



Published in final edited form as:

*Epidemiology*. 2017 May ; 28(3): 338–345. doi:10.1097/EDE.0000000000000623.

## Measurement error correction for predicted spatiotemporal air pollution exposures

**Joshua P. Keller,**

Department of Biostatistics, University of Washington, Seattle, WA, USA

**Howard H. Chang,**

Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, USA

**Matthew J. Strickland,** and

School of Community Health Sciences, University of Nevada Reno, Reno, NV, USA

**Adam A. Szpiro**

Department of Biostatistics, University of Washington, Seattle, WA, USA

### Abstract

**Background**—Air pollution cohort studies are frequently analyzed in two stages, first modeling exposure then using predicted exposures to estimate health effects in a second regression model. The difference between predicted and unobserved true exposures introduces a form of measurement error in the second stage health model. Recent methods for spatial data correct for measurement error with a bootstrap and by requiring the study design ensure spatial compatibility, i.e., monitor and subject locations are drawn from the same spatial distribution. These methods have not previously been applied to spatiotemporal exposure data.

**Methods**—We analyzed the association between fine particulate matter (PM<sub>2.5</sub>) and birth weight in the U.S. state of Georgia using records with estimated date of conception during 2002–2005 (n=403,881). We predicted trimester-specific PM<sub>2.5</sub> exposure using a complex spatiotemporal exposure model. To improve spatial compatibility, we restricted to mothers residing in counties with a PM<sub>2.5</sub> monitor (n=180,440). We accounted for additional measurement error via a non-parametric bootstrap.

**Results**—Third-trimester PM<sub>2.5</sub> exposure was associated with lower birth weight in the uncorrected (–2.4g per 1 µg/m<sup>3</sup> difference in exposure; 95% Confidence Interval [CI]: –3.9, –0.8) and bootstrap-corrected (–2.5g, 95% CI: –4.2, –0.8) analyses. Results for the unrestricted analysis were attenuated (–0.66g, 95% CI: –1.7, 0.35).

**Conclusions**—This study presents a novel application of measurement error correction for spatiotemporal air pollution exposures. Our results demonstrate the importance of spatial

---

Corresponding Author: Joshua P. Keller, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, jkelle46@jhu.edu, (443) 287-8782.

**Availability of Code and Data:** Code and simulated data for implementing the methods described in this manuscript are available as an eAppendix on the journal website.

**Conflict of Interest:** None declared.

**Human Subjects Approval:** This study was approved by the IRB of Emory University, #45413.

compatibility between monitor and subject locations and provide evidence of the association between air pollution exposure and birth weight.

### Keywords

air pollution; bootstrap; low birth weight; measurement error; particulate matter; spatiotemporal modeling

---

### Introduction

Air pollution epidemiology studies of ambient outdoor pollution frequently make use of predicted exposures through a two-stage procedure. First, a spatial or spatiotemporal model is developed using monitoring data from regulatory networks to predict exposures. Next, health outcomes are regressed on the predicted exposure and confounder variables in a second model. The use of predicted exposures in the health model can introduce measurement error in coefficient estimates.<sup>1-8</sup>

Traditional measurement error is classified into classical error, in which the measured value is an error-prone version of true exposure, and Berkson error, in which the measured value is a smoothed version of true exposure.<sup>9</sup> For predicted spatial exposures in air pollution epidemiology, two similar types of measurement error arise. Reduced variability in exposure predictions, due to smoothing in the exposure regression model, induces Berkson-like error that can underinflate standard errors and, for some exposure models, induce bias in the point estimate.<sup>2,5,8</sup> Classical-like error arises from having a finite number of monitors for building an exposure model, which results in variability in the exposure model parameter estimates and can induce bias and impact standard errors.<sup>2</sup>

Regression calibration<sup>9</sup> is a standard approach to measurement error correction, which uses validation data to estimate the relationship between the true exposure and the error-prone measurement and then corrects bias in regression coefficients estimated from the primary data. Although mathematically one can view the two-stage exposure and health modeling paradigm as a specific form of regression calibration, with the monitor data used as external validation data for calibrating geographic covariates,<sup>10</sup> the spatial or spatiotemporal nature of the data and practical sample size considerations imply that further correction is needed for the residual classical-like and Berkson-like error.<sup>11</sup>

The paradigm for modeling the exposure surface determines the appropriate method for accounting for the induced measurement error. If the exposure surface is assumed to be random with fixed monitor and cohort locations, such as in the classical formulation of kriging problems,<sup>12</sup> then a parametric bootstrap, or a similar approximate procedure, can be used to correctly remove bias and adjust standard errors.<sup>2</sup> A second approach to measurement error correction under this paradigm is the spatial simulation extrapolation method developed by Alexeeff *et al.*,<sup>7</sup> which introduces additional measurement error into simulated datasets and then calculates a back-transformed value that corresponds to no measurement error.

An alternative paradigm considers the exposure surface as fixed, with the locations of monitors and subjects providing the sampling variability.<sup>5,8</sup> In studies with long-term exposures, the fixed surface paradigm better reflects the frequentist sampling nature of the study, in that replicate experiments would involve different subjects at different locations, and possibly different monitor locations, but not a different long-term pollution concentration surface. Szpiro and Paciorek outlined the conditions necessary for consistent estimation of a health association using predicted exposures under the fixed surface paradigm with unlimited monitoring data.<sup>5</sup> An important criterion is *spatial compatibility*, which means that the monitor and cohort locations are samples from identical, or at least very similar, spatial distributions. This presents a challenge to air pollution epidemiology studies in which monitoring data are limited by regulatory siting procedures while cohort data are collected independently. Szpiro and Paciorek suggest excluding exposure or outcome data as one way to improve spatial compatibility.<sup>5</sup>

Improving spatial compatibility reduces bias from Berkson-like error, but bias may still exist from classical-like error with finite monitoring data. For both types of error, corrections must still be made for incorrect estimates of the standard error.<sup>5</sup> Under the fixed surface paradigm, a non-parametric bootstrap that resamples monitors and subjects provides an appropriate method to estimate bias and corrected standard errors.<sup>5</sup>

While these measurement error correction methods are well developed for long-term spatial exposures, they have not been applied to spatiotemporal exposures that vary across space and in time. In the spatiotemporal setting, the fixed surface paradigm extends to a setting in which the three-dimensional exposure ‘surface’ is fixed, meaning that pollutant concentrations do not change in repeat experiments, but the location, time, and duration of observations may change due to monitor placement in space and time.

