Potential for Timing Bias When Estimating Critical Windows to Air Pollution in Children’s Health

Ander Wilson1*, Yueh-Hsiu Mathilda Chiu2, Hsiao-Hsien Leon Hsu3, Robert O. Wright2, Rosalind J. Wright2, and Brent A. Coull3
1 Colorado State University, 2 Icahn School of Medicine at Mount Sinai, 3 Harvard T. H. Chan School of Public Health
ander.wilson@colostate.edu | anderwilson.github.io

Background
Evidence supports an association between maternal exposure to air pollution during pregnancy and children's health outcomes. Recent interest focuses on identifying critical windows of vulnerability. We use a simulation study to compare popular and easy to implement methods to estimate critical windows and highlight the potential for timing bias in estimates obtained by separate analyses of exposure in each trimester.

Data
We use data from the Asthma Coalition on Community, Environment, and Social Stress (ACCESS) project. ACCESS is a prospective, longitudinal study originally funded to recruit n=500 mother-child pairs recruited between August 2002 and January 2007 in the Boston, MA, USA area. We use data from full-term birth with complete data.

TAE = Trimester Average Exposure

Analysis of Fat Mass and BMI z-score highlights inconsistencies between estimates

Regression Methods
Separate TAE approach: Three separate regression models to estimate the association between the three TAEs and the outcome:

\[ Y_i = \beta_j + \alpha_j TAE_{ij} + z_i^T \gamma_j + \epsilon_{ij}, \]

Joint TAE approach: One regression model to estimate the association between all three TAEs and the outcome:

\[ Y_i = \beta_0 + \alpha_1 TAE_{1i} + \alpha_2 TAE_{2i} + \alpha_3 TAE_{3i} + z_i^T \gamma + \epsilon_i, \]

DLM approach: Smoothly varying exposure effect where \( \alpha_t \) is the regression coefficient for week \( t \). Model \( \alpha_t = \alpha(t) \) as a natural cubic spline with \( df \) chosen via generalized cross-validation:

\[ Y_i = \beta_0 + \sum_{t=1}^{37} \alpha_t PM_{it} + z_i^T \gamma + \epsilon_i \]

• \( TAE_{ij} \) is the trimester average exposure for trimester \( j \), \( j = 1,2,3 \)
• \( z_i \) is a vector of baseline covariates
• \( \epsilon_{ij} \) are independent and identically distributed mean zero residuals
• \( \beta, \alpha, \) and \( \gamma \) are unknown regression coefficients

Conclusions
• Estimates using separate models for each TAE were biased and identified incorrect windows.
• This bias arose from seasonal trends in PM\(_{2.5}\) that induce correlation between TAEs.
• Including all TAEs into one model reduced bias.
• DLM was unbiased, added flexibility to identify critical windows, and is preferred whenever possible.