataset $(x_{111}, x_{211}, y_{11}),\ldots,(x_{11T_1}, x_{21T_1}, y_{1T_1})$ for the first study.
4. Poisson regressions were fit to estimate the log relative risk of each exposure in a two-exposure model and in single exposure models.
5. The above was repeated 10 times and hierarchical models were used to estimate the effect of each exposure, using the standard, intercept and slope methods. This completed a single simulation of a 10-study hierarchical model.
6. The above set of simulations was repeated 500 times for each scenario.

Based on this overall strategy, we investigated 13 scenarios, denoted as (a)–(m), by varying either the value of regression coefficients, the amount of heterogeneity in these coefficients across city, the correlation between the two exposures, and the amount of heterogeneity of the slopes characterizing these correlations across city. Table 2 lists settings of each of these choices for each scenario.

Figure 1 shows the median and empirical 5th and 95th percentiles for the simulated estimators for scenarios (a)–(k). The first, second and third intervals for each scenario represent results for the standard, intercept, and slope estimator, respectively. The upper panel of the plot shows results for $\widehat{\beta}_1$, whereas the lower panel shows results for $\widehat{\beta}_2$. The general pattern illustrates the well-known phenomenon that the standard estimator can incur upward bias for the coefficient associated with the exposure with less measurement error. Although this bias is larger than Monte Carlo error for several scenarios, the magnitude of this bias is modest. This accords with the results of Zeger and co-workers (2000), who found relatively pathological patterns were needed to produce large upward biases.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

