Author's response to reviews

Title: Exposure Measurement Error in PM2.5 Health Effects Studies: A Pooled Analysis of Eight Personal Exposure Validation Studies

Authors:

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 H. Suh (h.suh@neu.edu)

Version: 2 Date: 6 December 2013

Author's response to reviews: see over
Response to Reviewers

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

General Response:

We would like to thank our reviewers, Drs. Ito and Brunekeef, for their thoughtful comments and suggestions. We agree with the reviewers and have revised our manuscript accordingly.

Specific responses to the reviewers’ comments are presented below.

Responses to Dr. Ito:

“(1) Acknowledge the implications of focusing on sulfate as the PM2.5 of ambient origin

The authors use sulfate as the marker of PM2.5 of outdoor origin. This is, in one aspect, a reasonable choice in that there is no major source of sulfate indoors. However, the peculiar nature of sulfate must be acknowledged upfront and in interpreting the results, especially since the motivation of this study is the investigation of the impact of exposure error on the long-term health effects of PM2.5, which, in cohort studies in the US, may be influenced by relative presence of sulfate (secondary formation from sulfur dioxide emitted from coal-fired power plants) vs. other local combustion sources such as traffic. Sulfate is a regional pollutant and highly spatially uniform within a city. Because its size is in the sub-micron range, its penetration efficiency indoors is also high. Compared to primary PM2.5 constituents from local combustion sources (e.g., elemental carbon), within a city, sulfate likely exhibits the least amount of spatial misalignment for exposure estimation. Other primary PM2.5 constituents would have larger error between personal exposures and the measurements at the community monitor because the primary PM2.5 constituents have larger within-city variation, which also means that the distance to nearest monitor would matter. Thus, the calibration coefficients with the focus on sulfate would yield higher values compared to those from the PM2.5 from local sources. This needs to be acknowledged upfront in “Estimation of Personal Exposure of Ambient Origin” and also in the limitation discussion (there are many limitations that the authors listed, but this one is not specifically acknowledged).”

We have revised our manuscript to describe limitations of the use of the sulfate tracer method, both in the Methods (p. 5, 4$^{th}$ paragraph) and the Discussion (end of p. 13 and beginning of p. 14) sections.

In the Methods section, we added a description of the sources of ambient sulfate. We write: “The majority of SO$_4$ 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
Because of negligible indoor sources and its similar spatial homogeneity as PM$_{2.5}$, SO$_4$ can serve as a tracer for PM$_{2.5}$ of ambient origin in locations where SO$_4$ comprises a large part of the PM$_{2.5}$ mass [49,50], …

In addition, we have revised our manuscript to describe its appropriateness as a tracer for ambient PM$_{2.5}$. In the Discussion, for example, we explain that the sulfate tracer method performs best when regional sources are the predominant contributors to the PM mass, with less expected error. We write: “To estimate personal PM$_{2.5}$ of ambient origin we used the SO$_4$ tracer method. In cities where SO$_4$ comprises a large fraction of the total ambient PM$_{2.5}$ mass, as in the northeastern US [58], the SO$_4$ tracer method has been shown to perform well [49]. In places, however, where ambient PM$_{2.5}$ mass is strongly influenced by local sources, such as traffic, ambient SO$_4$ would not act as good tracer, given that the spatial and size distributions of SO$_4$ may differ from those of PM$_{2.5}$. Since 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 PM$_{2.5}$ of ambient origin, which was estimated using the SO$_4$ 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$_4$ 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$_4$ concentrations [58]. Even there, however, secondary sulfate was found to comprise 38% of the total PM$_{2.5}$ mass [59] and in our data, the ratio of ambient SO$_4$ over PM$_{2.5}$ in Atlanta was, on average, similar to the ratios in the three northeastern cities (Table A1).”

