

Exposure Measurement Error in PM_{2.5} Health Effects Studies: A Pooled Analysis of Eight Personal Exposure Validation Studies

Marianthi-Anna Kioumourtzoglou^{*1}, Donna Spiegelman^{2,3}, Adam A. Szpiro⁴, Lianne Sheppard^{4,5}, Joel D. Kaufman⁵, Jeff D. Yanosky⁶, Ronald Williams⁷, Francine Laden^{1,2}, Biling Hong², Helen Suh⁸

¹Department of Environmental Health, Harvard School of Public Health, Boston, MA

²Department of Epidemiology, Harvard School of Public Health, Boston, MA

³Department of Biostatistics, Harvard School of Public Health, Boston, MA

⁴Department of Biostatistics, University of Washington, Seattle, WA

⁵Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA

⁶Department of Public Health Sciences, Pennsylvania State University College of Medicine, Hershey, PA

⁷U.S. Environmental Protection Agency, Research Triangle Park, NC

⁸Department of Health Sciences, Northeastern University, Boston, MA

Email: Marianthi-Anna Kioumourtzoglou^{*} - marianthi.anna@mail.harvard.edu; Donna Spiegelman - stdls@hsph.harvard.edu; Adam A. Szpiro - aszpiro@u.washington.edu; Lianne Sheppard - sheppard@uw.edu; Joel D. Kaufman - joelk@uw.edu; Jeff D. Yanosky - jdy10@psu.edu; Ronald Williams - williams.ronald@epa.gov; Francine Laden - francine.laden@channing.harvard.edu; Biling Hong - bhong@hsph.harvard.edu; Helen Suh - h.suh@neu.edu;

^{*} Corresponding author

Abstract

Background

Exposure measurement error is a concern in long-term $PM_{2.5}$ health studies using ambient concentrations as exposures. We assessed error magnitude by estimating calibration coefficients as the association between personal $PM_{2.5}$ exposures from validation studies and typically available surrogate exposures.

Methods

Daily personal and ambient $PM_{2.5}$ and when available sulfate, measurements were compiled from nine cities, over 2 to 12 days. True exposure was defined as personal exposure to $PM_{2.5}$ of ambient origin. Since $PM_{2.5}$ of ambient origin could only be determined for five cities, personal exposure to total $PM_{2.5}$ was also considered. Surrogate exposures were estimated as ambient $PM_{2.5}$ at the nearest monitor or predicted outside subjects' homes. We estimated calibration coefficients by regressing true on surrogate exposures in random effects models.

Results

When monthly-averaged personal $PM_{2.5}$ of ambient origin was used as the true exposure, calibration coefficients equaled 0.31 (95%CI:0.14, 0.47) for nearest monitor and 0.54 (95%CI:0.42, 0.65) for outdoor home predictions. Between-city heterogeneity was not found for outdoor home $PM_{2.5}$ for either true exposure. Heterogeneity was significant for nearest monitor $PM_{2.5}$ for both true exposures, but not after adjusting for city-average motor vehicle number for total personal $PM_{2.5}$.

Conclusions

Calibration coefficients were <1 , consistent with previously reported chronic health risks using nearest monitor exposures being under-estimated when ambient concentrations are the exposure of interest. Calibration coefficients were closer to 1 for outdoor home predictions, likely reflecting less spatial error. Further research is needed to determine how our findings can be incorporated in future health studies.

Keywords

exposure measurement error, fine particles, fine particles of ambient origin, monitoring data, spatio-temporal models

Background

Exposure measurement error is a limitation of epidemiologic studies of fine particles ($\text{PM}_{2.5}$) [1–3], which generally assess exposures using ambient concentrations measured at centrally located monitors. The impact of error on observed health risks can be substantial, potentially distorting associations and interactions between covariates and outcomes, reducing the power to detect effects, and leading to invalid inference [3–5].

In time series studies, use of measurements from ambient monitors, even in absence of any instrumental error, has been shown to introduce both a Berkson error component, a result of using aggregated instead of individual exposure data, and a classical error component, a result of the difference between the aggregated exposure data and the true ambient $\text{PM}_{2.5}$ concentrations [5]. Berkson error would not bias the health effect estimates, but would lead to an increased variance, while classical error, conversely, can lead to bias [5,6]. It has been shown that in presence of multiple monitoring sites in a city, using across-monitor averages, or population-weighted averages, would lead to less bias in time series studies [6,7]. Furthermore, less bias is expected when the pollutant of interest is spatially homogeneous, such as $\text{PM}_{2.5}$ [6,8].

To minimize error, $\text{PM}_{2.5}$ exposures would ideally be measured using personal monitors, with analytical methods used to apportion these measurements into $\text{PM}_{2.5}$ constituents and sources. Such methods, however, are both expensive and intrusive and thus not feasible for studies conducted over long time periods with many subjects. Recently, for application to cohort studies, researchers have used statistical models to predict exposures outside participant residences [9–14], thus accounting for spatial variation in ambient concentrations. While an improvement, such models still do not account for all sources of exposure variability, such as activity patterns, which can lead to biased results [15]. It can be argued, nevertheless, that such “biases” are the result of different target parameters of interest for the health effects of ambient concentration vs. personal or ambient source exposure [16].

Further, spatial smoothing models can contribute a Berkson-like error component that results from smoothing the exposure surface and a classical-like component from variability in estimating model parameters [17–19]. The classical-like error can induce bias both towards and away from the null [18] with increased variability [18,20,21]. Even in absence of other error sources, nevertheless, health effects estimated using outdoor concentrations will be attenuated proportionally to the $\text{PM}_{2.5}$ infiltration factor, the factor describing how much of the personal $\text{PM}_{2.5}$ was generated outdoors, penetrated the building

envelope and remained airborne [5]; if the exposure of interest is outdoor pollutant concentration rather than infiltrated personal exposure, however, it has been argued that this attenuation should not be regarded as a manifestation of measurement error [19].

Several previous acute effects studies have adjusted for exposure measurement error, showing that use of surrogate exposures tends to bias the health effect estimates towards the null [22–24]. For long-term $\text{PM}_{2.5}$ effects, a limitation in understanding the impact of measurement error on estimated health risks is the paucity of long-term personal exposure data [25–28]. We compiled exposure data from nine studies to estimate calibration coefficients for $\text{PM}_{2.5}$ of ambient origin and total personal $\text{PM}_{2.5}$ for cases when ambient concentrations or spatial models are used to assess exposures. In light of the complexity of measurement error in air pollution, the time scale of our validation data, and the uncertainty in our estimated calibration coefficients, our aim was to estimate and characterize calibration coefficients for $\text{PM}_{2.5}$ but not to recommend their use to adjust health effect estimates in epidemiology studies directly. Our group is currently developing statistical methods to account for these limitations.

Methods

Personal Exposure Datasets

We included data from studies of personal $\text{PM}_{2.5}$ exposures based on the following criteria: i) the study had to be conducted in the United States, ii) during or after 1999, to ensure availability of $\text{PM}_{2.5}$ concentration measurements at a EPA monitor located nearby, and iii) we had to be able to obtain the raw data, vs. the published summary statistics, from the investigators who originally conducted the study.

Measurements of personal and ambient $\text{PM}_{2.5}$ and, when available, sulfate (SO_4^{2-}), were compiled from nine cities located throughout the United States (Table 1) [29–41]. A brief description of the validation studies is presented in the Appendix.

