Published in final edited form as:

Epidemiology. 2015 March ; 26(2): e28. doi:10.1097/EDE.0000000000000231.

Confounding Adjustment and Exposure Prediction in Environmental Epidemiology: Additional Insights

Francesca Dominici and

Dept. of Biostatistics, Harvard School of Public Health, Boston, MA

Matthew Cefalu*

Rand, Santa Monica, CA

In our recent paper,¹ we discuss the interplay between confounding adjustment and exposure prediction in environmental epidemiology. We focus on a two-stage approach where the exposure is predicted in the first-stage and the estimated exposure from the first-stage is used as a known covariate in a second-stage health-effect regression model. In this two-stage approach, there is always uncertainty about the exact set of confounders that must be included into the health-effect regression model. We argue that: (1) the exposure prediction stage can be beneficial or detrimental when the goal of the study is health effect estimation; (2) it is dependent on which covariates are included in the health-effects regression model to control confounding. Several special cases are discussed, and all analytic results are left to the eAppendix.

Szpiro and colleagues²⁻⁴ have been conducting a similar line of research. Specifically, they (1) demonstrate that more accurate exposure prediction does not necessarily improve health-effect estimation,² (2) provide an analytic solution for correction of measurement error with spatially misaligned data,³ and (3) discuss confounding and measurement error in the context of air pollution epidemiology.⁴

In our paper in *EPIDEMIOLOGY*, we inadvertently omitted their latest contribution,⁵ which is closely related to our work. Szpiro and Paciorek⁵ propose an analytic framework for describing the complex measurement error that is induced when a two-stage approach is used in environmental epidemiology. In Section 2.3, they provide two conditions that need to be satisfied in the first-stage exposure prediction model to guarantee consistent estimation of a health-effect in the second-stage.

According to Szpiro and Paciorek,^{5p504} Condition 1 is that the spatial distribution of the covariates must be the same at the monitoring locations and at the subject locations. We made the stronger assumption that the regression coefficients from the exposure prediction model are known exactly. This can be interpreted as a combination of infinite monitoring data, along with Condition 1.⁵

Szpiro and Paciorek^{5p504} further propose Condition 2, which states that the covariates and basis functions used in the exposure prediction model must include the spatially structured

*Corresponding Author: Matthew Cefalu, Rand, 1776 Main St., Santa Monica, CA 90401, m.s.cefulu@gmail.com.

components of the covariates in the health-effect model. This condition is related to our results in the section titled “Confounding Bias Due to Exposure Prediction Under Exposure Prediction Model Misspecification.”^{1p585} We report that under a correctly specified health-effects regression model, using a predicted exposure will bias the health-effect estimate unless (1) the exposure prediction model is correctly specified, (2) the covariates used in the exposure prediction are uncorrelated with the confounders in the health-effect model, or (3) all the true confounders are included in the exposure prediction model. This third condition is a simplified version of Condition 2 of Szpiro and Paciorek.^{5p504}

One important difference between our results and the results of Szpiro and Paciorek⁵ is that they assume the health-effect of interest is the coefficient from whatever health-effect regression model is specified. In other words, all of their results are conditional on a pre-specified health-effect regression model that does not necessarily fully adjust for confounding. In our work, instead, we report our results under a situation where we target the health effect from a correctly specified health-effect regression model. This difference allowed us to consider situations where both confounder selection and exposure prediction are needed.

In observational studies, there is inherent uncertainty in how to properly adjust for confounding, and we have explored the interplay between exposure prediction and confounding adjustment. Our results, along with those of Szpiro and Paciorek,⁵ highlight the fact that exposure prediction and confounding adjustment need to be considered simultaneously.

We are thankful to Drs. Szpiro and Paciorek for their important contributions on this topic and thankful to the Editor for giving us the opportunity to tie together these two important contributions in the literature.

References

1. Cefalu M, Dominici F. Does Exposure Prediction Bias Health-Effect Estimation? The Relationship Between Confounding Adjustment and Exposure Prediction. *Epidemiology*. 2014; 25(4):583–590. [PubMed: 24815302]
2. Szpiro AA, Paciorek CJ, Sheppard L. Does more accurate exposure prediction necessarily improve health effect estimates? *Epidemiology*. 2011; 22(5):680–685. [PubMed: 21716114]
3. Szpiro AA, Sheppard L, Lumley T. Efficient measurement error correction with spatially misaligned data. *Biostatistics*. 2011; 12(4):610–623. [PubMed: 21252080]
4. Sheppard L, Burnett RT, Szpiro AA, Kim SY, Jerrett M, Pope CA, Brunekreef B. Confounding and exposure measurement error in air pollution epidemiology. *Air Quality, Atmosphere & Health*. 2012; 5(2):203–216.
5. Szpiro AA, Paciorek CJ. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics*. 2013; 24(8):501–517. [PubMed: 24764691]