Icahn School of Medicine at Mount Sinai

A Novel LASSO type Bayesian Weighted Quantile Sum Regression for highly correlated multi-group mixture analysis:

current context, importance and future directions

Vishal Midya
Post Doctoral Fellow
Department of Environmental Medicine and Public Health
Icahn School of Medicine, Mount Sinai

Background

- **Weighted quantile sum (WQS)** approaches have been quite popular to model exposure mixtures.

- The underlying philosophy is that even when the individual concentrations of single exposures are all below the regulatory level, the cumulative exposure might be substantially higher or may be associated with health endpoints.

- Quite recently, a **Bayesian Hierarchical Model** (Colicino et al., 2020) was proposed as a counterpart of WQS. This model, called, **Bayesian Weighted quantile sum (BWQS)** regression enjoys the flexibility and stability of usual Bayesian procedures.

- There is also another **Bayesian Hierarchical Model** called **Bayesian Group Index Regression** (BGIR) by Wheeler et al., 2021, which is very similar to BWQS (some subtle differences in prior specification).

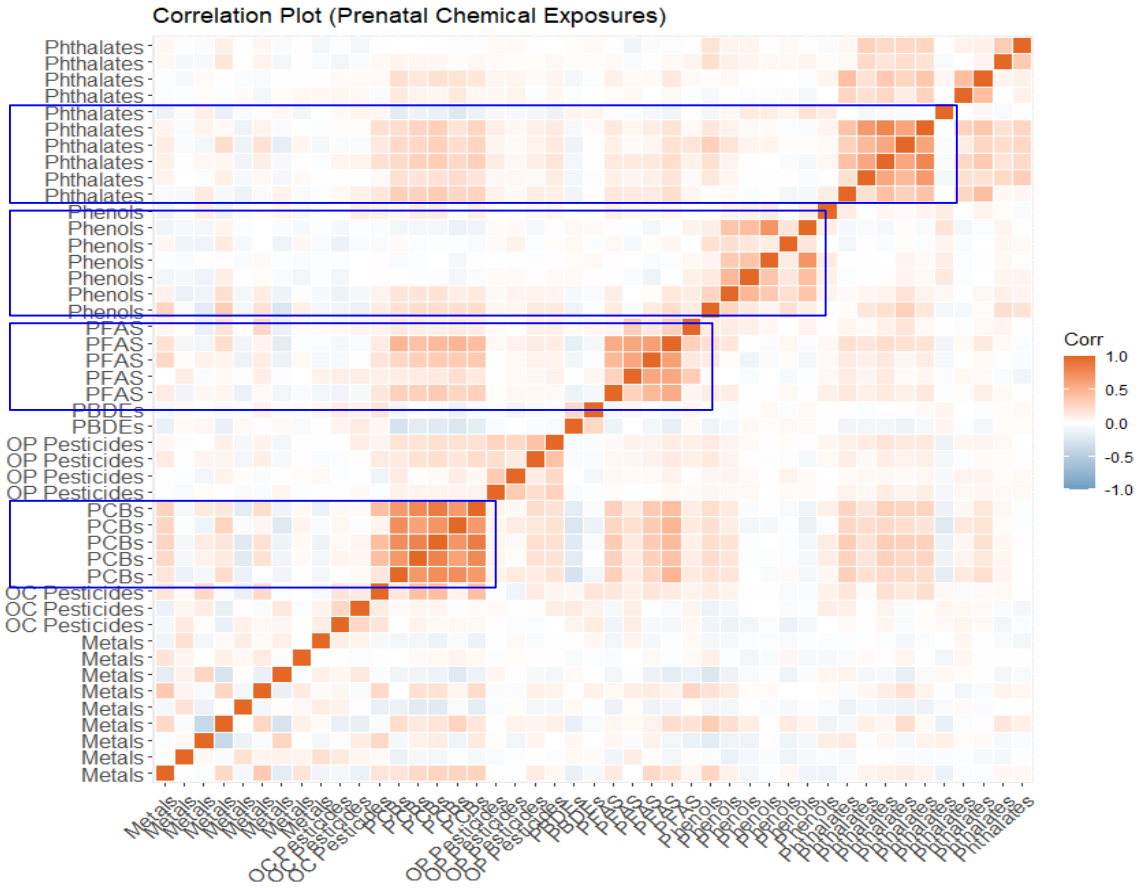
What's yet to be done

These Bayesian Hierarchical Models offer a lot of **flexibility** but that might lead to data being **severely overfitted**, particularly when the exposures are **highly correlated** in the **presence of multiple groups**

How to address this problem ?