In each study, daily personal $\text{PM}_{2.5}$ exposure data were collected following panel study sampling designs. The number of subjects per study ranged between 15–201, with sampling session durations ranging from 2 to 12 days (median: 7 days). For each subject, we estimated monthly average personal exposures and used these in our analyses.

All subjects were non-smokers and were monitored in multiple seasons. Study subjects included the elderly, patients with myocardial infarction, children, and adults. All subjects younger than 18 years were excluded

from the analysis, since long-term air pollution health studies are often focused on adult mortality.

Surrogate Exposures

For each subject, we calculated two monthly $\text{PM}_{2.5}$ surrogate exposures. First, we determined monthly ambient $\text{PM}_{2.5}$ concentrations from the nearest US Environmental Protection Agency (EPA) Air Quality System monitor (nearest monitor), restricting the maximum allowed monitor-residence distance to 30 mi [42]. Monthly concentrations were estimated using all available data within the month, i.e. not only the days used for the monthly averages of the personal exposures.

Second, we estimated monthly outdoor $\text{PM}_{2.5}$ concentrations outside each subject’s residence, at the latitude and longitude of each subject’s residence, using a nationwide expansion of a geographic information system (GIS)-based spatio-temporal model [14, 43]. This model predicts monthly $\text{PM}_{2.5}$ concentrations using a generalized additive model that fits monitoring data from governmental and research networks together with GIS-based covariates, including population density, distance to nearest roads, elevation, urban land use, $\text{PM}_{2.5}$ point-source emissions and weather variables.

Estimation of Personal Exposures of Ambient Origin

We assume the true exposure metric is personal exposures to $\text{PM}_{2.5}$ of ambient origin, which reflect $\text{PM}_{2.5}$ from sources relevant to epidemiological studies of ambient air pollution [26, 44, 45]. This quantity cannot be measured directly.

To estimate personal $\text{PM}_{2.5}$ of ambient origin, we used ambient SO_4^{2-} measurements, which were available for four cities (Atlanta, Baltimore, Boston and Steubenville). The majority of SO_4^{2-} is formed in the atmosphere through secondary reactions via either gas-phase or gas/particle phase oxidation [46] and is generally associated with coal combustion and coal-fired power plant emissions [47, 48]. Because of negligible indoor sources and its similar spatial homogeneity as $\text{PM}_{2.5}$, SO_4^{2-} can serve as a tracer for PM of ambient origin in locations where SO_4^{2-} comprises a large part of the $\text{PM}_{2.5}$ mass [49, 50], with personal to ambient SO_4^{2-} ratio approximating the fraction of ambient $\text{PM}_{2.5}$ that infiltrates indoors and remains airborne:

$$\text{PM}_{\text{pers. of ambient origin}} = \frac{\text{SO}_4^{2-} \text{ pers.}}{\text{SO}_4^{2-} \text{ ambient}} \text{PM}_{\text{ambient}}$$

In Seattle, for which personal SO_4^{2-} data were not available, personal $\text{PM}_{2.5}$ of ambient origin was

estimated as the weighted average of the indoor PM_{2.5} of ambient origin (estimated using the corresponding calculated home infiltration efficiency) and ambient PM_{2.5}, with the proportion of time each subject spent indoors and outdoors as weights [51].

Since personal exposures to PM_{2.5} of ambient origin could only be estimated in five cities, we also assessed error using total personal PM_{2.5} exposure. For this measure, calibration coefficients will be less accurate, since total personal PM_{2.5} exposures also include indoor- and personally-generated PM_{2.5}, which are independent from ambient PM_{2.5} [45].

Calibration Coefficients

The calibration coefficients were estimated as the fixed regression coefficients (γ_1) from linear mixed effects models, of monthly averaged “true” on surrogate exposures, accounting for within-city correlated observations and repeated measures within subject:

$$X_{ijk} = (\gamma_0 + g_{1i} + g_{2ij}) + (\gamma_1 + g_{3i})Z_{ijk} + \gamma_2 \text{Season}_{ijk} + \varepsilon_{ijk}, \quad (1)$$

where X_{ijk} are the “true” (either personal PM_{2.5} of ambient origin or total personal PM_{2.5}) and Z_{ijk} the surrogate exposures (either nearest ambient PM_{2.5} monitor or spatio-temporal model predictions) for $j=1, \dots, J_i$ subjects within city $i=1, \dots, I$, and $I=5$ or 9 , with $k=1, \dots, K_{ij}$ repeated measures, $g_{1i} \sim \mathcal{N}(0, \sigma_{\text{city}}^2)$, $g_{2ij} \sim \mathcal{N}(0, \sigma_{\text{subject}}^2)$, $g_{3i} \sim \mathcal{N}(0, \sigma_{\text{CF-city}}^2)$ and $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_W^2)$.

We explored the sensitivity of our results to assumptions about the covariance structure for repeated measures within subjects. Results are reported assuming compound symmetry covariance, with results similar for autoregressive covariance structure or when allowing heteroscedasticity. We also allowed for random seasonal effects by city, but our results were materially unchanged (results not shown).

Calibration coefficients equal to 1 suggest no bias, while coefficients <1 suggest an attenuated effect estimate. The p-values (as p-value_1) presented with the estimated calibration coefficients correspond to the hypothesis that $\gamma_1 = 1$ and were obtained using $\frac{(\hat{\gamma}_1 - 1)^2}{\text{var}(\hat{\gamma}_1)} \sim \chi_1^2$.

Potential effect modification by season was assessed, as the association between personal exposures and ambient concentrations differs by season [29, 33, 37, 38]. Stratified calibration coefficients are presented when the estimated interaction term for season was significant, with October–March as winter and April–September as summer. Statistical significance was assessed at the 0.05 level.

Between-City Heterogeneity

To assess potential between-city heterogeneity in the calibration coefficients, we tested the hypothesis H_0 : $\sigma_{\text{CF-city}}^2=0$, comparing Model 1 to Model 2, where Model 2 is the same as Model 1 without the random slope for cities (g_{3i}):

$$X_{ijk} = (\gamma_0 + g_{1i} + g_{2ij}) + \gamma_1 Z_{ijk} + \gamma_2 \text{Season}_{ijk} + \varepsilon_{ijk} \quad (2)$$

We used a likelihood ratio test (LRT) for this comparison, with $\text{LRT} \sim 50 : 50$ mixture of χ_0^2 and χ_1^2 and $\text{p-value}=0.5$ if $\hat{\sigma}_{\text{CF-city}}^2 = 0$ and $\text{p-value}=0.5 \times (1 - \chi_1^2 \times (\text{LRT}))$ otherwise [52].

We used step-wise selection to identify city-specific variables explaining any observed between-city heterogeneity in the calibration coefficients. In presence of significant heterogeneity, we added to Model 1 candidate city-specific variables together with interaction terms between the candidate variable and the surrogate exposure (Model 3). The candidate variables were kept in the model if the interaction term was significant.

$$X_{ijk} = (\gamma_0 + g_{1i} + g_{2ij}) + (\gamma_1 + g_{3i})Z_{ijk} + \gamma_2 \text{Season}_{ijk} + \gamma_3 \text{CityVariables}_i + \gamma_4 Z_{ijk} \text{CityVariables}_i + \varepsilon_{ijk} \quad (3)$$

Candidate city-specific variables were identified from previous studies showing their importance to the personal-ambient relationship, including air conditioning use, unemployment, race, public transport [53–55] and traffic [54] (Table A4). City-specific variables were obtained from the U.S. Census Bureau (Census 2000, <http://www.census.gov>), the American Housing Survey (<http://www.census.gov/housing/ahs>), the National Climatic Data Center (www.ncdc.noaa.gov) and the Bureau of Labor Statistics (www.bls.gov).