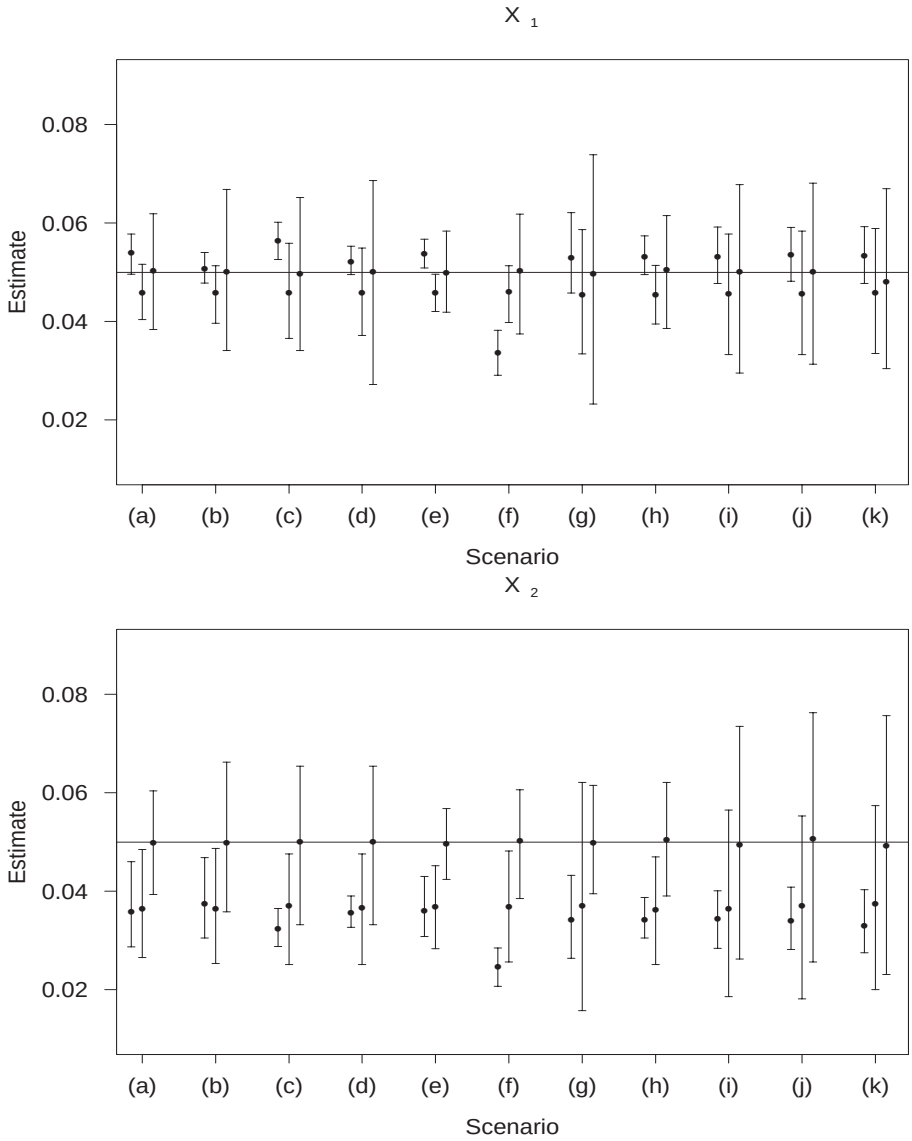

Fig. 1. 5th, 50th, 95th percentiles of simulated estimates from simulation scenarios (a)–(k) in Table 2. For each scenario, the first, second and third intervals represent the standard, intercept and slope estimates, respectively.

In contrast, the bias incurred by the intercept estimate is still biased toward the null for the better-measured exposure. The slope estimates, however, are unbiased for both exposures. These improvements come at the price of a modest increase in the variance for the intercept estimate, and a somewhat larger increase in the variance for the slope estimate. The main exception to this rule is the case of a negative correlation among the two exposures, in which the standard estimator incurs significant downward bias for both coefficients.

Reducing the correlation between exposures from 0.8 to 0.6 (going from scenario (a) to scenario (b),

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

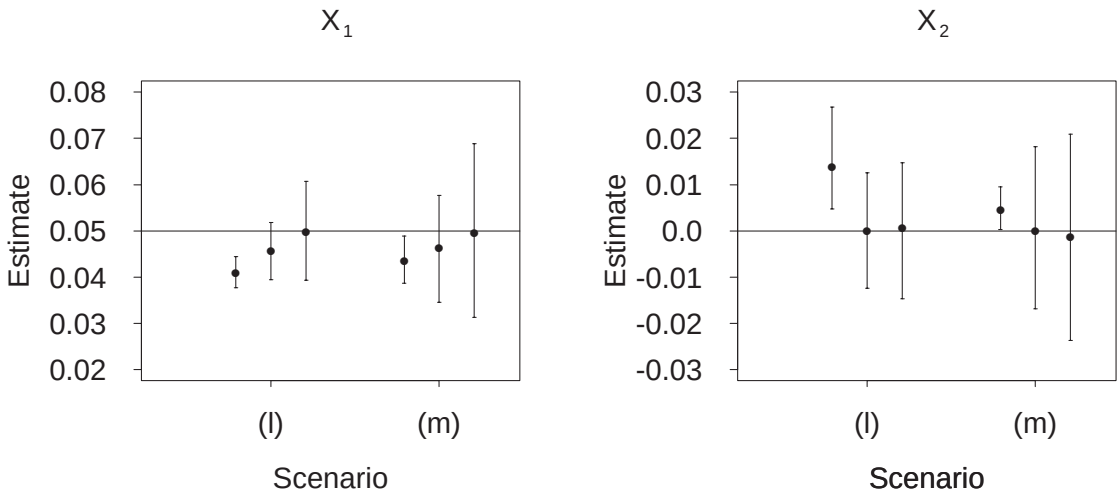

Fig. 2. 5th, 50th, and 95th percentiles of simulated estimates from simulation scenarios (l)–(m) in Table 2. For each scenario, the first, second and third intervals represent the standard, intercept and slope estimates, respectively.