Introduce statistical regularization or penalty terms

Motivation for this method



What's the Novelty?

- We propose adding **Penalty term or statistical regularization** through a **Bayesian LASSO type modelling approach** within the framework of BWQS for multi group mixture-outcome association.

- In particular, we implemented **Bayesian Group LASSO and Elastic Net** to model the **variances** of each group mixture association term.

- The groups can be designed to be **uncorrelated in** or **share covariance terms** depending on the regularization (**Bayesian Fused LASSO**).

What's the mathematical innovation here ?

Modelling the Laplace (double-exponential) distribution as a scale mixture of a normal distribution with an exponential mixing density (Kyung et al., 2010)

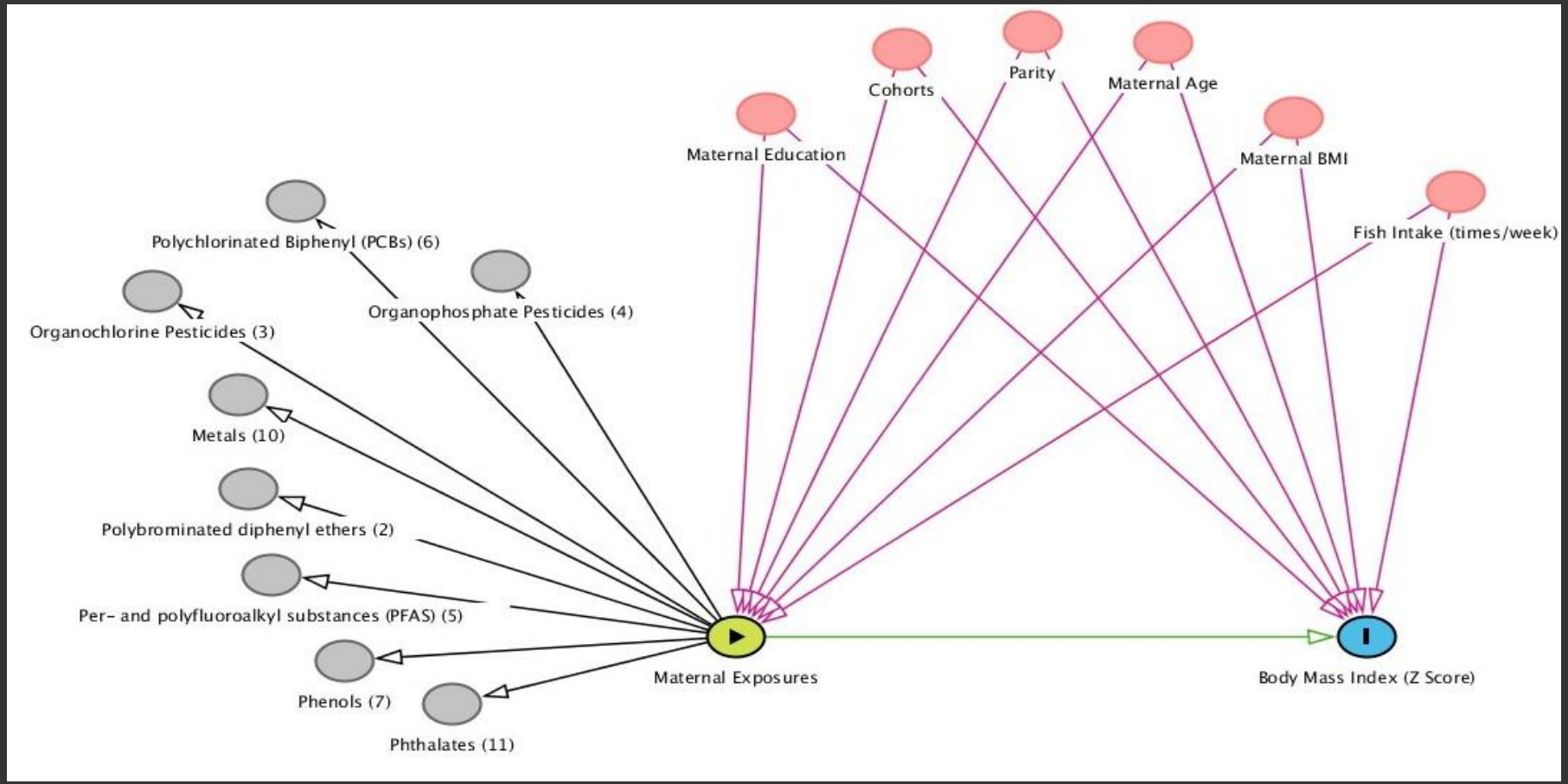
That's great but does this method work in real life ?