Leave-one-out cross-validation techniques were employed to validate the variable selection process [56, Chapter 7.10]. By omitting one city at a time ($-i$), we re-fit Model 3, using data from the remaining $I - 1$ cities, allowing for a different set of variables to be selected each time. We then predicted the city-specific calibration coefficient for the omitted city using the estimated model parameters together with the selected variable(s) of the omitted city, i.e. $\hat{\gamma}_{1i-} = \hat{\gamma}_{1(-i)} + \hat{\gamma}_{4(-i)} \text{CityVariables}_i$. We also estimated city-specific calibration coefficients ($\hat{\gamma}_{1i}$) employing city-specific mixed effects models (Model 4). Finally, we compared the predicted to the observed city-specific calibration coefficients obtained from the city-specific models.

$$X_{ijk} = (\gamma_{0i} + g_{2ij}) + \gamma_{1i} Z_{ijk} + \gamma_{2i} \text{Season}_{ijk} + \varepsilon_{ijk} \quad (4)$$

We assessed the cross-validated results by the correlation between the predicted ($\hat{\gamma}_{1i-}$) and observed ($\hat{\gamma}_{1i}$) calibration factors, the relative bias $\frac{\hat{\gamma}_{1i-} - \hat{\gamma}_{1i}}{\hat{\gamma}_{1i}}$ and the absolute bias $|\frac{\hat{\gamma}_{1i-} - \hat{\gamma}_{1i}}{\hat{\gamma}_{1i}}|$, both averaged over all cities.

Sensitivity Analyses

To assess the robustness of our results, we assessed potential effect modification by subpopulation: seniors (subjects older than 65 years old) and subjects with COPD, myocardial infarction (MI), and coronary heart disease (CHD).

Sensitivity analyses were also performed to assess the effect of imperfectly matched monthly ambient and personal exposures. We calculated calibration coefficients for monthly ambient levels estimated using only those days for which personal exposure measures were available. Since the EPA does not collect data daily at all locations, we allowed subjects to be matched to the nearest monitor with available data for that day. This sensitivity analysis could only be performed for the nearest ambient monitor concentrations, as the outdoor home model predictions were calculated at the monthly level only.

In addition, we calculated calibration coefficients for total personal $\text{PM}_{2.5}$ exposures using the identical data as used to calculate calibration coefficients for personal $\text{PM}_{2.5}$ of ambient origin.

All statistical analyses were conducted using SAS software (Version 9.3, SAS Institute Inc, Cary, NC).

Results

Summary statistics and ambient-personal correlations are presented in Table 2 and Table A2, respectively. By-city summary statistics are presented in Table A1, and the relationship between exposure to $\text{PM}_{2.5}$ of ambient origin and ambient $\text{PM}_{2.5}$ concentrations is presented in Figure A1. On average, total personal $\text{PM}_{2.5}$ was higher than both concentrations at the nearest ambient monitor and outdoor home predictions. Concentrations at the ambient monitors were strongly correlated with outdoor home model predictions (Spearman $r_s = 0.86$). $\text{PM}_{2.5}$ of ambient origin contributed 62%, on average, to the total personal $\text{PM}_{2.5}$.

Calibration Coefficients

The results from the linear mixed effects model (Model 1) for both personal PM_{2.5} of ambient origin and total personal PM_{2.5} are presented in Table 3.

When the nearest ambient monitor was used as the surrogate exposure, the calibration coefficient for personal PM_{2.5} of ambient origin was estimated as 0.31 ((95%CI:0.14, 0.47), p-value₁<0.0001), when adjusted for seasonal effects. We found no significant seasonal effect modification (p-value=0.71). The season-adjusted calibration coefficient was higher for outdoor home model predictions, as compared to nearest monitor PM_{2.5}, equaling 0.54 (95%CI:0.42, 0.65, p-value₁<0.0001). We found significant effect modification by season for outdoor home model predictions (p-value=0.006), with season-stratified calibration coefficients higher during winter (0.60 (95%CI:0.36, 0.64)) than summer (0.50 (95%CI:0.42, 0.78)).

Total personal PM_{2.5} exposure calibration coefficients were higher than those for personal PM_{2.5} of ambient origin (Table 3). For total personal PM_{2.5} exposures, the season-adjusted calibration coefficient for the nearest ambient monitor was 0.56 (95%CI:0.24, 0.88, p-value₁=0.007). Effect modification by season was significant (p-value=0.041), with higher season-stratified calibration coefficients during summer (0.78 (95%CI:0.36, 1.19)) than winter (0.48 (95%CI:0.12, 0.83)). The corresponding calibration coefficient, using outdoor home model predicted PM_{2.5} as the surrogate exposure, was higher, 0.81 (0.49, 1.12, p-value₁=0.234). There was no significant seasonal effect modification.

Between-City Heterogeneity

For both personal PM_{2.5} of ambient origin and total personal PM_{2.5} calibration coefficients, we found no statistically significant evidence of heterogeneity across cities for outdoor home model predictions (p-values=0.11 and 0.17, respectively) and therefore results from Model 2, instead of Model 1, can be used. For personal PM_{2.5} of ambient origin and total personal PM_{2.5}, calibration coefficients equaled 0.56 (0.44, 0.68) and 0.79 (0.54, 1.04), respectively. Since no between-city heterogeneity was detected, no further adjustment to these calibration coefficients was done.

Significant between-city heterogeneity (p-value=0.003) was detected in the calibration coefficients for personal PM_{2.5} of ambient origin, when the nearest monitor was used as the surrogate exposure, with estimated city-specific calibration coefficients ranging between 0.0-0.71 (Figure 1(a)). The observed between-city heterogeneity was explained by two variables: the city's average number of residents in a

housing unit and the city's 30-year average of annual heating degree days, an indicator of the typical number of heating days in a year (p-value=0.50 for the test for residual heterogeneity). Cross-validation showed, however, that these variables were not robust predictors of the between-city variation in the calibration coefficient (Figure A2).

Significant between-city heterogeneity in the calibration coefficient was also detected for total personal PM_{2.5} when the nearest monitor was used as the surrogate measure (p-value=0.008). Using Model 4, estimated city-specific coefficients ranged between 0.0-1.78 (Figure 1(b)). Step-wise selection found that some of the observed between-city heterogeneity was explained by the average number of vehicles per housing unit in each city (p-value=0.221 for the test for residual heterogeneity). The effect of the city average vehicles per housing unit on the relationship between total personal PM_{2.5} and nearest ambient monitor PM_{2.5} concentrations was -2.53 (SE: 0.82), implying that as the average number of vehicles per housing unit increases, the calibration coefficient decreases for cities with larger numbers of vehicles per housing unit. For instance, if the average number of vehicles per housing unit in a city increased by 0.1, then the calibration coefficient for that city would decrease by 0.25. The selection of this variable was confirmed in the cross-validation, as it was consistently selected when cities were omitted one by one (Figure A2). The correlation between the predicted calibration coefficients from each city and the observed by-city coefficients was 0.62 (p-value=0.05), the mean percent relative bias was estimated -0.76% and the mean percent absolute bias 149%.