Table 3. *Settings for simulation scenarios of Poisson health data with three covariates. Parameters for each city were generated uniformly from a given interval. All scenarios used a baseline count of 22*

Scenario	Number of cities	Range for β_1	Range for β_2	Percentage of X_3	Range for $\rho_{x,i}$	Range for γ_{1i}	$\rho_{u,i}$
n	10	0.05	0.05	0.25	0.8	8-fold	0.0
o	10	0.05	0.05	0.50	0.8	8-fold	0.0
p	10	[0.035,0.065]	[0.035,0.065]	0.25	0.8	8-fold	0.0
q	20	[0.035,0.065]	[0.035,0.065]	0.50	0.8	8-fold	0.0

and from scenario (c) to (d), for instance) essentially eliminates the risk of overestimating the effect of X_1 using any of the three methods. However, the downward bias in X_2 remains substantially larger than for X_1 when we use the two-pollutant and intercept estimates. Again, the slope estimate is unbiased. In scenarios having a narrower range of slopes between the two exposures, the intercept estimate avoids upward bias and the slope estimates are unbiased, but these estimates have larger variances than those where the range of slopes was greater. Results from scenarios (i)–(k) demonstrate that, in the presence of city-to-city heterogeneity in health effects, the same general relationships among the three estimators remained the same, with only an increase in the variability of the estimates. Scenario (k) also suggests that, as in the normal response case, the intercept and slope estimates incur small downward bias in the presence of heterogeneous measurement error correlation, although this bias is less than that incurred by the full model.

Figure 2 illustrates the performance of the three estimators when only one of the exposures is causally related to the outcome. In both cases, the two-exposure model shows upward bias in the non-causal exposure. In fact, the 95% confidence interval does not include 0, so there is a high chance of falsely attributing an association to the non-causal variable. This risk is eliminated by the use of either the intercept or slope estimates.

We also examined several other scenarios not presented in Figures 1 and 2. For instance, it is possible

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

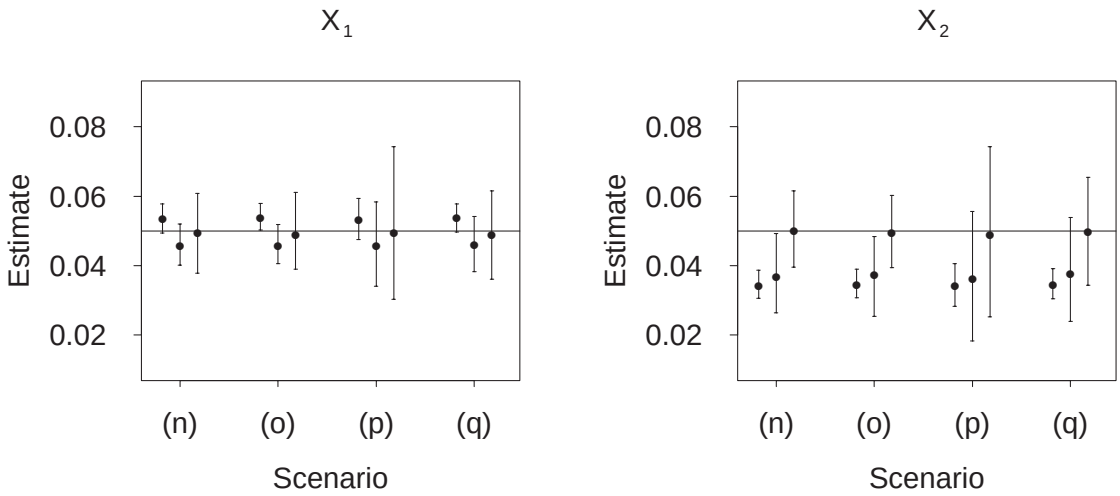

Fig. 3. 5th, 50th, and 95th percentiles of simulated estimates from simulation scenarios (n)–(q) in Table 2. For each scenario, the first, second and third intervals represent the standard, intercept and slope estimates, respectively.