In this paper, we extend these measurement error correction methods to an analysis of ambient fine particulate matter (PM<sub>2.5</sub>; particles less than 2.5 μm in aerodynamic diameter) and birth weight in a cohort from the U.S. state of Georgia. Exposure to ambient PM<sub>2.5</sub> during pregnancy has been associated with low birth weight in many studies, although the estimates reported in the literature are heterogeneous.<sup>13–16</sup> In birth outcome studies, estimates of time-varying exposures are needed for specific pregnancy windows using reported gestational age and birth date. Other than the work of Alexeeff *et al.*,<sup>7</sup> who applied spatial simulation extrapolation to spatial exposure models to correct for measurement error in a Massachusetts birth cohort, published studies of birth weight and PM<sub>2.5</sub> have not accounted for exposure measurement error. Here, we extend bootstrap-based measurement error correction methods to spatiotemporal exposures at the trimester scale and compare the results obtained from applying different correction approaches.

## Methods

### Study Population

We obtained vital record data for births from the Office of Health Information and Policy (OHIP), Georgia Division of Public Health. The study cohort included all singleton, full-term (gestational age ≥ 37 weeks) births without major structural congenital birth defects and

with conception date between January 1, 2002 and December 31, 2005 to mothers ages 15 to 44 (n=442,436). Gestational age was defined according to date of last menstrual period, and date of conception was estimated by adding 14 days to the date of last menstrual period. We limited the study cohort to births to non-Hispanic white, non-Hispanic black, and Hispanic mothers. Each record was geocoded to one of the 4,788 Census 2000 block groups in Georgia by OHIP using maternal residential address at delivery. After removing records with incomplete covariate information, 403,881 birth records were available for analysis (see Figure 1). This study was approved by the Institutional Review Board of Emory University.

## Exposure Prediction

To predict trimester-specific ambient  $PM_{2.5}$  exposures for each maternal record, we fit a spatiotemporal model for  $PM_{2.5}$  with 2-week temporal and 1 km grid-cell spatial resolution. The  $PM_{2.5}$  modeling region included all of Georgia and portions of surrounding states (see Figure 2). We divided the monitoring domain into 598,320 1km grid cells to accommodate gridded covariate information.

We obtained daily 24-hour average  $PM_{2.5}$  concentrations from the EPA Air Quality System (<http://www.epa.gov/ttn/airs/airsaqs/>) for 74 Federal Reference Method monitors (see Figure 2). We aggregated the daily measurements to two-week average concentrations, dropping two-week averages derived from fewer than four measurements. For modeling, monitor values were assigned to the center of the grid cell containing the monitor. We obtained measurements of average elevation within each grid cell from the national elevation database (<http://ned.usgs.gov>). We computed the average percent forest cover using data from the National Land Cover Database (<http://www.epa.gov/mrlc/nlcd-2001>) and the total road length within each grid cell from ESRI Streetmap road data (Environmental Systems Research Institute Inc., Redland, CA). We calculated the distance to emissions point sources using data from the 2002 National Emissions Inventory report. Two-week average values of temperature, wind speed, and relative humidity were calculated at each grid cell for the entire study period, as described by Hu *et al.*<sup>17</sup>

The hierarchical spatiotemporal model for outdoor ambient  $PM_{2.5}$  can be written as

$$C(s, t) = \beta_0(s) + \beta_1(s)f(t) + \sum \gamma_k M_k(s, t) + \nu(s, t),$$

where  $\beta_0(s)$  is a long-term spatial mean,  $\beta_1(s)f(t)$  is a time trend with spatially-varying coefficients,  $M_k(s, t)$  are spatiotemporal covariates, and  $\nu(s, t)$  is a space-time residual term.<sup>18–21</sup> This model is implemented in the SpatioTemporal package in R.<sup>22</sup> The long-term mean component included percent forest cover and distance to emissions sources as covariates and a spatially correlated error structure, parameterized as an exponential covariance with range and sill. The time trend was estimated from observations at the Air Quality System monitors.<sup>21,23</sup> The trend coefficients had a mean depending upon elevation, forest cover, local and highway road length and a spatially correlated error structure.

We evaluated the predictive accuracy of the model using leave-one-out cross validation, a procedure that involves leaving out one monitor location at a time and refitting the model

using all remaining monitors. We quantified cross-validation performance via mean-squared error-based  $R^2$ ,<sup>18,19</sup> root mean-squared error, and mean absolute error. We computed these measures with two-week values, which incorporate spatial and temporal variability, and with non-overlapping twelve-week averages, to better reflect the trimester exposures of interest.

We predicted two-week average concentrations of ambient  $PM_{2.5}$  at grid cell locations across the modeling domain for the period January 1, 2002 to December 31, 2006. Ambient concentrations at each census block group were assigned based upon the value of the grid cell containing the block group, and so each birth record linked to a single block group had the same exposure. Trimester average exposures were computed for each record by assigning trimester start and end dates to nearest two-week modeling period and averaging predicted exposures for the corresponding block group between these dates.

### Epidemiologic Analysis

The spatial distribution of births (Figure 1) encompasses a much broader area of the state than is covered by the  $PM_{2.5}$  monitors (Figure 2), although most monitors are in counties with high birth counts. The siting guidelines for  $PM_{2.5}$  monitors favor areas of high population density,<sup>24</sup> but also include criteria addressing areas of high pollution levels and locations near important sources. Therefore, we do not regard the spatial distributions of monitor and subject locations as equivalent. To improve spatial compatibility between the distributions of monitor locations and maternal residences, we restricted our primary analysis to births to mothers residing in a county with a monitor, which left 180,440 records for analysis. As a sensitivity analysis, we also fit the health model using records from all counties, which we expect to suffer from measurement error due to spatial incompatibility between the subjects and the monitors.

We modeled birth weight as a continuous outcome in a linear regression model with trimester  $PM_{2.5}$  exposure as the covariate of interest. Following the work of Darrow *et al.*,<sup>25</sup> we included indicators for maternal education (less than 9<sup>th</sup> grade, 9<sup>th</sup> through 12<sup>th</sup> grade, high school diploma or equivalent, some college or higher), maternal race, reported tobacco and alcohol use during pregnancy, and county as potential confounders. We included natural cubic splines for the potential confounders census block group percent below poverty (3 degrees of freedom [df]), conception date (16 df), and maternal age (3 df) and its interaction with maternal race. Due to their strong relationship with birth weight, we included indicators for infant sex, gestational week, and being firstborn.

We conducted a sensitivity analysis that varied the extent of temporal adjustment in the birth weight model by increasing the degrees of freedom in the spline for conception date. We explored possible non-linearity in the association between  $PM_{2.5}$  and birth weight by representing  $PM_{2.5}$  via a spline term in the birth weight model.