Sensitivity Analyses

Results from our sensitivity analyses are presented in the Appendix. Briefly, we observed no significant effect modification by subpopulation. We found significant effect modification by age, with subjects younger than 65 years of age having lower calibration coefficients than their older counterparts (Table A3).

Further, we found that estimated calibration coefficients were similar irrespective of the method used to calculate monthly ambient concentrations at the nearest monitor. When all days within the month were used in the calculation, the calibration coefficient for personal PM_{2.5} of ambient origin was 0.31 (95%CI:0.14, 0.47), vs. 0.35 (95%CI:0.26, 0.43) when monthly ambient concentrations were calculated using only those days with personal monitoring.

Discussion

We estimated calibration coefficients for studies of the association of long-term $\text{PM}_{2.5}$ health effects with ambient air pollution exposures, considering both estimated personal exposures to $\text{PM}_{2.5}$ of ambient origin as the exposure metric and personal exposures to total $\text{PM}_{2.5}$ as a second, albeit imperfect, exposure metric. Our goal was to assess and quantify error resulting from use of surrogate exposures and characterize the impact of different surrogate exposures on error. As discussed in the introduction, nevertheless, the estimated error could be from a variety of sources, and it has been argued that not all of these are properly characterized as measurement error [19].

Using estimated monthly personal $\text{PM}_{2.5}$ of ambient origin from five cities as the true exposure measure, we estimated a calibration coefficient of 0.54 (95%CI:0.42, 0.65) when outdoor home model predictions were used as the surrogate exposure, with no city-specific heterogeneity. This calibration coefficient suggests that when the parameter of interest is the health effect of ambient source pollution, the observed effect could be half the true estimate when outdoor home model predictions are used as the exposure metric in a linear health model, in absence of other potential bias sources. The lack of observed between-city variability likely reflects the use of the spatio-temporal model, which incorporates variables that may explain much of the between-city variability, such as population density, urban land use and distance to nearest road.

The estimated calibration coefficient for nearest ambient monitor concentrations as the exposure metric was lower (0.31 (95%CI:0.14, 0.47) compared to 0.54 (95%CI:0.42, 0.65) for outdoor model concentrations), reflecting the fact that nearest monitor concentrations do not account for as much spatial variability in ambient concentrations as the outdoor home model predictions. We also detected statistically significant between-city heterogeneity. Factors explaining between-city variability in the calibration coefficient, nevertheless, could not be reliably identified. This inability to explain the city-specific heterogeneity likely reflects the small number of cities included in our analysis.

When total $\text{PM}_{2.5}$ was used as the true exposure measure, calibration coefficients of 0.56 (95%CI:0.24, 0.88) and 0.81 (95%CI:0.49, 1.12) were found for nearest ambient monitor $\text{PM}_{2.5}$ and outdoor home model predictions, respectively. These results are consistent with those reported in Setton et al. (2011) [15], who reported an attenuation ranging between 0.70 to 0.84 for scenarios when mobility was not considered and only $\text{PM}_{2.5}$ predictions at the subjects' residences were included in the health model. As noted above, however, these calibration coefficients were calculated using total personal $\text{PM}_{2.5}$, an imperfect measure of

true exposure to ambient-generated pollutants.

As was the case with personal $\text{PM}_{2.5}$ of ambient origin, we detected significant between-city heterogeneity in total $\text{PM}_{2.5}$ calibration coefficients only when nearest monitor were used as the surrogate exposure. For nearest monitor $\text{PM}_{2.5}$, between-city heterogeneity was explained with the city average number of vehicles per housing unit. Results showed that error increases with vehicles per housing unit. A possible explanation for this association is provided by the strong negative correlation between the number of average vehicles per housing unit and population density ($r = -0.86$) and the strong positive correlation with the percentage of the detached homes in a study area ($r = 0.88$) as shown in Figure A3. These correlations suggest that in less dense cities, residents need to travel longer distances, possibly increasing the impact of pollutant spatial variability. These results are also in agreement with Setton et al. (2011) [15], who found increasing bias with increasing distance spent away from home. Selection of number of vehicles per housing unit to explain between-city heterogeneity could also reflect varying $\text{PM}_{2.5}$ composition, with local sources, such as traffic, likely comprising a larger portion of $\text{PM}_{2.5}$ mass in cities with more vehicles per housing unit, than regional sources. $\text{PM}_{2.5}$ of local sources is more spatially heterogeneous and more error is, therefore, expected when it comprises a large fraction of the total ambient $\text{PM}_{2.5}$. The fact, however, that our estimated city-specific calibration coefficients ranged between 0.0-1.9 complicates our interpretation of the overall estimate of 0.56 (95%CI:0.24, 0.88) and of the observed association with housing and transportation characteristics, suggesting that one average calibration coefficient may not adequately describe error from use of ambient monitor measurements across the United States.

Environmental tobacco smoke (ETS) may also contribute, at least partially, to the observed between-city heterogeneity. In all studies in our analyses, subjects were selected as non-smokers, living in non-smoking homes. Although this inclusion criterion would minimize potential exposure to ETS, it is possible that participants living in cities with more ETS would also have higher personal $\text{PM}_{2.5}$ exposures, thereby potentially contributing to between-city heterogeneity in the calibration coefficients. We, however, were not able to incorporate ETS exposures in our analysis, as some studies did not report ETS exposure information.

Our findings are consistent with two studies by Avery et al. (2010) [57], who found a median correlation coefficient of 0.54 between total personal $\text{PM}_{2.5}$ exposures and concentrations at a centrally located monitor, and strong between-city heterogeneity (p-value<0.0001). Although their reported median

correlation coefficient between total personal $\text{PM}_{2.5}$ exposures and outdoor home concentrations was similar, between-city heterogeneity in this association was lower (p-value=0.05). The weaker evidence of heterogeneity for outdoor home $\text{PM}_{2.5}$ concentrations is consistent with our suggestion that between-city heterogeneity in calibration coefficients is explained by variables included in outdoor home model predictions; this is one explanation for why we found heterogeneity only for nearest monitor but not outdoor home exposures.

Our study is limited by several factors. First, the data available to validate the exposure metrics of interest were limited to a small number of cities and participants, especially for personal $\text{PM}_{2.5}$ of ambient origin. Also, only a small number of days in a month were available in some cities to estimate monthly averages. These small numbers contributed to uncertainty in our data and estimates and potentially prohibited detection of any potential between-city heterogeneity for outdoor home predictions and the identification of factors explaining observed between-city heterogeneity in calibration coefficients when the nearest monitor $\text{PM}_{2.5}$ concentrations were the exposure surrogate. Further, the cities included in our analyses may not be representative of all US cities, and thus our estimated calibration coefficients might not be generalizable to other cities. Moreover, the association between personal exposures and ambient concentrations might vary over years. Since our studies were conducted over a one to two year time span (Table 1), we were not able to assess the contribution of longer term personal-ambient trends to total error.