that there are some biases that are hard to detect using small effect sizes. If we increase the effects to 0.50 per standard deviation change for both exposures, the results are essentially unchanged. We also considered confounding by additional covariates. Motivated by the examples in Section 5, we constructed a seasonally varying covariate. For each simulated data set, we generated a new variable X_{3it} from a $\mathrm{N}(0.5 \times \cos(2t/365), 0.5)$ distribution, $t = 1, \ldots, 2000$, and added either $0.25 \times X_{3it}$ or $0.5 \times X_{3it}$ (denoted by 'percentage of X_3') to X_{1it} and X_{2it}. We then simulated mortality counts from a Poisson model with means

$$\lambda_{it} = 22 \times \exp(0.05 \times X_{1it} + 0.05 \times X_{2it} + 0.05 \times X_{3it}). \tag{4.8}$$

The standard model estimates were computed from models with all three covariates in the first stage. The slope and intercept estimates were computed by estimating models for the outcome that depended on the seasonal covariate (X_3) and one of the other two exposures. The estimated regression coefficient for that exposure was regressed against the regression coefficient relating the two exposures, as before. Table 3 shows the simulation settings for the three-covariate case. Figure 3 demonstrates that the general patterns observed in the two-covariate case hold in this case as well, with the slope estimates appearing unbiased. In addition, in the three-covariate setting in which only one of the estimates is causal, the estimate for the effect of X_2 remains biased upward, while the potential for a false positive finding is avoided using either of the alternative estimators.

Finally, a key assumption of the proposed two-stage approach is that the attentuation factor α is constant across cities. We ran an additional simulation to check the effect of violations of this assumption on the performance of the estimators in the Poisson setting. In particular, otherwise using scenarios (h) and (j) from Table 2, we constructed two new scenarios [denoted (h′) and (j′)] with city-specific α_i by multiplying σ_{u1} and σ_{u2} by a city-specific variate uniformly generated from [0.7, 1.0]. Figure 4 shows that this violation has minimal effect on the performance of all three estimators in these settings, as results are essentially identical to those from scenarios (h) and (j) in Figure 1.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

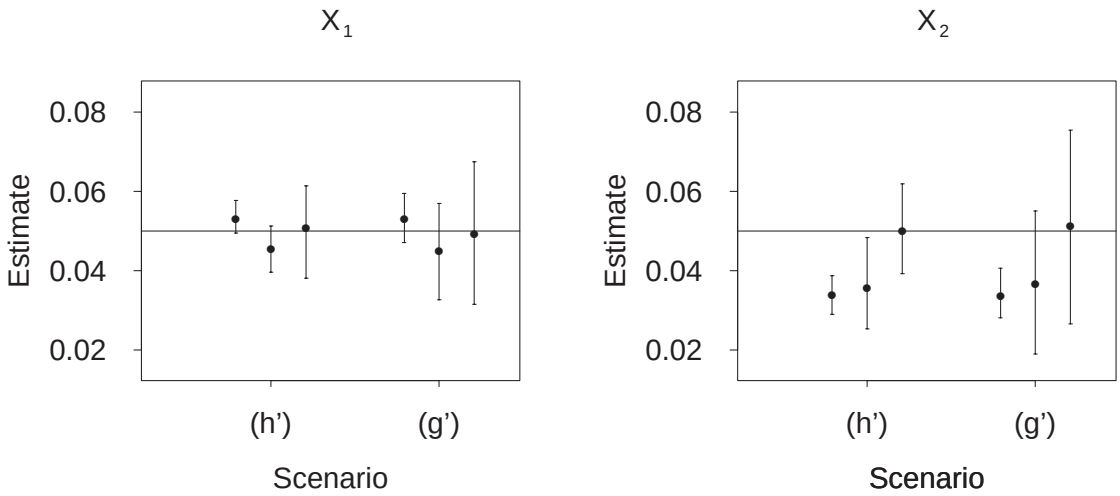

Fig. 4. 5th, 50th and 95th percentiles of simulated estimates from simulation scenarios (h′)-(j′) assuming heterogeneous α_i over cities. For each scenario, the first, second and third intervals represent the standard, intercept and slope estimates, respectively.

