ESTIMATING ACUTE AIR POLLUTION HEALTH EFFECTS IN THE MULTI-ETHNIC STUDY OF ATHEROSCLEROSIS (MESA)

Adam A. Szpiro, *University of Washington, USA*
Lianne Sheppard, *University of Washington, USA*
Sara D. Adar, *University of Michigan, USA*
Joel D. Kaufman, *University of Washington, USA*

Background and Aims: There is growing interest in exploiting cohort data to estimate acute health effects of air pollution. Since exposure and outcome have seasonal trends, we need to control for temporal confounding. A methodology developed for time series studies is to include semiparametric spline terms in the regression. Methods have been proposed for selecting degrees of freedom (df), but existing literature does not address the implications for cohort data.
Methods: Our first objective is to adapt semiparametric regression methods to cohort studies, accounting for the possibility of multiple or no health observations on some days. We derive methods for choosing df and for computing standard errors. Our second objective is a more efficient alternative to semiparametric regression based on preadjusting the exposure, exploiting exposure data on days with no health observations. The health model can be fit by ordinary least squares (OLS) or by generalized least squares (GLS) using weights derived from a mixed model. We report simulation results and apply our methods to a previously published study of fine particulate matter and retinal arterial diameter in MESA, a population-based cohort of men and women aged 45-84 from 6 sites, funded by NHLBI.
Results: Sufficient df to account for temporal structure in the outcome consistently results in unconfounded effect estimates. This is not always the case if df are only sufficient to account for the trend in the exposure. Preadjusting the exposure and using GLS is more efficient than semiparametric regression for datasets with limited health observations compared to the required df.
Conclusions: Semiparametric regression methods from time series studies can be adapted for cohort data, as long as the data structure is considered in determining the number of df and in calculating standard errors. Efficiency can be improved by preadjusting the exposure and fitting the health model with GLS.