Journal of Exposure Science and Environmental Epidemiology (2007) 17, S20–S28
© 2007 Nature Publishing Group All rights reserved 1559-0631/07/$30.00
www.nature.com/jes

Controlling for confounding in the presence of measurement error in hierarchical models: a Bayesian approach

ALEXANDROS GRYPARIS[a], BRENT A. COULL[a] AND JOEL SCHWARTZ[b]

[a]*Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, Massachusetts 02115, USA*
[b]*Department of Environmental Health, Harvard School of Public Health, Landmark Center, Room 415, 401 Park Drive, Boston, Massachusetts 02115, USA*

A major concern in studies that address the health effects of air pollution is whether an observed association between concentrations of a pollutant and a health outcome is all, or in part, due to the correlation between that exposure and either a second pollutant or a confounder. The addition of exposure measurement error to such data complicates matters further. To account for measurement error when data come from a multi-city study, Schwartz and Coull (2003) proposed a two-stage estimator. These authors showed via both first principles and simulation that their approach yields unbiased estimates for the parameters of interest. However, these estimates have large variability. In this paper, we describe a fully Bayesian approach that yields estimators that are much more efficient than the existing two-stage measurement error correction yet still unbiased. The proposed approach can also incorporate additional exposures or confounders without requiring strict assumptions that are necessary in existing formulations of the model. We compare the properties of the Bayesian estimators to existing approaches via simulation.
Journal of Exposure Science and Environmental Epidemiology (2007) **17,** S20–S28; doi:10.1038/sj.jes.7500624

Keywords: *air pollution, health effects, multi-pollutant models, particulate matter, Poisson regression, WinBUGS.*

Introduction

The potential health effects of ambient air pollution are a major public health issue that has received a great deal of attention over the past several decades. A major concern in studies that address such questions is whether an observed association between one pollutant and one outcome is due, all or in part, to the correlation between that exposure and either a second pollutant or a confounder. A high correlation between two covariates may yield relatively imprecise estimates of the exposures of interest. For example, in air pollution studies, the correlation between concentrations of airborne particles and carbon monoxide can exceed 0.8.

The presence of measurement error in such data complicates matters further. It is well known (Carroll et al., 1995) that in a model with only a single covariate, the usual estimator for the regression coefficient will be biased toward the null. In a model with two covariates, both measured with error, in general, the standard ordinary least-squares (OLS) estimates of the regression coefficients tend to be attenuated, and the one corresponding to the variable measured with considerable error will be attenuated more than the coefficient estimate corresponding to the variable measured with less error. However, depending on the magnitude and direction of the correlations between the two exposures and between the measurement errors, the effect of one variable might be transferred to the estimate of the other variable (Zeger et al., 2000). Generally, this transfer occurs from a poorly measured variable to one with less error. That is, there could be upward bias in the estimated effect of the better-measured pollutant, but downward bias in the estimated effect of the more poorly measured pollutant. This issue is particularly important in a regulatory context, where interest focuses on identifying safe levels of pollutants with dangerous health effects. In a multi-city setting, if bias is expected in the individual substudies, this bias propagates through to the pooled estimates of association.

To account for measurement error in linear regression, several approaches have been proposed. These include regression calibration, simulation–extrapolation and the use of instrumental variables (Carroll et al., 1995). For data that arise from a multi-city design, Schwartz and Coull (2003) proposed a two-stage approach that yields unbiased estimates for the health effects of interest. A disadvantage of this two-stage approach, however, is that the method yields imprecise estimators. This imprecision results in a reduction of power, which for some realistic scenarios can be severe. This power reduction could be crucial in particulate matter epidemiology,

1. Address all correspondence to: Dr. A Gryparis, Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA.
Tel.: 617 432 1056.
Fax: 617 432 5619.
E-mail: alexandros@post.harvard.edu
Received 7 September 2007; accepted 12 September 2007

where the risks associated with exposure are relatively small. Thus, interest focuses on the development of more efficient health effect estimators that control for measurement error in pollution concentrations.

This paper describes such an extension and explores its performance via simulation. The structure of this paper is as follows. The next section reviews the approach proposed by Schwartz and Coull (2003) and proposes a one-stage Bayesian approach that yields more precise health effect estimates. The penultimate section describes the setting and the results of our simulation study and is followed by a concluding discussion in the last section. In the appendix, we provide the WinBUGS code we used for fitting our models.

Methods

Review of the Schwartz and Coull Approach

In the simplest case, Schwartz and Coull (2003) considered a linear model in which the only covariates are two correlated exposures that are measured with error. For day t, $t=1,\ldots,T_i$, in city i, $i=1,\ldots,N$, consider the true model

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \varepsilon_{it}, \tag{1}$$

where

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

$\varepsilon_{it} \sim N(0, \sigma_i^2)$ and ε_{it} is independent of $\mathbf{X}_{it}$. If $\mathbf{X}_{it}$ is measured without error, then fitting the model

$$X_{2it} = \gamma_{0i} + \gamma_{1i} X_{1it} + \eta_{it}, \tag{3}$$

yields parameter estimate $\widehat{\gamma}_{1i}$ with expectation

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

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

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

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

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

separately for each city. Then, from the distributional assumptions on $\mathbf{X}_{it}$ 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 a second model

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

Due to the measurement error and the exclusion of W_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 the true $\gamma_{1i}^* = \alpha\gamma_{1i}$ is known. 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 OLS, yielding an attenuated estimate of β_1 but an unbiased estimate of β_2 in the presence of independent measurement error $\mathbf{U}$. If we are interested in an estimator for β_1, we reverse the roles of X_{2it} and X_{1it}.

