Evaluating a Mixture Effect of Perinatal Environmental Exposures on Childhood BMI:

Weighted Quantile Sum (WQS) Regression

Chris Gennings, PhD

Professor

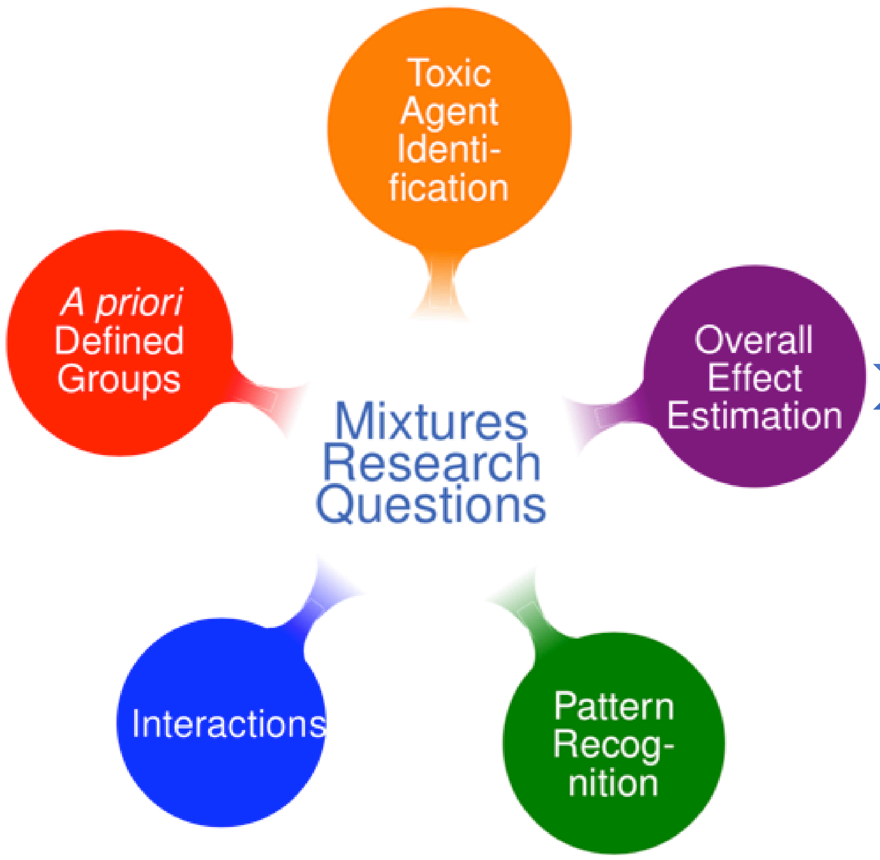
Environmental Medicine & Public Health

Icahn School of Medicine at Mount Sinai, New York, NY, USA

April 28, 2021

Icahn School of Medicine at Mount Sinai

Important research questions for environmental mixtures



Is there a "mixture effect"?

Figure by Marianthi-Anna Kioumourtzoglou

What is a “mixture effect”?

Relevant environmental exposures may result in the phenomenon of … “**something from nothing**” (Silva, Rajapakse, Kortenkamp (2002) Env Sci Tech)

- Individual components may be at exposures well below an effect level
- **Joint action** of the components produce significant effects
- Ignoring joint action of compounds may lead to significant underestimation of risk

How can we estimate a “mixture effect”?

One way of measuring a mixture effect is with Weighted Quantile Sum (**WQS) regression.***

WQSR measures the mixture effect using an empirically-weighted index of quantiles of components

Ensemble step

$$WQS = \sum_{j=1} \bar{w}_j q_j$$

Generally 40% of data

Training data

$$g(\mu) = \beta_0 + \beta_1 \sum_j w_j q_j + \sum_k \lambda_k z_k$$

Holdout validation data

Generally 60% of data

$$g(\mu) = \beta_0 + \beta_1 WQS + \sum_k \lambda_k z_k$$

Weighted indices may be estimated in both the positive and negative directions with constraints in the estimation step, then combined

