Repeated Holdout Validation for WQS

Eva M Tanner, PhD, MPH
Postdoctoral Researcher
Icahn School of Medicine at Mount Sinai
etanner@mssm.edu
https://github.com/evamtanner
Mentor: Chris Gennings

ISEE 2019 Workshop
Mixtures Analysis with Weighted Quantile Sum (WQS) Regression and its Extensions
8/25/19

Mount Sinai

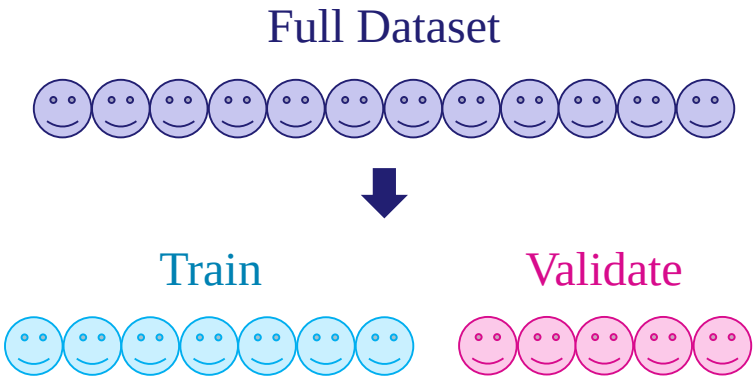
Outline

- Cross-validation Techniques
- Repeated Holdout Validation
- Application to SELMA Study
- R Tutorial

Cross-validation Techniques

What is Cross-Validation?

- Used in **predictive modeling** & machine learning for variable/model selection & to **evaluate model performance** (replicability of results)

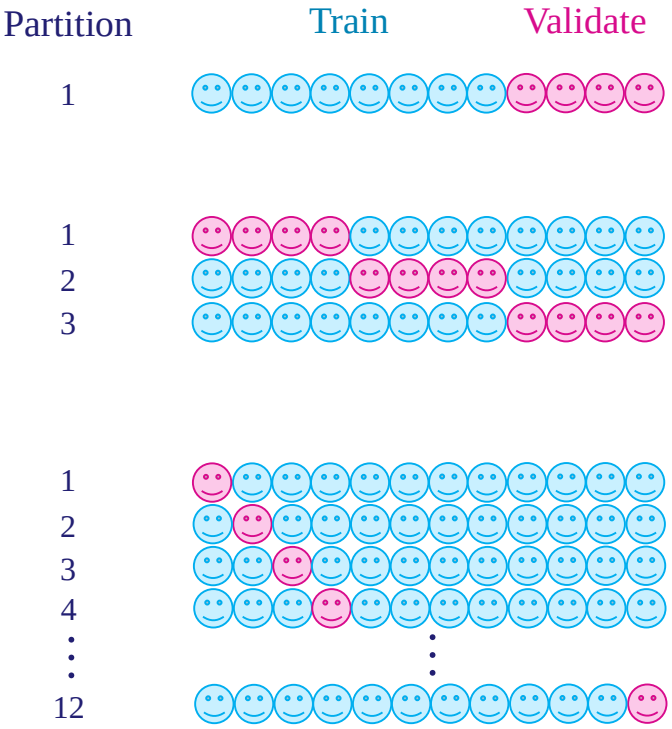


- Can also be used in explanatory (hypothesis driven) modeling to avoid fitting to noise & to assess generalizability

Shumueli (2010) DOI: 10.1214/10-STS330; Yarkoni & Westfall (2017) DOI: https://doi.org/10.1177/1745691617693393

Some Types of Validation

- Single Split
- K-fold
- Leave-One-Out (LOO)