### Non-Parametric Bootstrap

The restriction of the study cohort to subjects residing in a county with a monitor provides one correction for measurement error, by approximately matching the distribution of subject and monitor locations to improve spatial compatibility. To address remaining measurement

error, we implemented a non-parametric bootstrap that extends the method of Szpiro and Paciorek to spatiotemporal data.<sup>5</sup>

The non-parametric bootstrap resamples monitor locations to reflect variation in the predicted exposure surface derived from different monitor locations and resamples birth records to capture sampling variability in the epidemiologic analysis arising from different subjects. The procedure can be outlined as:

1. Resample 74 PM<sub>2.5</sub> monitors, together with their observations, with replacement.
2. Estimate spatiotemporal exposure model parameters using the sampled monitors.
3. Make new predictions of two-week average PM<sub>2.5</sub> concentrations at each grid cell center.
4. Resample 180,440 birth records with replacement, and compute the average trimester exposure for each record.
5. Fit a regression model using the sampled birth records and their new predicted exposures.

This procedure was repeated 5,000 times. If a monitor was selected more than once, we jittered its location by up to 300 meters to avoid model-fitting problems arising from exact colocation. eAppendix 1 provides code demonstrating this procedure.

We used the average PM<sub>2.5</sub> coefficient from the bootstrap datasets to estimate bias in the association estimate from the primary model.<sup>26</sup> The empirical standard error of the bootstrap point estimates was used to estimate the standard error of the bias-corrected estimate.

### Parameter Bootstrap

We conducted a sensitivity analysis to examine an alternative measurement error correction method that assumes a random spatiotemporal surface paradigm for PM<sub>2.5</sub> concentrations. For this analysis we implemented the parameter bootstrap, which is a computational approximation to the parametric bootstrap.<sup>2</sup> The parameter bootstrap holds locations and birth records fixed and simulates new exposures and monitor observations. This procedure can be outlined as:

1. Simulate new monitor observations and ‘true’ grid cell concentrations using the spatiotemporal exposure model with parameter estimates from the original data.
2. Estimate the spatiotemporal exposure model parameters using the simulated monitor observations.
3. Make new predictions of two-week average PM<sub>2.5</sub> concentrations at each grid cell center.
4. Simulate a new birth weight for each record from the ‘true’ exposure at the corresponding grid cell and all other covariates, using the coefficients from the original health model fit.
5. Fit a regression model using the simulated birth weights and all covariates as in the original analysis.

The bootstrap point estimates are used to bias-correct the original estimate and estimate its standard error in the same manner as the non-parametric bootstrap.

## Results

Mean birth weight for infants in the restricted cohort was 3,355 g (standard deviation [SD] of 465 g), compared to 3,368 g (SD: 466 g) in the statewide cohort. Mothers in the restricted cohort had a higher prevalence of self-reported tobacco use during pregnancy and were more likely to be non-Hispanic black than mothers in the statewide cohort (see Table 1).

The number of two-week observations at PM<sub>2.5</sub> monitoring sites during the study period ranged from 5 to 156, with a median of 89 observations. The mean of the two-week concentrations at each monitoring site ranged from 10.0 µg/m<sup>3</sup> to 18.1 µg/m<sup>3</sup>, with an average concentration of 14.2 µg/m<sup>3</sup> (SD: 1.4 µg/m<sup>3</sup>). The estimated PM<sub>2.5</sub> time trend showed regular seasonal variability across the study period, with slightly higher amplitude in later years (see eFigure 1). The estimated range parameters for the long-term mean, time trend coefficients, and residual term were 65km, 191km, and 659km, respectively, indicating a large amount of spatial smoothing in all model components. Full model coefficients are provided in eTable 1. Cross-validated model performance was excellent at the two-week scale (CV R<sup>2</sup> = 0.81, RMSE = 1.9 µg/m<sup>3</sup>, MAE = 1.4 µg/m<sup>3</sup>) and at the twelve-week scale (CV R<sup>2</sup> = 0.82, RMSE = 1.4 µg/m<sup>3</sup>, MAE = 1.0 µg/m<sup>3</sup>). eFigure 2 shows a scatterplot of predicted and observed two-week values at monitor locations.

Predicted PM<sub>2.5</sub> exposure was highest during the third trimester for the restricted cohort (mean of 14.6 µg/m<sup>3</sup>), since births were most frequent in the summer months, when PM<sub>2.5</sub> concentrations were also highest. In all trimesters, the mean exposure was higher for the restricted cohort than for the statewide cohort (see Table 2). Correlations between exposures in consecutive trimesters, without controlling for seasonality, were small (0.26 to 0.39), while exposures in the first and third trimester were negatively correlated. Differences in predicted exposure across space can be seen in Figure 3, which shows the long-term average prediction at each block group.

In births to mothers living in a county with a PM<sub>2.5</sub> monitor, a difference of 1 µg/m<sup>3</sup> in ambient PM<sub>2.5</sub> during the first, second, and third trimesters was associated with differences in birth weight of -1.5 g (95% Confidence Interval [CI]: -3.1, 0.1), -1.6 g (95% CI: -3.2, 0.1), and -2.4 g (95% CI: -3.9, -0.8), respectively, when no bootstrap correction was made (Table 3). The non-parametric bootstrap corrected estimates for each trimester were -1.5 g (95% CI: -3.2, 0.1), -1.6 g (95% CI: -3.3, 0.0), and -2.5 g (95% CI: -4.2, -0.8), respectively. The bootstrap-corrected standard error for the third trimester was 0.85 g, compared to 0.78 g in the uncorrected analysis. In contrast to the non-parametric bootstrap, the parameter bootstrap results showed little bias in the point estimates but much larger underinflation in the standard errors. For the third trimester, the parameter bootstrap corrected estimate was -2.4 g with a standard error of 0.92 g.

In the sensitivity analysis that included births from all counties, the estimated associations for all trimesters were attenuated. The estimates for second and third trimester exposure

were similar ( $-0.62$  [95% CI:  $-1.7, 0.45$ ] and  $-0.66$  [95% CI:  $-1.7, 0.35$ ] g per  $1 \mu\text{g}/\text{m}^3$ , respectively), and the association for the first trimester was weaker and had opposite sign ( $0.28$  g per  $1 \mu\text{g}/\text{m}^3$ ). The standard errors for these estimates were much smaller ( $0.51$  to  $0.55$ ) than those from the analysis of the restricted cohort. As a further sensitivity analysis, we applied the bootstrap correction methods to the statewide dataset. The results, provided in eTable 2, are similar to the uncorrected statewide estimates.

