Harvard University

Harvard University Biostatistics Working Paper Series

Year 2008 *Paper* 87

Measurement Error Caused by Spatial Misalignment in Environmental Epidemiology

Alexandros Gryparis[*] Christopher J. Paciorek[†] Ariana Zeka[‡]

Joel Schwartz[**] Brent A. Coull[††]

[*]Harvard University

[†]Harvard School of Public Health, paciorek@hsph.harvard.edu

[‡]Brunel University

[**]Harvard University, jschwrtz@hsph.harvard.edu

[††]Harvard University, bcoull@hsph.harvard.edu

Measurement error caused by spatial misalignment in environmental epidemiology

Alexandros Gryparis,[1,*] **Christopher J. Paciorek,**[1] **Ariana Zeka,**[2] **Joel Schwartz**[3] **and Brent A. Coull**[1]

[1] Department of Biostatistics, Harvard University, Boston, MA 02115, U.S.A.

[2] Institute for the Environment, Brunel University, West London, U.K.

[3] Department of Environmental Health, Harvard University, Boston, MA 02115, U.S.A.

September 8, 2008

SUMMARY. In many environmental epidemiology studies, the locations and/or times of exposure measurements and health assessments do not match. In such settings, health effects analyses often use the predictions from an exposure model as a covariate in a regression model. Such exposure predictions contain some measurement error, as the predicted values do not equal the true exposures. We provide a framework for spatial measurement error modeling, showing that smoothing induces a Berkson-type measurement error with non-diagonal error structure. From this viewpoint, we review existing approaches to estimation in a linear regression health model, including direct use of the spatial predictions and exposure simulation, and explore some modified approaches, including Bayesian models and out-of-sample regression calibration, motivated by measurement error principles. We then extend this work to the generalized linear model framework for health outcomes. Based on analytical considerations and simulation results, we compare the performance of all these approaches under several spatial models for exposure. Our comparisons underscore several important points. First, exposure simulation can perform very poorly under certain realistic scenarios. Second, the rel-

[*] *email:* alexandros@post.harvard.edu

ative performance of the different methods depends on the nature of the underlying exposure surface. Third, traditional measurement error concepts can help to explain the relative practical performance of the different methods. We apply the methods to data on the association between levels of particulate matter and birthweight in the greater Boston area.

Key words: spatial misalignment, measurement error, predictions, air pollution

1. Introduction and scientific motivation

Exposure assessment studies have shown that there exist important factors, such as traffic conditions, point sources of pollution, and urban building canyon effects, that induce spatial variability in pollution levels. With the advent of Geographic Information System (GIS)-based modeling, researchers have begun to focus on spatial variability in air pollution and its relationship with human health (Berhane et al., 2004; Zidek et al., 2004; Kunzli et al., 2005; Gryparis et al., 2007). Such spatial analyses have several advantages over studies that assign exposure readings from a central-site monitor to all study participants. First, spatial analyses do not assume that exposure is constant over the region of interest, thereby reducing exposure measurement error that would otherwise lead to a loss of power. Second, in the case of chronic diseases, analyses rely primarily on exposure heterogeneity induced by spatial variability. Finally, it is now widely recognized that air particulates are a complex mixture of multiple sources of pollution, with pollution from each source having a distinct chemical profile and perhaps different toxicity. Because pollutants from different sources have different spatial distributions, with regional pollutants (e.g., sulfates from coal-fired power plants) being more homogeneous over space and local sources (e.g., black carbon from traffic emissions) demonstrating higher spatial variability, incorporation of the spatial variability of local pollutants in a health effects analysis may help separate health effects from different sources.

In many such studies, the locations of the exposure data and those of the health data do not coincide. Standard regression methods cannot be applied to such misaligned data. To overcome this problem, several methods have been proposed. Most approaches involve directly using predictions from statistical exposure models that incorporate spatial structure (Shaddick and Wakefield, 2002; Kunzli et al., 2005; Gryparis et al., 2007). Higgins et al. (1997) used polynomial regression to generate covariate predictions when outcomes and covariates were misaligned in time. Waller and Gotway (2004) used kriging to predict exposures and used resampling to account for the uncertainty in using the predictions in place of the true values. They classified predicted exposures as high, medium or low, and fitted multiple health regressions using the simulated categorical exposures as covariates. Kunzli et al. (2005) assigned exposure values for subject-specific locations derived from a geostatistical model and used weighted least squares in the subsequent health effects model, with the weights specified as the inverse of the standard errors from the exposure kriging model. For this same problem, Madsen et al. (2008) considered both a generalized least squares estimator with a bootstrap-type variance estimator as well as a maximum likelihood approach that jointly fits the exposure and health models.

In this paper we evaluate and compare approaches to fitting linear and logistic health models with predicted exposures, including approaches specifically suggested for this setting as well as several modified approaches motivated by measurement error principles. We first use a very simple linear model to illustrate measurement error issues associated with spatially misaligned exposure and health point data. This simple structure is instructive in demonstrating the relative strengths and weaknesses of the various methods proposed for dealing with this type of data, which commonly arises in chronic and within-urban area studies of the health effects of air pollution. We then consider nonlinear models under the generalized linear model framework, focusing on logistic regression.

The structure of this paper is as follows. In Section 2 we introduce our notation. In Section 3 we discuss how smoothing converts classical measurement error to Berkson error and the implications of this. In Sections 4 and 5 we describe and analytically evaluate multiple approaches to the problem for continuous and binary outcomes, respectively. In Section 6 we present a simulation study to further compare the methods. In Section 7 we describe an application of the methods to data on the association between traffic particle levels and birthweights in the greater Boston area. We conclude with discussion in Section 8.

2. Modeling framework

To introduce our notation, let $\boldsymbol{X}$ be the vector of the true exposures and $\boldsymbol{W}$ be the vector of its error-prone, but not misaligned, measurements. Moreover, let $\boldsymbol{S}$ be the vector of smoothed estimates of $\boldsymbol{X}$ based on $\boldsymbol{W}$, $\boldsymbol{U} = \boldsymbol{W} - \boldsymbol{X}$ the vector of measurement errors, $\boldsymbol{V} = \boldsymbol{X} - \boldsymbol{S}$ the vector of the error after the smoothing procedure and $\boldsymbol{Y}$ the health response. Let $(\cdot)^*$ indicate values at locations without exposure observations; for example, $\boldsymbol{Y}^*$ is the vector of health observations at locations without exposure data.

In what follows, we assume that Y_i^* given X_i^* and $\mathbf{Z}_i^*$ are independent random variables having a distribution in the natural exponential family (McCullagh and Nelder, 1989). Let $\mu_i^* = E(Y_i^* | X_i^*, \boldsymbol{Z}_i^*)$. We assume the following model holds:

$$g\left(\mu_i^*\right) = \beta_0 + \beta_1 X_i^* + \boldsymbol{\beta}_z \mathbf{Z}_i^*, \quad i = 1, 2, ..., n_y, \tag{1}$$

$$W_i = X_i + U_i, \quad \text{where } U_i \sim N(0, \sigma_u^2), \quad i = 1, 2, ..., n_w, \tag{2}$$

where $g(x)$ is a monotonically increasing link function, β_1 and $\boldsymbol{\beta}_z$ are unknown parameters, and the measurement errors U_i are independent of Y_i^* given X_i^* and Z_i^*. In the above equation, X_i^* is the exposure (e.g., air pollutant level) at the residence of the i^{th} subject, over some biologically relevant period of interest, and $\boldsymbol{Z}_i^*$ is a $q \times 1$ vector of covariates measured without error. In this work we treat $\boldsymbol{X}$ as correlated in space, although in spatio-temporal settings $\boldsymbol{X}$ could also be serially correlated over time. W_i represents an exposure measurement, which may or may not differ from X_i, depending on whether or not instrument error is present. In the misalignment scenario, the variable $\boldsymbol{X} = (X_1, X_2, \ldots, X_{n_w})^T$ is measured with error by $\boldsymbol{W} = (W_1, W_2, \ldots, W_{n_w})^T$ at different points in space than the variable $\boldsymbol{Y}^* = (Y_1^*, Y_2^*, \ldots, Y_{n_y}^*)^T$. Hence, to estimate exposure we obtain smoothed predictions $\boldsymbol{S}^* = (S_1^*, S_2^*, \ldots, S_{n_y}^*)^T$ of the unobserved $\boldsymbol{X}^* = (X_1^*, X_2^*, \ldots, X_{n_y}^*)^T$ from an exposure model. Scientific interest then focuses on β_1, the regression coefficient relating exposure and health.

An important aspect is the nature of the error in the stochastic exposure process. We decompose the process as

$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{X}^* \end{pmatrix} = \boldsymbol{g} + \boldsymbol{\delta},$$

where $\boldsymbol{g}(\cdot)$ represents a smooth spatial surface, and $\boldsymbol{\delta}(\cdot)$ is additive uncorrelated error with variance σ_δ^2 that accounts for fine-scale heterogeneity in the exposure. In this case, the measurement error $\boldsymbol{U}$ represents instrument error. Unless multiple measurements at a given site and time are available, one cannot resolve the fine-scale heterogeneity $\boldsymbol{\delta}$ in the presence of $\boldsymbol{U}$, as the model cannot inform both σ_u^2 and σ_δ^2 (Cressie, 1993; page 59). In air-pollution studies we believe that most of the unexplained variability is fine-scale heterogeneity and not instrument error, such that $\sigma_u^2 \approx 0$ and $Var(\delta_i + U_i) \approx \sigma_\delta^2$.