“(2) Discuss the results in the context of within- vs. between-city variation of PM$_{2.5}$, their source types, and epidemiological studies of the long-term effects:

In Discussion section, the authors warn that we cannot use these calibration coefficients to incorporate in the health effects analyses and that they are currently working to investigate the issues. I understand that the authors would not want to outline exactly what they are working on next, but leaving the discussion like this can also give the impression that the value of the results presented in this analysis is not certain. I think the results are valuable, but the authors need to discuss the issues involving the use of the calibration coefficients in the context of the long-term epidemiological studies. For example, in the study of the cardiovascular effects of PM$_{2.5}$ in 36 US metropolitan areas (reference 42; not discussed in this context), they reported that the between-city effect appeared to be smaller than the within-city effect, which implies that the local combustion sources (which would have lower calibration coefficients) may be important. In a study of indoor vs. central monitor concentrations of PM$_{2.5}$ and its chemical constituents in New York City (Hsu et al., 2012; JESEE, doi:10.1038/jes.2012.23), they found that high correlations between indoor vs. central monitor measurements for sulfur but poor correlations for nickel (from residual oil burning for space heating in large buildings). Thus, the use of
calibration coefficients for epidemiological studies need to address the influence of local vs. regional pollutants which may have different toxicity. This is one of the major challenges. As the contributions from different source types to PM2.5 change in the future (e.g., for traffic, non-exhaust emissions such as brake wear may become important), the calibration coefficients may also change. A calibration coefficient of 0.5 would be an optimistic estimate when the contributions from regional pollution decline.”

We revised the discussion to mention that future research on exposure measurement error should focus on PM composition, as error associated with local PM is likely different from regional PM (top of p. 16): “Furthermore, future research on PM2.5-related measurement error should characterize measurement error for regional and local PM2.5 by focusing on PM2.5 composition, which changes both over space and time, suggesting that calibration coefficients will also change over space and time [6,8,48].”

In addition, we revised the discussion section to include local and regional sources, and to a lesser extent PM composition, (p. 12, 1st paragraph), as a contributor to between-city heterogeneity: “Selection of number of vehicles per housing unit to explain between-city heterogeneity could also reflect varying PM2.5 composition, with local sources, such as traffic, likely comprising a larger portion of PM2.5 mass in cities with more vehicles per housing unit, than regional sources. PM2.5 of local sources is more spatially heterogeneous and more error is, therefore, expected when it comprises a large fraction of the total ambient PM2.5.”

“Minor essential revisions:

(3) Mention sources of sulfate

While it may be obvious to the authors, it should be clarified where sulfate is coming from. While coal-fired power plants are not the only source of sulfate, I imagine that they are the major source of this constituent. I searched for “coal” in this manuscript, and it is not even mentioned.”

We have added the sources of ambient sulfate in the Methods section, as described in comment 1. Specifically (p. 5, 4th paragraph): “The majority of SO4 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 PM2.5, SO4 can serve as a tracer for PM2.5 of ambient origin in locations where SO4 comprises a large part of the PM2.5 mass [49,50], …”

“(4) Highlight and discuss the reason for between-city heterogeneity in calibration coefficients

Looking at Figure 1, I imagine that at least some of the observed heterogeneities
in calibration coefficients across cities must be due to the relative contributions from regional pollutants (sulfate and secondary organics; which are spatially uniform within city) vs. local combustions (spatially heterogeneous within city). For example, Atlanta, while its sulfate levels are not low (4.8 \text{ug/m3} in Table A1), there must be a major contribution of traffic pollution to PM2.5. The PM2.5 levels in Los Angeles also must be influenced by traffic air pollution (e.g., nitrate, primary and secondary organics). In the text (Result section), the authors describe that the heterogeneity of between-city calibration coefficient for PM2.5 was in part explained by the average vehicle per housing unit, indicating that the local combustion sources (as opposed to regionally uniform secondary aerosols) reduces the calibration coefficient. This point is not coming out clearly because, while a list of all the variables considered to explain between-city heterogeneity of calibration coefficients is provided in Table A4, the result of this analysis is only mentioned as text. This result can also be further expanded in Discussion to explain the importance of the impact of (spatially non-uniform) local sources on the calibration coefficient."