Final Estimate

β & p-value

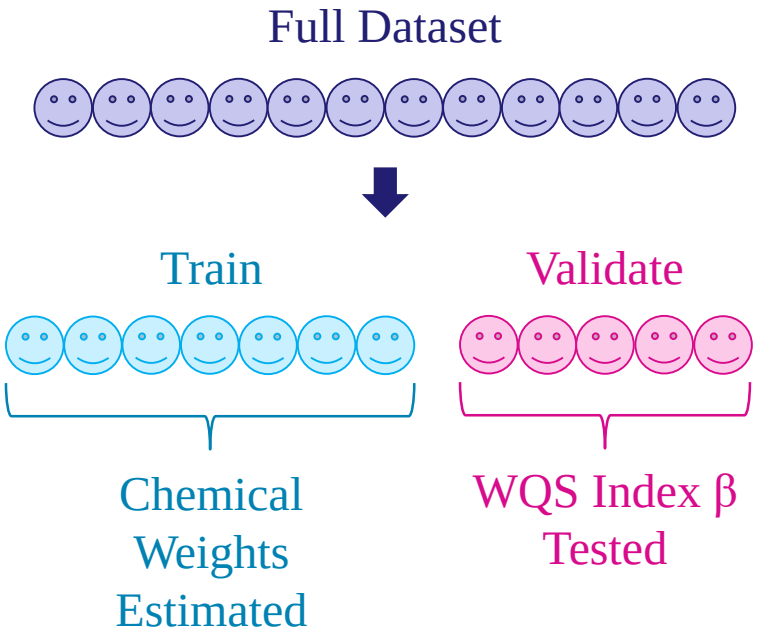
typically k = 5-10

$$\bar{\beta} = \frac{1}{k}\sum_{i=1}^{k}\beta_i$$

n = sample size

$$\bar{\beta} = \frac{1}{n}\sum_{i=1}^{n}\beta_i$$

Prior WQS Applications used a Single Split



$$Y = \beta_0 + \boldsymbol{\beta_1}\left(\sum_{i=1}^{c} \boldsymbol{w_i} q_i\right) + \text{covariates}$$

Outcome Y
Intercept β_0
WQS Index Estimate $\boldsymbol{\beta_1}$
c chemicals binned into quantile q_i with weight $0 \leq \boldsymbol{w_i} \leq 1$

In smaller sample sizes a single split can lead to **unrepresentative partitions** and **unstable estimates**

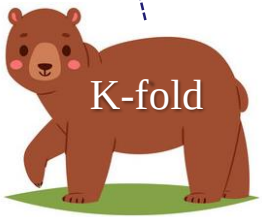
Carrico (2015) doi: 10.1007/s13253-014-0180-3

Which validation technique is just right for WQS applied to smaller (epi-relevant) sample sizes?

too unstable

Computationally intensive, still no inference

Repeated Holdout may be the sweet spot

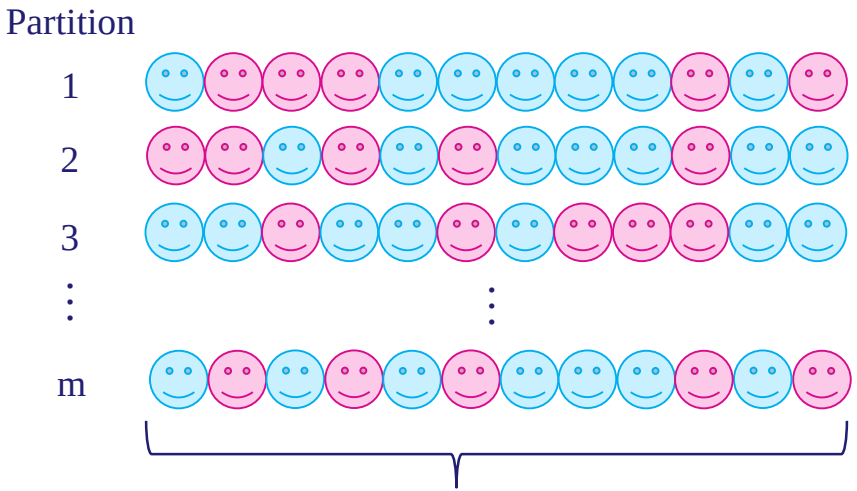


Averages WQS Index β, but no inference

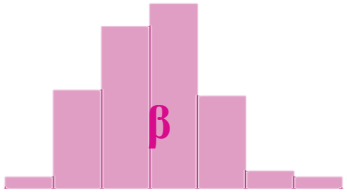
May just fit to within-sample noise

Repeated Holdout Validation

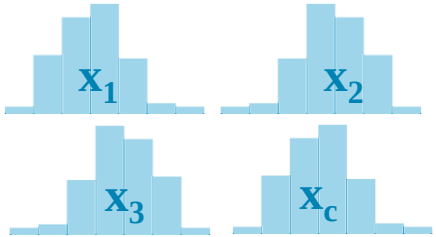
- Single Split + m Bootstraps
- Enables nonparametric inference of WQS Index estimate
  - β_{50} ($\beta_{2.5}$, $\beta_{97.5}$)
- Characterize uncertainty in selecting chemicals of concern