3. Smoothing-induced Berkson error

In this section we argue that the plug-in approach that uses smoothed predicted values of exposure as covariates in a health effects model is a form of regression calibration that produces a Berkson structure (Carroll et al., 1995) in the health model. Exposure estimates most often are generated using one of the many approaches to spatial smoothing, such as kriging and its extensions (Cressie, 1993), Gaussian process modeling and Bayesian smoothing (Gaudard et al., 1999, Banerjee et al., 2004), penalized regression splines (Kammann and Wand, 2003; Ruppert et al., 2003; Gryparis et al., 2007), and kernel smoothing (Hobert et al., 1997), among others. For concreteness, consider a Bayesian framework in which we place a Gaussian process prior on $\boldsymbol{X}(\cdot)$: $\boldsymbol{X}(\cdot) \sim GP\left[\boldsymbol{\mu}(\cdot), \boldsymbol{R}(\cdot)\right]$. Hence,

$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{X}^* \end{pmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{pmatrix}\right].$$

The interim posterior (before any health analysis) for the conditional distribution of $\boldsymbol{X}^*$ given $\boldsymbol{W}$ is:

$$\boldsymbol{X}^*|\boldsymbol{W} \sim N\left(\boldsymbol{\mu}_2 + \boldsymbol{R}_{21}(\boldsymbol{R}_{11} + \sigma_u^2\boldsymbol{I})^{-1}(\boldsymbol{W} - \boldsymbol{\mu}_1), \boldsymbol{R}_{22} - \boldsymbol{R}_{21}(\boldsymbol{R}_{11} + \sigma_u^2\boldsymbol{I})^{-1}\boldsymbol{R}_{12}\right). \quad (3)$$

In a measurement error framework, the interim posterior mean takes the form of a regression calibrator, representing the mean of the unobserved covariate given the observations $\boldsymbol{W}$. As all regression calibrators do, use of this estimator turns around the conditioning and yields a Berkson framework whereby the distribution of an unknown value X_i^* is centered around the posterior mean, as shown in (3). The term multiplying $\boldsymbol{W}$ in (3) is the spatial analogue of $\frac{\sigma_x^2}{\sigma_u^2+\sigma_x^2}$, which is the familiar correction factor in the simplest independent measurement error setting, but the covariance accounts for the spatial covariance structure. Assuming the covariance structure is known, then as in a standard Berkson model the OLS estimator based on a regression model using $E(X_i^*|\boldsymbol{W})$ as a covariate is unbiased. We can write

$$X_i^* = E(X_i^*|\boldsymbol{W}) + V_i^*, \quad (4)$$

where $\boldsymbol{V}^* = (V_1^*, V_2^*, \ldots, V_{n_y}^*)^T$ has mean zero and variance-covariance matrix $\boldsymbol{\Sigma}^*$ equal to the posterior variance given above. For a given degree of smoothing, other smoothers should give a similar decomposition. That is, if the data are really generated from a Gaussian process (GP) with known variance components and we use the BLUPs for our exposure predictions, then the Berkson error analogy holds exactly. However, in reality the data do not come from a Gaussian process with known variance components, and so this analogy does not hold exactly. Other smoothing techniques will create a structure analogous to Berkson error, in that the smoothed predictions will have a smaller variance than the observed data and the covariance of V_i^* and $E(X_i^*|\mathbf{W})$ will be small. Thus, the analytic results obtained in the simple GP setting for which exact results exist lend insights into the likely performance of predictions obtained by other smoothers generally. For instance, the kriging estimator is the Best Linear Unbiased Predictor (BLUP) of $\boldsymbol{X}^*$ and is equivalent to (3). Similarly, estimated smooths from regression splines, including spatial smoothing, are also BLUPs within a mixed-model framework (Ruppert et al., 2003). Thus, each approach will approximately produce a decomposition $\boldsymbol{X}^* = \boldsymbol{S}^* + \boldsymbol{V}^*$, in which $\boldsymbol{V}^*$ is orthogonal to $\boldsymbol{S}^*$, as in (4). For an empirical example of such structure, see Paciorek et al. (2008). In the case of spatial smoothing, with $\boldsymbol{X}^* = \boldsymbol{S}^* + \boldsymbol{V}^*$, note that the residual term, $\boldsymbol{V}^*$, does not have a diagonal covariance structure. The uncertainty in $\boldsymbol{X}^*$, as captured by the covariance matrix $\boldsymbol{\Sigma}^* = Var(\boldsymbol{X}^*|\boldsymbol{W})$, is spatially correlated and heteroscedastic. Note that $\boldsymbol{\Sigma}^*$ should include any component σ_δ^2 but not σ_u^2.

For the general health model (1), even when the variance components of the spatial exposure process $\boldsymbol{g}(\cdot)$ are known, standard approaches to estimation do not yield unbiased estimates of β_1 (Carroll et al., 1995). However, closed-form expressions for this bias under the most general form of this model are unavailable, except for certain special cases. We now focus on two such special cases in the following two sections.

4. Linear Health Effect Model

We now consider the special case of (1) when Y_i^* is normally distributed. Interest focuses on the linear regression model

$$Y_i^* = \beta_0 + \beta_1 X_i^* + \boldsymbol{\beta}_z \mathbf{Z}_i^* + \varepsilon_i, \quad \text{where } \epsilon_i \sim N(0, \sigma_\epsilon^2), \quad i = 1, 2, ..., n_y. \tag{5}$$

We assume the errors, ϵ_i, are independent of the measurement errors, U_i. The remainder of this section describes two existing approaches, a plug-in estimator and exposure simulation, as well as two approaches, regression calibration and Bayesian methods, drawn from the measurement error literature but not yet applied in the spatial setting. Section 6 compares these approaches in a simulation study. We also considered two other approaches, standard weighted least squares and an iterative generalized least squares method, but relegate discussion of these to Section A of the supplementary material to this article, available at *Biostatistics* online at `http://www.biostatistics.oxfordjournals.org`.

4.1 *Plug-in Approach*

The plug-in approach fits the exposure model for $[\boldsymbol{X}^*|\boldsymbol{W}]$ and uses the predictions $\boldsymbol{S}^*$ as a covariate in the health model, which is fitted using ordinary least squares (OLS). By using $\boldsymbol{S}^*$ instead of $\boldsymbol{X}^*$ in the health model, we induce correlation: $\boldsymbol{Y}^* = \beta_0+\beta_1\boldsymbol{X}^*+\boldsymbol{\epsilon} = \beta_0+\beta_1(\boldsymbol{S}^*+\boldsymbol{V}^*)+\boldsymbol{\epsilon} = \beta_0+\beta_1\boldsymbol{S}^*+\boldsymbol{\eta}$, where $\boldsymbol{\eta} = \beta_1\boldsymbol{V}^*+\boldsymbol{\epsilon}$. The new error term $\boldsymbol{\eta}$ no longer has a diagonal covariance matrix. Thus, although the OLS estimator for β_1 is unbiased, the variance estimator is incorrect, since it does not account for the correlated, heteroscedastic error structure (Carroll et al, 1995; page 63). To address this, one could use generalized least squares (GLS) to account for the induced covariance in the health model. Although it seems intuitive to use the uncertainty estimates from the exposure model (i.e., the elements of the diagonal in $\boldsymbol{\Sigma}^*$) as the weights for the health model via simple weighted least squares, these are not the correct weights under the induced Berkson model. Since $\boldsymbol{\eta} = \beta_1\boldsymbol{V}^* + \boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon} \sim N(0, \sigma^2_\epsilon \boldsymbol{I}_{n_y})$, it follows that the residual variance for the health model is given by: $\beta_1^2\boldsymbol{\Sigma}^* + \sigma^2_\epsilon\boldsymbol{I}_{n_y}$.

In practice the variance components or smoothing parameters are not known, and we must estimate the parameters that govern the degree of smoothing. If we oversmooth, the OLS estimator from the health model may be biased (Wakefield and Shaddick, 2006), with more bias occurring in situations in which it is difficult to estimate the appropriate amount of smoothing in the exposure model. Such scenarios include sparse monitoring data in a sub-region or exposures that are very heterogeneous in space. Bias can also occur if the residual, $\boldsymbol{V}^*$, is correlated with confounders, $\boldsymbol{Z}^*$, in the health model, such as might result from correlation of confounders and exposure at small spatial scales.

4.2 *Exposure Simulation Approach*

Some have proposed an exposure simulation approach as an attempt to correct the variance of the plug-in estimator. Under this approach, M samples $\boldsymbol{X}^*_{(t)} = \boldsymbol{S}^* + \boldsymbol{V}^*_{(t)}$, $t = 1, 2, ..., M$, are generated from the estimated distribution of exposure, given the data, and each of these M samples is used as a predictor to fit the health model. This yields M health effect estimates, $\widehat{\beta}_{1(t)}$, $t = 1, 2, ..., M$, which are then averaged to obtain an overall estimate. The corresponding variance of $\widehat{\beta}_1$ is equal to $Var(\widehat{\beta}_1) = Var(E(\widehat{\beta}_{1(t)})) + E(Var(\widehat{\beta}_{1(t)}))$. Although Waller and Gotway (2004) [page 406] used quantiles of simulated exposures, we consider direct use of the simulated exposures in the health model.

A simple analysis based on the Berkson framework we presented in Section 3 (supported by the results of our simulations in Section 6) indicates that rather than adjusting for the uncertainty induced by using the predicted exposures, such an approach induces bias in a linear health model. To illustrate, consider samples ${\boldsymbol{X}^*}_{(t)} = \boldsymbol{S}^* + \boldsymbol{V}^*_{(t)}$, $t = 1, 2, ..., M$. We have argued that use of $\boldsymbol{S}^*$ should result in nearly unbiased health effect estimators. By adding $\boldsymbol{V}^*_{(t)}$, one adds error to a variable that, if used on its own, yields unbiased estimates in the health model. Hence, this approach converts the problem back to the classical measurement error setting, where instead of using a covariate that yields unbiased results, measurement error in the covariate produces biased estimates. The size of the bias depends on the size of $\boldsymbol{\Sigma}^*$. Therefore, we recommend against this approach. One can also view this problem from a multiple imputation perspective, where an appropriate exposure simulation scheme should take into account all available data and hence resample from the posterior predictive distribution of $[\boldsymbol{X}^*|\boldsymbol{W}, \boldsymbol{Y^*}]$ (Rubin, 1987). By using the conditional distribution $[\boldsymbol{X}^*|\boldsymbol{W}]$, we discard the information for $\boldsymbol{X}^*$ in $\boldsymbol{Y}^*$, resulting in an incorrect imputation scheme.