5. APPLICATIONS

5.1 *Effects of particles on daily deaths*

The utility of these alternative approaches can be illustrated using two examples of interest. In 1997, the US Environmental Protection Agency promulgated a standard for airborne concentrations of particles less than 2.5 μm in aerodynamic diameter ($PM_{2.5}$, or fine particles) (EPA, 1997). This was based on evidence that serious health effects were seen for airborne particles at concentrations well below the previous standards (Pope *et al.*, 1995). However, that previous standard governed all particles with aerodynamic diameter less than 10 μm (PM_{10}). EPA cited recent evidence indicating that the smaller particles ($PM_{2.5}$) were more toxic, and focused the new regulations on them. One important study in reaching this conclusion examined the association between daily concentrations of both $PM_{2.5}$ and particles with sizes between 2.5 and 10 μm (usually referred to as coarse mode particles) and daily deaths in 6 US cities (Schwartz *et al.*, 1996). That study reported a second-stage analysis of models containing concentrations of particles in both size ranges; that is, it used the standard model for two covariates. There was a positive association between $PM_{2.5}$ and daily deaths (effect $= 0.0149$, se $= 0.00197$ for a 10 μg m^{-3} increase). However, the association with coarse model particles was negative and small (effect $= -0.00206$, se $= 0.00491$). This paper was criticized by others, who argued that greater measurement error in coarse mode particle measurements could have transferred some of the effect to the fine particles, whereas in reality they were more nearly equipotent. This had clear implications for which sources should be controlled to reduce risk. The slopes between $PM_{2.5}$ and coarse mass vary by a factor of four across the six cities. Thus, using the slope estimate approach, we have re-examined this issue. Table 4 shows the original estimates for the effects of $PM_{2.5}$ and coarse mode particles on daily deaths, and, as an alternative, the slope estimates for each pollutant, which should be unbiased by measurement error. These slope estimates suggest that both original estimates were biased toward the null by their measurement error, rather than one being biased upward. The effect estimate for $PM_{2.5}$ is increased, and the effect estimate for coarse particles grows more negative. This may reflect the association of course particles with windy days, which can reduce combustion pollutants.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

Table 4. *Estimated effect and associated standard errors of a 10 $\mu g\ m^{-3}$ increase in fine and coarse particles on daily deaths in six US cities*

Particle measure	Standard estimate (S.E.)	Slope estimate (S.E.)
$PM_{2.5}$	0.0149 (0.00197)	0.0342 (0.00287)
Coarse Mass	−0.00206 (0.00491)	−0.0235 (0.00616)

Table 5. *Estimated effect and associated standard errors of an increase in gaseous pollutants on daily deaths in 10 US cities using the slope estimator*

Pollutant	$\hat{\beta}$ (S.E.)	Range of PM_{10} Slopes
SO_2 (10ppb)	0.00259 (0.00384)	0.08 – 1.24
Ozone (10ppb)	−0.00227 (0.00312)	−0.22 – 1.07
Carbon Monoxide (1ppm)	−0.00711 (0.00492)	0.013 – 0.080

5.2 *Effect of gaseous air pollutants on daily deaths*

Another recent air pollution study analyzed the association between particulate air pollution and daily deaths in 10 US cities (Schwartz, 2000) with daily PM_{10} monitoring. Positive associations were seen for PM_{10}. No results were reported for gaseous air pollutants, however. Such associations can be quickly assessed using a slope estimate relating each of the other pollutants to PM_{10}. Moreover, the original paper estimated effects in each location separately for the warm and cold seasons. This provides twenty estimates for the second stage, and increases the variability in the slopes between the pollutants, which vary substantially by season in many locations. While somewhat noisier than two-pollutant model estimates, the freedom from bias in the presence of measurement error makes these estimates attractive. These estimates are shown in Table 5, which also shows the range of slopes between PM_{10} and the gaseous pollutants. There is little evidence for any association of SO_2 or ozone with daily deaths, and only very weak evidence for a protective effect of carbon monoxide.