* Carrico et al 2014, JABES

Motivation/Characterization of WQS regression features

- Binning concentrations into quantile scores
  - Reduces impact of outliers
- Ensemble step
  - Improves sensitivity of detecting components correctly related to the outcome (Simulation studies with environmentally-relevant correlation patterns)
  - Two versions: bootstrap samples (Carrico et al, 2015) of observations and random subset (Bello, 2014; Curtin et al 2020) selection of components (allows for p>>n)
- Directionality constraints
  - improve ill-conditioning due to complex correlation patterns typical of environmental exposures
  - Avoids concern around reversal paradox
  - Once indices are constructed, they can be put in a common regression model for exploration of the response surface
- Repeated holdout WQS regression (Tanner et al, 2020)
  - Addresses generalizability of results across validation holdout sets
- Checking linearity assumption
  - LOESS plot of covariate-adjusted outcome variable with WQS index
  - Goodness-of-fit test based on significance of quadratic WQS term
- Stratified interaction WQS analyses
- gWQS R package





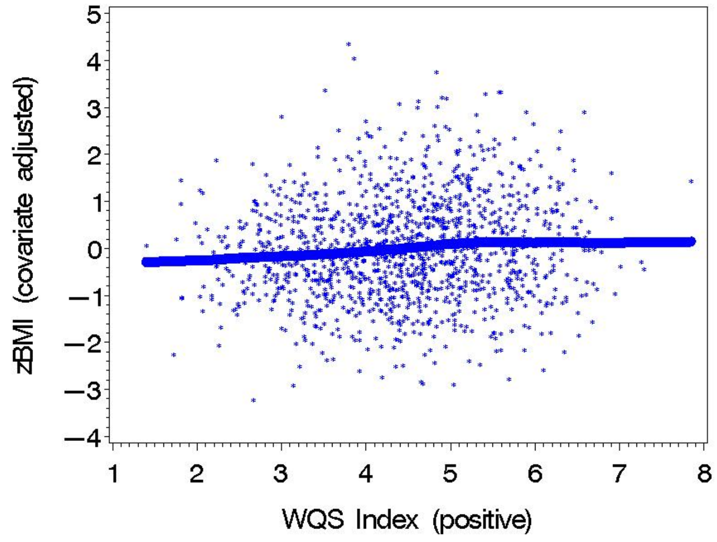
Illustration using Helix Challenge Data

- **Data** provided based on the HELIX project (N=1300) across 6 cohorts
- **Response**: zBMI at 6-7 years of age
- **Covariates**: Models adjusted for following set of covariates: cohort identifier; maternal BMI, education level, parity, native indicator; child birth weight, KIDMED (nutrition index)
- **Exposures**: both prenatal and postnatal biomonitoring concentrations from 38 chemicals/metabolites (metals, organochlorines (includes sum of 5 PCBs), PFASs, phenols, phthalates (includes sum of DEHP metabolites), OP pesticides, PBDEs)

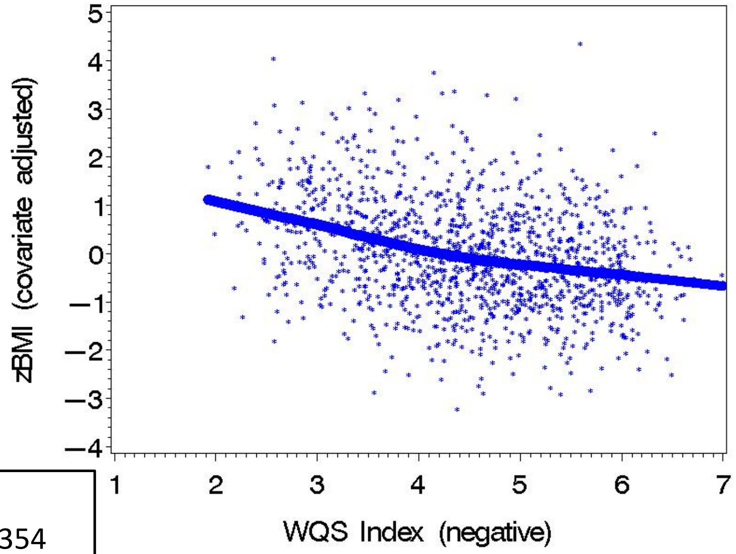
Analysis Strategy and Preliminary Results