Second, personal $\text{PM}_{2.5}$ of ambient origin was estimated rather than measured. As a result, estimated exposures did not take into account the uncertainty related to their prediction when estimating the calibration coefficients. Moreover, given data availability, we were not able to estimate the contribution of instrumental to total error. Both personal and ambient measurements are prone to instrumental error, presence of which is likely to introduce classical error [5]. In our setting, however, personal exposures are the outcome variable in the regression and therefore random error in these exposures is not expected to introduce error in the estimated calibration coefficients. Furthermore, personal exposures are on average measured with high precision and accuracy [29,30].

To estimate personal $\text{PM}_{2.5}$ of ambient origin we used the SO_4^{2-} tracer method. In cities where SO_4^{2-} comprises a large fraction of the total ambient $\text{PM}_{2.5}$ mass, as in the northeastern US [58], the SO_4^{2-} tracer method has been shown to perform well [49]. In places, however, where ambient $\text{PM}_{2.5}$ mass is strongly influenced by local sources, such as traffic, ambient SO_4^{2-} would not act as good tracer, given that the

spatial and size distributions of SO_4^{2-} may differ from those of $\text{PM}_{2.5}$. Since $\text{PM}_{2.5}$ from local sources is more spatially heterogeneous, larger spatial misalignment would be expected in these cities and, hence, more measurement error. For these cities, we would expect the calibration coefficients for personal $\text{PM}_{2.5}$ of ambient origin, which was estimated using the SO_4^{2-} ratio, to be overestimated and the error to be underestimated, a factor likely contributing to the observed between-city heterogeneity. In our study, we only had SO_4^{2-} data in four cities, three of which are in the northeastern US (Baltimore, Boston and Steubenville). The fourth city was Atlanta, which has been shown, on average, to have lower SO_4^{2-} concentrations [58]. Even there, however, secondary sulfate was found to comprise 38% of the total $\text{PM}_{2.5}$ mass [59] and in our data, the ratio of ambient SO_4^{2-} over $\text{PM}_{2.5}$ in Atlanta was, on average, similar to the ratios in the three northeastern cities (Table A1).

In addition, we estimated the outdoor home predictions using a specific spatio-temporal model. This model has been validated and shown to perform very well [14, 43]. We would therefore expect that our findings for outdoor home predictions could be extended to similarly performing spatio-temporal models and could be qualitatively used for predicted concentrations obtained from other spatio-temporal models.

Moreover, we were not able to disentangle how specific error types would impact the health effect estimates obtained using either of the surrogate exposures. We did not assume models addressing specific error structures and our approach assesses overall error from use of surrogate exposures, combining the multiple error types that are likely present [5, 18].

Furthermore, our study is not able to determine how much of the estimated calibration coefficient reflects infiltration of particles from outdoor to indoor environments, as compared to other sources of the difference between personal exposure and outdoor concentration metrics [5, 19]. Infiltration, however, does not appear to explain all of the observed error found in our analysis, since the average estimated calibration coefficients for personal $\text{PM}_{2.5}$ of ambient origin were <0.64 (the approximated penetration efficiency using the SO_4^{2-} ratio), consistent with additional contributing error sources.

Additionally, personal exposures were measured for each participant for periods less than one month. We would expect this temporal mismatch to introduce both Berkson, through the errors in the true exposures that were randomly selected within a month, and classical, through the errors in the temporal misalignment of the surrogate exposures, error components. Through sensitivity analyses, comparing

PM_{2.5} concentrations measured at the nearest monitor using all data within a month with that measured on days when personal data were also available, we showed the point estimates to be very similar, but the confidence intervals for the calibration coefficients estimated using the temporally mismatched data were wider. Since outdoor home model predictions were only available at the monthly level, we were unable to quantitatively assess the effect of this temporal mismatch on the estimation of the calibration coefficients. Monthly concentrations at the nearest ambient monitor, however, were very strongly correlated with outdoor home model predictions ($r = 0.86$). In any event, randomly temporally mismatched data relating personal exposures to outdoor home predictions may also lead to increased uncertainty, but likely no bias, in the calibration coefficients.

Finally, our goal was to assess exposure measurement error in long-term PM_{2.5} exposures. As described earlier, personal exposure studies are infeasible for long periods and, given current data availability, we were only able to conduct our analyses using monthly averages. Many long-term PM_{2.5} studies use exposure metrics based on functions of monthly averages (e.g. 12-month moving average [11] or cumulatively-updated monthly average [60]), and we therefore believe that our findings provide useful information in the interpretation of chronic health effect estimates.

We compiled data from 9 cities across the United States for our analyses and calculated calibration coefficients that may be informative for interpreting risk estimates in nationwide studies of long-term PM_{2.5} health effects. For instance, differential measurement error could be partially responsible for the higher effects reported by Puett et al. (2009) [11], who used PM_{2.5} predictions outside the participant’s homes, as compared to the effects found by Krewski et al. (2005) [61], who used metropolitan area means of PM_{2.5} concentrations at ambient monitors.

To our knowledge, this is the first study to assess error due to two different, widely used, surrogate exposures, using personal exposure data from multiple US cities. Further, we identified variables explaining the heterogeneity in the calibration coefficients across cities, with the variances of the reported calibration coefficients potentially reflecting this heterogeneity.

At this time, we do not recommend using the calibration coefficients reported here to directly adjust health effect estimates in epidemiology studies. Given the observed between-city heterogeneity the complex, time-varying nature of the exposures and the lack of information on individual characteristics, which would

be included as confounders in health models, standard error correction methods such as ordinary regression calibration could still yield biased estimates [62,63]. Our group is currently developing methods to account for the above limitations in order to correctly adjust health effect estimates obtained using surrogate exposures. Furthermore, future research on PM_{2.5}-related measurement error should characterize measurement error for regional and local PM_{2.5} by focusing on PM_{2.5} composition, which changes both over space and time, suggesting that calibration coefficients will also change over space and time [6,8,48].

Conclusions

With our study we were able to assess the ability of two widely used surrogate exposures to reflect personal exposures: ambient concentrations measured at centrally located monitors, as well as outdoor home predictions. Our estimated calibration coefficients are consistent with previously reported chronic health risks using nearest monitor exposures being under-estimated when ambient concentrations were the exposure of interest. For outdoor home predictions, our results suggest less error.

Abbreviations

CHD: Coronary Heart Disease; COPD: Chronic Obstructive Pulmonary Disease; EPA: Environmental Protection Agency; MI: Myocardial Infarction; PM_{2.5}: Fine particulate matter; SO₄²⁻: Sulfate

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MAK was responsible for design, conduct, analysis, interpretation of findings and writing the manuscript; DS for conception and design; AAS for interpretation of findings and manuscript review; LS and JDK for compilation of data and interpretation; JDY for providing the spatio-temporal model predictions and manuscript review; RW for compilation of data and manuscript review; FL for interpretation; BH for analysis; and HHS for conception, design, compilation of data and drafting the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This article was developed under STAR Fellowship Assistance Agreement (FP-9172890-01) from the US EPA. The views expressed here are those of the authors and do not necessarily reflect the views and policies of the US EPA. This manuscript has been subjected to U.S. EPA review and approved for publication. Additionally, this work was funded by NIH (ES 09411) and NIEHS (T32 ES 007069 and R01 ES017017).

