Analysis of Environmental Chemical Mixtures using Weighted Quantile Sum (WQS) Regression

Chris Gennings, PhD
Environmental Medicine and Public Health
Icahn School of Medicine at Mount Sinai
New York, NY USA

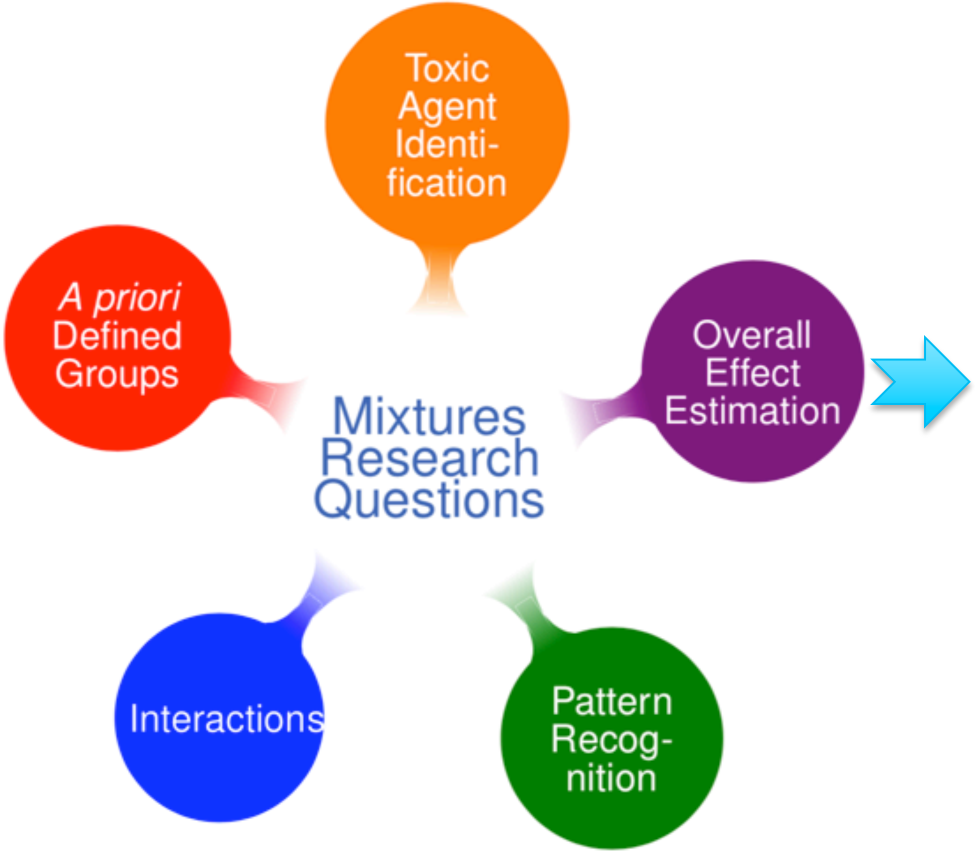
CHE Webinar 2020 Workshop
May 20, 2020

Mount Sinai

Important research questions for environmental mixtures

Figure by Marianthi-Anna Kioumourtzoglou

Important research questions for environmental mixtures



Is there a "mixture effect"?

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 do we measure the mixture effect?
  - **WQS regression*** measures the mixture effect using an empirically-weighted index of quantiles of components under the assumption of additivity
    - The index is a 1st degree approximation in the presence of interaction

* Carrico et al 2014, JABES

Weighted Quantile Sum (WQS) regression for identifying mixtures of exposures linked to health effects

Training Data (40%)

$$g(\mu) = \beta_0 + \beta_1 \sum_{j=1}^{c} w_j q_j + \sum_{k=1} \gamma_k z_{ik}$$

Final WQS index is a weighted average across the ensemble step samples using a ‘signal function’

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

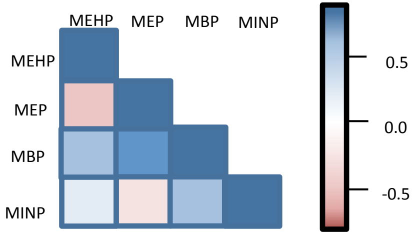
Ensemble Step

$$\bar{w}_j = \frac{1}{B} \sum_{b=1}^{B} w_{j(b)} f\left(\hat{\beta}_{1(b)}\right)$$