The above formulation requires that we know the true γ_{1i}^*. In practice, we estimate these parameters for each city. To get good estimates, we need a large sample size within each city. That is often the case in air pollution studies, and hence the estimated $\widehat{\beta}_2$ will be relatively precise. In contrast, in the presence of measurement error, the standard approach based on fitting the two-pollutant model

$$Y_{it} = \beta_0^* + \beta_1^* W_{1it} + \beta_2^* W_{2it} + \varepsilon_{it} \tag{4}$$

will yield biased estimates of $\widehat{\beta}_1$ and $\widehat{\beta}_2$, even for large sample sizes. In addition, even though Schwartz and Coull (2003) used a multivariate normal formulation to make the estimation strategy clear, normality is not required as long as the linear models (1) and (3) hold.

Schwartz and Coull (2003) also considered the performance of the two-stage estimator in the presence of correlated measurement error. If the measurement error correlation remains constant across cities, then the two-stage estimator will be unbiased. Furthermore, the authors also considered the case of a heterogeneous effects model, where

$$Y_{it} = \beta_{0i} + \beta_{1i} X_{1it} + \beta_{2i} X_{2it} + \varepsilon_{it}, \tag{5}$$

and

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

with $\beta = (\beta_0, \beta_1, \beta_2)^T$. They showed that their approach is consistent under such a scenario.

Bayesian Approach

As in Schwartz and Coull (2003), we also start from the model defined by (1) and (2). The parameters of interest are the coefficients that represent associations between the outcome and the two exposure concentrations, e.g. β_1 and β_2. Now suppose that X_{hit}, $h=1,2$, is measured with additive measurement error, and that the measurement errors are uncorrelated.

As noted in the Introduction, interest focuses on improving the efficiency of the estimators proposed in Schwartz and

Coull (2003), while maintaining unbiasedness. To do this, we fit the full model and jointly estimate all the unknown parameters simultaneously. This will increase efficiency and thus the power to detect effects. The Bayesian methodology is a natural way to fit complex, hierarchical models with many parameters.

Assume the model defined by (1) and (2) holds, and instead of X_{hit} we observe $W_{hit} = X_{hit} + U_{hit}$, $h = 1,2$, where the measurement errors $\mathbf{U}_1$ and $\mathbf{U}_2$ are uncorrelated. It is easy to show using a method of moments (MOM) approach that all of the parameters in the model are estimable. The available data for each city i, $i = 1,2,\ldots, N$, consists of the outcome $\mathbf{Y}_i$ and the two error-prone exposures $\mathbf{W}_i = \mathbf{W}_{1i}, \mathbf{W}_{2i}$. Hence, for identifiability it is sufficient to show that all the parameters of our full Bayesian model can be expressed as a function of the moments of observed quantities. Let $\mathbf{Y}^T = (\mathbf{Y}_1^T, \mathbf{Y}_2^T, ..., \mathbf{Y}_N^T)$, $\mathbf{W}_1^T = (\mathbf{W}_{11}^T, \mathbf{W}_{12}^T, ..., \mathbf{W}_{1N}^T)$ and $\mathbf{W}_2^T = (\mathbf{W}_{21}^T, \mathbf{W}_{22}^T, ..., \mathbf{W}_{2N}^T)$. If we integrate out the unobserved $\mathbf{X}$, then it is easy to show that for each city i, we have $E(\mathbf{W}_{1i}) = \mu_{1,i}$ $E(\mathbf{W}_{2i}) = \mu_{2,i}$ $Cov(\mathbf{W}_{1i}, \mathbf{W}_{2i}) = \sigma_{x1}\sigma_{x2}\rho_{x,i}$ $Cov(\mathbf{Y}_i, \mathbf{W}_{1i}) = \beta_1\sigma_{x1}^2 + \beta_2\sigma_{x1}\sigma_{x2}\rho_{x,i}$ $Cov(\mathbf{Y}_i, \mathbf{W}_{2i}) = \beta_2\sigma_{x2}^2 + \beta_1\sigma_{x1}\sigma_{x2}\rho_{x,i}$ $Var(\mathbf{Y}_i) = \sigma_i^2 + \beta_1^2\sigma_{x1}^2 + \beta_2^2\sigma_{x2}^2 + 2\beta_1\beta_2\sigma_{x1}\sigma_{x2}\rho_{x,i}$

These equalities, along with the fact that $E(\mathbf{Y}_i) = \beta_0 + \beta_1\mu_{1,i} + \beta_2\mu_{2,i}$ and $Var(\mathbf{W}_1) = \sigma_{x1}^2 + \sigma_{u1}^2$ and $Var(\mathbf{W}_2) = \sigma_{x2}^2 + \sigma_{u2}^2$, enable one to express all unknown model parameters as functions of the moments of observed variables as long as β_1 and β_2 are not equal to zero. This indicates that, in such cases, the model parameters are identifiable.

To fit our Bayesian approach we choose conjugate, non-informative but proper prior distributions for the parameters of the model. Specifically, we use: $P(\beta_0) = P(\beta_h) = P(\mu_{1,i}) = P(\mu_{2,i}) = \text{Normal}(0, 10^3)$, $P(\rho_{x,i}) = \text{Uniform}\,(-1, 1)$, $P(\sigma_i^2) = P(\sigma_{x_h}^2) = P(\sigma_{u_h}^2) = \text{Inverse}-\text{Gamma}(0.01, 0.01)$ for $i = 1,2,\ldots, N$ and $h = 1,2$. If one has available prior information for any of the model parameters (i.e. from previous studies), this can be easily incorporated in the Bayesian scheme by using appropriate informative prior distributions for those parameters.