4.3 *Out-Of-Sample Regression Calibration Estimator (RC-OOS)*

In many cases it may be that spatial smoothing results in $\boldsymbol{S}^*$ being a biased estimate of the unknown expectation $E(\boldsymbol{X}^*|\boldsymbol{W})$. Departure from the simple $\boldsymbol{X} = \boldsymbol{S} + \boldsymbol{V}$ measurement error model may occur for several reasons, including poor estimation of the smoothing parameters and sparse exposure data in regions with health data.

Let $(\cdot)^{**}$ indicate values at locations where exposure is observed, but held out of the main model fitting for assessing model prediction. Hence $\boldsymbol{X}^{**}$ is the vector of exposure measures that are held out from the exposure model, $\boldsymbol{S}^{**}$ the smoothed estimates that correspond to these locations based on the remaining exposure data and $\boldsymbol{Z}^{**}$ is the matrix of covariates measured without error that correspond to these locations. We use the held out data to fit a calibration of $\boldsymbol{X}^{**}$ to $\boldsymbol{S}^{**}$ and $\boldsymbol{Z}^{**}$. We assume a simple measurement error model, in the spirit of Carroll et al. (1995) [page 8], of the form:

$$X_i^{**} = \gamma_0 + \gamma_1 S_i^{**} + \boldsymbol{\gamma}_z^T \boldsymbol{Z}_i^{**} + \epsilon_{x,i}, \tag{6}$$

where $E(\epsilon_{x,i}) = 0$ and $Var(\epsilon_{x,i}) = \sigma_x^2$. By fitting model (6), we obtain parameter estimates $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_0, \widehat{\gamma}_1, \widehat{\boldsymbol{\gamma}}_z^T)^T$, which we use to calibrate the predicted exposures $\boldsymbol{S}^*$ at the locations of interest. Hence, this method is an out-of-sample regression calibration (RC-OOS) approach, which has been described by Thurston et al. (2003) in the context of a study design that corrects for measurement error by incorporating external validation data. In our case we use it to correct for possible bias in our predictions. Define the matrices:

$$\boldsymbol{\gamma} = \begin{pmatrix} 1 & \gamma_0 & \mathbf{0}_{1\times q} \\ 0 & \gamma_1 & \mathbf{0}_{1\times q} \\ \mathbf{0}_{q\times 1} & \boldsymbol{\gamma}_z & \boldsymbol{I}_{q\times q} \end{pmatrix} \text{ with } \boldsymbol{\gamma}^{-1} = \begin{pmatrix} 1 & -\gamma_0/\gamma_1 & \mathbf{0}_{1\times q} \\ 0 & 1/\gamma_1 & \mathbf{0}_{1\times q} \\ \mathbf{0}_{q\times 1} & -\boldsymbol{\gamma}_z/\gamma_1 & \boldsymbol{I}_{q\times q} \end{pmatrix}.$$

Then, using (6), we have that $E(X_i^*) = \gamma_0 + \gamma_1 S_i^* + \boldsymbol{\gamma}_z^T \boldsymbol{Z}_i^*$. We replace $(\gamma_0, \gamma_1, \boldsymbol{\gamma}_z)$ by $(\widehat{\gamma}_0, \widehat{\gamma}_1, \widehat{\boldsymbol{\gamma}}_z)$ and then calculate new estimated exposures, $\widehat{X}_i^* = \widehat{\gamma}_0 + \widehat{\gamma}_1 S_i^* + \widehat{\boldsymbol{\gamma}}_z \boldsymbol{Z}_i^*$, which we plug into the health model. This is equivalent to estimating $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\gamma}}^{-1} \widehat{\boldsymbol{\beta}}_{plug-in}$, where $\boldsymbol{\beta} = (\beta_0\ \beta_1\ \boldsymbol{\beta}_z^T)^T$ and $\widehat{\boldsymbol{\beta}}_{plug-in}$ is the estimate from the plug-in model. It can be shown

(Thurston et al., 2003) that the corrected $\boldsymbol{\beta}$ estimate is equal to

$$\widehat{\boldsymbol{\beta}} = \widehat{\gamma}^{-1}(\boldsymbol{D}^{*T}\boldsymbol{D}^{*})^{-1}\boldsymbol{D}^{*}\boldsymbol{Y}, \tag{7}$$

where $\boldsymbol{D}^* = [\mathbf{1}|\boldsymbol{S}^*|\boldsymbol{Z}^*]$ and $\mathbf{1}$ is a vector of 1's. The variance of $\widehat{\boldsymbol{\beta}}$ can be derived using either the sandwich method or a Taylor series expansion (Thurston et al., 2003). Both methods result in

$$Var(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\gamma}^{-1}[\beta_1^2\sigma_x^2(\boldsymbol{D}^{**T}\boldsymbol{D}^{**})^{-1} + \sigma_\epsilon^2(\boldsymbol{D}^{*T}\boldsymbol{D}^{*})^{-1}](\boldsymbol{\gamma}^{-1})^T. \tag{8}$$

In practice, especially if the number of exposure locations is not large enough to support holding out a sizable subset of locations, one could consider cross-validation to estimate the predicted exposures S_i^{**}, although simulation studies we conducted (results not shown) indicate that this approach does not perform as well as out-of-sample validation.

4.4 *Bayesian Approaches*

In the fully Bayesian approach, one fits a joint model for the health and the exposure data. A fully Bayesian measurement error model adjusts in a natural way for the extra uncertainty associated with using the predicted exposure values in the health model and provides us with a correct variance estimate (Berry et al., 2002). Also, heteroscedasticity and correlation among exposure values are naturally incorporated in a Bayesian model through the uncertainty in $\boldsymbol{X}^*$. The fully Bayesian model samples from the distribution $[\boldsymbol{X^*}, \boldsymbol{\beta}|\boldsymbol{Y^*}, \boldsymbol{W}, \boldsymbol{Z^*}]$. Thus, in this model, when we update the unobserved exposure $\boldsymbol{X}^*$ we use information from the health data $\boldsymbol{Y}^*$ along with that from the proxy $\boldsymbol{W}$, which results in a proper multiple imputation scheme (Little, 1992).

In practice, we expect the number of exposure monitoring locations to be relatively small compared to the locations from which we have health data. In such cases the health data could be very influential in determining the exposure predictions (Shaddick and Wakefield, 2002; Wakefield and Shaddick, 2006). If there are outliers in the health outcomes, especially if they correspond to locations for which we do not have adequate exposure information, then they could strongly affect the exposure surface. Other forms of model misspecification in either the exposure or health model could also result in the exposure surface being overly influenced by the health observations, a situation similar to that in Yucel and Zaslavsky (2005) and for which the approach of 'cutting feedback' has been suggested (Rougier, 2008). Poor estimation of the exposure surface could in turn affect estimates of the health effects.

An alternative to the fully Bayesian approach is a two-stage Bayesian approach. In this approach, the first stage model is the exposure model $[\boldsymbol{X}^*|\boldsymbol{W}] \propto [\boldsymbol{W}|\boldsymbol{X}^*][\boldsymbol{X}^*]$, and the second stage model is the health model $[\boldsymbol{X^*}, \boldsymbol{\beta}|\boldsymbol{W}, \boldsymbol{Y^*}, \boldsymbol{Z^*}] \propto [\boldsymbol{Y}^*|\boldsymbol{X^*}, \boldsymbol{W}, \boldsymbol{Z^*}, \boldsymbol{\beta}][\boldsymbol{X}^*|\boldsymbol{W}][\boldsymbol{\beta}]$, where we use the interim posterior from the exposure model $[\boldsymbol{X}^*|\boldsymbol{W}]$ as a prior distribution for $\boldsymbol{X}^*$ in the health model. The main difference between the two Bayesian approaches is that, in the two-stage Bayesian approach, we use a normal distribution for the interim posterior of $\boldsymbol{X}^*$ and numerically estimate its covariance matrix, whereas the fully Bayesian approach uses the exact version of this distribution, by virtue of fitting the models jointly. The difference between this approach and the plug-in model is that the prior for $\boldsymbol{X}^*$ in the health model, which is the posterior for $\boldsymbol{X}^*$ from the exposure model, accounts for the uncertainty in $\boldsymbol{X}^*$, including correlation and heteroscedasticity. We note that this two-stage approach does not cut feedback between the health observations and the exposure estimates, since the prior distribution for the exposure values is updated in the second stage. When the exposure model is complicated or when one is interested in running multiple epidemiological models, either with different sets of covariates for a single outcome or for multiple outcomes, this two-stage approach has the advantage that one does not have to re-fit the exposure model when running multiple health effect analyses.

5. Generalized Linear Models for Binary Health Outcomes

Interest may also focus on use of exposure predictions from spatially misaligned exposure data in generalized linear models for discrete outcomes (e.g., a binary or a count variable). Again consider the Berkson error structure $X_i^* = E(X_i^*|\boldsymbol{W}) + V_i^*$. Unlike the linear regression case, even under the correct amount of smoothing (e.g., if the variance components of the spatial exposure process are known), model fitting under this Berkson error structure does not yield unbiased estimates of β_1 (Carroll et al., 1995). Although it is difficult to obtain analytical expressions for the bias, closed-form expressions are available for certain special cases.