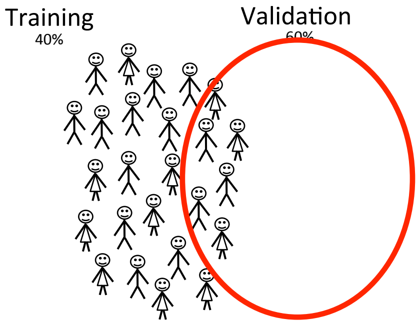
Validation Data (60%)

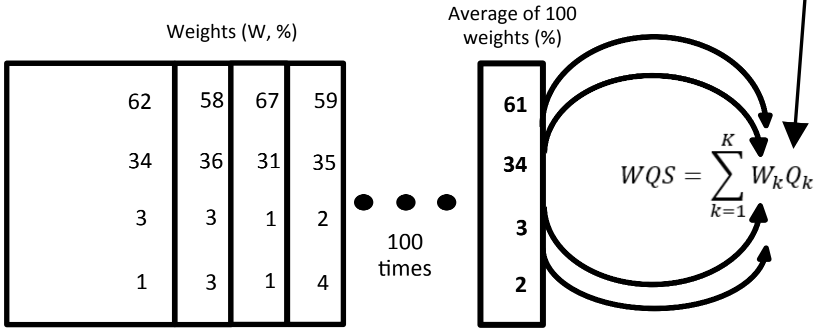
$$g(\mu) = \beta_0 + \beta_1 WQS + \sum_{k=1} \gamma_k z_{ik}$$

Weighted Quantile Sum (WQS) Regression

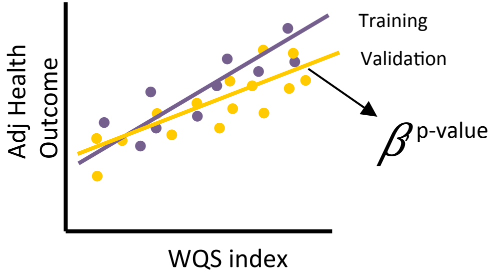


Subject ID	Concentration of MEHP (ng/ ml)	Rank (Q) of MEHP
1	3.4	4
2	1.2	2
3	10.3	9





$$WQS = \sum_{k=1}^{K} W_k Q_k$$



WQS regression: Ensemble step

- Bootstrap samples of ***observations***
  - Why?
  - How many samples?
  - Distribution of weights

- Random subset of ***components***
  - Subsets of size, say, $\sqrt{c}$ for c components
  - 1000 random subsets
  - Average across full set
  - Works when c > N

Two Strategies

WQS bootstrap: Carrico et al, JABES, 2014;
WQS random subset: Curtin et al, Comm in Stat – Sim & Comp, 2019

Analysis of environmental chemical mixtures and NHL risk in the NCI-SEER NHL Study (Czarnota et al, EHP 2015)

- Population based case-control study of NHL in four NCI-SEER centers: Detroit, Iowa, LA, Seattle
- Study design described in Colt et al 2004
  - Eligible Cases:
    - 20-74 yrs old; diagnosed with a first primary NHL between 1998-2000; N=1321
    - 2248 potentially eligible: 14% died before interview; 6% not located; 1% moved; 3% refused
    - 1728 remaining cases – 1321 (76%) participated
  - Controls:
    - >64 yrs old were selected from Medicare files; <65 yrs old were selected with random digit dialing
    - Frequency matched to cases by sex, age, race and study site
    - 2409 potentially eligible 2046 were located and contacted; 1057 (52%) participated
- 27 chemicals measured in house dust (from vacuum cleaners and where > half of their rugs were owned for more than 5 years)
  - 5 PCBs; 7 PAHs; 15 pesticides
- N=1180 subjects with complete dust analysis results and covariate values (43% controls; 57% cases)
- Covariates: sex, age at diagnosis (cases) or selection date (controls); race (White vs Not white); education level (<12, 12-15, >=16 yrs); and study site in overall model

27 chemicals and NHL risk (continued)

Table 1. Characteristics of the study population overall and by study site [*n* (%)].