The assumption of independent measurement error is a strong assumption, which ensures identifiability. It may be approximately true in some applications. In the specific case of air pollution epidemiology, there is 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, Maryland. 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. To investigate how these measurement errors vary across studies, we compared these correlations to the ones observed in recently collected data from Boston, Massachusetts. These 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 suggest that the homogeneous measurement error correlation assumption may be reasonable, although more studies are necessary to confirm this observation.

In addition to the predictors of interest, most models must control for confounding. In the model considered so far, error-contaminated exposures $\mathbf{W}$ are the only covariates in the model. Hence, it is necessary to extend our model to accommodate variables measured without error. It is well known that measurement error in one variable results in biased coefficient estimators for all variables that are included in the model, even for those that are measured without error (Judge et al., 1985, p. 708). Our proposed approach can easily accommodate more variables in the model, without any extra assumptions. Using the MOM approach outlined above, one can easily show that all the parameters of the model are identifiable in this expanded model as well.

In the two-stage approach proposed by Schwartz and Coull (2003), if one is interested in adjusting for covariates measured without error, one must make the additional assumption that the correlations between the confounder and each of the exposures are constant across cities. Otherwise, if these correlations vary significantly across cities, this approach yields biased estimates, and this bias can be severe. The Bayesian joint modeling approach makes no such assumption. Thus, this approach will outperform the two-stage approach when this assumption is violated. To examine the performance of the methods under such a scenario, we include in our simulation study a setting where the correlations between a confounder and the two error-prone exposures vary significantly across cities. We report on the results in the next section.

Similarly, the fully Bayesian approach accommodates models that contain more than two error-prone exposures. In this setting, no extra assumptions on how the correlations among exposures vary across cities are required. Again this is not true for the two-stage approach, which requires that the pairwise correlations between each of the two pollutants involved in the two-stage process and a third pollutant are constant across cities. Furthermore, Schwartz and Coull (2003) noted that a simple two-stage strategy that considers only two exposures at a time yields health effect estimates that are not unique. For example, if the pollutants of interest are $\mathbf{W}_1$, $\mathbf{W}_2$, and $\mathbf{W}_3$, then one can obtain an unbiased (assuming constant correlations among the pollutants across cities) slope estimate of β_2 by using either $\mathbf{W}_1$ or $\mathbf{W}_3$ as the covariate in the first stage of this approach. The Bayesian approach uses all the available data simultaneously and the resulting estimates are unique.

Zeka and Schwartz (2004) applied the Schwartz and Coull approach to data from the National Morbidity and

Mortality Air Pollution Study. These data come from 90 cities and consist of five pollutants. Zeka and Schwartz (2004) attempted to recover the independent effect of each pollutant, controlling for the other four pollutants. In this application, the five pollutants were paired with each other, and hence for each pollutant, the authors obtained four different (and, in principle, unbiased) estimates relating that pollutant to daily mortality. An issue that arises is how best to combine the four health effect estimates for each pollutant, which could improve the precision of the overall estimate of each health effect. Although there are statistical methods of doing so (e.g. using weighted averages), the Bayesian approach is more attractive for this purpose since it uses all of the data to produce a unique health effect estimate for each pollutant.

Our proposed model can easily be applied to cases in which the outcome is not continuous. In the case where the response is categorical (e.g. a binary or count variable), or heavily skewed, generalized linear models are appropriate. For instance, a Poisson regression with a log link is a common model for analyzing mortality counts. We assume that $Y_{it} \sim \text{Poisson}\ (\lambda_{it})$, where

$$\lambda_{it} = \exp(\beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it}).$$

To fit the linear models considered in this section, we used the software package WinBUGS 1.4. This software implements Monte Carlo Markov chain (MCMC) methods for simulating the posterior distributions of all model parameters, given the data and the prior distributions for the parameters. The approach, based on Gibbs sampling, successively samples from the full conditional distribution of each parameter, given values of all other parameters in the model. Although the model we considered is a complex, hierarchical model, when the response is continuous it consists only of series of linear regressions, and WinBUGS is an efficient tool in such cases. We also used WinBUGS to implement the hierarchical generalized linear model (e.g. Poisson regression). Model fitting takes longer in this case since our formulation is no longer a series linear models. For both the linear and Poisson regression models, we conducted a simulation study to examine the properties of our proposed Bayesian approach, as implemented in WinBUGS. More details can be found in the next section.

To avoid poor mixing for the correlation coefficients $\rho_{x,i}$, which is a consequence of the constrained parameter space (e.g. $[-1, 1]$) for these parameters, we used the variance stabilizing transformation of Fisher, whereby we assume $z_i = 0.5(1+\rho_{x,i})/(1-\rho_{x,i})$ follows a normal distribution. This transformation removes the restriction on the model parameter space, making the MCMC sampling more efficient.

Results

To compare the performances of the various approaches to fitting models (1) and (2), we performed a simulation study. We considered the standard OLS estimator that does not correct for measurement error, the two-stage approach of Schwartz and Coull, and the Bayesian approach to model fitting. Our first simulation setting corresponds to a linear regression model. For each city $i, i = 1,2,\ldots, 6$, we generated the true covariates $\mathbf{X}_i = (\mathbf{X}_{1i}, \mathbf{X}_{2i})$ from a multivariate normal distribution, with $\sigma^2_{x_1} = \sigma^2_{x_2} = 1$ and the correlation $\rho_{1:2,i}$ between the two exposure variables $\mathbf{X}_{1i}$ and $\mathbf{X}_{2i}$ varying uniformly among specified intervals. We assumed independent $\mathbf{U}_{1i}$ and $\mathbf{U}_{2i}$ and tried different sets of values for $\sigma^2_{u_1}$ and $\sigma^2_{u_2}$. The outcome was randomly generated from a Normal distribution with mean $\beta_1 X_{1i} + \beta_2 X_{2i}$, and variance equal to one. Because of the complex, hierarchical setting, our simulations run slowly in WinBUGS. As a result, we present the results from only $S = 20$ simulated data sets for each setting.