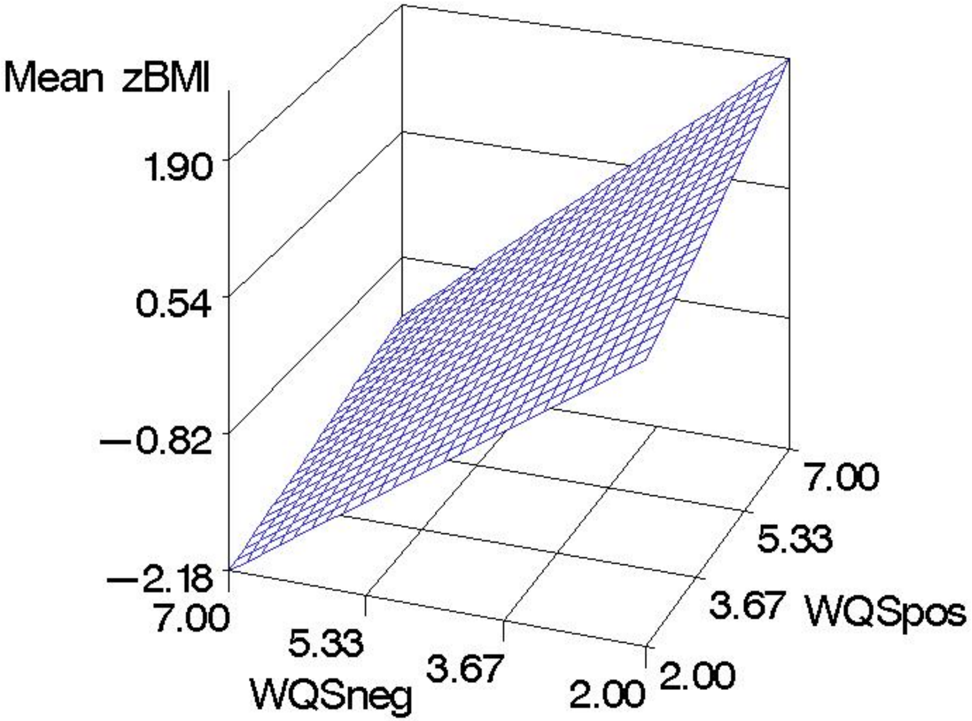
- **Single chemical analyses:**
  - 48 of 76 (63%) negative estimates (12, $p<0.05$); 37% positive estimates (4, $p<0.05$)
- **WQS regression analysis per cohort**
  - All six cohorts had significant WQS index in negative direction ($p<0.001$); none were significant in the positive direction
- **Negatively constrained WQS regression**
- **Positively constrained WQS regression**
- **Repeated holdout in negative direction**
- **Stratified Interaction WQS regression (negative direction)**

Positively and Negatively constrained WQS regression results



Significance of WQS weighted indices
WQS positive direction: p=0.142
WQS negative direction: p<0.001