First, for simplicity, suppose the V_i^* are uncorrelated and homoscedastic. For a probit model for binary responses, the model based on the mean of the estimated exposure given the observed data is

$$\mathrm{pr}\,(Y_i^* = 1|\mathbf{W}, \mathbf{Z}_i^*) = \Phi\left[\frac{\beta_0 + \beta_1 E(X_i^*|\boldsymbol{W}) + \boldsymbol{\beta}_z \mathbf{Z}_i^*}{\left(1 + \beta_1^2 \sigma_v^2\right)^{1/2}}\right], \tag{9}$$

where σ_v^2 is the variance of V_i^*. Therefore, the plug-in estimator obtained by fitting $\mathrm{pr}\,(Y_i^* = 1|\mathbf{W}, \mathbf{Z}_i^*) = \Phi\left[\beta_0 + \beta_1 E(X_i^*|\boldsymbol{W}) + \boldsymbol{\beta_z} \mathbf{Z}_i^*\right]$ can yield bias, although the denominator on the right hand side of (9) suggests that this bias will be small unless both σ_v^2 and β_1 are relatively large. Bias expressions in the analogous logistic model are typically approximated using the approximate relationship between the logistic and probit links (Carroll et al. 1995, Equation 7.16). When the V_i^* are heteroscedastic and correlated (as is the case for spatially misaligned data), even in the probit case, the marginal distribution of $[\mathbf{Y}^*|\mathbf{W}, \mathbf{Z}^*]$ involves an intractable multivariate probit integral. See Ochi and Prentice (1984) for a discussion of this issue for an equicorrelated multivariate probit model, and Chib and Greenberg (1998) and De Iorio and Verzilli (2007) for Bayesian approaches to this problem.

6. Simulations

To compare the different methods we performed a simulation study. For each scenario, we used N=500 simulated datasets. For each data set, we used the geocodes of the $n_w = 82$ monitoring stations used in a recent Boston study (Gryparis et al., 2007) as the fixed exposure locations. We generated our exposure measurements, $\boldsymbol{W}$, with no instrument error $\boldsymbol{U}$, using $\boldsymbol{W} = \boldsymbol{X} = \boldsymbol{g} + \boldsymbol{\delta}$, with $\boldsymbol{g} \sim N(\mu\mathbf{1}, \boldsymbol{R}(\rho, \nu))$, where for

$\boldsymbol{R}$ we used the Matérn correlation function. Specific parameter values depended on the exposure scenario, which we describe shortly. For the local heterogeneity $\boldsymbol{\delta}$, we assumed a mean zero normal distribution with *i.i.d.* errors, $\sigma_\delta^2 \boldsymbol{I}_{n_w}$. We considered both continuous and binary outcomes, discussed separately in the following two sections.

6.1 *Continuous Outcomes*

As noted in Section 4.4, the number of subjects on which health outcomes are measured is typically larger than the number of the exposure locations. Hence, for the linear model, we set $n_y = 200$. For the distribution of the health data, $\boldsymbol{Y}^*$, we assume $\boldsymbol{Y}^* \sim N(\beta_0 + \beta_1 \boldsymbol{X}^*, \sigma_\epsilon^2 \boldsymbol{I}_{n_y})$. We set $\beta_0 = 0$ and $\beta_1 = 1$ for all scenarios except the last, in which we set $\beta_0 = \beta_1 = 0$ in order to check the type I error of each approach. We also ran simulations assuming incrementally smaller values of β_1, but the relative performances of the various approaches remained the same as that reported below and the relative bias changed little with different effect sizes (not shown). The assumption of independent health errors implies that the only component responsible for spatial autocorrelation of the health outcome is the exposure.

We considered four exposure scenarios. Scenario A corresponds to a very smooth surface, Scenario B a moderately smooth surface, while Scenario C (the roughest surface) is much more heterogeneous and therefore quite challenging to estimate. Figure 1 shows one realization of the true exposure surface, $\boldsymbol{X}$, obtained from each of the above scenarios. Scenario D is the same as Scenario C, except exposure is not causally related to health ($\beta_1 = 0$). More details on the simulations are given in Section B of the online supplementary material.

[Figure 1 about here.]

We used the methods described in Section 4 to fit the datasets generated under the above scenarios. First, we applied the plug-in approach, estimating the smooth exposure surface using the *spm* function in the SemiPar package (Wand, 2008) in R. This function uses a mixed model representation of penalized regression splines, described in more detail in Section C of the online supplementary material. The degrees of freedom for the spatial component was chosen by the default method, REML. Second, we considered the exposure simulation approach; we fitted the exposure model using a Bayesian framework, and then we sampled 100 realizations from the posterior distribution of the exposure. We then fitted 100 health models, and used as the health effect estimate the mean of the parameter of interest, $\widehat{\beta}_1 = \sum_{t=1}^{100} \widehat{\beta}_{1(t)}$. For each dataset we used the normal approximation, $\widehat{\beta}_1 \pm 1.96\sqrt{Var(\widehat{\beta}_1)}$, to calculate the confidence interval, with $Var(\widehat{\beta}_1)$ defined in Section 4.2. Next, we fitted the fully Bayesian and two-stage Bayesian approaches, integrating the unobserved exposure $(\boldsymbol{X}, \boldsymbol{X}^*)$ out of both models to improve mixing. For all Bayesian approaches, we report results for the most common choice for prior distributions, which are vague but proper Inverse-Gamma(0.01,0.01) priors for all variance components and $N(0, 1000)$ priors for all regression coefficients. Because the vague inverse gamma prior has some undesirable characteristics (Gelman, 2006), we also ran the simulations using Unif(0,1000) priors for the variance components. This change produced a negligible effect on the results, and so we do not report them here. We examined convergence of the algorithms using both graphical and formal approaches (Cowles and Carlin, 1996) for a random subsample of the 500 datasets. We also applied the RC-OOS approaches. For the latter, we used a simulated external dataset with 40 observations to estimate $\boldsymbol{\gamma}$. For each simulated dataset and for each approach, we calculated estimates of β_1 and the model-based standard error. We report the estimated bias, average model-based standard error, the Monte Carlo standard deviation, the mean square error, and the coverage of the 95% confidence or credible intervals.

Tables 1 shows the results from the 500 simulations. These results show that when the exposure is relatively smooth (Scenario A), all methods perform reasonably well. The bias of the plug-in estimator increases as the exposure surface becomes more heterogeneous, and the resulting confidence intervals do not provide satisfactory coverage due to the fact that this estimator does not account for the uncertainty associated with exposure estimation. In the more challenging scenarios, the exposure simulation approach performs very poorly. The resulting estimator is highly biased, and its MSE is large. The RC-OOS approach performs relatively well under all scenarios considered. This estimator incurs small bias, and the resulting confidence intervals yield good coverage probabilities for the true parameter. We note that for this scenario, we excluded one dataset from the results, for which we had an extremely low estimate for β_1.

The fully Bayesian approach performed very well. This is the only approach presented where the exposure model and the health model are fitted simultaneously and hence there is feedback between the health and the exposure data. Since we have sparse exposure data, $n_y > n_w$ and $\sigma_\epsilon^2 >> \sigma_u^2$, some influential health observations could produce anomalies in which the estimate of the spatial surface is spurious, driven solely by the health model, as discussed in Section 4.4. In our simulations though, we did not observe any such distortion, and the fully Bayesian model performed very well, even for the roughest exposure surface. In addition, the two-stage Bayesian fits approximated the full Bayes results very well for all scenarios.

[Table 1 about here.]

6.2 *Binary Outcomes*

Due to the lack of closed-form results for generalized linear models for discrete responses, we extended the simulation study to this setting. Because geo-referenced binary outcomes (e.g., mortality, low birthweight) are more common than geo-referenced count data in the PM epidemiology settings we encounter, we set up the simulation study to examine the methods in a logistic regression model for binary health outcomes. The overall simulation strategy is the same as that used for the linear model, with the actual simulations differing in several ways. First, the health effects model is $\text{logit}\,(\pi_i) = \beta_0 + \beta_1 X_i$, where $\pi_i = \text{pr}(Y_i^* = 1)$, with $\beta_0 = 0$ and $\beta_1 = 0.30$. Second, we assume that there are 7000 study subjects, rather than 200, since there is inherently less information contained in a single binary outcome as compared to a continuous outcome. We note that, although 7000 subjects may seem large, this number of subjects is typically much less than that encountered in applications involving Boston area binary outcomes (e.g., Maynard et al., 2007). Third, we considered only the plug-in, exposure simulation, and regression calibration approaches in the logistic setting. We expect the Bayesian approach would perform well in the nonlinear setting as well, but MCMC sampling in this setting, in which the spatial term cannot be marginalized out of the model, can be difficult to implement effectively (Christensen et al., 2006; Paciorek, 2007). Our initial efforts to implement the Bayesian logistic model in a straightforward MCMC scheme showed poor mixing, and because the development of a carefully-tailored sampling strategy is beyond the scope of this paper, we do not pursue the approach further here. Fourth, because the linear regression simulations provided insight on the degradation of the estimators as a function of spatial heterogeneity, we ran the simulations only for Scenarios A and C, representing a smooth and spatially heterogeneous exposure surface, respectively. In this setting, for the out-of-sample RC estimator, we used formulas for the standard regression calibration estimator and associated standard error provided in Thurston et al. (2003), who derived these estimators for the broad class of generalized linear models. The variance formula is the generalized analogue to (8), incorporating the weight matrix $\mathbf{W} = \text{Diag}\left[\pi_i\left(1-\pi_i\right)\right]$ associated with binary responses.

[Table 2 about here.]

Table 2 presents the results of this simulation study. The patterns in this table are similar to those exhibited by the linear regression results. While simple regression calibration is known to give biased estimates in nonlinear model settings, the magnitude of this bias is relatively small in the scenarios considered, which agrees with closed form results (9) and recent investigations of regression calibration in the standard logistic regression measurement error setting (Thoresen and Laake, 2000).