We extended our discussion of factors affecting between-city heterogeneity to include particulate composition, which may vary across cities. We also clarify that PM sources, through their impact on the relationship of local (specifically traffic) and regional sources with the city-average number of vehicles per housing unit. Please also see our response to comment #2. Specifically (p. 12): "Selection of number of vehicles per housing unit to explain between-city heterogeneity could also reflect varying PM$_{2.5}$ composition, with local sources, such as traffic, likely comprising a larger portion of PM$_{2.5}$ mass in cities with more vehicles per housing unit, than regional sources. 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 PM$_{2.5}$."

“(5) Additional information regarding the spatio-temporal model is needed

A spatio-temporal model was used to estimate monthly outdoor PM2.5 concentrations outside each subject’s residence. The model involves prediction based on regulatory and research monitors’ data as well as GIS-based covariates including population density, distance to roads, land-use, point source emissions, etc. Although the citations are provided, very little information is provided in terms of what variables were important in explaining the variation of PM2.5 in these cities. In particular, it would be good to describe how much variation the indicators of local source combustions, especially traffic (i.e., the PM2.5 constituents that would bring the calibration coefficients down), contributes to PM2.5 in these cities. If such quantitative information can be provided, then it may be connected to the result that the heterogeneity in calibration coefficients were in part explained by the average vehicle per housing unit.”

We agree with Dr. Ito that it would be of interest to describe how much the variation of the indicators of local sources contributes to the predicted PM$_{2.5}$ in
the cities in our study. We are unable to quantify any such contribution at this point, however, given partial R² values, which are used to estimate the contribution of model predictors to PM$_{2.5}$ predictions, were not available. Further complicating our ability to estimate the explanatory contribution of model predictors was the fact that our GIS-based spatio-temporal model was a GAMM model, which allowed us to model predictors using non-linear relationships with PM$_{2.5}$, and the models estimated PM$_{2.5}$ on the log scale. Finally, the model was fit in the entire US, so extracting quantitative information for just the cities in our study is not straightforward.

Previous findings from Yanosky et al. (2009), nonetheless, showed percent changes in PM$_{2.5}$ concentrations across the range of several model covariates, finding for example that with increasing proximity to A1 roads, the percent change in PM$_{2.5}$ concentrations also increases. While the reported effects of the model covariates may vary by location, they provide some evidence of their impact on PM$_{2.5}$ concentrations.

“Discretionary revisions:

- Figure A1 is potentially interesting, but it is hard to distinguish between the cities with just these symbols. Perhaps enlarge the figures (2 rows x 1 column in a full page rather than 1 row x 2 columns taking up only half of the page) and/or may use colors. Also, add a few sentences in the text describing notable features, if any.”

We have changed the size of the graphs, so now they are 2 rows x 1 column in a full page.
Response to Dr. Brunekreef:

“Spatial variation in (long term, let's say annual) air pollution is usually rather smaller than temporal variation in the short run, typically daily averages used in time series studies. Validation studies of the spatial, long term exposures used in many epidemiology studies is therefore both needed and challenging. Quite a few studies of the temporal correlations have been published, usually focusing on daily averages as in much of the epidemiology. In contrast, no personal exposure studies have measured for a whole year, so that the validation signal (personal long term PM) itself is subject to error which is usually not well quantified. The authors in this paper have calculated monthly means, and have done some sensitivity work by using the exact dates of the personal measurements only. I think they should do more with the data at hand to quantify the error associated with not having true long term personal PM data. First of all, the monthly means will vary over the year(s) so that part of the reported correlations are temporal not spatial. Can the authors estimate how large the temporal contribution is? Second, an alternative approach, used in many LUR studies which also suffer from incomplete long term coverage is to calculate unbiased annual means with either the ratio or the difference method, using the signal from one or more continuously operating ambient monitors as input for the calibration. As the authors have the raw data, I think it is both feasible and very useful to do this.”