6. Discussion

We have demonstrated that there are alternative approaches to estimating the effect of two exposures on an outcome, which, in the almost inevitable presence of measurement error and the context of multiple studies or subgroups, may have some advantages over traditional estimators. The requirement for multiple studies is not much of a limitation if we expect the two exposures to be moderately well correlated. Multiple studies, or large datasets (which in principle can be subdivided) would be necessary in any event to obtain stable estimates. We have shown analytically that in the case of Gaussian data, the slope estimator is unbiased and the intercept estimator is free from upward bias as long as the attentuation factor arising from measurement error and the measurement error correlation are constant across studies. Thus, the methods do not require the measurement error for different exposure to be independent. Limited data from two cities suggests these assumptions may be plausible in air pollution epidemiology. Future research should focus on confirming these assumptions in a larger sample of cities, as well as developing ways to diagnose variations in measurement error across locations.

For non-normal responses, simulations suggest that the slope estimator is unbiased and the intercept estimate is free from upward bias, even in the presence of other correlated covariates and true hetero-

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

geneity. This is notable, and suggests the approach could have important applications in epidemiology. The ability of either the intercept and slope estimates to reduce or eliminate the risk of overestimating the effect of one exposure, and particularly to avoid falsely concluding that one exposure is associated with the outcome, may be quite valuable. Further, simulations suggest that the proposed estimators perform well even when the assumption of homogeneous attenuation factors is violated. The proposed estimators pay a price in precision, however. The slope estimates become more variable as the correlation between the exposures falls. Thus, in the case of mildly correlated variables, the methods should be used with caution. This difference in precision diminishes when the number of cities grows to be approximately 20. Thus, given the substantial amount of exposure misclassification that exists in many epidemiological studies, it appears to be a valuable tool in the epidemiologist's portfolio.

One additional benefit of the method is the reduced likelihood of a false positive finding for a non-causal exposure in the presence of measurement error. Recent studies of air pollution and daily death have examined as many as five different pollutants. It is unlikely that all of these are causally related to mortality. The simulations demonstrate that the risk of a falsely positive association for the non-causal pollutant is substantially reduced using either of the two alternative models. This value is illustrated by two applications. The first shows that, quite opposite from being upwardly biased, the coefficient of $PM_{2.5}$ appears to have been downwardly biased in the original report of Schwartz *et al.* (1996). While debate over this issue has generated a substantial number of exposure assessment studies, this alternative approach appears straightforward. The finding that gaseous air pollutants are not associated with daily deaths in a multi-city study after adjustment for PM_{10} accords with the results recently reported by Samet *et al.* (2000) using the standard approach in 20 cities. These results replicate those in a different sample of cities. More importantly, they add some assurance that Samet and co-workers' results were not driven by differential measurement error among the pollutants.

ACKNOWLEDGMENTS

This research was partially supported by EPA grant R827353 and NIEHS grant ES 00002. The authors thank Alexandros Gryparis, the associate editor, and a referee for their helpful comments.

REFERENCES

BERKEY, C. S., HOAGLIN D. C., MOSTELLER, F. AND COLDITZ, G. A. (1995). A random-effects regression model for meta-analysis. *Statistics in Medicine* **14**, 395–411.

CARROLL, R. J., RUPPERT, D. and STEFANSKI, L. A. (1995) *Measurement Error in Nonlinear Models*. London: Chapman & Hall.

DANIELS, M. J., DOMINICI, F., SAMET, J. M. AND ZEGER, S. L. (2000). Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. *American Journal of Epidemiology 152*, 397–406.

DERSIMONIAN, R. AND LAIRD N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–88.

