MR-PATH: A Latent Mixture Model for Heterogeneous Causal Mechanisms in Mendelian Randomization

Daniel Iong, Qingyuan Zhao, Yang Chen

Department of Statistics, University of Michigan, Ann Arbor and Statistical Laboratory, University of Cambridge

Background

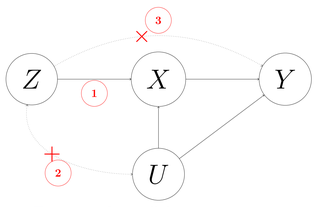


Figure: Core IV assumptions

- **Mendelian Randomization (MR)**: an instrumental variables (IV) method that uses genetic variants (Z) as instruments to estimate the causal effect of a modifiable risk exposure (X) on a disease outcome (Y) in the presence of confounders (U).
- Most robust MR methods rely on the **"effect homogeneity" assumption**: the risk exposure has the same causal effect for every individual. This assumption may be unrealistic when we use MR to study complex biological systems involving multiple mechanisms.

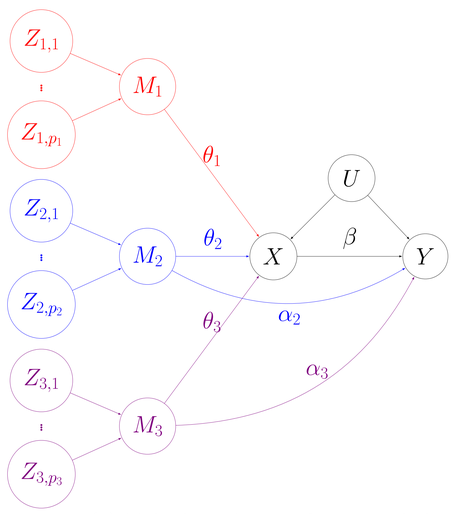
Our contributions

- **"Mechanistic heterogeneity"**: A novel concept we formalized to describe effect heterogeneity in MR due to multiple causal mechanisms.
- **MR-PATH**: A mixture model for mechanistic heterogeneity in MR.
- **Statistical inference for MR-PATH**: Monte-Carlo EM algorithm, approximate confidence intervals, modified BIC criterion

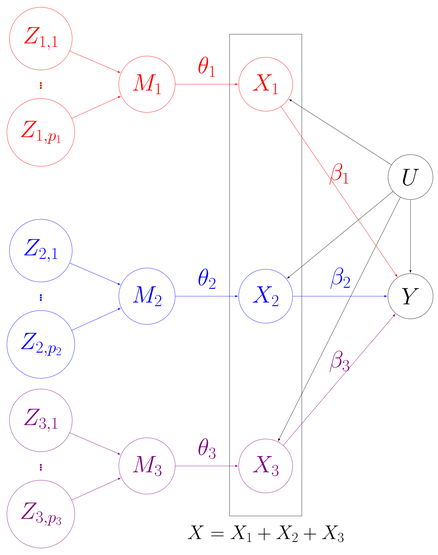
Contact Information

Email: daniong@umich.edu
Phone: (510) 816-8686

Mechanistic Heterogeneity in MR



(a) Multiple pathways of horizontal pleiotropy.



(b) Multiple mechanisms for the exposure X.

Model

$$\begin{pmatrix}\hat{\theta}_{X_i}\\ \hat{\theta}_{Y_i}\end{pmatrix} \overset{\text{indep.}}{\sim} N\left(\begin{pmatrix}\theta_{X_i}\\ \beta_i\theta_{X_i}\end{pmatrix}, \begin{pmatrix}\sigma^2_{X_i} & 0\\ 0 & \sigma^2_{Y_i}\end{pmatrix}\right) \quad i = 1, \ldots, p.$$

$$Z_i \sim \text{Categorical}\ (\pi_1, \ldots, \pi_K)$$
$$\beta_i | Z_i = k \sim N(\mu_k, \sigma_k^2), \quad k = 1, \ldots, K.$$

Observed Data

- $\hat{\theta}_{X_i}$,$\hat{\theta}_{Y_i}$: Observed SNP-exposure/SNP-outcome effects
- σ_{X_i}, σ_{Y_i}: Corresponding standard errors

Latent variables

- θ_{X_i}: True SNP-exposure effects
- β_i: SNP-specific causal effects

Parameters

- π_k: Proportion of SNPs in cluster K
- μ_k: Average causal effect of cluster K
- σ_k^2: Variance of cluster K

Statistical Inference

EM algorithm

- E-step is not tractable ⇒ approximate with importance sampling ⇒ **Monte-Carlo EM (MC-EM)**

Approximate Confidence Intervals

- Compute observed information matrix using Louis (1982) ⇒ invert to get standard errors.

Model Selection

- Modified BIC criterion from Ibrahim et al. (2008).

Software

The **MRPATH** R package is available at
`https://github.com/danieliong/MRPATH`

Preprint

A preprint of this work is available at
`https://arxiv.org/abs/2007.06476`

Application: Is HDL the good cholesterol?

- Observational studies have found a consistent **inverse association** between HDL cholesterol (HDL-C) and coronary heart disease (CHD)

  ⇒ **HDL hypothesis**: HDL protects from atherosclerosis. (*HDL is the "good" cholesterol.*)

- However, the HDL hypothesis has received scrutiny after
  - Several clinical trials raising HDL cholesterol showed, at best, modest cardiovascular benefit.
  - MR studies reached conflicting conclusions.
- **Our hypothesis**: There are multiple causal mechanisms between HDL-C and CHD.