Characteristic	All sites (*n* = 1,180)	Detroit (*n* = 202)	Iowa (*n* = 340)	Los Angeles (*n* = 292)	Seattle (*n* = 346)
Case status					
Control	508 (43)	75 (37)	147 (43)	125 (43)	161 (47)
Case	672 (57)	127 (63)	193 (57)	167 (57)	185 (53)
Age[a] (years)	60 ± 11.2	58 ± 11.3	61 ± 11.4	60 ± 11.2	59 ± 10.8
Sex					
Male	631 (54)	114 (56)	181 (53)	163 (56)	173 (50)
Female	549 (47)	88 (44)	159 (47)	129 (44)	173 (50)
Race					
White	1,033 (88)	164 (81)	336 (99)	213 (73)	320 (92)
Non-white	147 (12)	38 (19)	4 (1)	79 (27)	26 (8)
Education					
< 12 years	106 (9)	24 (12)	33 (10)	30 (10)	19 (5)
12–15 years	741 (63)	123 (61)	244 (72)	171 (59)	203 (59)
≥ 16 years	333 (28)	55 (27)	63 (19)	91 (31)	124 (36)

[a]Continuous variable summarized using mean ± SD.

27 chemicals and NHL risk (continued)

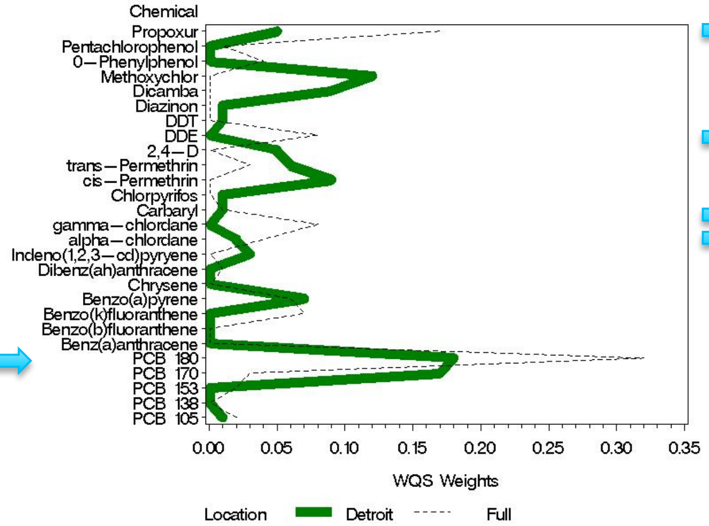
WQS regression:

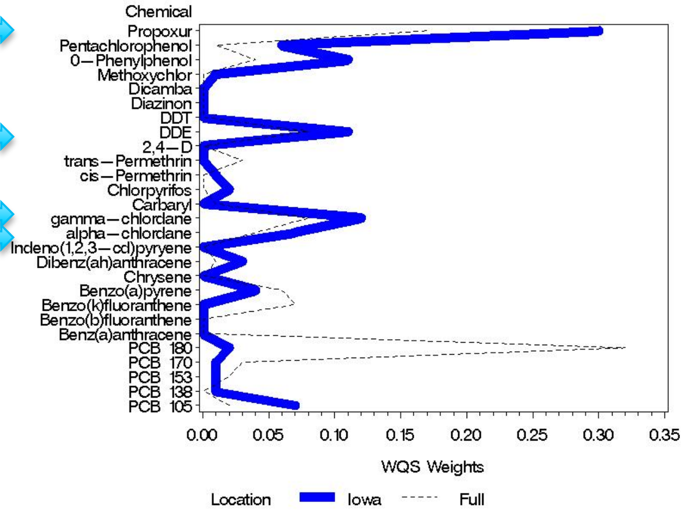
- Chemical concentrations quartiled using some imputation of analyte values (due to LOD, etc; Colt et al)
- Primary analysis based on one randomly selected imputation dataset
- B=100 bootstrap samples
- Due to the subset analyses, the data were not split as the estimated weights were not stable

Table 2. Associations between NHL and the weighted quantile sum regression index in the study population and in each study site.