Sensitivity analyses for increased temporal adjustment did not yield substantively different estimates for the association of interest (see eTable 3). There was no evidence that the  $\text{PM}_{2.5}$ —birth weight association was nonlinear.

## Discussion

We have presented an approach for correcting for measurement error induced by using predicted spatiotemporal pollutant exposures in regression models. Using Georgia birth records, we demonstrated these methods in an analysis of the association between birth weight and ambient fine particulate matter. To improve spatial compatibility between the monitor and subject location distributions, we restricted the analysis to births to mothers residing in a county with a monitor. We observed a strong association between third trimester exposure and reduced birth weight. This association was robust to further correction for measurement error from a non-parametric bootstrap.

The large difference in point estimates between the primary analysis of the restricted cohort and the sensitivity analysis of the statewide cohort highlights the importance of accounting for spatial compatibility in air pollution epidemiology studies. However, the small number of monitors makes it difficult to be certain of the true extent of spatial incompatibility and our success in correcting it by restricting the cohort. Theoretical results of Szpiro and Paciorek<sup>5</sup> established the importance of spatial compatibility when using unpenalized spatial splines, but this work suggests it is an important condition for spatiotemporal, kriging models as well.

Within the restricted cohort analysis, the non-parametric bootstrap showed different amounts of measurement error between trimesters. The slight bias in the first trimester estimate is within the range of simulation error expected for the number of bootstrap replications. Although the magnitudes of the absolute bias in the second and third trimester estimates differed ( $0.08$  g and  $0.14$  g, respectively), their relative biases were quite similar ( $5.1\%$  and  $5.9\%$ , respectively). Bias due to classical-like measurement error has been shown to be multiplicative for unpenalized spatial spline exposure models under the fixed-surface paradigm,<sup>5</sup> and these results suggest this may also hold for spatiotemporal kriging models.

The contrasting results from the parameter bootstrap and non-parametric bootstrap corrections demonstrate the importance of the assumed underlying framework for the exposure surface across space. The trimester-length averaging period is long enough that we believe long-term geographical and commercial features and seasonal meteorology determine the average concentrations and therefore the fixed spatiotemporal surface paradigm is appropriate. Under this paradigm, we do not assume that the statistical model



can fully represent the underlying spatiotemporal surface. Resampling of the monitor locations and times leads to spatially- and temporally-varying bias, but the finite variability in the underlying surface limits variation between predicted exposures derived from different monitor locations and times. Thus, we expect the non-parametric bootstrap to partition measurement error into both bias and increased variability of the point estimate. In contrast, the random surface paradigm assumes a correctly specified exposure model that varies across studies due to different realizations of a spatiotemporal process. We expect that the parameter bootstrap, which simulates new observations from the model, to identify additional variability in the point estimate but little bias.

The birth weight results of this analysis are consistent with previously reported findings that corrected for exposure measurement error. In an analysis of a Massachusetts birth cohort from 2008, Alexeeff *et al.* found that a 1  $\mu\text{g}/\text{m}^3$  difference in ambient  $\text{PM}_{2.5}$  exposure during the third trimester was associated with 3.5 g lower birth weight.<sup>7</sup> This difference increased in magnitude to 4.9 g lower birth weight when they applied their spatial simulation extrapolation measurement error correction method. Unlike our approach, they corrected exposures estimated from averages of independent monthly spatial exposure models instead of correcting predictions from a spatiotemporal model.

Although our correction methods address measurement error introduced from the use of predicted exposures, they do not account for all sources of measurement error, particularly factors that influence the relationship between ambient level and personal exposure. For example, maternal residential mobility may introduce measurement error when exposures are assigned based upon address at delivery,<sup>27,28</sup> although a recent study suggests this impact may be limited.<sup>29</sup> For mothers who do not move during pregnancy, the assignment of each birth record to a location at the resolution of census block group potentially introduces exposure misclassification because this location may not well reflect the daily activity pattern of the mother. Although the spatiotemporal modeling framework provides more spatially refined estimates of exposure than using assignment to values at a central site, the 1km grid cannot capture fine-scale gradients below this resolution. However, particulate matter concentrations ( $\text{PM}_{2.5}$  and coarser fractions) tend to be more spatially homogeneous than primary pollutants such as  $\text{NO}_2$ , so this may not substantively impact inference.<sup>30–32</sup> Less daily mobility in the final weeks prior to birth may reduce this source of measurement error in the third trimester analysis. In general, these sources of measurement error could affect both point estimates and standard errors, and are difficult to correct for without further individual-level information.

We believe we have adequately accounted for residual temporal confounding, as the results did not change substantively when additional degrees of freedom were added to the temporal adjustment; however, some residual within-county spatial confounding might be present. The regression model for birth weight does not include several potentially confounding maternal factors for which information was not available, including illicit drug use, stress, and socioeconomic differences beyond education level and census tract-level poverty measures.<sup>33</sup> We controlled for gestational age in the analyses because preterm birth is a complex disease with multiple causes that exhibits strong socioeconomic and spatial patterning at the population level.<sup>34</sup> As reduced birth weight shares many of these same

causes, we hope to have limited the potential for bias from spatial confounding by controlling for gestational age. Furthermore, gestational age is very strongly correlated with birth weight, and controlling for gestational age greatly improves the precision of the association estimates. The restriction of the cohort to full-term, live births without structural defects could potentially introduce some collider stratification bias in an estimate of the association between particulate matter and birth weight. However, most of the variability in gestational age is caused by factors other than ambient particulate matter concentrations, so the magnitude of this potential collider bias would likely be small.

While the restriction to births to mothers residing in counties with a monitor reduces measurement error by improving spatial compatibility, it also results in the analysis being performed on a subset of the population. Regulatory monitors tend to be sited near urban areas, and the restricted cohort differed from the statewide cohort in racial composition and tobacco use. As such, the effect estimates from the restricted cohort analysis likely better represent the associations in urban areas than for Georgia as a whole. Methods exist for generalizing these results to the full state population,<sup>35,36</sup> but they rely on strong assumptions about the relationships between PM<sub>2.5</sub>, birth weight, and measured covariates that we do not make here. The tradeoff between reducing measurement error and analyzing only a subset of the larger population is an important challenge facing birth cohort studies. Supplemental air quality monitoring provides one approach to mitigating this challenge,<sup>37</sup> but can be prohibitively difficult for large cohorts. Satellite-based measurements of surrogate exposure are becoming more widely used in areas with limited monitoring, but do not directly solve the problem of limited monitoring data since satellite measurements require ground-based monitoring data for calibration.<sup>17,38</sup> Furthermore, other sources of measurement error discussed above, such as variability in daily activity patterns, would remain.