Table 1 presents the different scenarios we fit in the linear regression setting. Scenarios (a) and (b) correspond to linear regression models containing two error-prone exposures only. Scenario (c) adds a confounder measured without error. Here, $\rho_{1,i}$ and $\rho_{2,i}$ represent the correlation between the confounder and $\mathbf{X}_1$ or $\mathbf{X}_2$, respectively, in city i. We note that we do not keep these correlations constant for all cities.

To get good starting values for our sampler, we used the MOM described in the section on "Methods". For each data set, we used 20,000 iterations for burn in, and we ran the chain for an additional 80,000 iterations. We used vague, non-informative priors for the hyperparameters, as defined in the previous section. For all models, we examined convergence of the Markov chain using both graphical and formal approaches (Cowles and Carlin, 2004).

To compare these methods, we calculated estimates of β_j, $j = 1,2$. For each method, we estimated the bias $E\left(\widehat{\widehat{\beta}_j - \beta_j}\right) \approx S^{-1} \sum_{h=1}^{S} \left[(\widehat{\beta}_j)_h - \beta_j\right]$, the mean square error (MSE) $E\left\{\left[\widehat{(\widehat{\beta}_j)_h - \beta_j}\right]^2\right\} \approx S^{-1} \sum_{h=1}^{S} \left\{\left[(\widehat{\beta}_j)_h - \beta_j\right]^2\right\}$, the percentiles of the simulated distributions of each estimator, and the Monte Carlo standard deviation of each estimator.

Table 2 presents selected results from our simulations in the linear regression setting, and Figures 1 and 2 present the simulated distributions of each estimator graphically. Results show that the Bayesian approach outperforms the two-stage approach. The simulated standard deviations of the Bayesian

Table 1. Different scenarios for linear regression simulation.

Scenario	σ_{u_1}	σ_{u_2}	β_1	β_2	$\rho_{x,i}$	$\rho_{1,i}$	$\rho_{2,i}$
(a)	0.3	0.6	1	1	8-fold	—	—
(b)	0.5	1.0	1	1	8-fold	—	—
(c)	0.3	0.6	1	1	8-fold	8-fold	8-fold

posterior mean estimator are much smaller than the two-stage counterparts. This results in a big reduction in MSE, in some cases by more than 80%. Thus, joint estimation of the parameters increases the efficiency of the parameter estimates. Agreeing with results in Zeger et al. (2000) and Schwartz and Coull (2003), estimates from the uncorrected approach are biased, with the coefficient estimator for the more poorly measured variable (e.g. X_2) incurring more bias.

In the section on "Methods," we noted that unlike the two-stage approach, the joint approach does not require the correlations among exposures and confounders to be constant among cities. The results for Scenario (c) illustrate this point. In this scenario, the Schwartz and Coull (2003) estimator of β_2 incurs more bias than the standard estimator, and the interval defined by the range of the 2.5th and 97.5th percentiles of its distribution does not include the true value. This is due to the fact that the correlations between the confounder and the two exposures vary across cities. As shown in Table 2, the MSE of the Bayesian estimator is more than 30 times smaller than that for the two-stage estimator in this case.

To evaluate the efficiency gains obtained by the Bayesian approach in the Poisson setting, we performed another simulation study. For each city $i, i = 1,2,\ldots, N$, we generated the true covariates $\mathbf{X}_i = (\mathbf{X}_{1i}, \mathbf{X}_{2i})$ from a multivariate normal distribution, with $\sigma^2_{x_1} = \sigma^2_{x_2} = 1$ and correlation $\rho_{1:2,i}$ among the two exposure variables $\mathbf{X}_{1i}$ and $\mathbf{X}_{2i}$ varying uniformingly among specified intervals. We considered different values for β_j, $\sigma^2_{u_1}$, $\sigma^2_{u_2}$, and $\rho_{u,i}$. For the outcome, we simulated a random Poisson count using the two exposures.

Table 3 presents the different scenarios considered. We used a baseline Poisson mean of either 22 or 5. For each variable, we considered a log relative risk of either 0 or 0.05 per standard deviation in exposure. We used $N = 6$ cities in all scenarios, except Scenario (h), which is based on $N = 20$ cities. Scenario (f) is the heterogeneous model described in (5). Scenario (i) allows for correlated measurement errors, with this measurement error correlation varying 2-fold across cities. Again, due to the computational burden of the Bayesian approach, for each setting we present the results from only $S = 20$ simulated data sets.

