# DO MORE ACCURATE EXPOSURE PREDICTIONS ALWAYS IMPROVE HEALTH EFFECT ESTIMATES?

**Adam A. Szpiro,** *University of Washington, USA*
**Christopher J. Paciorek,** *UC Berkeley, Harvard University, USA*
**Lianne Sheppard,** *University of Washington, USA*

**Background and Aims:** Air pollution concentrations are typically not measured at subject locations, so monitoring data are used to predict exposures by land-use regression (LUR) or related methods. It is often assumed that minimizing prediction error improves health effect estimation. We show this is not always correct..
**Methods:** We simulate exposures at subject and monitor locations by a LUR model with three geographic covariates, assuming only monitoring data are observed. We simulate health outcomes from a linear disease model. We fit a LUR model to the monitoring data and use predicted subject exposures to estimate the health effect. We calculate average out-of-sample R^2 as a measure of prediction accuracy and the bias and standard deviation of effect estimates. We compare the correctly specified LUR model to a misspecified version that omits the third geographic covariate. We vary the numbers of subjects and distribution of monitor locations.
**Results:** In all scenarios, the average out-of-sample R^2 at subject locations for the correctly specified LUR model is 0.70, apparently better than the corresponding value of 0.50 for the misspecified model. Health effect estimates are nearly unbiased, but their standard deviations vary between scenarios. With 10,000 subjects and uniformly high variability of geographic covariates in monitoring data, the standard deviation of health effect estimates is 0.12 for the correctly specified model and 0.21 for the misspecified model. If the third geographic covariate is less variable across monitor locations, the standard deviation is 0.23 for the correctly specified model and 0.16 for the misspecified model; i.e., the “better” exposure model leads to less certain effect estimates.
**Conclusions:** Improved air pollution exposure predictions do not necessarily lead to better estimation of health effects. Exposure models should be evaluated based on utility for estimating health effect parameters, in contrast to the current practice of optimizing prediction accuracy.