Distribution of validated WQS Index estimates



Distribution of weights for c chemicals



Application to SELMA Study

Endocrine Disrupting Chemicals (EDCs) = Xenobiotics able to Interfere with Hormone Action

PHENOLS

PERSISTENT ORGANIC POLLUTANTS (POPs)

PLASTICIZERS

PESTICIDES

Prenatal EDC Exposure Impacts Child Neurodevelopment

- POPs, organophosphate & pyrethroid pesticides, phthalates, & BPA associated with
  - Altered infant brain development
  - Lower cognitive functioning
  - Neurobehavioral changes

- LIMITATION: Single chemicals evaluated in isolation

- GOAL: Evaluate impact of EDCs mixture on child IQ

Berghuis (2015) doi:10.1007/s00204-015-1463-3; Braun (2017) doi:10.1016/j.neuro.2017.07.027; Eskenazi (2018) doi:10.1289/EHP2129; Factor-Litvak (2014) doi:10.1371/journal.pone.0114003; Rauh (2011) doi:10.1289/ehp.1003160

Swedish Environmental Longitudinal Mother and Child, Asthma and Allergy (SELMA)

EDCs measured in blood & urine (N=2325)

Phthalates associated with Language Delay (N=963)

Cognitive assessment using WISC-IV (N=718)