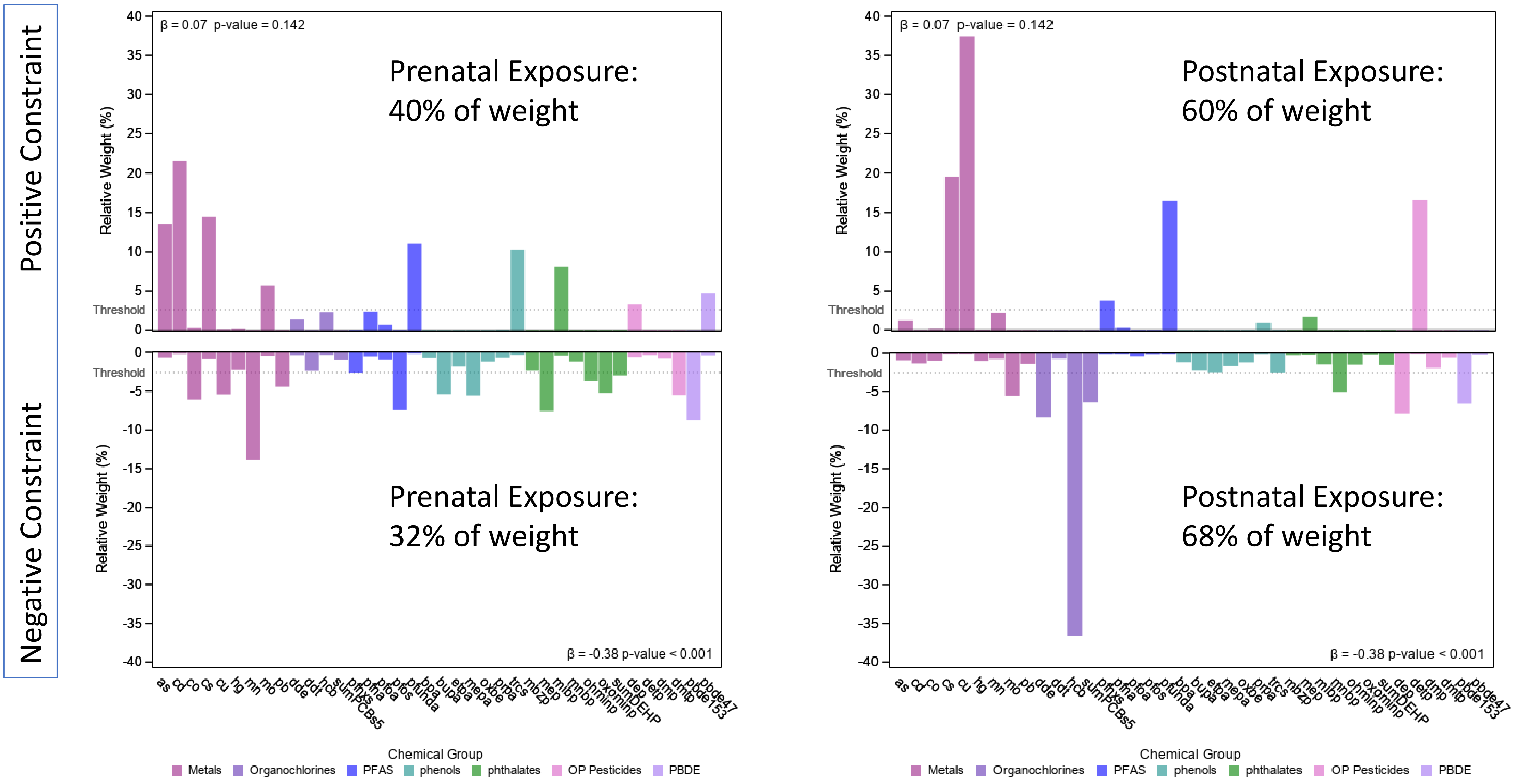




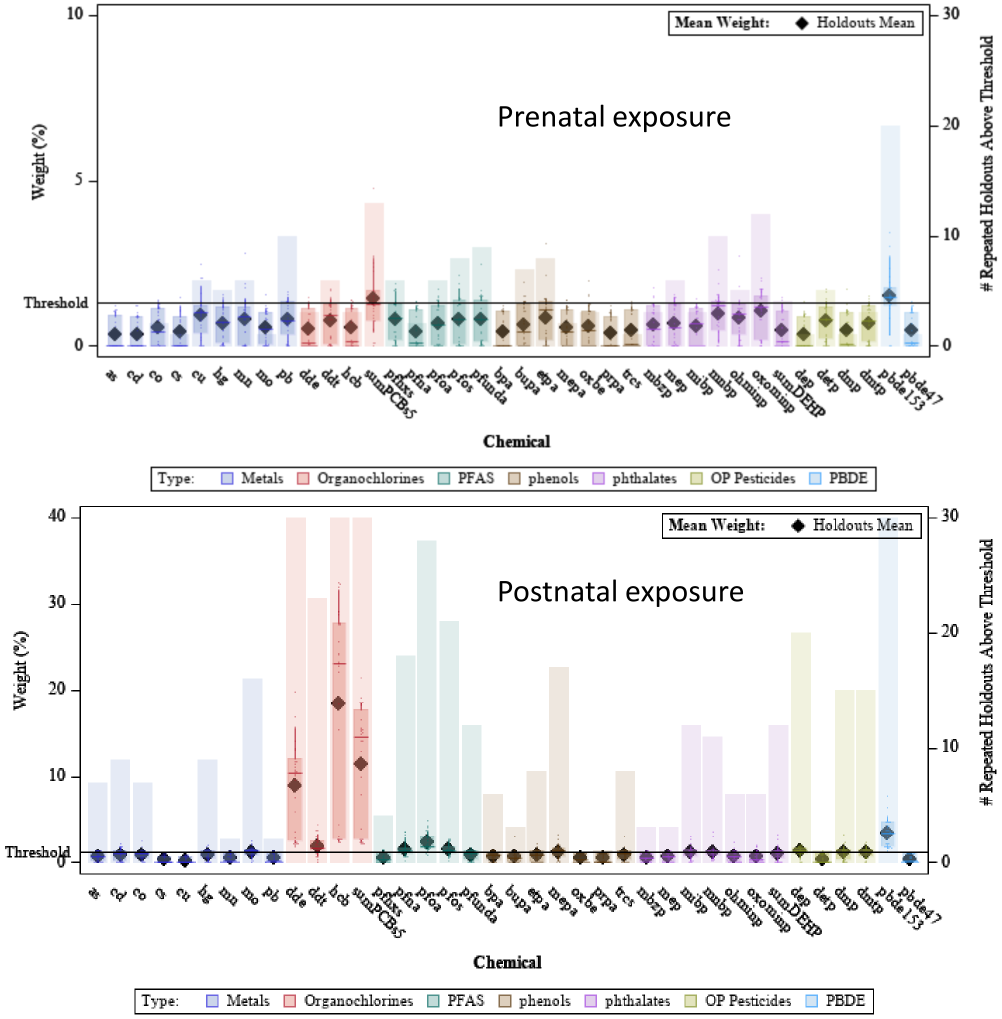
Test of Linearity,
Goodness of fit, p=0.354

