Published in final edited form as:

Environmetrics. 2013 December ; 24(8): 529–530. doi:10.1002/env.2242.

Measurement error in air pollution epidemiology: Guidance for uncertain times

Roger D. Peng

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

It is a great pleasure to provide a discussion of Adam Szpiro's and Chris Paciorek's (SP) comprehensive and clarifying paper on measurement error in air pollution epidemiology. I see this paper as a culmination of much methodological work, some conducted by Szpiro and Paciorek themselves, that has been conducted over the past decade regarding the estimation of associations between air pollution concentrations and health outcomes. Not only do SP provide insight into how measurement error arises in cohort studies of air pollution and health, they clearly specify its impacts on the estimation of the health effects parameters and propose a novel solution for accounting for the various types of measurement error.

For those unfamiliar with the general area of air pollution epidemiology, it may be worthwhile to review briefly the context in which this paper sits in order to clarify its significance. A key point to note is that a fundamental problem of essentially all studies of the health effects of air pollution is that the relevant exposure is not directly measured. With the exception of experimental studies, where exposures are controlled, and smaller-scale observational studies where direct monitoring of personal or ambient air pollution concentrations can be accomplished, studies requiring air pollution exposure estimates require some kind of surrogate. One advantage of conducting air pollution epidemiological studies is that there are extensive monitoring networks sited in major metropolitan areas that provide publicly accessible data. However, measurement error is the price one must pay when it comes to linking that data with health outcomes in a human population.

Broadly speaking, there are at least two targets of exposure estimation one can usefully imagine. The first is a person's personal exposure to air pollution, which is a combination of outdoor pollution levels and indoor pollution levels, say in their home or work place. The second is simply a person's exposure to outdoor concentrations, which is commonly the true target of estimation, even if it is not explicitly stated. While a person's personal exposure might appear to be the best exposure measure, health effects estimated using this measure may be challenging to interpret because modifying a person's personal exposure may involve multiple types of interventions. On the other hand, outdoor exposure, which represents a part of a person's overall exposure, has a more clear tool for intervention in the formulation of outdoor air quality standards by regulatory agencies such as the Environmental Protection Agency. This paper by SP focuses on the estimation of outdoor concentrations, for example outside a person's home, estimated from outdoor monitoring networks. This focus raises some unique measurement error problems and SP do a nice job of extending the traditional measurement error framework into this setting.

In the development of our understanding of measurement error in air pollution studies, a key recent contribution came from Gryparis et al. (2009). They showed that prediction of exposures can introduce a Berkson-like error component which is not like typical Berkson error because of the spatial correlation between study subjects. However, that paper did not directly address the issue of possible error introduced by the estimation of the prediction model itself. Later, Szpiro et al. (2011) noted that estimation of the prediction model introduces a classical-like error component. Paciorek (2010) laid some of the groundwork for proper estimation of health effects in cross-sectional studies while adjusting for potential unmeasured confounding. This work combines the previous work into a single overarching framework and proposes a method to address both the Berkson-like and classical-like problems at once.

One tension that arises in many areas of biomedical science is between the need to have a single estimate of whatever parameter is of interest and the recognition that there is no perfect model that accounts for all sources of uncertainty and bias. Air pollution epidemiology is certainly not immune to this problem. Almost all would agree that there is no perfect model and in fact there are likely many unmeasured (or unmeasurable) factors that could confound the association between pollution exposure and a health outcome. However, most might also agree that there is often a need to come up with a single “best” estimate that can be used as a summary while accounting for known sources of uncertainty and bias. This debate played out previously in the literature on time series analyses of air pollution and health data which primarily are used to estimate short-term or acute health effects (Dominici et al., 2004; Peng et al., 2006). The time series literature roughly complements the work that is covered here by SP. In time series studies, the advent of multi-site studies and the use of hierarchical models allowed for the estimation of a robust summary estimate (averaged across communities) while also allowing for reasonable sensitivity analysis with respect to the site-specific models. SP have produced in this paper an analogous development, essentially a method that accounts for known sources of uncertainty and bias, providing a “best” estimate, while also allowing for reasonable sensitivity analysis.

In Table 2 of SP’s analysis of the MESA Air cohort data, it is interesting to note that the range of variation found across the different approaches in estimating the health effect parameter β is relatively small. The most naive model produces an estimate of 0.66 (SE: 0.62) while perhaps the best model (5 degrees of freedom with bias correction + bootstrap) produces an estimate of 0.69 (SE: 0.67). Across the 9 different models, if one were to convert the estimates and standard errors into approximate p -values, they would range from 0.11 to 0.16. Even the most optimistic investigator would struggle to claim that any of those estimates were significantly different from zero. Therefore, if one’s primary question is to the effect of “Is there an association between NO_x and LVMI”, it seems clear that the answer is no. SP’s analysis gives us comfort that our qualitative conclusion is not seriously biased and that we would not draw an incorrect conclusion from the analysis. For the purpose of producing a single best estimate, I am struck by the similarity between the result from the optimal model and the result of the naive estimate. Although the simulations indicate that there is a potential for serious bias and undercoverage, in the analysis of the real data, the

range of outcomes does not seem to match as closely that which was found in the simulation.

Ultimately, it is useful to separate SP's work into two parts. There is the framework which decomposes the various sources of error that are induced in the common practice of estimating health effects in cross-sectional studies. The delineation of this framework is critical and provides new insights into model development for such studies. The practical advantages of SP's methodological development (bias correction with nonparametric bootstrap) will likely depend on how estimates of health effects are interpreted and for what they will subsequently be used. If the estimates are used to answer the basic question of "Is there an effect?" then I would argue one is more interested in conducting a thorough sensitivity analysis that presents a range of estimates based on assumptions about unmeasured confounding, model misspecification, and many other factors. Here, I do not think modest differences in the standard errors such as those shown in Table 2 would dramatically alter any conclusions drawn regarding this qualitative question. However, in a scenario where one might be feeding a health effects estimate into a subsequent analysis, such as a risk analysis where health impacts might be calculated, then it seems far more important to have a set procedure that obtains the "best" estimate on which to make inference. It is here that I think SP's approach will be quite applicable and should draw attention. Given that SP's method does not appear to be unreasonable in terms of computational demands, there is not a substantial tradeoff to consider when applying it and it should be feasible to use if appropriate software were available.

Overall, I have found SP's analysis framework enlightening and it has helped me understand the impact of various modeling decisions in the estimation of health effects in a cross-sectional setting. I think this is a key contribution to the field and will result in a number of readily implementable and practical methodological recommendations.

Acknowledgments

This work was supported by awards R01ES019560 and R21ES020152 from the National Institute from Environmental Health Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Environmental Health Sciences or the National Institutes of Health.

References

- Dominici F, McDermott A, Hastie T. Improved Semiparametric Time Series Models of Air Pollution and Mortality. *Journal of the American Statistical Association*. 2004; 99:938–948.
- Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull BA. Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics*. 2009; 10:258–274. [PubMed: 18927119]
- Paciorek CJ. The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Stat Sci*. 2010; 25:107–125. [PubMed: 21528104]
- Peng RD, Dominici F, Louis TA. Model choice in time series studies of air pollution and mortality (with discussion). *Journal of the Royal Statistical Society, Series A*. 2006; 169:179–203.
- Szpiro AA, Sheppard L, Lumley T. Efficient measurement error correction with spatially misaligned data. *Biostatistics*. 2011; 12:610–623. [PubMed: 21252080]