Conception

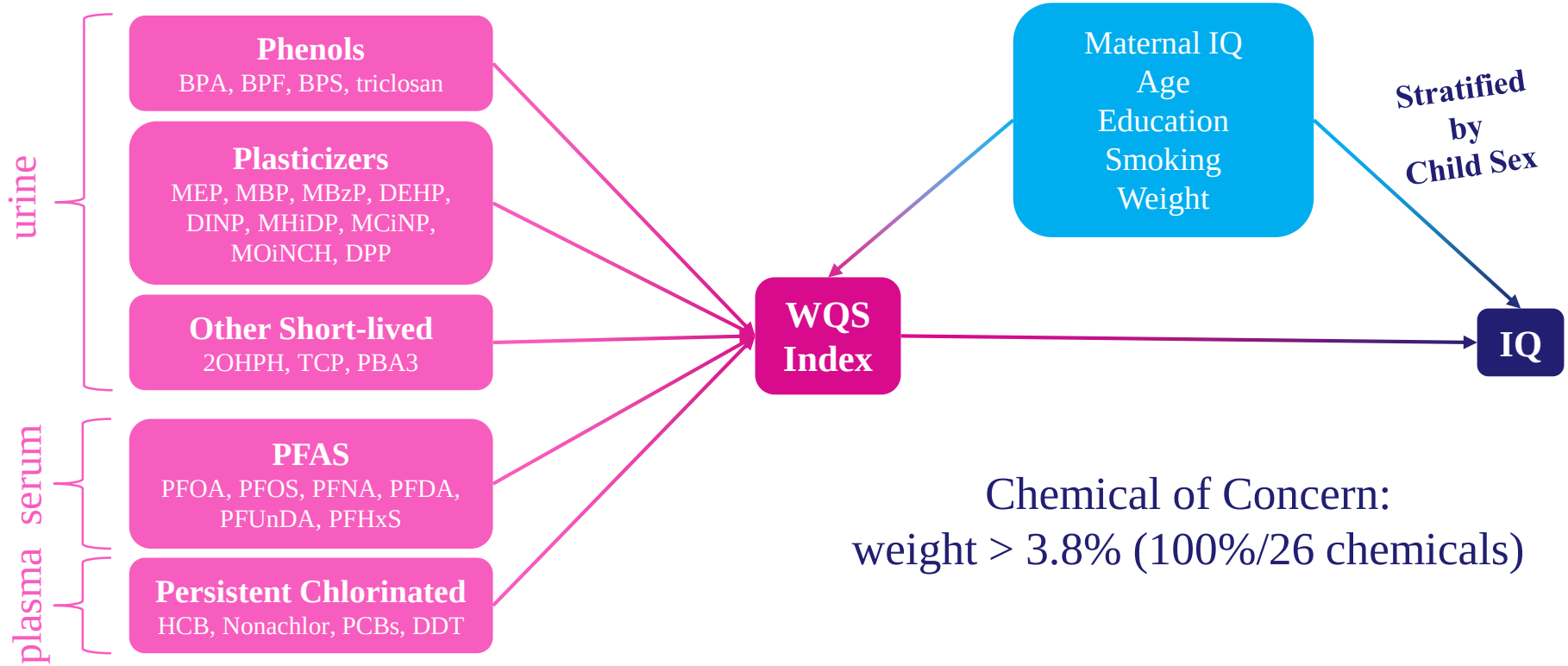
Week 10

Birth

7 years

Bornehag (2018) doi:10.1001/jamapediatrics.2018.3115

Evaluate EDCs in Relation to IQ using WQS Regression



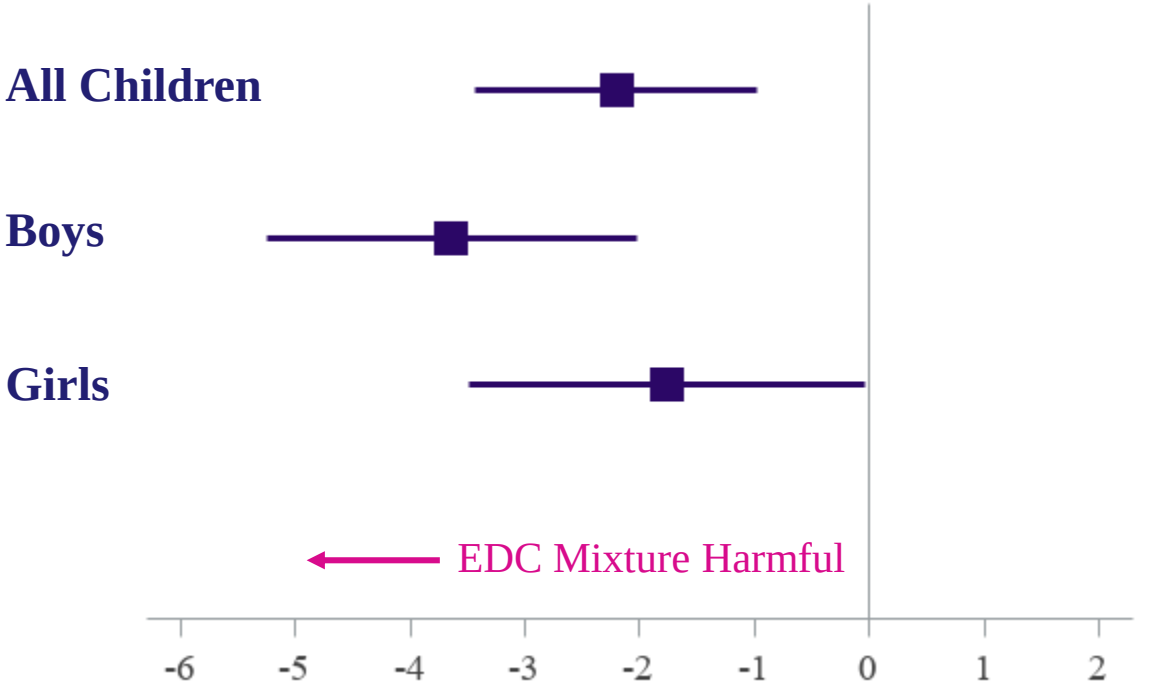
Chemical of Concern:
weight > 3.8% (100%/26 chemicals)