7. Traffic particles and Birthweight in the greater Boston area

In this section we illustrate the relative performance of the various methods considered in this article by analyzing the association between traffice-related particulate matter generated by motor vehicles and birthweight in the greater Boston area. Because black carbon (BC) and elemental carbon (EC) particles are well-known markers of traffic pollution, we use output from a previously developed exposure model for BC and EC particles (Gryparis et al., 2007), and assess the association between these predictions and all birthweights in the greater Boston area over the period of January 1, 1996 - December 31, 2002.

Briefly, the exposure predictions are derived from a validated spatio-temporal model for 24-hour measures of traffic exposure based on individual exposure data and ambient monitoring sites from over 82 locations in the Boston area. Predictions are based on meteorological conditions and other characteristics (e.g., weekday/weekend) of a particular day, as well as measures of the amount of traffic activity (e.g., GIS-based measures of cumulative traffic density within 100 meters, population density, distance to nearest major roadway, percent urbanization) at a given location. The model allowed these factors to affect exposure levels in a potentially nonlinear way via nonparametric regression terms. It also used the mixed model representation of thin plate splines, described in Section C of the online supplementary material, to capture additional spatial variation unaccounted for after including all relevant spatial predictors in the model. The model was fitted using a Bayesian MCMC approach. Results of this analysis suggest that there exists significant spatial variability in these concentrations in the Boston area. For instance, the spatial variability in exposure varies by a factor of approximately three, and the concentrations are highest in the downtown Boston area and along the I-95 and I-90 interstates.

[Figure 2 about here.]

Our health data come from a study population that initially included all live births in

Eastern Massachusetts for the counties of Bristol, Essex, Middlesex, Norfolk, Plymouth, Suffolk, and Worcester. The data were obtained from the Massachusetts Birth Registry for the period between January 1, 1996 and December 31, 2002. The population of the selected counties covered about 83% of the state's population and about 53% of the state's area. From a total number of births of 477,495, we restricted our study to singleton births (95.8% of all births), born between 20 and 45 weeks of gestation and with birth weight between 200 grams and 5500 grams. Of these births we excluded those that could not be correctly assigned an address (4.9%) and those that were not within the Interstate-495 beltway, which corresponded to the study region for which we had exposure predictions (51%). In total we analyze data on 219,060 births. The address of the mother at the time of birth was geocoded by a private firm and was reassessed by us for accuracy and completeness. Figure 2 shows the locations of the residences of the study subjects and their positioning relative to the 82 exposure monitors. The study and the use of birth data were approved by the Massachusetts Department of Public Health and the Human Subjects Committee of the Harvard School of Public Health.

In this analysis, we first fitted linear regression models for birthweight in grams. In this huge sample size setting, Bayesian approaches are computationally demanding and thus infeasible to apply in a reasonable amount of time. Accordingly, we use the naive plug-in, exposure simulation, and the out-of-sample regression calibration approaches to analyze the data. We also applied standard weighted least squares, but relegate reporting of this result to Section B of the online supplementary material. Because our health outcome is a pregnancy outcome, we use as our exposure metric nine-month averages of 24-hour predicted black carbon levels, corresponding to the gestational period for each birth. We note that, although the time scales of our prediction model (daily) and our exposure covariate (nine months) do not coincide, use of the estimate $\widehat{\gamma}$ to correct the naive plug-in estimator is still valid due to the linear assumption in the validation relationship (6). To account for well-known confounding factors of birthweight, we included the following covariates on biologic grounds: maternal age, maternal race, gestational age, amount of cigarette smoking during pregnancy, chronic conditions of the mother or of pregnancy, mother having previous preterm birth, mother having previous infant weighing $>$ 4000 grams, gender, year of birth, maternal education, Kotelchuck Index of adequacy of prenatal care, and Census Tract (CT) median income. We include education and CT median income to account for both individual as well as contextual effects of socioeconomic status, a well-known important predictor of birthweight, on the outcome.

[Table 3 about here.]

Table 3 presents the estimated coefficients and estimated 95% confidence intervals for all terms included in the model based on the RC-OOS fit. This approach yields moderate evidence of an association between birthweight and predicted BC concentrations. To put the magnitude of this estimate into perspective, we compare it to the estimated coefficients for other factors well-known to affect birthweight. We estimate an interquartile range change in BC (IQR=0.20 $\mu g/m^3$) is associated with a decrement in birthweight roughly equivalent to a tenth of the difference between high school and college educated women.

[Table 4 about here.]

Table 4 presents the results from the three different analyses, showing that the relative performance of the various methods follows the patterns suggested by both the analytical results and simulation studies presented in Sections 5 and 6, respectively. Compared to the naive plug-in approach, exposure simulation grossly attenuates the estimated health effect. Based on the held-out data, our proposed regression calibration correction approach yields $\widehat{\gamma}_0 = 0.20$ (S.E.=0.04) and $\widehat{\gamma}_1 = 0.84$ (S.E.=0.07). Although out-of-sample regression calibration detects an association between birthweight and estimated BC particle levels at the 95% confidence level whereas the naive approach does not, the magnitudes of these estimates are relatively close in this case, suggesting that the performance of the plug-in approach that simply uses the exposure estimates is not too bad. However, one would not have known this before performing this measurement error correction.

Because we controlled for a host of well-known confounding factors that explain a large amount of the spatial pattern in birthweights, the regression model assumes independent errors. We checked the appropriateness of this assumption by constructing a semi-variogram plot (Waller and Gotway 2004; Section 8.2) based on the model residuals. This plot (not shown) showed that the semivariance of differences between pairs of residuals is approximately a constant function of distance between each pair, suggesting that the independence assumption is valid for these data.

We re-ran the above analyses two additional times, using estimated location-specific BC concentrations during the time periods corresponding to the first trimester and the third trimester of each pregnancy. Interestingly, the effect estimates from the third trimester model were similar in magnitude to those in Table 4, whereas the effect estimates from the first trimester model were all approximately half those in Table 4. This may occur because the third trimester is the more important period for weight gain of a developing fetus.

Finally, we also ran logistic regression models relating the probability of an infant having a low birthweight for their gestational age to the same exposure predictions used in the linear models for birthweight. As suggested by our simulations in this setting, the differences between the estimates from the different approaches were smaller than those observed in the linear setting, and none of the analyses showed strong evidence of an association between this binary outcome and estimated BC levels (results not shown).

8. Discussion

Taken together, the simulation results suggest that several approaches to analyzing spatially misaligned point data may be appropriate, depending on the amount of spatial heterogeneity in the exposure surface and the amount of data. For moderate sample sizes, a Bayesian approach to estimation is computationally feasible and seems to possess relatively good frequentist properties. The two-stage Bayesian approach allows one to break the joint model down into its two components. Simulation results suggested that this approach approximates the fully Bayesian results quite well. Thus, this two-stage estimator is attractive whenever either the exposure or health model is complicated, in which case designing well-mixing MCMC algorithms for the full model may be difficult, or when one is interested in running multiple epidemiological models but wants to avoid fitting the exposure model multiple times. Alternatively, one could consider the out-of-sample regression calibration approach. It is much easier to implement computationally, but is less statistically efficient, than the Bayesian approaches. These two features make it more attractive than the Bayesian approaches in large sample settings, since the Bayesian approaches can be computationally expensive and the inefficiency of the calibration estimators is not as much of a concern in this setting. The calibration parameters can be precisely estimated, which should improve the MSE compared to that seen in our simulations. Thus, the two approaches that work well in all of our simulation settings, the Bayesian and calibration approaches, are complementary, in terms of data settings for which each might be preferred.

Our results provide insight regarding existing findings in covariate-response misalignment problems. In a setting where the response and a covariate were misaligned over time, Higgins et al. (1997) noted that the plug-in estimator incurred little bias. The unknown smooth trends in the covariate over time were relatively smooth, so these results are the temporal analogue of our results based on a spatially smooth surface. Zhu et al. (2003) considered Bayesian approaches for spatial data that involve both misalignment and change-of-support, with interest focusing on relating monitoring data to zip-code level disease counts. They noted that a fully Bayesian approach performed well in this setting. Interestingly, these authors also showed via simulation that the exposure simulation approach performed similarly to the fully Bayesian approach, with the estimates of the exposure simulation approach being only slightly biased. Due to the differences between this problem and the one we consider here, there could be multiple reasons for this difference in findings. One possibility is that calculating exposure at the zip-code level of aggregation yields relatively smooth exposure surfaces, for which any approach seems to perform adequately.

In short, we used a simple linear model setting to illustrate measurement error issues associated with point-level, spatially misaligned exposure and health data and ran simulations for linear and logistic models. Of course, in practice, more complicated models may be necessary, and future research will focus on extending the methods considered here to such settings. Examples include settings involving health outcomes in complex spatio-temporal models, health effects models exhibiting spatially correlated residuals, and heavy-tailed prediction errors likely to arise for some exposures. One might also consider the impact of different spatial configurations and numbers of exposure monitors and health observations, as well as strategies for optimal monitoring design in such settings.

9. Acknowledgments

We thank the Associate Editor and the referees for comments that greatly improved this article. This research was supported by NIEHS grants ES012044 (AG, BAC), ES009825 (JS, BAC), ES007142 (CJP), and ES000002 (CJP), and EPA grant R-832416 (JS, BAC). It has not been formally reviewed by any sponsor and the views expressed in this document are solely those of the authors.

References

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data.* New York: Chapman & Hall.

Berhane, K., Gauderman, W. J., Stram, D. S. and Thomas, D. C. (2004). Statistical issues in studies of the long-term effects of air-pollution: the Southern California Children's Health Study. *Statistical Science* **19**, 414–449.

Berry, S. M., Carroll, R. J. and Ruppert, D. (2002). Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association* **97**, 160–169.

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

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika* **85**, 347–361.

Christensen, O. F., Roberts, G. O. and Sköld, M. (2006). Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics* **15**, 1–17.

Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative study. *Journal of the American Statistical Association* **91**, 883–904.

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* New York: John Wiley & Sons.