References

1. Armstrong BA: **THE EFFECTS OF MEASUREMENT ERRORS ON RELATIVE RISK REGRESSIONS**. *American Journal of Epidemiology* 1990, **132**(6).
2. Armstrong BA: **Exposure measurement error: consequences and design issues**. In *Exposure assessment in occupational and environmental epidemiology*, Oxford University Press 2004.
3. Bateson TF, Coull BA, Hubbell B, Ito K, Jerrett M, Lumley T, et al: **Panel discussion review: session three – issues involved in interpretation of epidemiologic analyses – statistical modeling**. *Journal of Exposure Science and Environmental Epidemiology* 2007, **17**:S90–S96.
4. Thomas D, Stram D, Dwyer J: **Exposure measurement error: Influence on exposure-disease relationships and methods of correction**. *Annu. Rev. Publ. Health* 1993, **14**:69–93.
5. Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, et al: **Exposure Measurement Error in Time-Series Studies of Air Pollution: Concepts and Consequences**. *Environmental Health Perspectives* 2000, **108**(5).
6. Goldman G, Mulholland J, Russell A, Gass K, Strickland M, Tolbert P: **Characterization of Ambient Air Pollution Measurement Error in a Time-Series Health Study using a Geostatistical Simulation Approach**. *Atmos Environ* 2012, **57**:101–108.
7. Lee MS, Magari S, Christiani DC: **Cardiac autonomic dysfunction from occupational exposure to polycyclic aromatic hydrocarbons**. *Occup. Environ. Med.* 2011, **68**(7):474–478.
8. Sarnat S, Klein M, Sarnat J, Flanders W, Waller L, Mulholland J, Russell A, Tolbert P: **An examination of exposure measurement error from air pollutant spatial variability in time-series studies**. *J Expo Sci Environ Epidemiol* 2010, **20**(2):135–146.
9. Hoek G, Brunekreef B, Goldbohm S, Fischer P, van den Brandt PA: **Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study**. *The Lancet* 2002, **360**(9341):1203–1209.
10. Jerrett M, Burnett RT, Ma R, Pope C, Krewski D, Newbold KB, Thurston G, Shi Y, Finkelstein N, Calle EE, Thun MJ: **Spatial analysis of air pollution and mortality in Los Angeles**. *Epidemiology* 2005, **16**(6):727–36.
11. Puett RC, Hart JE, Yanosky JD, Paciorek CJ, Schwartz J, Suh H, Speizer FE, Laden F: **Chronic fine and coarse particulate exposure, mortality and coronary heart disease in the Nurses' Health Study**. *Environmental Health Perspectives* 2009, **117**(11):1697–701.
12. Szpiro A, Sampson P, Sheppard L, Lumley T, Adar S, Kaufman J: **Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies**. *Environmetrics* 2010, **21**:606–631.
13. Sampson P, Szpiro A, Sheppard L, Lindstrom J, Kaufman J: **Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data**. *Atmospheric Environment* 2011, **45**(36):6593–6606.
14. Yanosky JD, Paciorek CJ, Suh HH: **Predicting Chronic Fine and Coarse Particulate Exposures Using Spatiotemporal Models for the Northeastern and Midwestern United States**. *Environ Health Perspect* 2009, **117**(4):522–529.
15. Setton E, Marshall JD, Brauer M, Lundquist KR, Hystad P, Keller P, Cloutier-Fisher D: **The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates**. *Journal of Exposure Science and Environmental Epidemiology* 2011, **21**:42–48.

16. Sheppard L: **Acute air pollution effects: consequences of exposure distribution and measurements.** *Journal of Toxicology and Environmental Health, Part A* 2005, **68**:1127–1135.
17. Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull BA: **Measurement error caused by spatial misalignment in environmental epidemiology.** *Biostatistics* 2009, **10**:258–74.
18. Szpiro AA, Sheppard L, Lumley T: **Efficient measurement error correction with spatially misaligned data.** *Biostatistics* 2011, **12**(4):610–23.
19. Sheppard L, Burnett RT, Szpiro A, Kim S, Jerrett M, Pope CA, Brunekreef B: **Confounding and exposure measurement error in air pollution epidemiology.** *Air Qual Atmos Health* 2012, **5**(2):203–216.
20. Kim SY, Sheppard L, Kim H: **Health Effects of Long-term Air Pollution: Influence of Exposure Prediction Methods.** *Epidemiology* 2009, **20**(3):442–450.
21. Szpiro AA, Paciorek CJ, Sheppard L: **Does More Accurate Exposure Prediction Necessarily Improve Health Effect Estimates?** *Epidemiology* 2011, **22**(5):680–685.
22. Dominici F, Zeger SL, Samet JM: **A measurement error model for time-series studies of air pollution and mortality.** *Biostatistics* 2000, **1**(2):157–175.
23. Strand M, Vedal S, Rhodes C, Dutton SJ, Gelfand EW, Rabinovitch N: **Estimating effects of ambient PM_{2.5} exposure on health using PM_{2.5} component measurements and regression calibration.** *Journal of exposure science and environmental epidemiology* 2006, **16**:30–38.
24. Van Roosbroeck S, Li R, Hoek G, Lebret E, Brunekreef B, Spiegelman D: **Traffic-Related Outdoor Air Pollution and Respiratory Symptoms in Children; The Impact of Adjustment for Exposure Measurement Error.** *Epidemiology* 2008, **19**(3):409–416.
25. Eftim S, Dominici F: **Multisite time-series studies versus cohort studies: methods, findings and policy implications.** *Journal of Toxicology and Environmental Health, Part A* 2005, **68**:1191–1205.
26. Sheppard L, Slaughter JC, Schildcrout J, Liu LJ, Lumley T: **Exposure and Measurement Contributions to Estimates of Acute Air Pollution Effects.** *Journal of Exposure Analysis and Environmental Epidemiology* 2005, **15**:366–376.
27. Thomas D: **Why do estimates of the acute and chronic effects of air pollution on mortality differ?** *Journal of Toxicology and Environmental Health, Part A* 2005, **68**:1167–1174.
28. Haneuse S, Wakefield J, Sheppard L: **The interpretation of exposure effect estimates in chronic air pollution studies.** *Stat Med.* 2007, **26**(16):3172–87.
29. Sarnat SE, Coull BA, Schwartz J, Gold DR, Suh HH: **Factors Affecting the Association between Ambient Concentrations and Personal Exposures to Particles and Gases.** *Environmental Health Perspectives* 2006, **114**(5):649–654.
30. Koutrakis P, Suh HH, Sarnat JA, Brown KW, Coull BA, Schwartz J: *Characterization of Particulate and Gas Exposures of Sensitive Subpopulations Living in Baltimore and Boston.* Boston, MA: Health Effects Institute 2005. [Research Report 131].
31. Liu LJ, Box M, Kalman D, Kaufman J, Koenig J, Larson T, et al: **Exposure assessment of particulate matter for susceptible populations in Seattle.** *Environ Health Perspect.* 2003, **111**(7):909–18.
32. Meng QY, Turpin BJ, Korn L, Weisel CP, Morandi M, Colome S, et al: **Influence of ambient (outdoor) sources on residential indoor and personal PM_{2.5} concentrations: analyses of RIOPA data.** *J Expo Anal Environ Epidemiol* 2005, **15**:17–28.
33. Sarnat JA, Koutrakis P, Suh HH: **Assessing the Relationship between Personal Particulate and Gaseous Exposures of Senior Citizens Living in Baltimore, MD.** *J. Air & Waste Manage. Assoc.* 2000, **50**:1184 – 1198.
34. Suh HH, Koutrakis P, Chang LT: *Characterization of the composition of personal, indoor, and outdoor particulate exposures.* Sacramento, CA: California Air Resource Board 2003. [Final Report].
35. Suh HH, Koutrakis P, Ebelt SE: *Detailed Characterization of Indoor and personal Particulate matter concentrations.* Sacramento, CA: California Air Resource Board 2004. [Final Report].
36. Suh HH, Zanobetti A: **Exposure error masks the relationship between traffic-related air pollution and heart rate variability.** *JOEM* 2010, **52**(7).