Given the availability of our data, we are unable to quantify the contribution of the temporal component to the total error. Although data are not available for a whole year, all studies sampled personal exposures in multiple seasons, in an effort to address this issue. We conducted sensitivity analyses to assess the impact of this temporal component and showed that it leads to increased variability in the calibration coefficients, but likely no bias.

Furthermore, all studies were conducted within one to two years of each other. As a result, we are unable to estimate the long-term trends in the personal-ambient relationship and thus to error. We have acknowledged this limitation in our Discussion (p. 13, 2nd paragraph), as we agree with Dr. Brunekreef that the temporal contribution could potentially be important: We write: “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.”

Although incomplete long coverage is a limitation of our study, our goal is to assess the error associated with use of PM concentrations measured at ambient monitors instead of personal PM exposures. Using, therefore, data from ambient monitors to calculate annual personal means, while useful in other settings, does not agree with our goal and assumptions.

“Personal measurements of PM are difficult technically, and (as all measurements) associated with a purely technical measurement error which may
be substantial compared to the between-subject variations of the quasi-long term measurement data themselves – especially when the remaining temporal component has been completely removed from the data. Would be useful to have a discussion of this, and an estimate of how important this might be."

Although we were not able to estimate the contribution of technical error on our calibration coefficients, we agree with Dr. Brunekreef that this might play an important role and have thus included a paragraph in the Discussion (p. 13, 3rd paragraph): “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].”

“Especially in the RIOPA centers, the total personal PM2.5 sd was much larger than the nearest monitor or model predicted PM2.5 sds. One suspects again that this may be related to insufficient filtering out of the temporal component, as RIOPA had 4 hours of personal measurements only.”

We agree that the large standard deviations in the RIOPA centers likely reflect the shorter monitoring durations (48 hours as compared to the 7 or 12 d in other studies). We discuss the temporal mismatch that may result from using some days in the month to estimate monthly averages in the Discussion section of our paper (bottom of p.14 and top of p.15). To elaborate on this discussion, we have added a paragraph discussing the limitation introduced by the lack of information about the personal-ambient relationship for PM2.5 over more years, which limits our ability to estimate the effect of time on our estimated calibration coefficients. (Please see our response to comment #1.)

“It could also be related to indoor sources but the tables and figures give us no information on smoking/ETS which should be the biggest contributor in most cases. Not having data on smoking/ETS in table A4 seems to severely limit the analysis of factors which might explain differences in calibration factors between cities.”

We agree that ETS may contribute in part to the observed between-city heterogeneity. All studies in our analyses required that the subjects were non-smokers and lived in non-smoking homes, as ETS may overwhelm the contribution of ambient PM2.5 (and all other PM2.5 sources) to personal PM2.5. Since people do not spend all of their time at home, however, it is possible that ETS still contributed to personal PM2.5 exposures. Unfortunately, we have no data on potential ETS exposures. We describe this limitation and its possible impacts on between-city heterogeneity in our revised Discussion. We write: (p. 12, 2nd paragraph): “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 PM$_{2.5}$ exposures, thereby potentially contributing to between-city heterogeneity in calibration coefficients. We, however, were not able incorporate ETS exposures in our analysis, as some studies did not report ETS exposure information.”

“Another obvious candidate is the ambient and personal distribution sd: one can easily imagine that in populations with little variation in one or more of these distributions, correlations between ambient and personal PM are decreased because of a larger role of measurement error.”

Dr. Brunekreef correctly here describes the behavior of correlation coefficients, which is a function of the variances of both the personal exposures and the ambient concentrations.

Although different cities may have different variances of personal exposures and/or ambient concentrations, when we regress the personal exposures on the ambient concentrations, the estimated slope asymptotically converges to a value that is invariant to the amount of variance in either personal exposures or ambient concentrations for each city. What changes, as the within-city variance in either variable changes, is the variance of the estimated slope, which will increase as the variance of the dependent variable increases or the variance of the independent variable decreases. Differences in the within-city variances of either personal exposures or ambient concentrations, hence, cannot account for differences in the slopes between cities, i.e. between-city heterogeneity.