Table 4 presents the results from the Poisson simulations, and Figures 3 and 4 summarize the results graphically. Once again, we see big reductions in the variability of the health effect estimates, and the corresponding decreases in MSE, from the Bayesian approach. The uncorrected estimators are

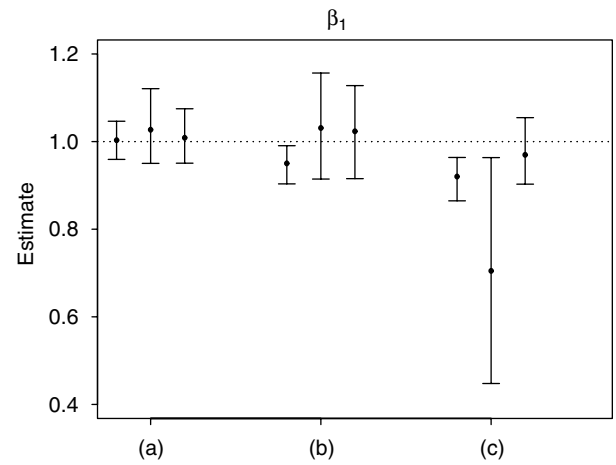

Figure 1. 2.5th, 50th, and 97.5th percentiles of simulated estimates for β_1 in linear regression simulation Scenarios (a–c). For each scenario, the first, second, and third intervals represent the standard, Schwartz and Coull (2003), and Bayesian estimates for β_1.

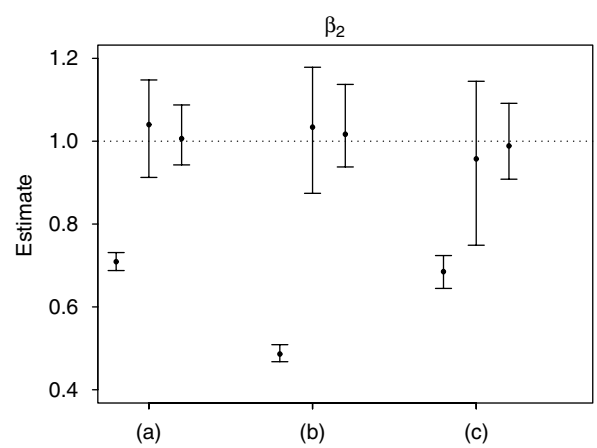

Figure 2. 2.5th, 50th, and 97.5th percentiles of simulated estimates for β_2 in linear regression simulation Scenarios (a–c). For each scenario, the first, second, and third intervals represent the standard, Schwartz and Coull (2003), and Bayesian estimates forβ_2.

Table 2. Bias, Monte Carlo standard deviation and MSE for the parameters β_1 and β_2 in a linear regression model.

Scenario	Parameter	Bias (S.C.)	Bias (B)	SD (S.C.)	SD (B)	MSE (S.C.)	MSE (B)
(a)	β_1	0.027	0.009	0.055	0.038	0.004	0.001
(a)	β_2	0.040	0.006	0.075	0.043	0.007	0.002
(b)	β_1	0.031	0.023	0.076	0.059	0.006	0.004
(b)	β_2	0.034	0.017	0.102	0.064	0.011	0.004
(c)	β_1	−0.295	−0.003	0.132	0.054	0.104	0.004
(c)	β_2	−0.043	0.011	0.105	0.053	0.013	0.003

The results are for the approach in Schwartz and Coull (2003) (S.C.) and for the Bayesian approach (B)

biased, and the estimators corresponding to the variable with more measurement error (e.g. X_2) incur more bias. Of particular interest from a regulatory standpoint are the results from Scenario (g). Under this scenario, only $\mathbf{X}_1$ is causally related with the outcome. Because of the correlation between the two exposures, it is possible that some of the effect of $\mathbf{X}_1$ may be transferred to $\mathbf{X}_2$. Our results show that the Bayesian approach performs very well, reducing the estimated MSE as expected. Under this scenario, some of the parameters in the joint model are not identifiable, since $\beta_2 = 0$. Specifically, in this case, there is no information in the likelihood to separately identify both $\sigma^2_{x_2}$ and $\sigma^2_{u_2}$. However, their sum is identifiable. This is shown in Figure 5. Although the estimates for $\sigma^2_{x_2}$ and $\sigma^2_{u_2}$ are unstable in the Markov chain, the chain based on their sum performs well. The sample correlation between $\sigma^2_{x_2}$ and $\sigma^2_{u_2}$ in this case was -0.97. Besag et al. (1995) and Waller et al. (1997) noted that samplers operating on overparametrized spaces (i.e. having a subset of parameters identified by neither the likelihood nor the prior) are perfectly legitimate, provided that their samples are used only to summarize the posterior distribution of identifiable functions of the parameters. There is no notion of "convergence" for unidentified parameters, and their presence in the sampling order will have no negative effect on the convergence of the functions of interest. Hence, posterior

Table 3. Different scenarios for Poisson regression, with (f) being the heterogeneous model.

Scenario	Baseline count	Cities	σ_{u_1}	σ_{u_2}	β_1	β_2	$\rho_{x,i}$	$\rho_{u,i}$
(d)	22	6	0.3	0.6	0.05	0.05	6-fold	0
(e)	5	6	0.3	0.6	0.05	0.05	6-fold	0
(f)	22	6	0.3	0.6	0.05	0.05	6-fold	0
(g)	22	6	0.3	0.6	0.05	0	6-fold	0
(h)	22	20	0.3	0.6	0.05	0.05	6-fold	0
(i)	22	6	0.3	0.6	0.05	0.05	6-fold	2-fold

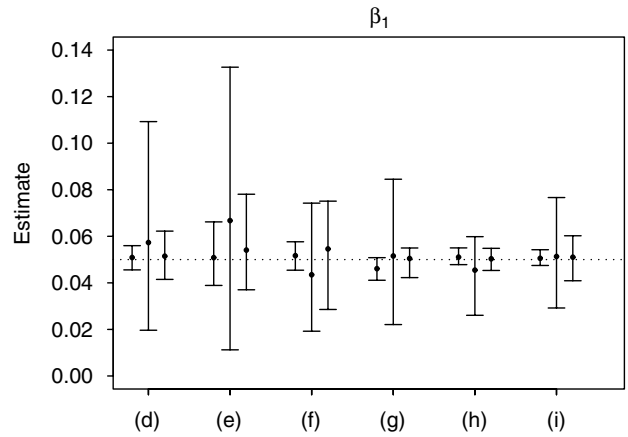

Figure 3. 2.5th, 50th, and 97.5th percentiles of simulated estimates for β_1 in Poisson regression simulation Scenarios (a–c). For each scenario, the first, second, and third intervals represent the standard, Schwartz and Coull (2003), and Bayesian estimates for β_1.