Yes it does!

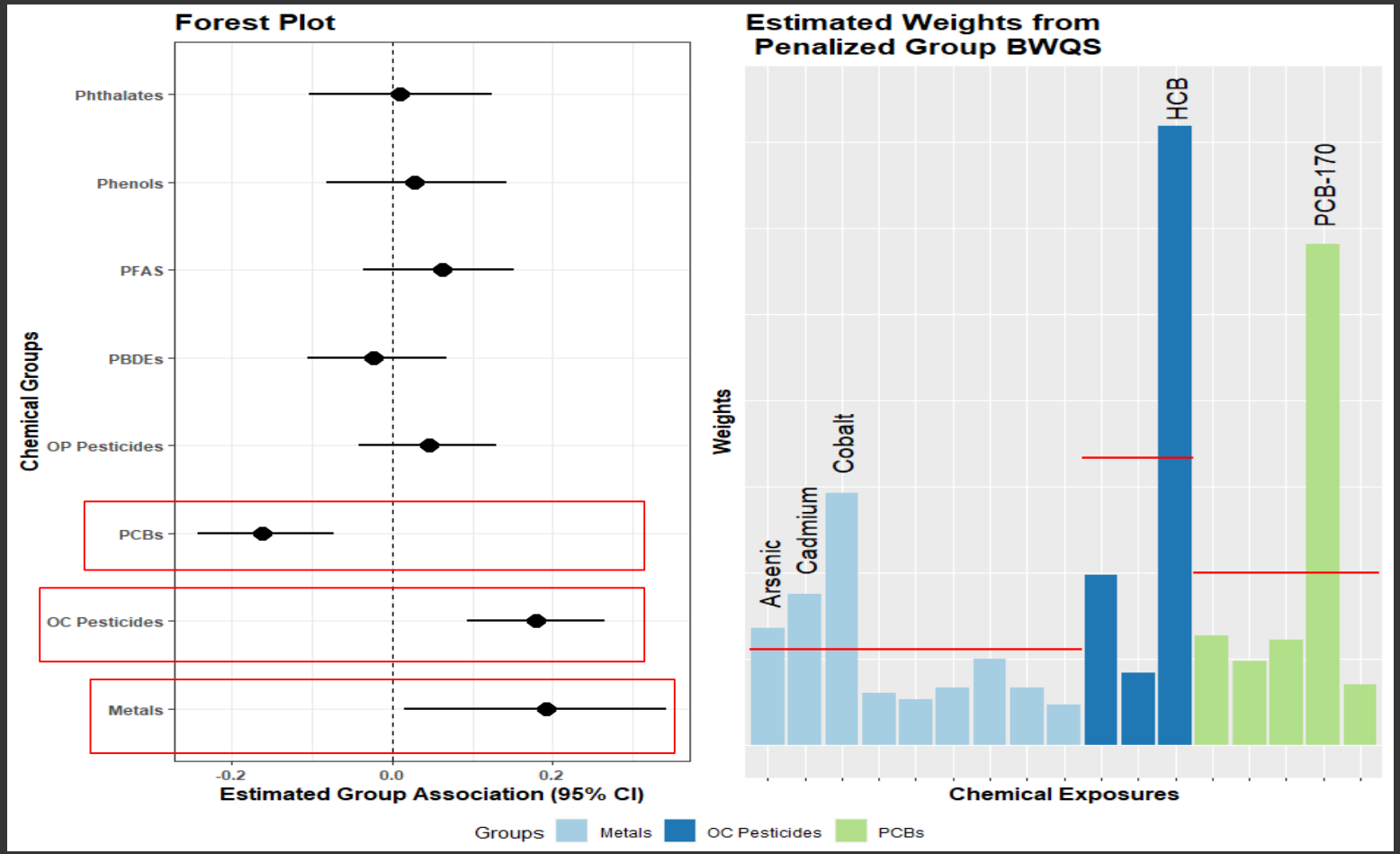
We tested the LASSO type Group BWQS method addressing the aim

To evaluate the association between prenatal environmental chemical exposures and child BMI at age ~8 years

Conceptual Framework using DAG



Estimated Group Associations and corresponding Weights



But are there other Bayesian methods to handle similar problems (multiple groups and regularization) ?