Analysis Challenges in SELMA

- Single Split
  - WQS Index β and selected chemicals changed depending on random seed
- No Split (training/testing on same data)
  - Stable estimates, but lacked rigor of validation step
- Repeated holdout a viable solution

Impact of Prenatal EDC Mixture on IQ at Age 7

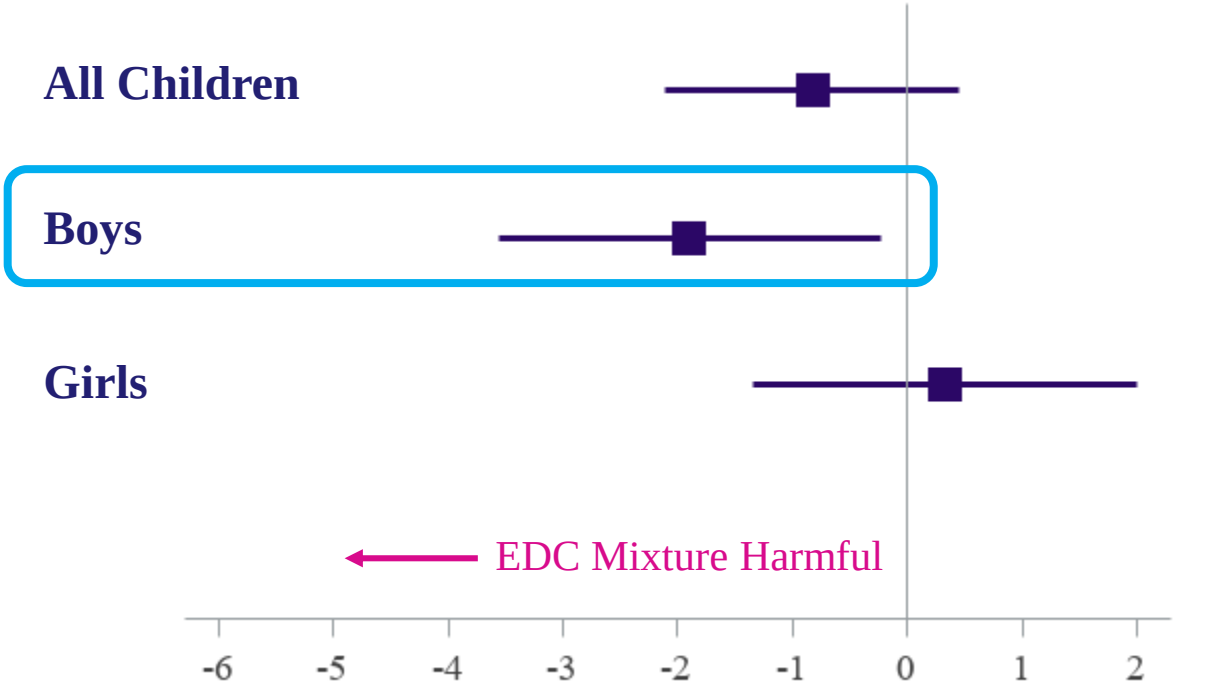
No Split



Difference in IQ Score per IQR Change in WQS Index

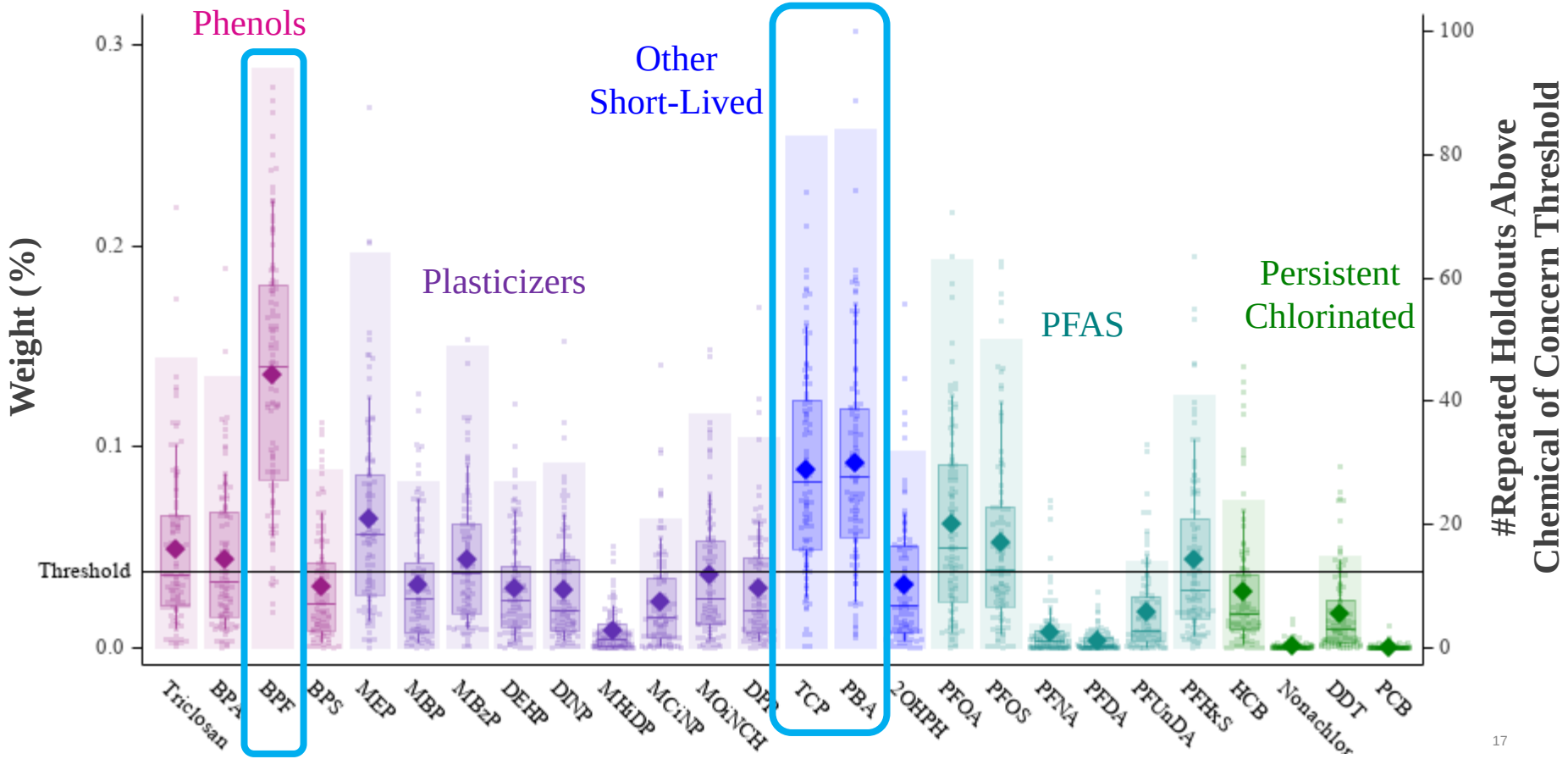
Impact of Prenatal EDC Mixture on IQ at Age 7

Repeated Holdout Validation



Difference in IQ Score per IQR Change in WQS Index

Chemicals of Concern Identification & Uncertainty: Boys



Summary

- WQS with single training-validation splits may lead to unrepresentative partitions & unstable results in finite samples
  - Test this by rerunning WQS with different random seed
- Training/testing on same data not necessarily wrong, but may reflect within-sample noise & results may not generalize
- Repeated Holdout Validation applied to WQS allows
  - Inference of WQS Index β
  - Characterizes weight uncertainty
- Number holdouts required depends on number needed to approximate a ~N sampling distribution in that sample

Questions?

Acknowledgements

- **Chris Gennings**
- **Carl-Gustaf Bornehag**
- **Co-authors:**
  - Maria Unenge Hallerbäck
  - Sverre Wikström
  - Christian Lindh
  - Hannu Kiviranta
- **Colleagues:**
  - Jonathan Heiss
  - Anu Joshi
- **SELMA participants**

- **Funding:**
  - EDC-MixRisk (#634880) European Union's Horizon 2020 Research and Innovation Programme
  - NIEHS PRIME: #R01ES028811-01

Repeated Holdout for WQS
R Tutorial
Available HERE