In summary, we extended bootstrap correction methods to spatiotemporal exposures to further reduce measurement error and improve inference about health effects of environmental exposures. We analyzed a restricted cohort to reduce measurement error, and we demonstrated the importance of spatial compatibility between subject and monitor locations. Our results support an association between third-trimester ambient fine particulate matter exposure and reduced birth weight.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

### Sources of Financial Support:

This publication was made possible by grants RD-834799 and RD-83479601-0 from the United States Environmental Protection Agency (US EPA). Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the US EPA. Further, US EPA does not endorse the purchase of any commercial products or services mentioned in the publication. Support was also provided by NIH/NIEHS through R01ES009411, R01ES020871, R21ES024894, R21ES022795, and the Biostatistics, Epidemiologic, and Bioinformatic Training in Environmental Health Training Grant T32ES015459.

## References

1. Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull BA. Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics*. 2009; 10(2):258–274. DOI: 10.1093/biostatistics/kxn033 [PubMed: 18927119]
2. Szpiro AA, Sheppard L, Lumley T. Efficient measurement error correction with spatially misaligned data. *Biostatistics*. 2011; 12(4):610–23. DOI: 10.1093/biostatistics/kxq083 [PubMed: 21252080]
3. Basagaña X, Aguilera I, Rivera M, et al. Measurement error in epidemiologic studies of air pollution based on land-use regression models. *Am J Epidemiol*. 2013; 178(8):1342–1346. DOI: 10.1093/aje/kwt127 [PubMed: 24105967]
4. Lopiano KK, Young LJ, Gotway Ca. Estimated generalized least squares in spatially misaligned regression models with Berkson error. *Biostatistics*. 2013; 14:737–751. DOI: 10.1093/biostatistics/kxt011\_online [PubMed: 23568241]
5. Szpiro AA, Paciorek CJ. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics*. 2013; 24:501–517. DOI: 10.1002/env.2233 [PubMed: 24764691]
6. Bergen S, Sheppard L, Sampson PD, et al. A National Prediction Model for PM<sub>2.5</sub> Component Exposures and Measurement Error-Corrected Health Effect Inference. *Environ Health Perspect*. 2013; 121(9):1017–1025. DOI: 10.1289/ehp.1206010 [PubMed: 23757600]
7. Alexeeff SE, Carroll RJ, Coull B. Spatial measurement error and correction by spatial SIMEX in linear regression models when using predicted air pollution exposures. *Biostatistics*. 2016; 17(2): 377–389. DOI: 10.1093/biostatistics/kxv048 [PubMed: 26621845]
8. Bergen S, Szpiro AA. Mitigating the impact of measurement error when using penalized regression to model exposure in two-stage air pollution epidemiology studies. *Environ Ecol Stat*. 2015; 22:601–631. DOI: 10.1007/s10651-015-0314-y
9. Carroll, RJ., Ruppert, D., Stefanski, LA., Crainiceanu, CM. *Measurement Error in Nonlinear Models*. 2. Boca Raton: Chapman & Hall/CRC; 2006.
10. Spiegelman D. Regression calibration in air pollution epidemiology with exposure estimated by spatio-temporal modeling. *Environmetrics*. 2013; 24(8):521–524. DOI: 10.1002/env.2249
11. Szpiro AA, Paciorek CJ. Rejoinder. *Environmetrics*. 2013; 24(8):531–536. DOI: 10.1002/env.2254
12. Cressie, N. *Statistics for Spatial Data*. Hoboken, NJ: John Wiley & Sons; 1993. Revised Ed
13. Davdand P, Parker J, Bell ML, et al. Maternal exposure to particulate air pollution and term birth weight: A multi-country evaluation of effect and heterogeneity. *Environ Health Perspect*. 2013; 121(3):367–373. DOI: 10.1289/ehp.1205575
14. Shah PS, Balkhair T. Air pollution and birth outcomes: A systematic review. *Environ Int*. 2011; 37(2):498–516. DOI: 10.1016/j.envint.2010.10.009 [PubMed: 21112090]
15. Stieb DM, Chen L, Eshoul M, Judek S. Ambient air pollution, birth weight and preterm birth: A systematic review and meta-analysis. *Environ Res*. 2012; 117:100–111. DOI: 10.1016/j.envres.2012.05.007 [PubMed: 22726801]
16. Fleischer NL, Merialdi M, van Donkelaar A, et al. Outdoor air pollution, preterm birth, and low birth weight: Analysis of the world health organization global survey on maternal and perinatal health. *Environ Health Perspect*. 2014; 122(4):425–430. DOI: 10.1289/ehp.1306837 [PubMed: 24508912]
17. Hu X, Waller LA, Lyapustin A, et al. Estimating ground-level PM<sub>2.5</sub> concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sens Environ*. 2014; 140:220–232. DOI: 10.1016/j.rse.2013.08.032
18. Keller JP, Olives C, Kim S-Y, et al. A unified spatiotemporal modeling approach for predicting concentrations of multiple air pollutants in the Multi-Ethnic Study of Atherosclerosis and Air Pollution. *Environ Health Perspect*. 2015; 123(4):301–309. DOI: 10.1289/ehp.1408145 [PubMed: 25398188]
19. Lindström J, Szpiro AA, Sampson PD, et al. A flexible spatio-temporal model for air pollution with spatio-temporal covariates. *Environ Ecol Stat*. 2014; 21:411–433. DOI: 10.1007/s10651-013-0261-4 [PubMed: 25264424]