Yes!

Hierarchical Bayesian Kernel Machine Regression (BKMR)

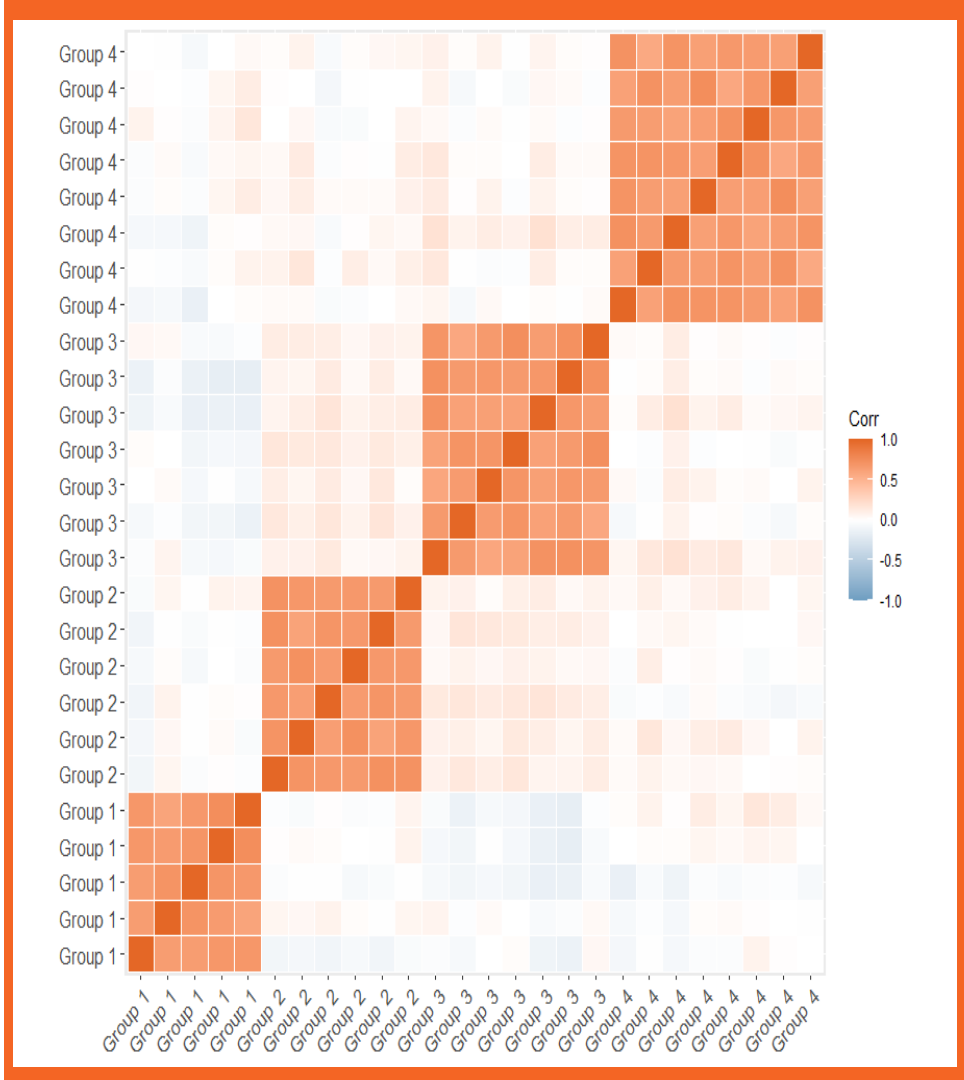
a hierarchical variable selection approach to identify important mixture components and account for the correlated structure of the mixture.

Then what are the potential advantages/limitations of using either method ?