De Iorio, M. and Verzilli, C. J. (2007). A spatial probit model for fine-scale mapping of disease genes. *Genetic Epidemiology* **31**, 252–260.

Gaudard, M., Karson, M., Linder, E. and Sinha, D. (1999). Bayesian spatial prediction. *Environmental and Ecological Statistics* **6**, 147–171.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–533.

Gryparis, A., Coull, B. A., Schwartz, J. and Suh, H. H. (2007). Semiparametric latent variable regression models for spatio-temporal modeling of mobile source particles in the greater Boston area. *Journal of the Royal Statistical Society, Series C* **56**, 183–209.

Higgins, K. M., Davidian, M. and Giltinan, D. M. (1997). A two-step approach to measurement error in time-dependent covariates in nonlinear mixed-effects models, with application to igf-1 pharmacokinetics. *Journal of American Statistical Association* **92**, 436–448.

Hobert, J. P., Altman, N. S. and Schofield, C. L. (1997). Analyses of fish species richness with spatial covariate. *Journal of the American Statistical Association* **92**, 846–854.

Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society, Series C* **52**, 1–18.

Kunzli, N., Jerrett, M., Mack, W. J., Beckerman, B., LaBree, L., Gilliland, F., Thomas, D., Peters, J. and Hodis, H. N. (2005). Ambient air pollution and atherosclerosis in Los Angeles. *Environmental Health Perspectives* **113**, 201–206.

Little, R. J. A. (1992). Regression with missing X's. A review. *Journal of the American Statistical Association* **87**, 1227–1237.

Madsen, L., Ruppert, D. and Altman, N. S. (2008). Regression with spatially misaligned data. *Environmetrics* **19**, 453–467.

Maynard, D., Coull, B. A., Gryparis, A. and Schwartz, J. (2007). Mortality risk associated with short-term exposure to traffic particles and sulfates. *Environmental Health Perspectives* **115**, 751–755.

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

Ochi, Y. and Prentice, R. L. (1984). Likelihood inference in a correlated probit regression model. *Biometrika* **71**, 531–543.

Paciorek, C., Yanosky, J., Puett, R., Laden, F. and Suh, H. (2008). Practical large-scale spatio-temporal modeling of particulate matter concentrations. *Annals of Applied Statistics* **Under review**.

Paciorek, C. J. (2007). Computational techniques for spatial logistic regression with large datasets. *Computational Statistics and Data Analysis* **51**, 3631–3653.

Rougier, J. (2008). Comment on article by Sans'o et al. *Bayesian Analysis* **3**, 45–56.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression.* Cambridge, UK: Cambridge University Press.

Shaddick, G. and Wakefield, J. (2002). Modelling daily multivariate pollutant data at multiple sites. *Journal of the Royal Statistical Society, Series C* **51**, 351–372.

Thoresen, M. and Laake, P. (2000). A simulation study of measurement error correction methods in logistic regression. *Biometrics* **56**, 868–872.

Thurston, S. W., Spiegelman, D. and Ruppert, D. (2003). Equivalence of regression calibration methods in main study/external validation study designs. *Journal of Statistical Planning and Inference* **113**, 527–539.

Wakefield, J. and Shaddick, G. (2006). Health-exposure modeling and the ecological fallacy. *Biostatistics* **7**, 438–455.

Waller, L. A. and Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data.* New York: John Wiley & Sons.

Wand, M. P. (2008). The semipar package reference manual. Technical report, Available at `http://cran.r-project.org/web/packages/SemiPar/SemiPar.pdf`.

Yucel, R. M. and Zaslavsky, A. M. (2005). Imputation of binary treatment variables with measurement error in administrative data. *Journal of the American Statistical Association* **100**, 1123–1132.

Zhu, L., Carlin, B. P. and Gelfand, A. E. (2003). Hierarchical regression with misaligned spatial data: relating ambient ozone and pediatric asthma ER visits in Atlanta.

Environmetrics **14**, 537–557.

Zidek, J., Shaddick, G., White, R., Meloche, J. and Chatfield, C. (2004). Using a probabilistic model (pCNEM) to estimate personal exposure to air pollution. *Environmetrics* **16**, 481–493.

A. Weighted least squares and generalized least squares

Here we describe two additional approaches to the problem of measurement error in exposure predictions induced by spatial misalignment.

A.1 *Weighted Least Squares (WLS)*

An approach that downweights anomalous exposure predictions having high uncertainty, which might be influential in the health model, is weighted least squares (WLS) with the weights based on the uncertainty estimates from the exposure model (e.g., Kunzli et al., 2005). Although this approach seems intuitive may be useful for downweighting estimates with larger error, these are not the correct weights in our modeling framework, as discussed in Section 3. As shown there, the correct covariance is $\beta_1^2\boldsymbol{\Sigma}^* + \sigma_\epsilon^2\boldsymbol{I}_{n_y}$. As an example of when this approach could perform poorly, when all values have large, but similar, uncertainty, this approach will give similar results to that from OLS without adjusting for the correlation. In practice though, there may exist some problematic spatial regions where the exposure has not been estimated well (e.g., at locations far from data, which give large prediction errors), in which case this method may improve upon the plug-in estimator.

A.2 *GLS based on the exposure covariance estimate (GLS)*

As mentioned in Section 3, if one is interested in directly using the predictions from an exposure model, one should use GLS with covariance $\beta_1^2\boldsymbol{\Sigma}^* + \sigma_\epsilon^2\boldsymbol{I}_{n_y}$. This is essentially the "Krige and Regress" estimator of Madsen et al. (2008), except for the implementation differences described below. Like the fully Bayesian approach, GLS accounts for the structure in the uncertainty in $\boldsymbol{X}^*$. To apply this approach, one needs a good estimate of $\boldsymbol{\Sigma}^*$, the prediction error variance at the health locations. This estimate can be obtained from the exposure model. Since we use $\boldsymbol{S}^* = \hat{E}(\boldsymbol{X}^*|\boldsymbol{W})$, we want an estimate of $\hat{\mathrm{Var}}(\boldsymbol{X}^*|\boldsymbol{W})$. This conditional variance is difficult to compute (**?**), so we use the approximation considered by Ruppert et al. (2003, page 103):

$$\widehat{\boldsymbol{\Sigma}^*} = \boldsymbol{C}^*\widehat{Cov}\left(\left[\begin{array}{c}\widehat{\boldsymbol{\beta}}_{\boldsymbol{w}}\\ \widehat{\boldsymbol{b}}_{\boldsymbol{w}}\end{array}\right]|\boldsymbol{b}_w\right)(\boldsymbol{C}^*)^T + \widehat{\sigma}_\delta^2\boldsymbol{I}_{n_y}.$$

Since β_1 appears in both the mean and the covariance in the health data, one cannot use a typical GLS approach. Madsen et al. (2008) chose to use an initial estimate of β_1 for the covariance matrix. We choose to maximize the likelihood of the health model using a direct optimization routine, such as *nlm* or *optim* in R. These functions require initial values for all unknown parameters, for which an obvious choice is the plug-in estimates. To estimate the standard error of the estimated coefficients, one can use the standard likelihood approach and invert the information matrix.

We emphasize that our GLS approach maximizes the likelihood of the health model treating an estimate of the covariance structure as fixed and known. Note that this approach is not a joint maximum likelihood approach to fitting the health and exposure models simultaneously (Madsen et al., 2008). Such an approach would be the frequentist analogue of the fully Bayesian approach. We applied that joint approach as well in our simulations and found results similar to those from the fully Bayesian approach (not shown). This finding is in contrast to the simulation results presented of Madsen et al. (2008), who found that the joint ML approach yields coverage of only 45% and suggested that the conditions required for the asymptotic variance estimator do not hold in their spatial setting. One possible reason for this difference is the fact that Madsen et al. (2008) considered relatively large residual correlation among second-stage outcomes, motivated by an ecological application in which the outcome represented an environmental variable (log chloride concentration in streams). In contrast, motivated by health outcomes that are likely to be much less spatially correlated, we considered independent residuals in the second-stage outcome and did not see any evidence that this assumption was violated in the Boston birthweight data.

A.3 *WLS and GLS simulation results*

In addition to the primary methods described in Section 6, we considered the WLS and GLS approaches in our simulations. For the WLS weights we used the inverse of the prediction variances from the penalized spline model used in the plug-in approach.

Table 5 duplicates Table 1 in the paper with added rows for the WLS and GLS methods. When the exposure is relatively smooth (Scenario A), all methods, including WLS and GLS, perform reasonably well. In the other scenarios, the WLS approach provides a slightly improved fit over the plug-in estimator, mostly by decreasing the bias, but overall it performs quite similarly to the plug-in approach. The GLS approach, which accounts for heteroscedasticity and correlation in $\boldsymbol{X}^*$, performs reasonably well. Under Scenario C, it decreases the bias of the plug-in estimator substantially and attains a coverage of 86%. However, we note that in Scenario C, this approach had numerical difficulties in the estimation procedure that resulted in very small (< 0.01) estimates for the variance of the health model for 3% of the datasets. With regard to type I error reflected in Scenario D, all the approaches perform well, with the exception of the WLS approach, for which the estimated type I error is 0.094, almost double the nominal type I error of 0.05. This occurs because the WLS approach uses incorrect weights when $\beta_1 = 0$.

[Table 5 about here.]

A.4 *WLS in the application*

We also used WLS in the birthweight application, in which it gave a seemingly untrustworthy estimate of -55.25 for the health effect coefficient, well away from the estimates from the three approaches (Section 7), with very large standard error of 52.07 and 95% confidence interval of (-157.31, 46.81). This may have occurred because the weighting strategy systematically downweights suburban locations relative to urban locations (see Figure 2 in the paper), without taking into account the spatial structure of these weights. This may be a form of selection bias.