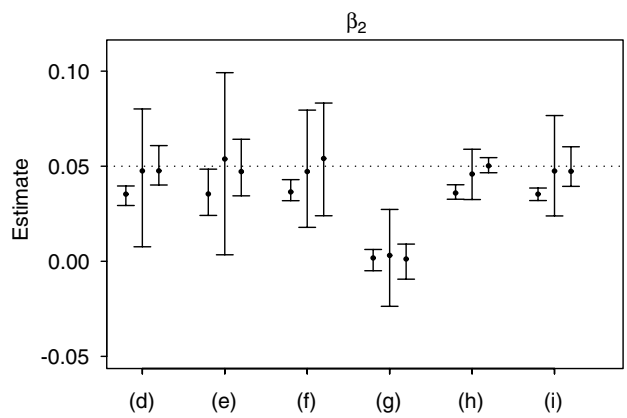

Figure 4. 2.5th, 50th, and 97.5th percentiles of simulated estimates for β_2 in Poisson regression simulation Scenarios (a–c). For each scenario, the first, second, and third intervals represent the standard, Schwartz and Coull (2003), and Bayesian estimates forβ_2.

Table 4. Bias, Monte Carlo standard deviation and MSE for the parameters β_1 and β_2 in a Poisson regression model.

Scenario	Parameter	Bias (S.C.)	Bias (B)	SD (S.C.)	SD (B)	MSE (S.C.)	MSE (B)
(d)	β_1	0.007	0.001	0.031	0.006	9.8e−04	3.3e−05
(d)	β_2	−0.002	−0.002	0.021	0.006	4.2e−03	3.3e−05
(e)	β_1	0.017	0.004	0.036	0.011	1.3e−04	1.3e−04
(e)	β_2	0.004	−0.003	0.027	0.009	0.001	8.3e−05
(f)	β_1	−0.007	0.007	0.020	0.011	3.5e−04	1.4e−04
(f)	β_2	−0.003	0.004	0.019	0.012	3.2e−04	2.8e−04
(g)	β_1	0.052	0.000	0.017	0.004	0.003	1.5e−05
(g)	β_2	−0.047	0.001	0.015	0.005	0.002	2.8e−05
(h)	β_1	−0.005	0.000	0.010	0.003	1.2e−04	6.7e−06
(h)	β_2	−0.004	0.000	0.008	0.002	7.2e−05	4.8e−06
(i)	β_1	0.001	−0.001	0.014	0.006	2.0e−04	3.4e−05
(i)	β_2	−0.003	−0.003	0.014	0.007	2.0e−04	5.0e−05

The results are for the approach in Schwartz and Coull (2003) (S.C.) and for the Bayesian approach (B).