DOMINICI, F., SAMET, J. M. AND ZEGER, S. L. (2000). Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy (with discussion). *Journal of the Royal Statistical Society, Series A* **163**, 263–302.

ENVIRONMENTAL PROTECTION AGENCY (1997). *National Ambient Air Quality Standards for Particulate Matter; Part II. Federal Register*. **62**.

FAIRLEY, D. (1990). The relationship of daily mortality to suspended particulates in Santa Clara County, 1980–1986. *Environmental Health Perspectives* **89**, 159–68.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017

FULLER, W. (1987). *Measurement Error Models*. New York: Wiley.

JUDGE, G. G., GRIFFITHS, W. E., HILL, R. C., LUTKEPOHL, H. AND LEE, T.-C. (1985). *The Theory and Practice of Econometrics*. New York: Wiley.

KATSOUYANNI, K., TOULOUMI, G., SPIX, C., SCHWARTZ, J., BALDUCCI, F., MEDINA, S., ROSSI, G., WOJTYNIAK, B., SUNYER, J., BACHAROVA, L., *et al.* (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.

LIPFERT, F.W. AND WYZGA, R. E. (1997). Air pollution and mortality: the implications of uncertainties in regression modeling and exposure measurement. *Journal of Air & Waste Management Association* **47**, 517–523.

LYLES, R. H. AND KUPPER, L. L. (1997). A detailed evaluation of adjustment methods for multiplicative measurement error in linear regression with applications in occupational epidemiology. *Biometrics* **53**, 1008–1025.

MARCUS, A. H. AND KEGLER, S. R. (2001). Confounding in air pollution epidemiology: When does the two-stage regression identify the problem? *Environmental Health Perspectives* **109**, 1193–1196.

POPE, C. A. III, DOCKERY, D. W. AND SCHWARTZ, J. (1995). Review of epidemiological evidence of health effects of particulate air pollution. *Inhalation Toxicology* **7**, 1–18.

SAMET, J. M., DOMINICI, F., CURRIERO, F. C., COURSAC, I. AND ZEGER, S. L. (2000). Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. *New England Journal of Medicine* **343**, 1742–1749.

SARNAT, J. A., SCHWARTZ, J., CATALANO, P. J. AND SUH, H. H. (2001). Gaseous pollutants in particulate matter epidemiology: Confounders or surrogates? *Environmental Health Perspectives* **109**, 1053–1061.

SCHWARTZ, J. AND DOCKERY, D. W. (1992). Increased mortality in Philadelphia associated with daily air pollution concentrations. *American Review of Respiratory Disease* **145**, 600–604.

SCHWARTZ, J. (1993). Air pollution and daily mortality in Birmingham, Alabama. *American Journal of Epidemiology* **137**, 1136–1147.

SCHWARTZ, J., DOCKERY, D. W. AND NEAS, L. M. (1996). Is daily mortality associated specifically with fine particles? *Journal of Air & Waste Management Association* **46**, 2–14.

SCHWARTZ, J. (2000). The distributed lag between air pollution and daily deaths. *Epidemiology* **11**, 320–326.

SCHWARTZ, J. (2000). Assessing confounding, effect modification, and thresholds in the association between ambient particles and daily deaths. *Environmental Health Perspectives* **108**, 563–568.

TOSTESON, T. D., BUONACCORSI, J. P. AND DEMIDENKO, D. (1998). Covariate measurement error and the estimation of random effect parameters in a mixed model for longitudinal data. *Statistics in Medicine* **17**, 1959–1971.

ZEGER, S. L., THOMAS, D., DOMINICI, F., SAMET, J. M., SCHWARTZ, J., DOCKERY, D. W. AND COHEN, A. (2000). Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environmental Health Perspectives* **108**, 419–426.

[*Received August 8, 2002; first revision October 18, 2002; second revision January 21, 2003; accepted for publication January 27, 2003*]

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on January 6, 2017