20. Sampson PD, Szpiro AA, Sheppard L, Lindström J, Kaufman JD. Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data. *Atmos Environ*. 2011; 45(36):6593–6606. DOI: 10.1016/j.atmosenv.2011.04.073
21. Szpiro AA, Sampson PD, Sheppard L, Lumley T, Adar SD, Kaufman JD. Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. *Environmetrics*. 2010; 21:606–631. DOI: 10.1002/env
22. Lindström, J., Szpiro, AA., Sampson, PD., Bergen, S., Oron, AP. R package version 1.1.1. 2012. *SpatioTemporal: Spatio-Temporal Model Estimation*.
23. Fuentes, M., Guttorp, P., Sampson, PD. Using transforms to analyze space-time processes. In: Held, L.Finkenstadt, B., Isham, V., editors. *Statistical Methods for Spatio-Temporal Systems*. Chapman Hall/CRC; 2007. p. 77-150.
24. Georgia Department of Natural Resources. 2009 Ambient Air Surveillance Report. 2010.
25. Darrow LA, Klein M, Strickland MJ, Mulholland JA, Tolbert PE. Ambient air pollution and birth weight in full-term infants in Atlanta, 1994–2004. *Environ Health Perspect*. 2010; 119(5):731–737. DOI: 10.1289/ehp.1002785 [PubMed: 21156397]
26. Davison, AC., Hinkley, DV. *Bootstrap Methods and Their Applications*. New York, NY: Cambridge University Press; 1997.
27. Miller A, Siffel C, Correa A. Residential mobility during pregnancy: Patterns and correlates. *Matern Child Health J*. 2010; 14(4):625–634. DOI: 10.1007/s10995-009-0492-z [PubMed: 19568920]
28. Bell ML, Belanger K. Review of research on residential mobility during pregnancy: consequences for assessment of prenatal environmental exposures. *J Expo Sci Environ Epidemiol*. 2012; 22(5): 429–438. DOI: 10.1038/jes.2012.42 [PubMed: 22617723]
29. Pereira G, Bracken MB, Bell ML. Particulate air pollution, fetal growth and gestational length: The influence of residential mobility in pregnancy. *Environ Res*. 2016; 147:269–274. DOI: 10.1016/j.envres.2016.02.001 [PubMed: 26918840]
30. Yann S, Galineau J, Hulin A, et al. Health effects of ambient air pollution: Do different methods for estimating exposure lead to different results? *Environ Int*. 2014; 66:165–173. DOI: 10.1016/j.envint.2014.02.001 [PubMed: 24598283]
31. Brauer M, Lencar C, Tamburic L, Koehoorn M, Demers P, Karr C. A cohort study of traffic-related air pollution impacts on birth outcomes. *Environ Health Perspect*. 2008; 116(5):680–686. DOI: 10.1289/ehp.10952 [PubMed: 18470315]
32. Dionisio KL, Baxter LK, Chang HH. An empirical assessment of exposure measurement error and effect attenuation in bipollutant epidemiologic models. *Environ Health Perspect*. 2014; 122(11): 1216–1223. DOI: 10.1289/ehp.1307772 [PubMed: 25003573]
33. Hao H, Chang HH, Holmes HA, et al. Air pollution and preterm birth in the U.S. state of Georgia (2002–2006): Associations with concentrations of 11 ambient air pollutants estimated by combining Community Multiscale Air Quality Model (CMAQ) simulations with stationary monitor measurements. *Environ Health Perspect*. 2016; 124:875–880. DOI: 10.1289/ehp.1409651 [PubMed: 26485731]
34. Lorch SA, Enlow E. The role of social determinants in explaining racial/ethnic disparities in perinatal outcomes. *Pediatr Res*. 2016; 79:141–147. DOI: 10.1038/pr.2015.199 [PubMed: 26466077]
35. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. *J R Stat Soc Ser A Stat Soc*. 2011; 174(2):369–386. DOI: 10.1111/j.1467-985X.2010.00673.x
36. Lesko CR, Cole SR, Hall HI, et al. The effect of antiretroviral therapy on all-cause mortality, generalized to persons diagnosed with HIV in the USA, 2009–11. *Int J Epidemiol*. 2016; 45(1): 140–150. DOI: 10.1093/ije/dyv352 [PubMed: 26772869]
37. Cohen MA, Adar SD, Allen RW, et al. Approach to estimating participant pollutant exposures in the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Environ Sci Technol*. 2009; 43(13):4687–93. [PubMed: 19673252]

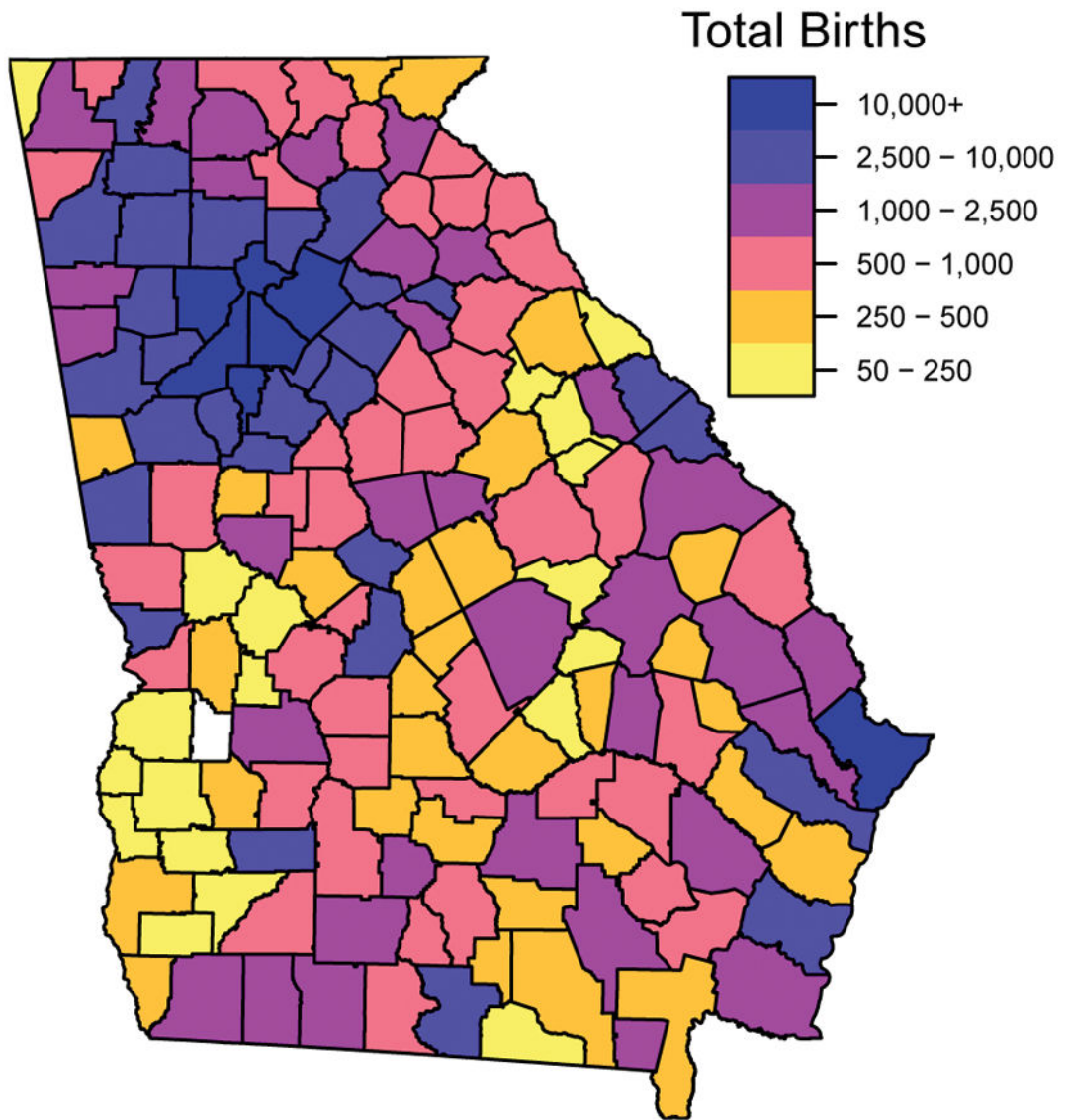
38. Kloog I, Koutrakis P, Coull BA, Lee HJ, Schwartz J. Assessing temporally and spatially resolved PM<sub>2.5</sub> exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos Environ.* 2011; 45(35):6267–6275. DOI: 10.1016/j.atmosenv.2011.08.066