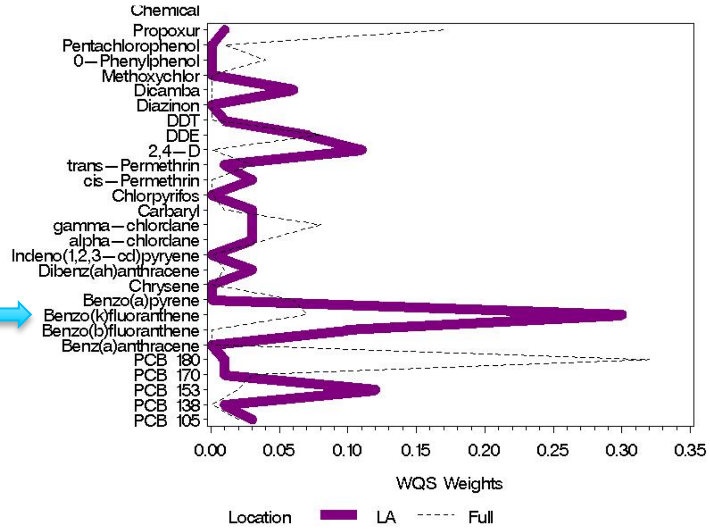
Parameter	*n*	OR[a] (95% CI)	*p*-Value
WQS_F	1,180	1.30 (1.08, 1.56)	0.006
WQS_D	202	1.71 (1.02, 2.92)	0.045
WQS_I	340	1.76 (1.23, 2.53)	0.002
WQS_L	292	1.44 (1.00, 2.08)	0.049
WQS_S	346	1.39 (0.97, 1.99)	0.071

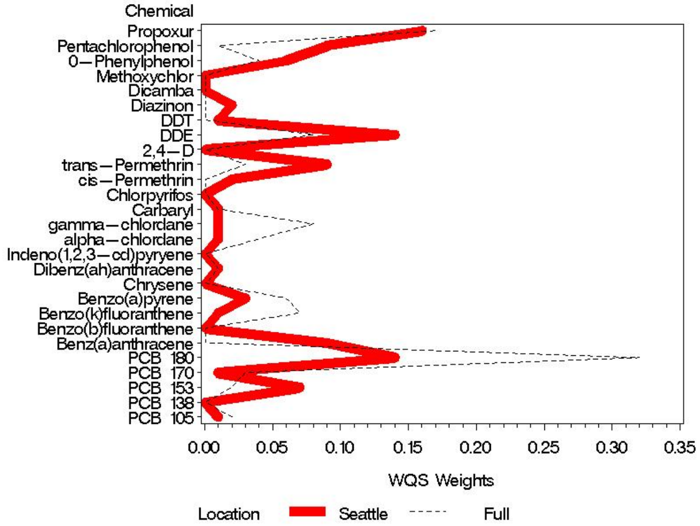
Abbreviations: WQS, weighted quantile sum index; F, full data set; D, Detroit; I, Iowa; L, Los Angeles; S, Seattle.
[a]Estimated ORs are associated with a unit increase in the WQS index. All models are adjusted for sex, race, education, and age. The model for the study population (i.e., the full data set) was also adjusted for study site.