37. Ward-Brown K, Sarnat JA, Suh HH, Coull BA, Spengler JD, Koutrakis P: **Ambient site, home outdoor and home indoor particulate concentrations as proxies of personal exposures.** *Journal of Environmental Monitoring* 2008, **10**:1041–51.
38. Ward-Brown K, Sarnat JA, Suh HH, Coull BA, Koutrakis P: **Factors influencing relationships between personal and ambient concentrations of gaseous and particulate pollutants.** *Science of the Total Environment* 2009, **407**:3754–65.
39. Weisel CP, Zhang J, Turpin BJ, Morandi MT, Colome S, Stock TH, et al: *Relationships of Indoor, Outdoor, and Personal Air (RIOPA): Part I. Collection Methods and Descriptive Analyses.* Boston, MA; Houston, TX: Health Effects Institute; Mickey Leland National Urban Air Toxics Research Center 2005. [Research Report 130].
40. Williams R, Suggs J, Rea A, Leovic K, Vette A, Croghan C, et al: **The Research Triangle Park particulate matter panel study: PM mass concentration relationships.** *Atmospheric Environment* 2003, **37**:5349–5363.
41. Williams R, Suggs J, Rea A, Sheldon L, Rhodes C, Thornburg J: **The Research Triangle Park particulate matter panel study: modeling ambient source contribution to personal and residential PM mass concentrations.** *Atmospheric Environment* 2003, **37**:5365–5378.
42. Miller KA, Siscovick DS, Sheppard L, Shepherd K, Sullivan JH, Anderson G, Kaufman JD: **Long-term exposure to air pollution and incidence of cardiovascular events in women.** *N Engl J Med.* 2007, **365**(5):447–58.
43. Paciorek C, Yanosky J, Puett R, Laden F, Suh H: **Practical large-scale spatio-temporal modeling of particulate matter concentrations.** *Ann Appl Stat* 2009, **3**:370–397.
44. Sarnat JA, Wilson WE, Strand M, Brook J, Wyzga R, Lumley T: **Panel discussion review: session one – exposure assessment and related errors in air pollution epidemiologic studies.** *Journal of Exposure Science and Environmental Epidemiology* 2007, **17**:S75–S82.
45. Wilson W, Suh H: **Fine particles and coarse particles: Concentration relationships relevant to epidemiologic studies.** *JAWMA* 1997, **47**:1238–1249.
46. Hazi Y, Heikkinen M, Cohen B: **Size distribution of acidic sulfate ions in fine ambient particulate matter and assessment of source region effect.** *Atmospheric Environment* 2003, **37**:5403–5413.
47. Hopke PK, Ito K, Mar T, Christensen WF, Eatough DJ, Henry RC, Kim E, Laden F, Lall R, Larson TV, Liu H, Neas L, Pinto J, Stölzel M, Suh H, Paatero P, Thurston GD: **PM Source Apportionment and Health Effects: 1. Intercomparison of Source Apportionment Results.** *Journal of Exposure Science and Environmental Epidemiology* 2006, **16**:275–286.
48. Hsu SI, Ito K, Kendall M, Lippmann M: **Factors affecting personal exposure to thoracic and fine particles and their components.** *J Expos Sci Environ Epidemiol* 2012, **22**:439–447.
49. Sarnat JA, Long CM, Koutrakis P, Coull BA, Schwartz J, Suh HH: **Using Sulfur as a Tracer of Outdoor Fine Particle.** *Environmental Science & Technology* 2002, **36**(24):5305–5314.
50. Wilson W, Mage D, Grant L: **Estimating separately personal exposure to ambient and nonambient particulate matter for epidemiology and risk assessment; why and how.** *J. Air & Waste Manage. Assoc.* 2000, **50**:1167–1183.
51. Koenig JQ, Mar TF, Allen RW, Jansen K, Lumley T, Sullivan JH, Trenga CA, Larson TV, Liu LJ: **Pulmonary effects of indoor- and outdoor generated particles in children with asthma.** *Environ Health Perspect* 2005, **113**:499–503.
52. Self SG, Liang KY: **Asymptotic properties of maximum-likelihood estimators and likelihood ratio tests under nonstandard conditions.** *Journal of the American Statistical Association* 1987, **82**(398):605–610.
53. Bell ML, Dominici F: **Effect modification by community characteristics on the short-term effects of ozone exposure and mortality in 98 US communities.** *Am J Epidemiol.* 2008, **167**(8):986–997.
54. Janssen NA, Schwartz J, Zanobetti A, Suh HH: **Air conditioning and source-specific particles as modifiers of the effect of PM₁₀ on hospital admissions for heart and lung disease.** *Environ Health Perspect.* 2002, **110**:43–9.

55. Smith RL, Xu B, Switzer P: **Reassessing the relationship between ozone and short-term mortality in U.S. urban communities.** *Inhalation Toxicology* 2009, **21**(S2):37–61.
56. Hastie T, Tibshirani R, Friedman J: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer; 2nd ed. 2009.
57. Avery CL, Mills KT, Williams R, McGraw KA, Poole C, Smith RL, Whitsel EA: **Estimating error in using ambient PM_{2.5} concentrations as proxies for personal exposures: a review.** *Epidemiology* 2010, **21**(2):215–223.
58. Bell ML, Dominici F, Ebisu K, Zeger SL, Samet JM: **Spatial and temporal variation in PM_{2.5} chemical composition in the United States for health effects studies.** *Environ Health Perspect* 2007, **115**(7):989–995.
59. Sarnat JA, Marmur A, Klein M, Kim E, Russell AG, Sarnat SE, Mulholland JA, Hopke PK, Tolbert PE: **Fine Particle Sources and Cardiorespiratory Morbidity: An Application of Chemical Mass Balance and Factor Analytical Source-Appportionment Methods.** *Environ Health Perspect* 2008, **116**(4):459–466.
60. Lipsett MJ, Ostro BD, Reynolds P, Goldberg D, Hertz A, Jerrett M, Smith D, Garcia C, Chang ET, Bernstein L: **Long-term exposure to air pollution and cardiorespiratory disease in the California teachers study cohort.** *Am J Respir Crit Care Med* 2011, **184**(7):828–35.
61. Krewski D, Burnett R, Jerrett M, Pope CA, Rainham D, Calle E, Thurston G, Thun M: **Mortality and Long-Term Exposure to Ambient Air Pollution: Ongoing Analyses Based on the American Cancer Society Cohort.** *Journal of Toxicology and Environmental Health, Part A: Current Issues* 2005, **68**(13–14):1093–1109.
62. Guo Y, Little RJ, McConnel DS: **On using summary statistics from an external calibration sample to correct for covariate measurement error.** *Epidemiology* 2012, **23**:165–174.
63. Rosner B, Spiegelman D, Willett WC: **Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error.** *Am J Epidemiol.* 1990, **132**(4):734–45.