Author Manuscript

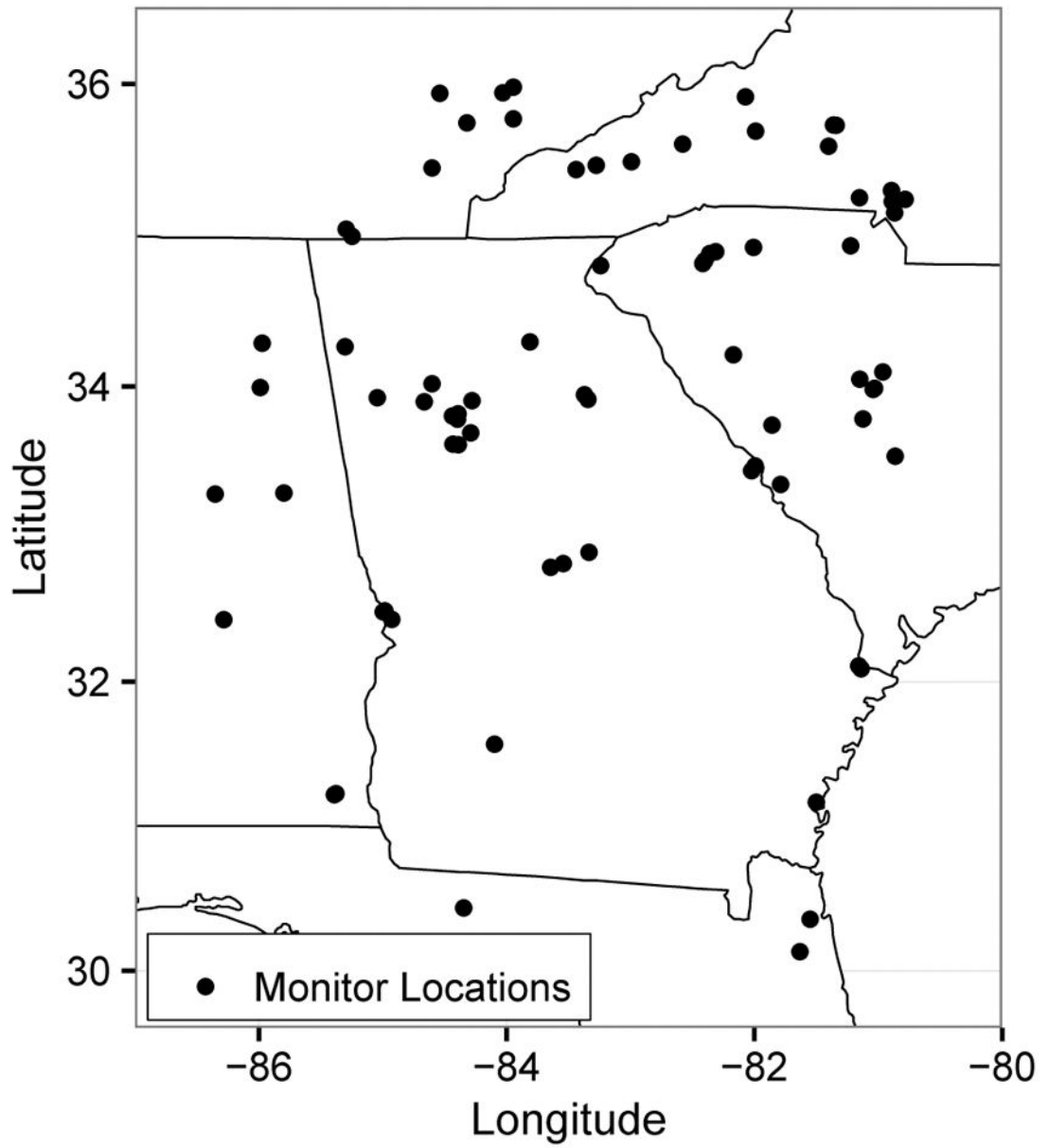
Author Manuscript

Author Manuscript

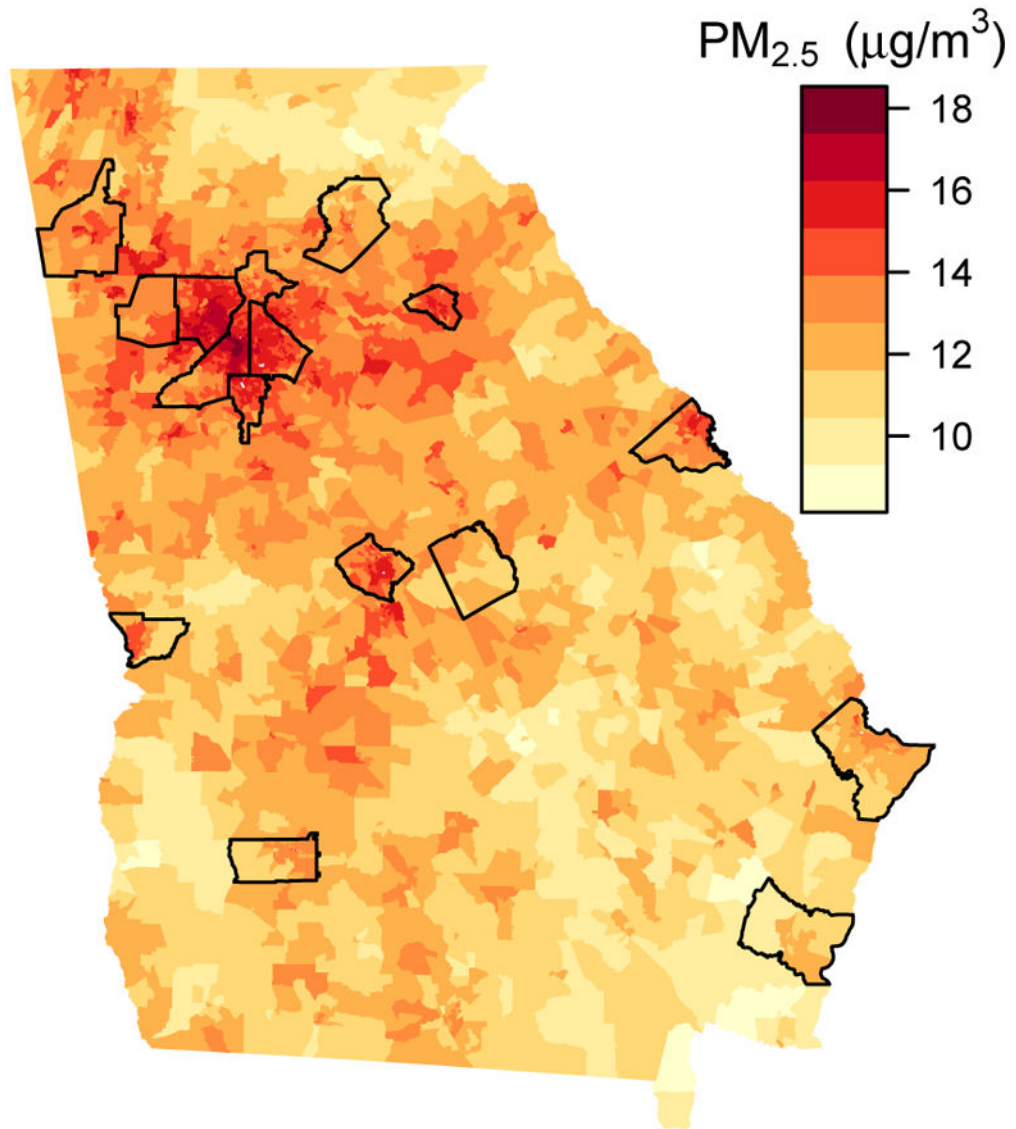
Author Manuscript



**Figure 1.**  
Total births by county, for estimated conception dates between 2002 and 2005.



**Figure 2.**  
Locations of Air Quality System monitors used for the exposure prediction model.



**Figure 3.** Predicted Long-term PM<sub>2.5</sub> concentrations at block groups in Georgia. Black lines outline the counties that contain a monitor.



**Table 1**

Maternal and infant characteristics [n(%)] for full-term births to mothers residing in a Georgia county with a PM<sub>2.5</sub> monitor and in all Georgia counties, January 1, 2002 to December 31, 2005.

Characteristic	Counties with Monitor (n=180,440)	All Counties (n=403,881)
Birthweight (g)		
<2,500	5,448 (3)	11,429 (3)
2,500–3,000	32,791 (18)	70,960 (18)
3,000–3,500	76,709 (43)	170,279 (42)
3,500	65,492 (36)	151,213 (37)
Female Sex	88,670 (49)	198,463 (49)
Firstborn	73,485 (41)	160,659 (40)
Maternal Race		
Non-Hispanic White	71,707 (40)	212,231 (53)
Non-Hispanic Black	78,201 (43)	128,795 (32)
Hispanic	30,532 (17)	62,855 (16)
Maternal Education		
Less than 9 <sup>th</sup> Grade	13,536 (8)	28,697 (7)
9 <sup>th</sup> through 12 <sup>th</sup> Grade	28,604 (16)	70,124 (17)
High School Degree	52,613 (29)	125,112 (31)
Some college or higher	85,687 (47)	179,948 (45)
Maternal Age (years)		
<20	19,512 (11)	46,811 (12)
20–34	137,571 (76)	310,806 (77)
35	23,357 (13)	46,264 (11)
Married	99,951 (55)	243,110 (60)
Reported Use during pregnancy		
Tobacco	8,443 (5)	31,090 (8)
Alcohol	1,359 (1)	2,4667 (1)
Census Tract Level Poverty (% residents below poverty level)		
<5%	47,626 (26)	106,339 (26)