B. Simulation details

To generate the Gaussian processes in the simulations we used the Fourier basis approximation in the spectralGP package (Paciorek, 2007) in R using the Matérn correlation function with the parameterization,

$$\frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{2\sqrt{\nu}\tau}{\rho\pi}\right)^{\nu} K_{\nu}\left(\frac{2\sqrt{\nu}\tau}{\rho\pi}\right),$$

with distance τ, spatial range ρ (correlation decay) and differentiability parameter $\nu > 0$. Note that ν dictates the differentiability of the surface, with large values corresponding to smoother surfaces.

The settings that we used for the simulations were:

- Scenario A:
  $\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{R}(1.6, 1))$, $\boldsymbol{\delta} \sim N(\boldsymbol{0}, \sigma_{\delta}^2 \boldsymbol{I}_{82})$, $\sigma_{\delta}^2 = 0.1^2$, $\sigma_{\epsilon}^2 = 0.8^2$

- Scenario B:
  $\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{R}(0.3, 2))$, $\boldsymbol{\delta} \sim N(\boldsymbol{0}, \sigma_{\delta}^2 \boldsymbol{I}_{82})$, $\sigma_{\delta}^2 = 0.2^2$, $\sigma_{\epsilon}^2 = 0.8^2$

- Scenario C:
  $\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{R}(0.3, 0.5))$, $\boldsymbol{\delta} \sim N(\boldsymbol{0}, \sigma_{\delta}^2 \boldsymbol{I}_{82})$, $\sigma_{\delta}^2 = 0.2^2$, $\sigma_{\epsilon}^2 = 0.8^2$

- Scenario D:
  $\boldsymbol{g} \sim N(\boldsymbol{0}, \boldsymbol{R}(0.3, 0.5))$, $\boldsymbol{\delta} \sim N(\boldsymbol{0}, \sigma_{\delta}^2 \boldsymbol{I}_{82})$, $\sigma_{\delta}^2 = 0.2^2$, $\sigma_{\epsilon}^2 = 0.8^2$ but we generated health data using $\boldsymbol{Y}^* \sim N(\boldsymbol{0}, \sigma_{\epsilon}^2 \boldsymbol{I}_{n_y})$

For Scenarios C and D, $\nu = 0.5$, giving the exponential correlation function, which corresponds to Gaussian processes with continuous, but not differentiable sample paths.

C. Mixed model spatial smoothing

In our simulations and application, we spatially smooth exposure using a mixed model representation of penalized regression splines (Ruppert et al., 2003). This approach is simple to implement, has low computational cost and is widely applicable. Consider the simple nonparametric regression model,

$$W_i = f(\boldsymbol{geog}_i) + U_i,\ 1 \le i \le n,\ \ U_i \sim N(0, \sigma_u^2),$$

where $\boldsymbol{geog}_i = (longitude, latitude)_i$. A mixed model representation of penalized regression splines for $f(\cdot)$ is:

$$f(\cdot) \equiv \boldsymbol{X} \equiv \boldsymbol{C}\boldsymbol{z}, \tag{10}$$

for a choice of basis functions and appropriate representation in terms of the parameters, $\boldsymbol{z}$ (Ruppert et al., 2003). The vector $\boldsymbol{z}$ consists of a fixed effects vector $\boldsymbol{\beta}$ of length p and random effects $\boldsymbol{b}$, with $b_i \sim N(0, \sigma_b^2)$, $i = 1, 2, ..., K$, where K is the number of knots and $\boldsymbol{C}$ is the corresponding design matrix. We use the thin plate spline generalized covariance to construct $\boldsymbol{C}$. Let $\tilde{\boldsymbol{z}} = (\frac{1}{\sigma_u^2}\boldsymbol{C}^T\boldsymbol{C}+\boldsymbol{B})^{-1}\frac{1}{\sigma_u^2}\boldsymbol{C}^T\boldsymbol{W}$ be the best linear unbiased predictor (BLUP) for $\boldsymbol{z}$, where

$$\boldsymbol{B} = \left[\begin{array}{cc} \mathbf{0}_{p\times p} & \mathbf{0}_{p\times K} \\ \mathbf{0}_{K\times p} & \frac{1}{\sigma_b^2}\mathbf{I}_K \end{array} \right].$$

Then the BLUP for $\boldsymbol{X}^*$, for known σ_u^2 and σ_b^2, is $\boldsymbol{S}^* \equiv \hat{E}(\boldsymbol{X}^*|\boldsymbol{W}) = \boldsymbol{C}^*\tilde{\boldsymbol{z}} = \boldsymbol{C}^*(\frac{1}{\sigma_u^2}\boldsymbol{C}^T\boldsymbol{C}+\boldsymbol{B})^{-1}\frac{1}{\sigma_u^2}\boldsymbol{C}^T\boldsymbol{W}$, which is a weighted average of the observed data $\boldsymbol{W}$. Note that $\boldsymbol{C}^*$ is the design matrix that corresponds to $\boldsymbol{X}^*$, for the same choice of basis functions and knots used in (10). The BLUP conditions on the available information, as in regression calibration, so that in this modeling framework the true covariate X_i^* is centered around around its BLUP, S_i^*. As in the Gaussian process framework, the smoothing reverses the conditioning, producing a Berkson structure, rather than the classical measurement error structure (Section 3). In implementation of the Bayesian approaches in the simulations, we used a Bayesian version of the mixed model representation with priors on the regression coefficients and variance components as described in Section 6.

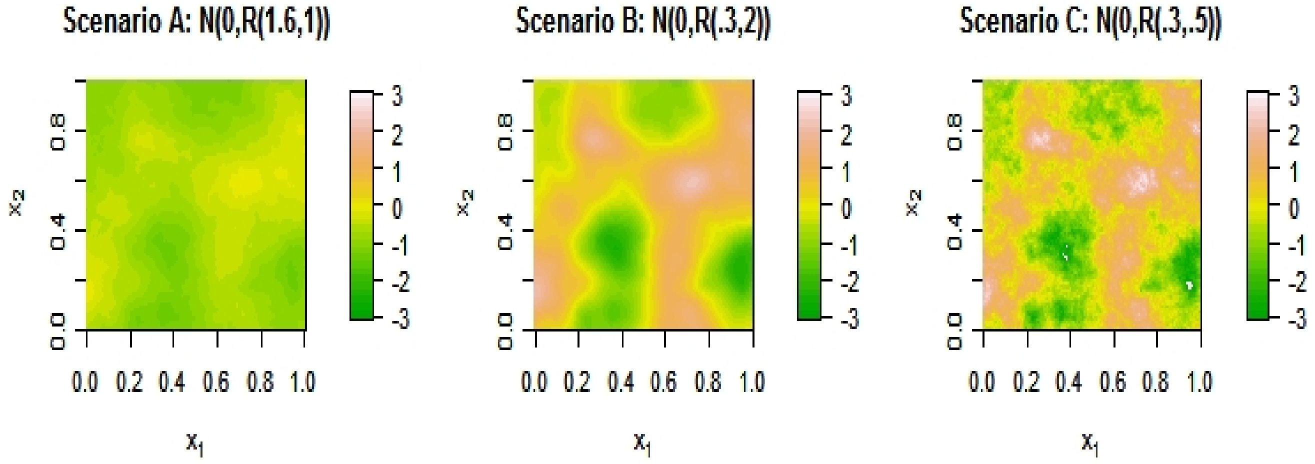

Figure 1. Realizations of the true smooth exposure surface $\boldsymbol{g}(\cdot)$ for simulation scenarios A, B and C, on the $[0, 1] \times [0, 1]$ grid.