* All covariates were either centered or at the reference level

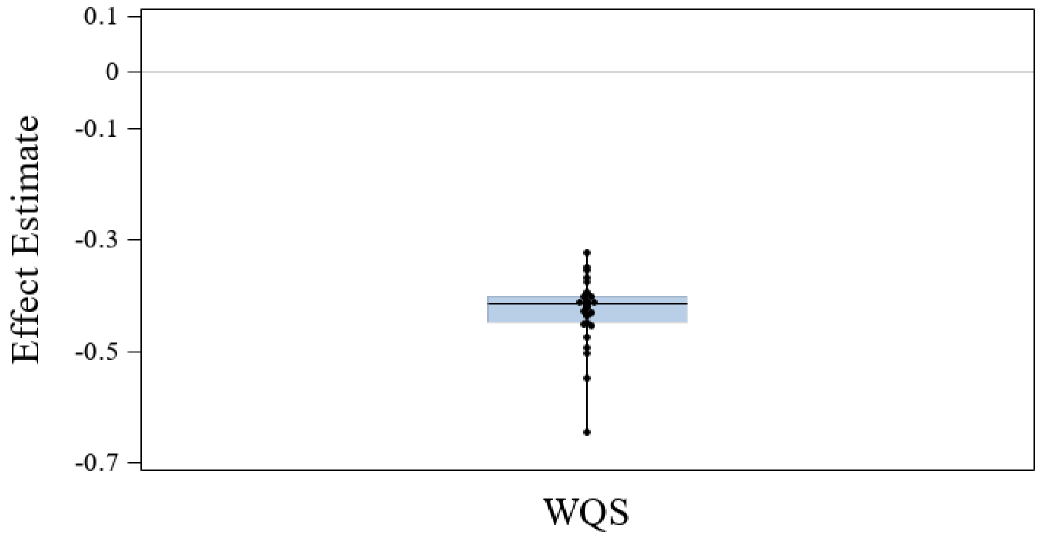
Positively and Negatively constrained WQS regression results