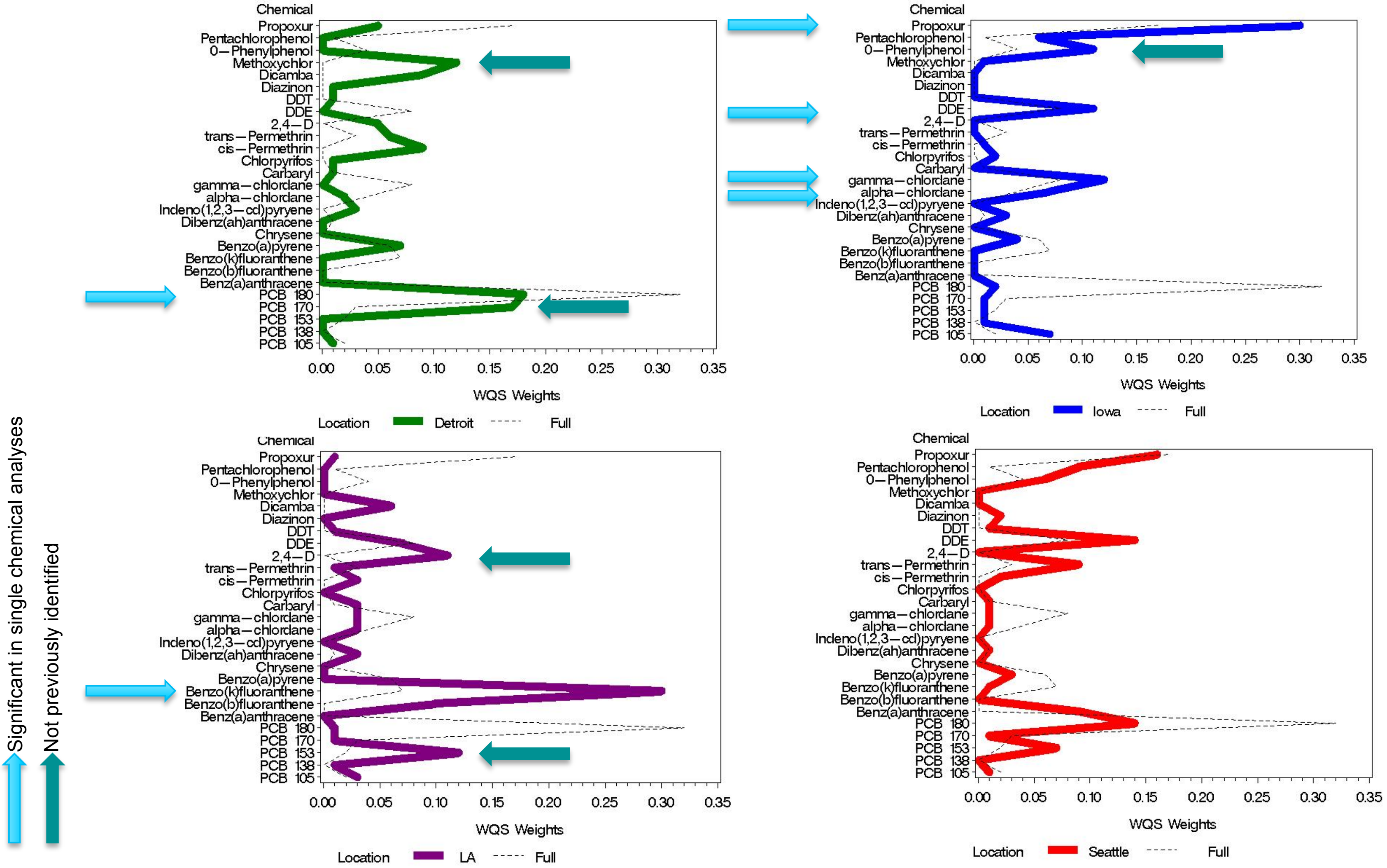




Significant in single chemical analyses






Chemical
Propoxur
Pentachlorophenol
0−Phenylphenol
Methoxychlor
Dicamba
Diazinon
DDT
DDE
2,4−D
trans−Permethrin
cis−Permethrin
Chlorpyrifos
Carbaryl
gamma−chlordane
alpha−chlordane
Indeno(1,2,3−cd)pyryene
Dibenz(ah)anthracene
Chrysene
Benzo(a)pyrene
Benzo(k)fluoranthene
Benzo(b)fluoranthene
Benz(a)anthracene
PCB 180
PCB 170
PCB 153
PCB 138
PCB 105
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35
WQS Weights
Location Detroit Full
Location Iowa Full
Location LA Full
Location Seattle Full
Significant in single chemical analyses
Not previously identified

Discussion

- We used WQS regression to model the association of a mix of 27 correlated environmental chemicals measured in house dust and risk of NHL in a case-control study of 4 study centers.
- We estimated weights associated with **increasing the risk of NHL**; constraints in the opposite direction are possible.
- The chemicals most heavily weighted in the site-specific mixture indices **varied by site**.
- In overall single chemical analyses, only PCB180 was significant and gamma-chlordane was marginally significant – indicating a **mixture effect** measured by the weighted index.
- Our findings also show that chemicals identified as important based on a site-specific WQS index may not be identified as important in an index derived from the full data set. Similarly, chemicals identified as important in the index developed from the full data set may not be identified as important in all site-specific indices.
  - These differences are due, in part, to different concentration ranges across sites and overall, to different sources of the chemicals across study sites, or differences in correlations with unmeasured exposures or other factors.
- Limitations:
  - the potential for exposure misclassification from the use of chemical concentrations in house dust as a measure of past exposures. But chemical concentrations in carpet dust may reflect integrated chemical exposure over the time.

Acknowledgments –
"A team is what we need…"

- **Caroline Carrico, PhD,** **original bootstrap WQS regression (2014, JABES)**
- **Ghalib Bello, PhD,** **lagged WQS regression (2017, Env Res; recently revised Gennings et al, 2020 Env Res)**

- **Paul Curtin, PhD,** **Random subset WQS regression – rsWQS (2020, Comm and Stat)**
- **Eva Tanner, PhD, MPH,** **Repeated holdout validation (2019, Methods X)**

- **Stefano Renzetti, MSc, PhD candidate,** **R package for gWQS**
- **Elena Colicino, PhD,** **Bayesian WQS Regression - BWQS**

Funding Sources:
T32ES0007334; R01ES028811; U2CES026555; P30ES023515