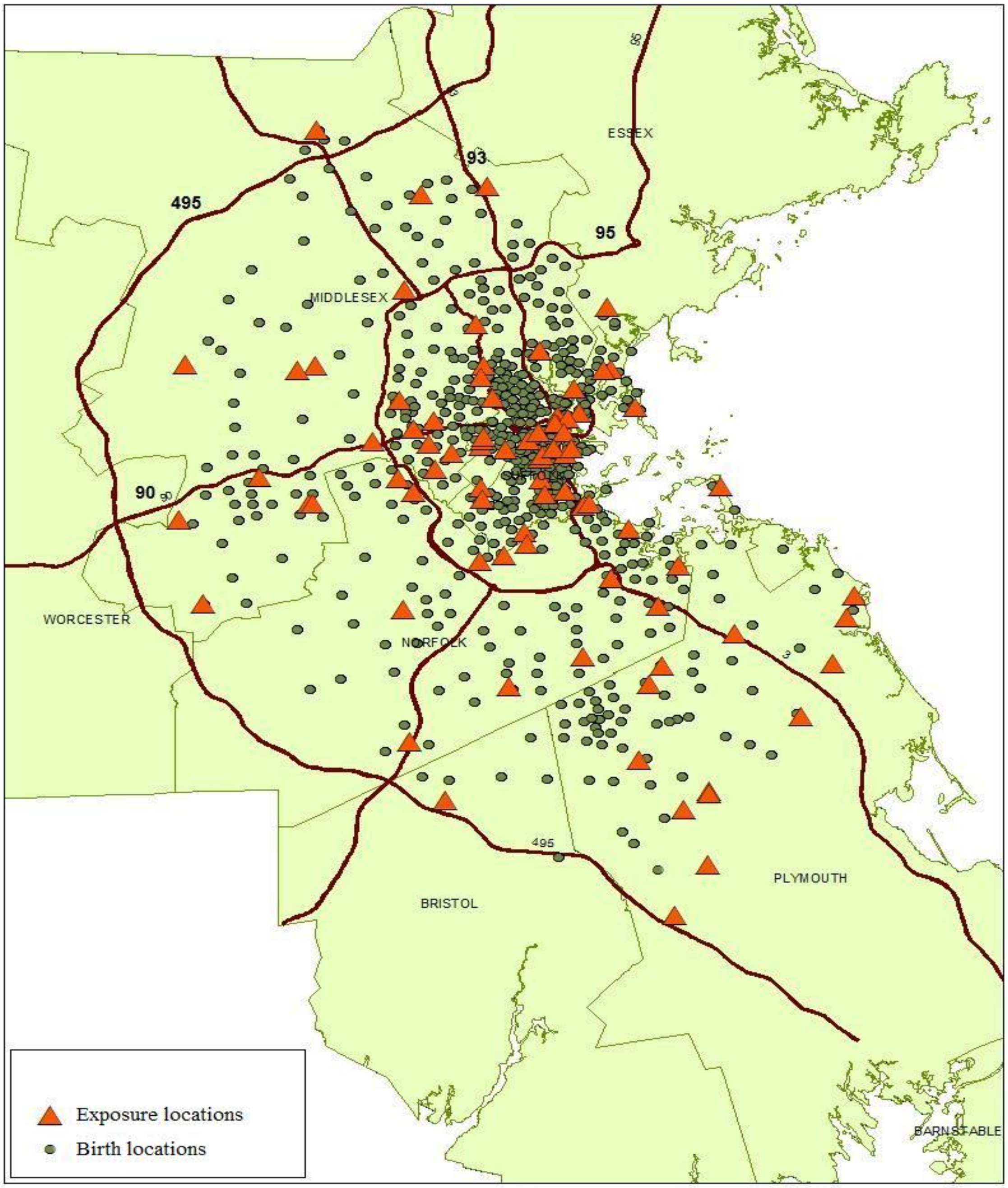

Figure 2. Map of the locations of the residences of the birtweight study subjects and their positioning relative to the 82 exposure monitors.

Scenario	Method	Bias	$E(se(\beta_1))$	$sd(\widehat{\beta_1})$	MSE	Coverage (%)
A	True exposure	-0.000	0.093	0.096	0.009	94.8
	Plug-in	0.004	0.105	0.122	0.015	91.6
	Exposure simulation	-0.068	0.118	0.119	0.019	91.2
	RC-OOS	0.006	0.122	0.122	0.015	96.4
	Fully Bayesian	0.002	0.109	0.122	0.015	92.8
	Two-stage Bayes	0.000	0.108	0.123	0.015	93.2
B	True exposure	0.002	0.059	0.059	0.003	95.2
	Plug-in	-0.085	0.091	0.149	0.029	69.8
	Exposure simulation	-0.254	0.116	0.126	0.080	42.2
	RC-OOS	0.036	0.197	0.251	0.064	95.6
	Fully Bayesian	0.011	0.107	0.151	0.023	86.4
	Two-stage Bayes	0.004	0.105	0.150	0.023	83.8
C	True exposure	0.004	0.058	0.058	0.003	95.2
	Plug-in	-0.140	0.130	0.211	0.064	63.4
	Exposure simulation	-0.591	0.141	0.146	0.371	0.4
	RC-OOS*	0.039	0.340	0.367	0.136	92.6
	Fully Bayesian	0.029	0.155	0.177	0.032	93.0
	Two-stage Bayes	0.039	0.1646	0.239	0.059	90.8
D	True exposure	0.003	0.059	0.062	0.004	93.4
	Plug-in	0.001	0.090	0.095	0.009	94.2
	Exposure simulation	0.000	0.068	0.054	0.003	98.8
	RC-OOS	0.001	0.111	0.115	0.013	95.6
	Fully Bayesian	0.000	0.159	0.140	0.019	94.0
	Two-stage Bayes	0.000	0.148	0.135	0.018	94.4

Table 1

Results of simulation study for $\widehat{\beta}_1$: Bias, average model-based standard error, Monte Carlo standard deviation, MSE, and coverage of 95% confidence or credible intervals, over 500 simulations, for Scenarios A-D. *One simulation with anomalous estimate omitted.

Scenario	Method	Bias	$E(se(\beta_1))$	$sd(\widehat{\beta_1})$	MSE	Coverage (%)
A	True exposure	-1.24	0.070	0.073	0.0054	95.0
	Plug-in	-0.55	0.094	0.102	0.0103	95.6
	Exposure simulation	-0.91	0.101	0.101	0.0102	95.6
	RC-OOS	-0.35	0.098	0.107	0.0114	100.0
C	True exposure	-1.23	0.030	0.029	0.0009	95.8
	Plug-in	-6.72	0.036	0.048	0.0027	81.8
	Exposure simulation	-13.2	0.042	0.043	0.0035	78.4
	RC-OOS	-1.22	0.046	0.050	0.0025	100.0

Table 2

Results of logistic regression simulation study for $\widehat{\beta}_1$: Bias, average model-based standard error, Monte Carlo standard deviation, MSE, and coverage of 95% confidence or credible intervals, over 500 simulations, for Scenarios A and C

Method	Estimate	SE	95% CI
Predicted BC	-9.46	4.38	(-18.05, -0.88)
Mother's Age	6.36	0.20	(5.97, 6.75)
Gest. Age	551.45	6.16	(539.37, 563.52)
Gest. Age Squared	-5.72	0.08	(-5.88, -5.55)
Num. Cigs.	-28.91	0.84	(-30.56, -27.26)
Num. Cigs. Squared	0.69	0.04	(0.61, 0.78)
Prev. Inf. $>$ 4000	480.10	11.56	(457.43, 502.77)
Prev. Preterm	-242.10	12.82	(-267.23, -216.97)
Maternal Cond.	-29.89	3.40	(-36.56, -23.23)
CT Med. Income (1000K)	0.15	0.04	(0.07, 0.24)
Maternal Educ. ($<$ 12 yrs.)	8.57	6.74	(-4.63, 21.77)
Maternal Educ. (12 - 16 yrs.)	1.00 (ref)	—	(—, —)
Maternal Educ. ($>$ 16 yrs.)	16.63	2.52	(11.70, 21.57)
Race (Caucasian)	1.00 (ref)	—	(—,—)
Race (African Amer.)	-131.01	3.64	(-138.15, -123.87)
Race (Asian)	-192.72	3.99	(-200.54, -184.90)
Race (Other)	-93.15	3.85	(-100.69, -85.61)
Sex (Male)	132.62	2.06	(128.58, 136.66)
Sex (Female)	1.00 (ref)	—	(—,—)
1996	19.37	3.96	(11.61, 27.14)
1997	16.52	4.36	(7.97, 25.06)
1998	23.73	3.85	(16.18, 31.27)
1999	17.02	3.78	(9.61, 24.43)
2000	10.49	3.77	(3.09, 17.89)
2001	3.36	3.75	(-3.98, 10.70)
2002	1.00 (ref)	—	(—,—)
K. Index (Inadequate)	-70.39	4.31	(-78.85, -61.94)
K. Index (Intermediate)	-51.16	4.36	(-59.71, -42.61)
K. Index (Appropriate)	1.00 (ref)	—	(—,—)
K. Index (Appropriate +)	-16.17	2.43	(-20.92, -11.41)

Table 3

Out-of-sample Regression Calibration Estimates for Greater Boston Birthweight Data

Method	Estimate (in grams)	SE	95% CI
Plug-in	-7.27	3.78	(-14.68 , 0.14)
Exposure simulation	-0.48	3.40	(-7.13 , 6.18)
RC-OOS	-9.46	4.38	(-18.05 , -0.88)

Table 4
Results for Greater Boston Birthweight Data

Scenario	Method	Bias	E(se(β_1))	sd($\widehat{\beta_1}$)	MSE	Coverage (%)
A	True exposure	-0.000	0.093	0.096	0.009	94.8
	Plug-in	0.004	0.105	0.122	0.015	91.6
	WLS	0.005	0.110	0.124	0.015	91.6
	Exposure simulation	-0.068	0.118	0.119	0.019	91.2
	GLS	0.005	0.110	0.120	0.014	93.2
	RC-OOS	0.006	0.122	0.122	0.015	96.4
	Fully Bayesian	0.002	0.109	0.122	0.015	92.8
	Two-stage Bayes	0.000	0.108	0.123	0.015	93.2
B	True exposure	0.002	0.059	0.059	0.003	95.2
	Plug-in	-0.085	0.091	0.149	0.029	69.8
	WLS	-0.049	0.089	0.135	0.021	79.2
	Exposure simulation	-0.254	0.116	0.126	0.080	42.2
	GLS	-0.022	0.103	0.144	0.021	82.4
	RC-OOS	0.036	0.197	0.251	0.064	95.6
	Fully Bayesian	0.011	0.107	0.151	0.023	86.4
	Two-stage Bayes	0.004	0.105	0.150	0.023	83.8
C	True exposure	0.004	0.058	0.058	0.003	95.2
	Plug-in	-0.140	0.130	0.211	0.064	63.4
	WLS	-0.096	0.130	0.204	0.050	72.0
	Exposure simulation	-0.591	0.141	0.146	0.371	0.4
	GLS	-0.020	0.169	0.215	0.047	85.6
	RC-OOS*	0.039	0.340	0.367	0.136	92.6
	Fully Bayesian	0.029	0.155	0.177	0.032	93.0
	Two-stage Bayes	0.039	0.1646	0.239	0.059	90.8
D	True exposure	0.003	0.059	0.062	0.004	93.4
	Plug-in	0.001	0.090	0.095	0.009	94.2
	WLS	-0.002	0.072	0.084	0.007	90.6
	Exposure simulation	0.000	0.068	0.054	0.003	98.8
	GLS	0.001	0.066	0.066	0.004	96.4
	RC-OOS	0.001	0.111	0.115	0.013	95.6
	Fully Bayesian	0.000	0.159	0.140	0.019	94.0
	Two-stage Bayes	0.000	0.148	0.135	0.018	94.4

Table 5

Results of simulation study for $\widehat{\beta}_1$: Bias, average model-based standard error, Monte Carlo standard deviation, MSE, and coverage of 95% confidence or credible intervals, over 500 simulations, for Scenarios A-D. *One simulation with anomalous estimate omitted.