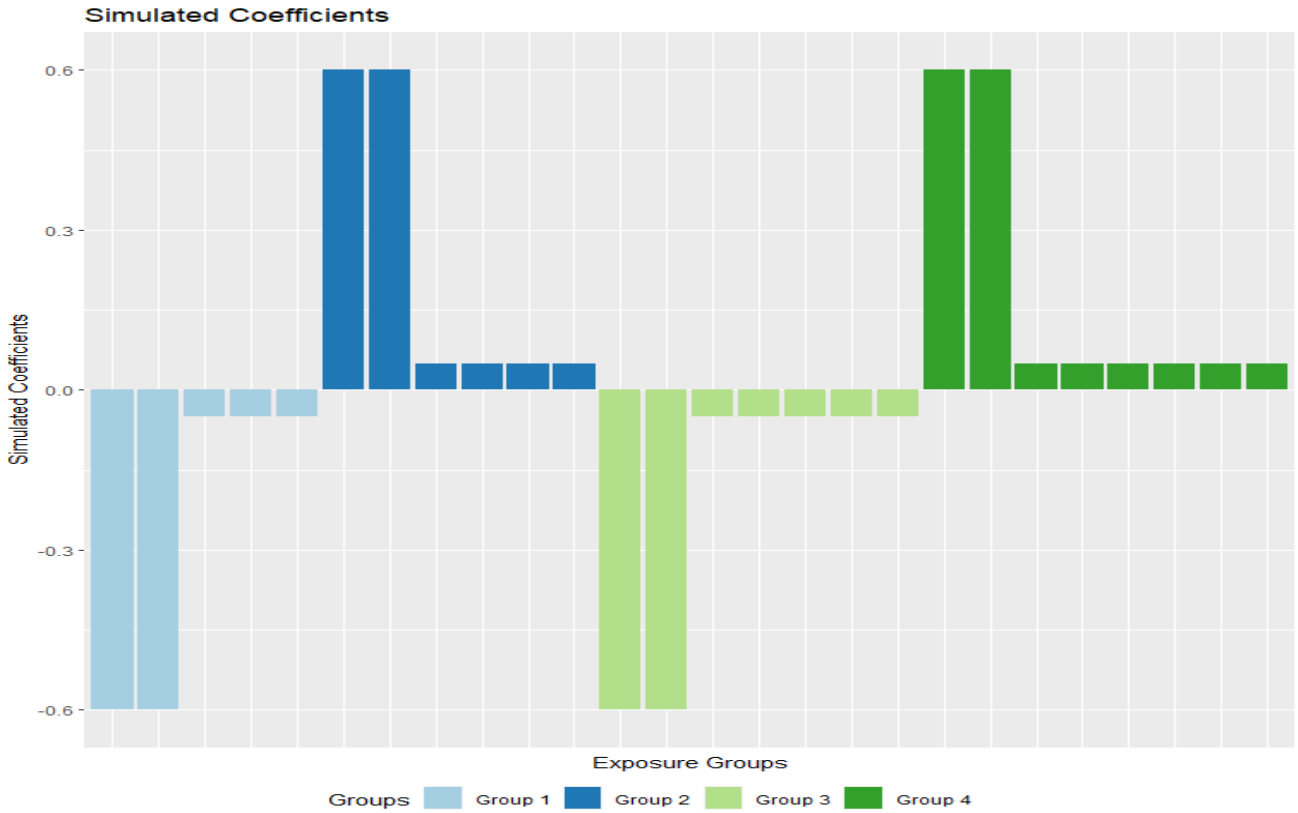
Let’s do some simulations!

Simulated Correlation Matrix for Exposures

The outcome is a linear combination of the Groups of exposures with pre-fixed coefficients