Tables

Table 1: Validation studies used in our analyses

Cities	Years	# Subjects	Sample Session Duration	Age Mean (S.D.)	References
Atlanta, GA	1999–2000	31	7 days	65.0 (13.5)	[36]
Baltimore, MD	1998–1999	35	12 days	70.8 (7.7)	[30, 33, 37, 38]
Boston, MA	1999–2000	56	7 or 12 days	62.3 (14.1)	
Los Angeles, CA	2000–2002	37	7 days	56.3 (13.9)	[34, 35]
RTP, NC	2000–2001	37	7 days	64.5 (7.8)	[40, 41]
RIOPA	1999–2001		48 hr		[32, 39]
Los Angeles, CA		73		46.1 (18.6)	
Elizabeth, NJ		57		48.2 (17.8)	
Houston, TX		62		48.5 (16.6)	
Seattle, WA	2000–2001	89	10 days	76.7 (6.5)	[31]
Steubenville, OH	2000	28	48 hr/we for 12 we	71.0 (10.0)	[29]

Table 2: Basic characteristics of exposure variables

	# person-months (# subjects)	Mean \pm SD ($\mu\text{g}/\text{m}^3$)
<u>All Year</u>		
Total Personal PM _{2.5}	919 (490)	24.54 \pm 18.97
Personal PM _{2.5} of ambient origin	261 (141)	9.71 \pm 4.32
Personal/Ambient SO ₄ ²⁻ ratio ^a	241 (131)	0.64 \pm 0.25
Model Predicted PM _{2.5}	1029 (502)	15.47 \pm 4.77
Monitor PM _{2.5}	1029 (502)	15.86 \pm 5.58
<u>Summer</u>		
Total Personal PM _{2.5}	429 (312)	23.92 \pm 16.92
Personal PM _{2.5} of ambient origin	130 (90)	10.95 \pm 4.12
Personal/Ambient SO ₄ ²⁻ ratio ^a	125 (86)	0.70 \pm 0.23
Model Predicted PM _{2.5}	493 (327)	15.67 \pm 4.84
Monitor PM _{2.5}	493 (327)	15.69 \pm 4.79
<u>Winter</u>		
Total Personal PM _{2.5}	490 (353)	23.94 \pm 19.67
Personal PM _{2.5} of ambient origin	131 (97)	8.73 \pm 4.27
Personal/Ambient SO ₄ ²⁻ ratio ^a	116 (87)	0.59 \pm 0.24
Model Predicted PM _{2.5}	536 (367)	15.36 \pm 4.71
Monitor PM _{2.5}	536 (367)	16.10 \pm 6.21

^a Not estimated for Seattle, WA

Table 3: Season-adjusted calibration factors for personal $PM_{2.5}$ of ambient origin and total personal $PM_{2.5}$.

	Monitor $PM_{2.5}$	Model Predicted $PM_{2.5}$
<i>Personal $PM_{2.5}$ of ambient origin</i>	5 cities (141 subjects)	
Estimate (95% CI) ^a	0.31 (0.14, 0.47)**	0.54 (0.42, 0.65)**
p-value for between-city heterogeneity	0.0034	0.1114
R_I^b	10.75%	0.96%
Estimate (95% CI) ^c	N/A	0.56 (0.44, 0.68)**
<i>Total Personal $PM_{2.5}$</i>	9 cities (490 subjects)	
Estimate (95% CI) ^a	0.56 (0.24, 0.88)**	0.81 (0.49, 1.12)
p-value for between-city heterogeneity	0.0084	0.1712
R_I^b	2.54%	1.28%
Estimate (95% CI) ^c	N/A	0.79 (0.54, 1.04)**

* p-value₁ < 0.05, ** p-value₁ < 0.01 for significant difference from 1

^a Results from Model 1 (including random slopes for cities, g_{3i})

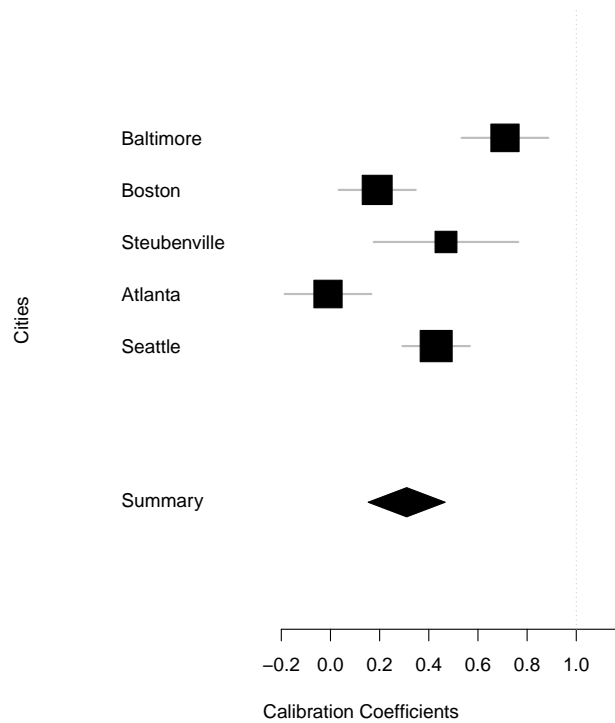
^b R_I : the proportion of variance explained by the between-cities heterogeneity

^c Results from Model 2, when no significant between-city heterogeneity was detected (without the random slopes for cities, g_{3i})

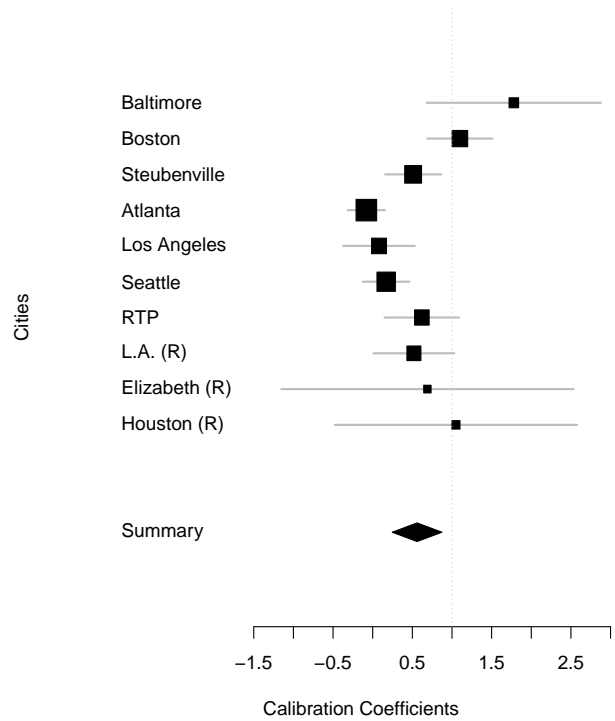
Figures

Figure 1

Forest plots of the by-city calibration coefficients for (a) personal $PM_{2.5}$ of ambient origin and (b) total personal $PM_{2.5}$ and nearest monitor concentrations. The size of the point used for the effect estimate is proportional to the precision of that calibration coefficient.



(a)



(b)

Figure 1: Forest plots of the by-city calibration factors for (a) personal $PM_{2.5}$ of ambient origin and (b) total personal $PM_{2.5}$ and nearest monitor concentrations. The size of the point used for the effect estimate is proportional to the precision of that calibration factor.