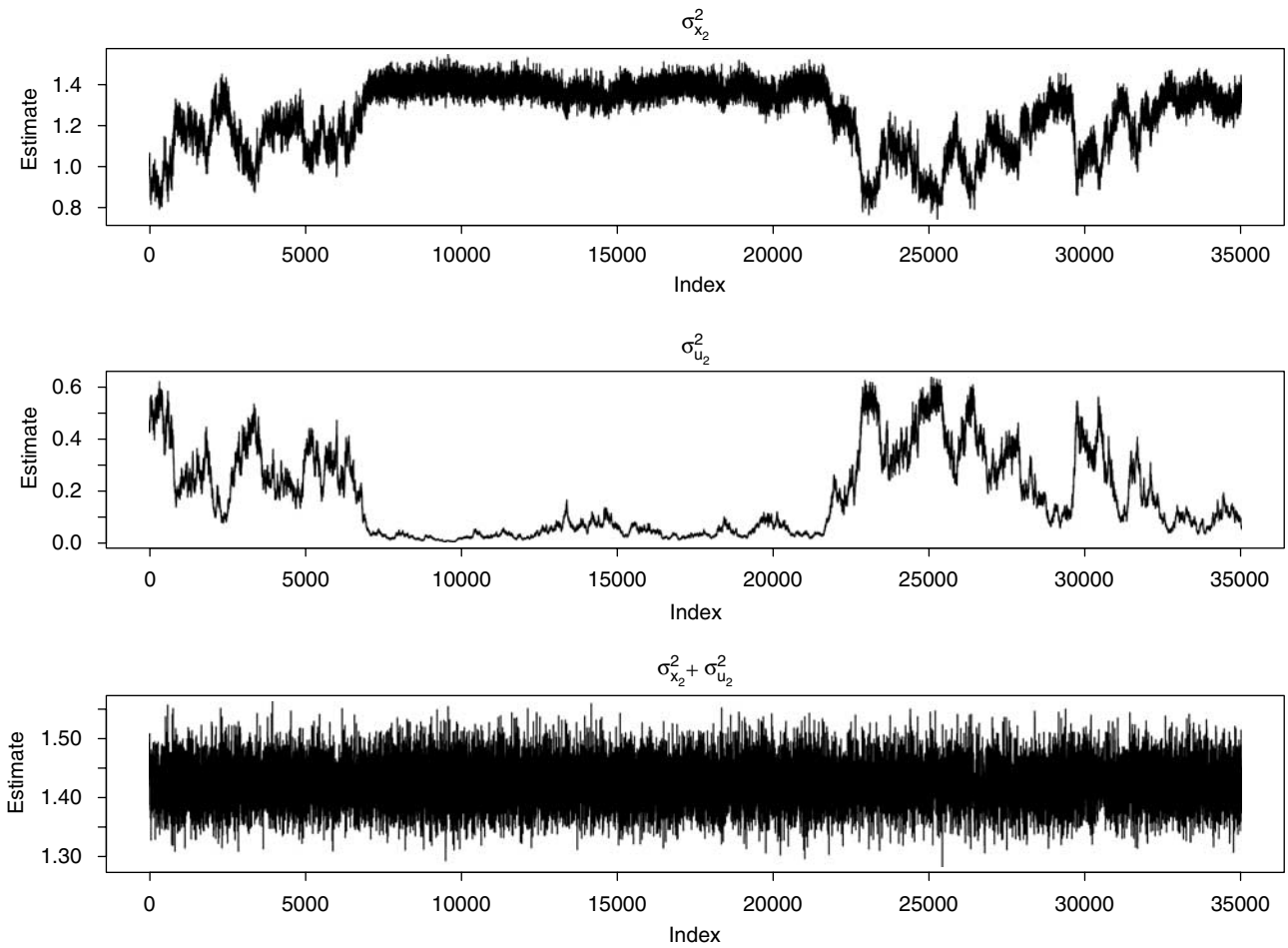

Figure 5. Plots of $\sigma^2_{x_2}$ and $\sigma^2_{u_2}$ and their sum from a fit of Poisson regression Scenario (g).

summaries from an MCMC are valid for identifiable parameters that are of interest (in our case, β_1 and β_2).

Discussion

In this paper, we extended the work of Schwartz and Coull (2003) by improving the efficiency of the measurement error estimators, while maintaining unbiasedness. We showed that, in this framework, all the parameters are identifiable as long as all variables have independent effects on the outcome, and that a joint estimation approach will yield estimates that are more precise than those from two-stage approach. Fitzmaurice et al. (2004) noted that this is generally true in hierarchical models. We found that the Bayesian machinery is very helpful in such complex, hierarchical settings and we used WinBUGS for model fitting. This approach yields samples from the posterior distribution of the parameters of interest.

Moreover, the joint modeling approach easily incorporates covariates measured without error. It is well known that if a variable is subject to measurement error, this error will not only affect the coefficient estimate for that variable, but will also affect the estimates of the other coefficients in the model, even those corresponding to variables measured without error. In this case, the direction of the bias is not easy in general to determine, since it depends on the association structure of the multiple covariates. The joint modeling approach uses information in the covariates measured without error to gain more information about the measurement error variance, ultimately yielding more precise estimators. Another advantage of the fully Bayesian approach is that it can incorporate smoothing in a semiparametric regression setting while controlling for measurement error (Berry et al., 2002; Carroll et al., 2004). Such smoothing techniques have been used extensively in environmental epidemiology, since these methods can adjust for nonlinear effects of temporal, meteorological, and spatial patterns.

Recent studies on the health effects of air pollution have examined the independent effect of as many as five pollutants. Zeka and Schwartz (2004) considered multiple two-pollutant models in such a setting. These authors then

pooled the resulting coefficients for a given pollutant from different models. In the case that one is interested in adjusting for two error-prone exposures and a confounder measured without error, the two-stage approach requires that one assumes the correlations between a confounder and each of the pollutants does not vary by city. These assumptions may not always be realistic, and our Bayesian approach extends naturally to settings involving more than two variables measured with or without error, without requiring extra assumptions. For instance, if one suspects additional potential confounders could be measured with error, this could be built into the model in a natural way by including these covariates in $\mathbf{X}_{it}$ as defined in (1) and (2). This is a very attractive feature of our proposed method, which can be useful in hierarchical studies focusing on multiple correlated pollutants or considering multiple confounders measured either with or without error.

Like all Bayesian approaches, our proposed method is computationally intensive. Part of this is due to the fact that the mixing of the Markov chains used to fit the model can be slow. We have found in previous work (Gryparis et al., 2006) that if one is not interested in estimating the true latent exposure itself (just the health effect associated with that exposure), mixing can be improved by integrating these latent variables out of the model. This practice increases the complexity of the model, since some parameters appear both in the mean and the variance structure of some posterior distributions. Further, the MCMC updates are more complicated, since updating of the model parameters must be done using Metropolis–Hastings, rather than simpler Gibbs sampling, steps. Thus, this practice complicates each updating step, but reduces the number of iterations required for convergence due to the better mixing properties of the chain. This computational complexity is manageable for a given data analysis, as long as the data set contains a small-to-moderate number of cities, say less than 20 or so. The computations presented difficulties in this paper mainly because we investigated its performance via a simulation study, where we applied the method repeatedly. One could extend the applicability of this approach to larger data sets by programming the Markov chain Monte Carlo fitting procedure in a programming language that is faster than WinBUGS, such as the R software package. We have found this to be helpful in other Bayesian settings (Gryparis et al., 2006).

We have shown that the proposed estimators are robust to covariate measurement error, assuming the model is correctly specified otherwise. That is, we assume that the important covariates are included in the model. Of course, the methods do not protect against the possibility that residual confounding still exists due to omission of an important covariate. We note however, that this is true of any regression model, and is not a disadvantage of the proposed procedure.