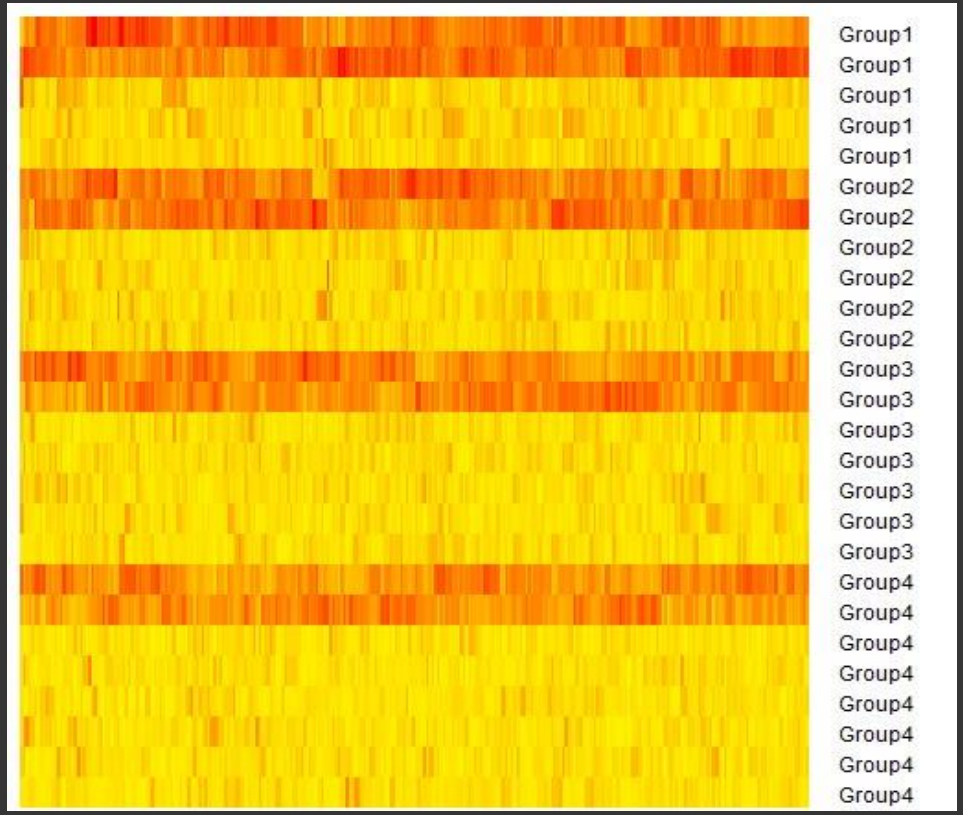


Scenario 1: Two Major Contributing Exposures in each group

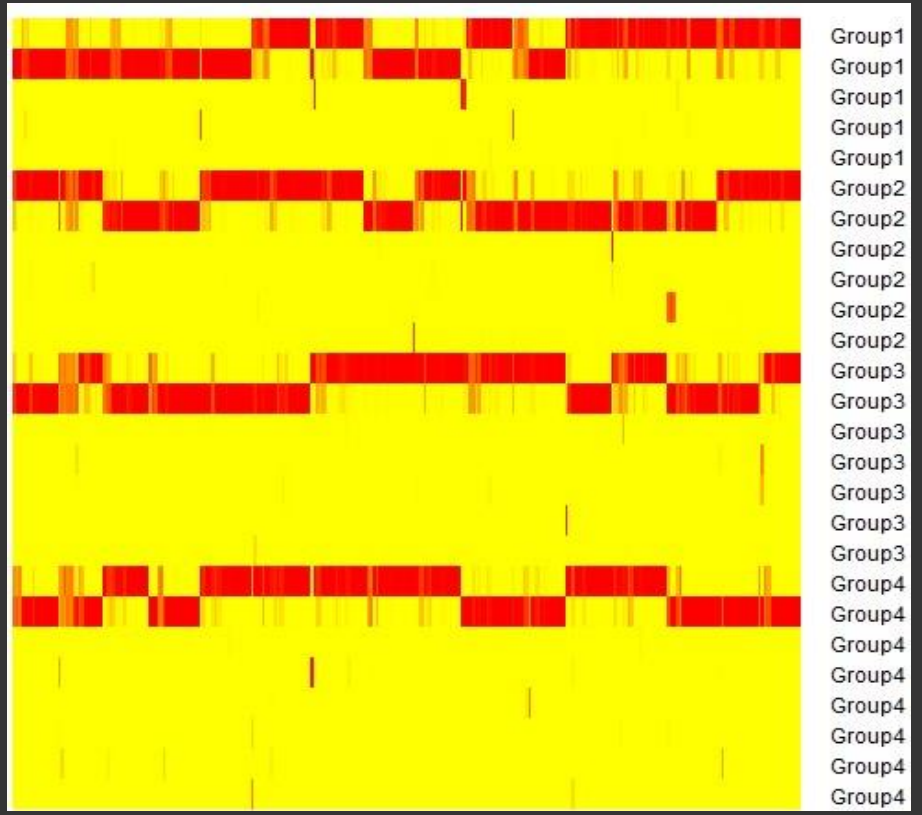


Heatmap of Estimated Weights

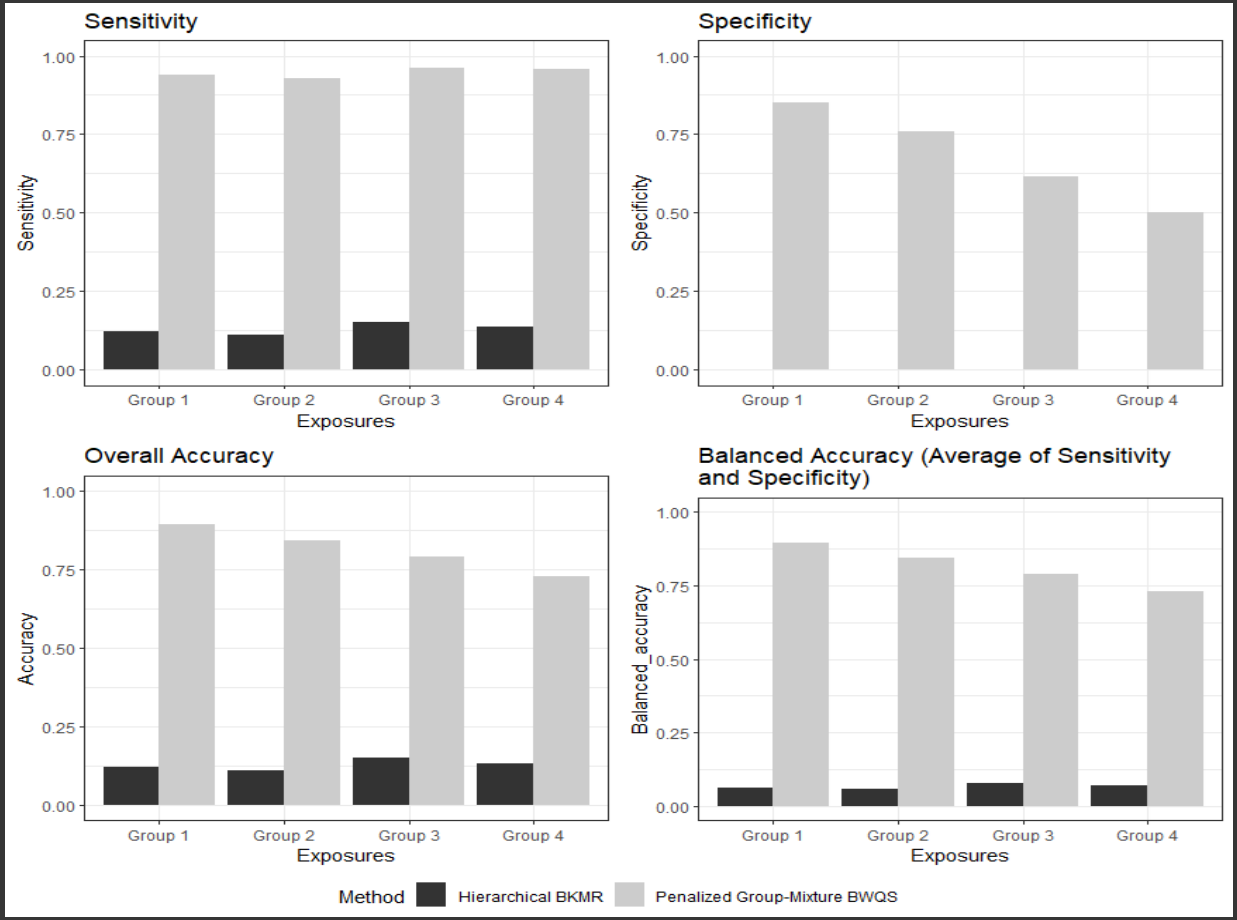
Penalized Group-Mixture BWQS



Hierarchical BKMR



Measures of the Performance in estimating correct weights



1) **True Positive:**

   Weights of ALL the major exposures > 1/group size

2) **True Negative:**

   0 < Weights of ALL the minor exposures < 1/group size

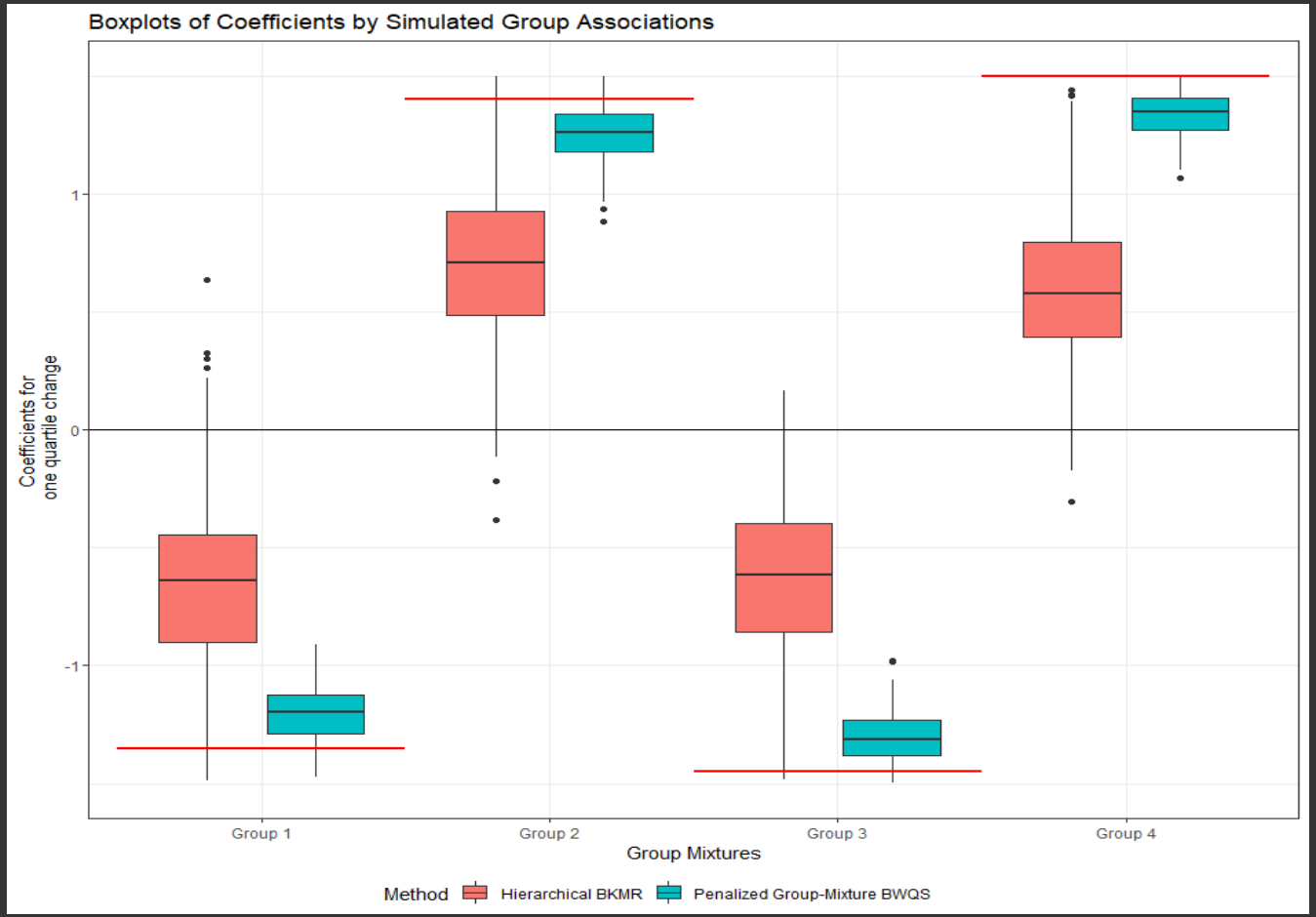
3) **False Positive:**

   Weight of AT LEAST ONE minor exposure > 1/group size

4) **False Negative:**

   Weight of AT LEAST ONE Major exposure < 1/group size

_ Performance in Estimating Group Association



Estimated Group associations for Group BWQS are more precise & less biased than Hierarchical BKMR

(May have important public health implications)

“It is important to always select the most appropriate statistical method for our research question so that we do not underestimate any true environmental health risk”

– Major Takeaways

1. Group BWQS is a useful approach for evaluating associations of multiple groups simultaneously **while taking into account overfitting and within group high correlation**

2. Group BWQS can be an advantageous method to use compared to other available methods (e.g., Hierarchical BKMR), especially when one suspects to have "**more than one OR no major contributing exposures within a group**"

3. Depending on the research question and the data structure, other exposure mixture approaches may be suitable (e.g., if there is only **one major contributing exposure** in a group, **both Group BWQS and Hierarchical BKMR may perform equivalently**). **But blind use of exposure mixture methods may lead to biased results!**

Future Directions : Using Group BWQS for **time varying data** (e.g., modelling prenatal exposures and post-natal exposures and adding **"interpretable" interaction terms**)

Acknowledgement

Icahn School of Medicine at Mount Sinai

Dr. Dania Valvi
Assistant Professor, Icahn School of Medicine at Mount Sinai, NY

Dr. Elena Colicino
Assistant Professor, Icahn School of Medicine at Mount Sinai, NY

Email : vishal.midya@mssm.edu