5%–10%	43,255 (24)	99,089 (25)
10%–20%	49,036 (27)	116,471 (29)
20%	40,523 (22)	81,982 (20)

**Table 2**  
Descriptive statistics for trimester-specific exposure predictions for the full and restricted birth cohort.

Cohort	Trimester	Mean $\pm$ SD	Min	Max	Correlation with Trimester	
					First	Second
Restricted <sup>a</sup>	First	14.4 $\pm$ 2.7	6.6	22		
	Second	14.5 $\pm$ 2.7	6.5	22	0.26	
	Third	14.6 $\pm$ 2.9	6.4	23	-0.40	0.32
Statewide	First	13.4 $\pm$ 2.8	4.9	22		
	Second	13.5 $\pm$ 2.8	5.0	22	0.34	
	Third	13.7 $\pm$ 3.0	4.8	23	-0.25	0.39

SD indicates standard deviation.

<sup>a</sup>Restricted cohort limited to births to mothers residing in a county with a PM<sub>2.5</sub> monitor.

Estimated difference in birth weight (in grams) associated<sup>a</sup> with 1  $\mu\text{g}/\text{m}^3$  higher ambient average  $\text{PM}_{2.5}$  during the specified trimester.

**Table 3**

Cohort	Bootstrap Correction	Trimester	Estimate	Std. Err	95% CI
Restricted <sup>b</sup>	None	1	-1.5	0.82	(-3.1, 0.1)
		2	-1.6	0.83	(-3.2, 0.1)
		3	-2.4	0.78	(-3.9, -0.8)
Non-Parametric	None	1	-1.5	0.84	(-3.2, 0.1)
		2	-1.6	0.85	(-3.3, 0.0)
		3	-2.5	0.85	(-4.2, -0.8)
Parameter	None	1	-1.6	1.01	(-3.5, 0.4)
		2	-1.6	0.98	(-3.5, 0.4)
		3	-2.4	0.92	(-4.2, -0.6)
Statewide	None	1	0.28	0.54	(-0.78, 1.3)
		2	-0.62	0.55	(-1.7, 0.45)
		3	-0.66	0.51	(-1.7, 0.35)

CI indicates confidence interval.

<sup>a</sup> All models are adjusted for gestational week, infant sex, being firstborn, maternal age, maternal race, maternal education, maternal tobacco, and alcohol use during pregnancy, county, and conception date.

<sup>b</sup> Restricted cohort limited to births to mothers residing in a county with a  $\text{PM}_{2.5}$  monitor.