In our simulation study, we used vague, non-informative but conjugate prior distributions in our Bayesian formulation. Another advantage of the Bayesian approach is that one can incorporate prior knowledge, by choosing appropriately the prior distributions of the hyperparameters. In practice, for the analysis of a given data set, incorporation of prior knowledge in the estimation procedure can add a great deal of information to the analysis.

Conceptually, one could also construct a frequentist approach to fitting the joint model defined by (1) and (2). Good starting values for all parameters would be necessary, and the simple MOM estimates may be useful for this purpose. We expect the joint likelihood approach to perform similarly to our proposed Bayesian approach that uses non-informative priors. One complication in the frequentist setting is that all model parameters must be identifiable, or an algorithm for obtaining the maximum likelihood estimates will not converge. A second difference between the Bayesian and frequentist approaches is that the Bayesian approach results in exact standard errors while a classical approach relies on large sample asymptotics.

Acknowledgements

We thank Thomas Bateson and Jee-Young Kim from the U.S. EPA for the invitation to participate in the workshop on which this special issue is based. We also thank two referees and the Editor for helpful comments that improved the manuscript. This research was partially supported by NIEHS Grants ES012044, EPA Grant R832416, and American Chemistry Council Grant 2843. It has not been formally reviewed by the American Chemistry Council and the views expressed in this document are solely those of the authors.

References

Berry S.M., Carroll R.J., and Ruppert D. Bayesian smoothing and regression splines for measurement error problems. *J Am Stat Assoc* 2002: 97: 160–169.

Besag J., Green P., Higdon D., and Mengersen K. Bayesian computation and stochastic systems (with discussion). *Stat Sci* 1995: 10: 3–66.

Carroll R.J., Crainiceanu C.M., Tosteson T.D., and Karagas M.R. Nonlinear and nonparametric regression and instrumental variables. *J Am Stat Assoc* 2004: 99: 736–750.

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

Cowles M.K., and Carlin B.P. Markov chain Monte Carlo convergence diagnostics: a comparative study. *J Am Stat Assoc* 2004: 91: 883–904.

Fitzmaurice G.M., Laird N.M., and Ware J. *Applied Longitudinal Analysis* 2004, Wiley, New York.

Gryparis A., Paciorek C.J., and Coull B.A. Measurement error caused by spatial misalignment in environmental epidemiology. *Harvard University Biostatistics Working Paper Series. Paper 59* 2006: http://www.bepress.com/harvardbiostat/paper59.

Judge G.G., Griffiths W.E., Hill R.C., Ltkepohl H., and Lee T.C. *The Theory and Practice of Econometrics* 1985, Wiley, New York.

Sarnat J.A., Schwartz J., Catalano P.J., and Suh H.H. Gaseous pollutants in particulate matter epidemiology: confounders or surrogates? *Environ Health Perspect* 2001: 109: 1053–1061.

Schwartz J., and Coull B.A. Control for confounding in the presence of measurement error in hierarchical models. *Biostatistics* 2003: 4: 539–553.

Waller L.A., Carlin B.P., Hong X., and Gelfand A.E. Hierarchical spatio-temporal mapping of disease rates. *J Am Stat Assoc* 1997: 92: 607–617.

Zeger S.L., Thomas D.C., Dominici F., Samet J.M., Schwartz J., Dockery D., and Cohen A. Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ Health Perspect* 2000: 108: 419–426.

Zeka A., and Schwartz J. Estimating the independent effects of multiple pollutants in the presence of measurement error: an application of a measurement-error-resistant technique. *Environ Health Perspect* 2004: 112: 1686–1690.

Appendix

Here, we provide the WinBUGS code that one can use to fit a linear health effects model using two correlated exposures. The following code can be easily modified to accommodate different scenarios (e.g. adding more covariates with or without measurement error, categorical outcomes, etc.).

```
model
{
for (o in 1:n.all)
{
mean[o] <- beta0 + beta1*x[o,1] + beta2*x[o,2]
w[o,1] ~ dnorm(x[o,1], tau.u1)
w[o,2] ~ dnorm(x[o,2], tau.u2)
y[o] ~ dnorm(mean[o],tau.y[city[o]])
w[o,1:2] ~ dmnorm(mu.x[city[o],],
tau.all[(2*city[o] 1):(2*city[o]),])
}
for (i in 1:n.city)
{
mu.x[i,1] ~ dnorm(0, 0.001)
mu.x[i,2] ~ dnorm(0, 0.001)
oriz[i] <- sigma2.x1*sigma2.x2*(1-pow(rho[i],2))
tau.all[(2*i-1),1] <- sigma2.x2/oriz[i]
tau.all[(2*i),2] <- sigma2.x1/oriz[i]
diag.el[i] <- -pow(sigma2.x1,.5)*pow(sigma2.x2,.5)*
rho[i]/oriz[i]
tau.all[(2*i-1),2] <- diag.el[i]
tau.all[(2*i),1] <- diag.el[i]
rho[i] ~ dunif(-.99,.99)
tau.y[i] ~ dgamma(0.01,0.01)
}
beta0 ~ dnorm(0,0.001)
beta1 ~ dnorm(0,0.001)
beta2 ~ dnorm(0,0.001)
tau.u1 ~ dgamma(0.01,0.01)
tau.u2 ~ dgamma(0.01,0.01)
tau.x1 ~ dgamma(0.01,0.01)
tau.x2 ~ dgamma(0.01,0.01)
sigma2.x1 <- 1/tau.x1
sigma2.x2 <- 1/tau.x2
}
```