Negatively constrained WQS regression results with 30 repeated holdouts



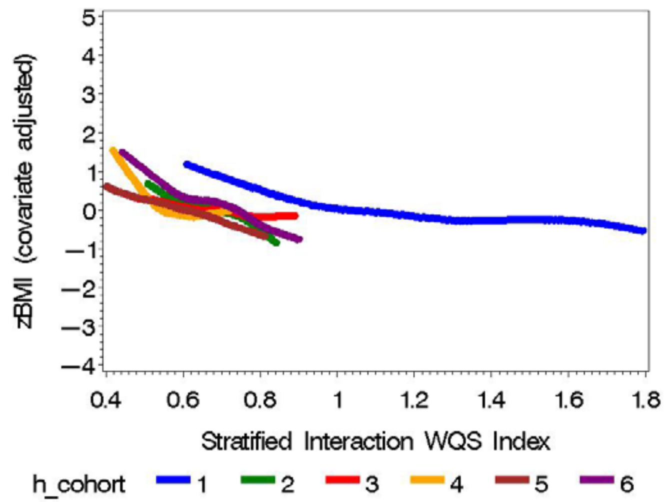
Notes: Bars correspond to right axis and indicate the number of times a chemical exceeded the concern threshold in 30 repeated holdouts. Data points, boxplots, and diamonds correspond to left axis. Data points indicate weights for each of the 30 holdouts. Box plots show 25th, 50th, and 75th percentiles, and whiskers show 10th and 90th percentiles of weights for the 30 holdouts. Closed diamonds show mean weights for the 30 holdouts.

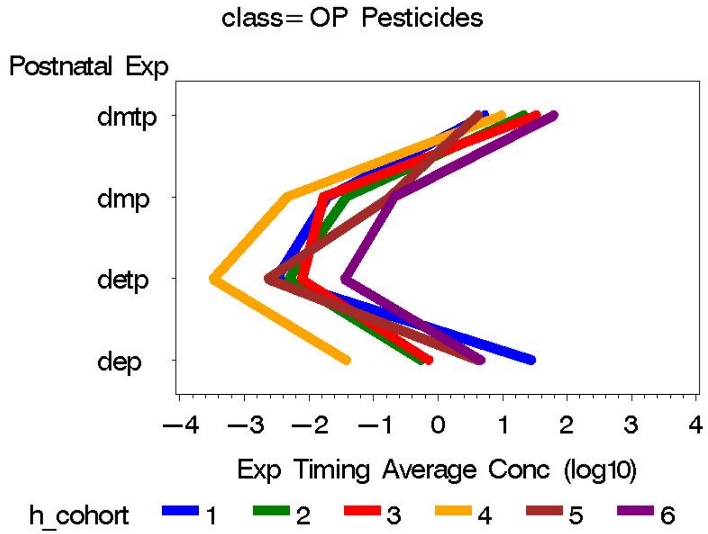


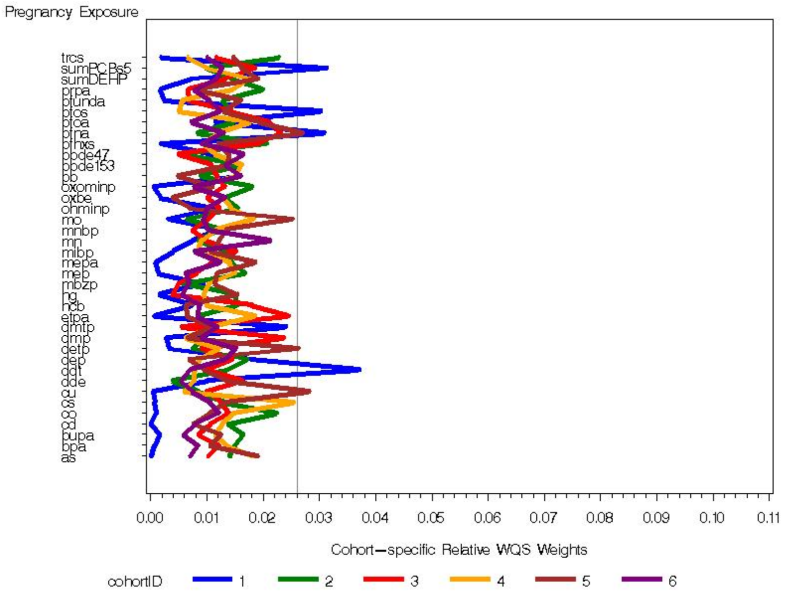
Stratified Interaction WQS regression results (negative direction)

