Workshop on Analyzing Mixtures in Environmental Health Studies: Bayesian Kernel Machine Regression

Brent Coull

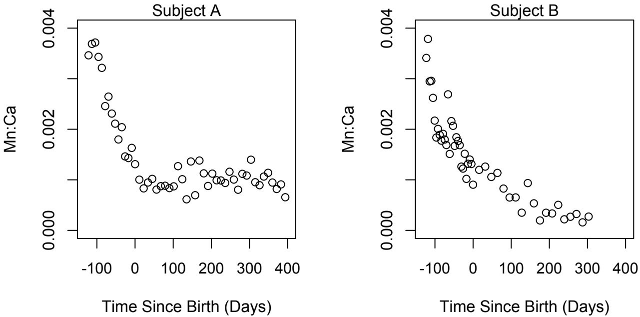
Harvard T.H. Chan School of Public Health
Departments of Biostatistics and Environmental Health

24 August 2018

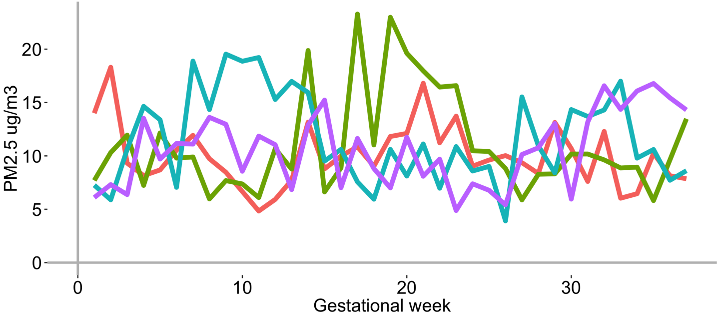
2015 NIEHS Workshop on Statistical Methods for Mixtures

- Great workshop!
- Compared practical performance and results of many of these methods on common datasets.
- Focus was on methods for a single set of measured exposures and single outcome.
- Modern epidemiology has moved beyond the single exposure, single outcome paradigm.
- I'll now discuss some of our efforts to fill these gaps.

Teeth Biomarkers



Estimated Daily Residential Air Pollution



Single Exposure: Lagged Regression

- **Distributed lag models**: Identify exposure time windows most associated with an outcome for a single exposure:

$$Y_i = \beta_0 + \sum_{t=1}^{T} \gamma_t z_{it} + \boldsymbol{\beta}\mathbf{x}_i + \epsilon_i$$

- Typically high correlation among z_{it} from multiple time windows.
  - Model γ_t as a function of t
  - Shrink γ_t from neighboring windows towards one another.

Mixtures: Lagged Kernel Machine Regression (LKMR)

Lack of methods for lagged regression for environmental mixture

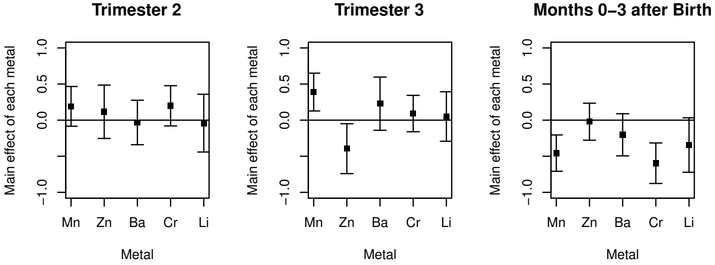
$$Y_i = \beta_0 + \sum_t h_t(z_{1it}, \ldots, z_{Mit}) + \boldsymbol{\beta}\mathbf{x}_i + \epsilon_i$$

$$Y_i = \beta_0 + \sum_t h_{i,t} + \boldsymbol{\beta}\mathbf{x}_i + \epsilon_i$$

- $h_t(\cdot)$: association between outcome and exposure to mixture measured at time window t

Liu et al. *Biostatistics* 2017

ELEMENT Cohort Pilot (n=81): LKMR Analysis of Tooth Metals and Age 8 WRAVMA



Longitudinal Outcome Data

- Association between mixture and outcome at time t:

$$Y_{it} = h(z_{1it}, ..., z_{Mit}) + \boldsymbol{\beta}\mathbf{x}_{it} + b_i + \epsilon_{it}$$

- Prior (e.g. prenatal) exposure to metals and neurocognitive trajectories:

$$Y_{it} = h_1(z_{1i}, ..., z_{Mi}) + h_2(z_{1i}, ..., z_{Mi}) * \text{age}_{ij} + \boldsymbol{\beta}\mathbf{x}_{it} + \boldsymbol{b}_i\mathbf{u}_{it} + \epsilon_{it}$$

Liu et al. *Statistics in Medicine* 2018, in press

Interaction Analyses (Hypothesis Testing)

In some grouped settings, often of interest to formally test for interaction between groups:

$$Y_i = h\left(\mathbf{z}_i, \mathbf{w}_i\right) + \boldsymbol{\beta}\mathbf{x}_i + \epsilon_i$$

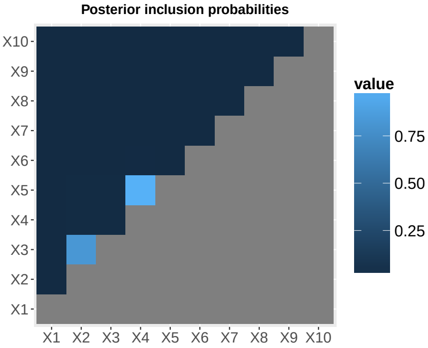
$$H_0 : h\left(\mathbf{z}_i, \mathbf{w}_i\right) = h_1\left(\mathbf{z}_i\right) + h_2\left(\mathbf{w}_i\right)$$

- Exposure timing (Prenatal, Postnatal)
- Nutrition $\times$ Environment
- Gene $\times$ Environment
- Psychosocial Stress $\times$ Environment
- ...

Zhu et al. *NIPS* 2017; manuscript in progress

Interaction Posterior Inclusion Probabilities

$$Y_i = h(z_{i1}, ..., z_{iM}) + \boldsymbol{\beta}\mathbf{x}_i + \epsilon_i$$



Antonelli et al. 2018, submitted
github.com/jantonelli111/NLinteraction

Conclusions

- Statistical methodology for assessing the health effects of environmental mixtures has recently matured.
- More work needs to be done to have a multi-purpose toolbox for a wide variety of common research questions and study designs.
- Our approach has been to embed BKMR into existing, popular modeling frameworks for environmental epidemiologic data.
- Analogous approaches could be employed with other approaches for quantifying mixture health effects.

Acknowledgments

Co-authors

- Jennifer F. Bobb
- Linda Valeri
- Maitreyi Mazumdar
- Birgit Claus Henn
- David Bellinger
- David Christiani
- Robert Wright
- Marianthi-Anna Kioumourtzoglou
- Joey Antonelli
- Shelley Liu
- Jeremiah Zhu
- Katrina Devick
- Kyu Ha Lee
- Jane Lee
- Chris Gennings
- Rosalind Wright
- John Godleski
- Howard Hu