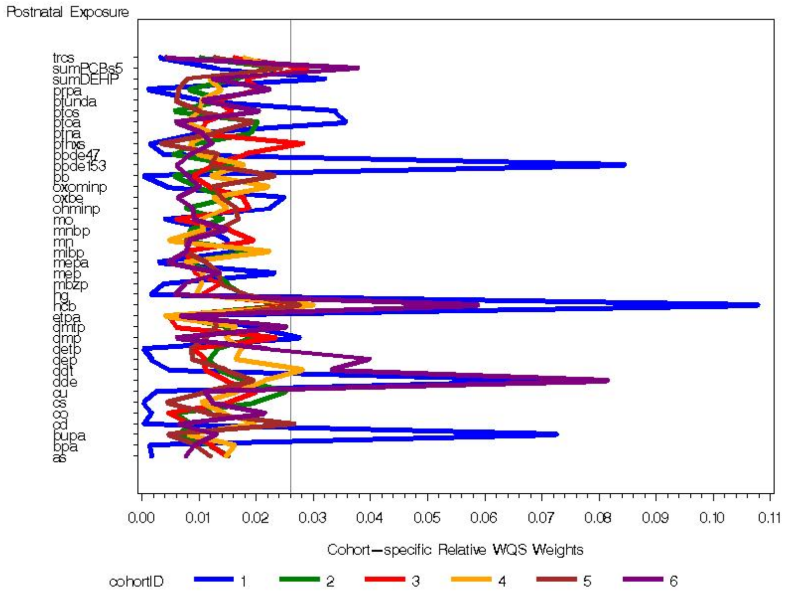
Results from the stratified interaction WQS regression analysis:

- Significant differences in beta coefficients associated with the stratified interaction WQS index. ($p<0.001$)
- The slope for cohort 1 is less steep than the others.
- Interpretation is complex due to varying exposures (i.e., patterns and averages) and weights across the cohorts.









Discussion limitations and future directions

- The focus of WQS regression is to **detect and characterize a mixture effect** among correlated exposures.
  - May include p>>n case with random subset WQSR (e.g., metabolomics; microbiome data)
  - Stratified interaction WQSR allows for strata-specific weights and beta coefficients
  - Repeated holdout validation addresses generalizability of results.

- **Limitations:**
  - Can not describe interactions among the components
  - Exposure curvilinearity is limited to low-degree polynomials (e.g., with quadratic term)
  - Splitting the data for estimation and validation allows for a strong result when significance is achieved in a holdout dataset; however, power is reduced based on the split.

- **Future directions:**
  - Lagged WQS regression allows for time-varying association between exposures and a later life health effect (Gennings et al, 2020 with LWQS R package recently published)

A team is what we need….

- **Caroline Carrico, PhD**, dissertation on WQS regression
- **Ghalib Bello, PhD**, dissertation included random subset WQS regression; and first version of lagged WQS regression
- **Stefano Renzetti, MSc, PhD candidate**, created the gWQS R package, with support from Paul Curtin
- **Paul Curtin, PhD**, developed the LWQS R package for lagged WQS regression based on the gWQS R package. Simulation study of random subset WQS regression
- **Eva Tanner, PhD, MPH**, Characterizing uncertainty in WQS weights with repeated holdout validation
- **Elena Colicino, PhD**, Bayesian Weighted Quantile Sum (BWQS) regression

FUNDING: T32ES007334, P30ES023515, U2CES026555






Thanks!