Bayesian Mendelian Randomization Analysis for Latent Exposures Leveraging GWAS Summary Statistics for Traits Co-Regulated by the Exposures

Yue Yu[1,*]; Andrew Lakkis[1,*] ; Jin Jin[1]

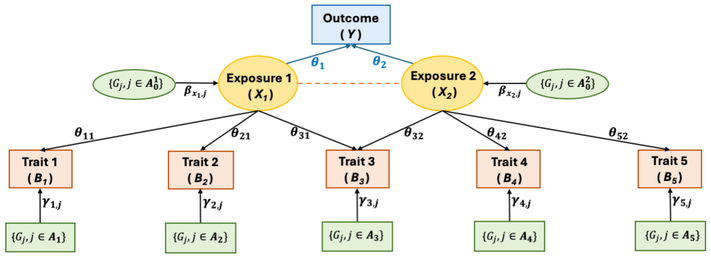
1. Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, USA; *These authors contributed equally

DEPARTMENT of BIOSTATISTICS EPIDEMIOLOGY & INFORMATICS

INTRODUCTION

We want to identify causal signals among a series of traits, but the exposures of interest (e.g., some biological functions or different disease subtypes) that regulate the observable traits are not directly observable.

In this study, we propose a Bayesian MR analysis framework, CaLMR, for jointly analyzing the direct effect of latent exposures on an outcome leveraging available GWAS summary-level association statistics for traits co-regulated by the exposures.



METHODS

Suppose the observed traits B_k *s,* latent exposures X_ls, and outcome Y are standardized, we parametrize three models

$$\text{Outcome Model: } Y = \sum_{l=1}^{L} \theta_l X_l + \epsilon_y, \quad \epsilon_y \sim N(0, \sigma^2_{\epsilon_y})$$

$$\text{Exposure Models: } X_l = \sum_{j=1}^{M} \beta_{x_l,j} G_j + \epsilon_{x_l}, \quad \epsilon_{x_l} \sim N(0, \sigma^2_{\epsilon_{x_l}})$$

$$\text{Biomarker Models: } B_k = \sum_{l=1}^{L} \theta_{kl} X_l + \sum_{j=1}^{M} \gamma_{k,j} G_j + \epsilon_{B_k}, \quad \epsilon_{B_k} \sim N(0, \sigma^2_{\epsilon_{B_k}})$$

where $\beta_{x_l,j} \sim N\left(0, h^2_{x_l}\right)$ and $\gamma_{k,j} \sim N\left(0, h^2_{\gamma_k}\right)$ are effects of the genetic IVs on the exposure X_l and observed trait B_k, respectively.

We constructed a MCMC algorithm to generate posterior samples based on the Regression with Summary Statistics (RSS) Likelihood and conjugate priors.

We conducted multi-exposure simulation studies to compare the proposed CaLMR (Multi) method with (1) CaLMR (Uni) (its single-exposure version), (2) MRLE (the only existing method designed for conducting MR for latent exposures), (3) MVMR PRESSO, and (4) MVMR IVW (multiple-exposure versions for MR PRESSO and IVW).

We then analyzed the causal effects of 4 factors of psychiatric disorders (compulsive, psychotic, neurodevelopmental, internalizing) on 18 disease outcomes using the five methods.

SIMULATION STUDIES

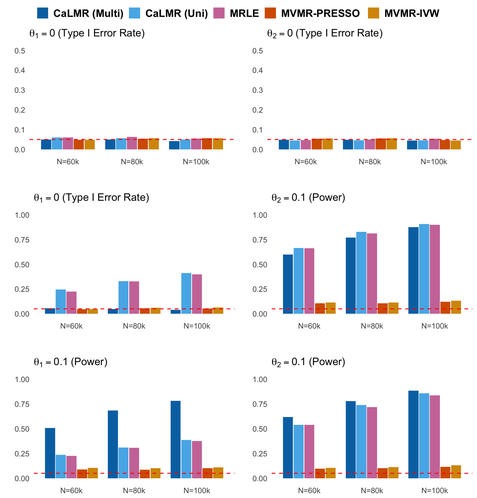
Basic Settings:

- L = 2 latent exposures, $cor(X_1, X_2) = -0.5$, K = 8 traits
- GWAS sample size $N_{B_k} = N_Y$ = 60k, 80k, 100k
- Assume K biomarkers and outcome Y come from distinct samples:
  $cor(B_n, B_m) = 0, n \neq m;\ cor(B_k, Y) = 0, k = 1, \ldots, K$
- $(\theta_1, \theta_2) = (0,0), (0,0.1), (0.1,0.1); \theta_{kl} = \sqrt{0.3}$

IV Selection:

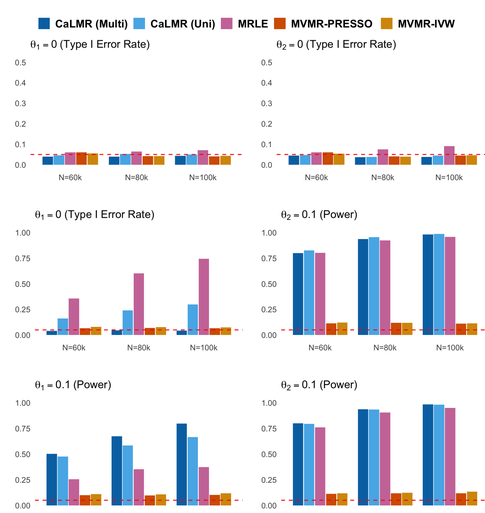
- Union of the SNPs that reach the significance threshold for at least two B_ks

➢ **Scenario One**: each exposure regulates 4 traits/biomarkers



Simulation – Scenario 1

➢ **Scenario Two**: each exposure regulates 5 traits/biomarkers, i.e., two biomarkers are regulated by two exposures simultaneously

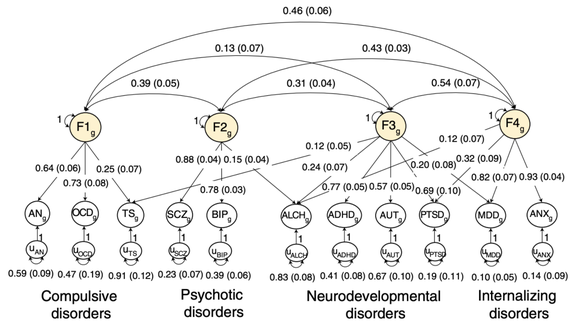


Simulation – Scenario 2

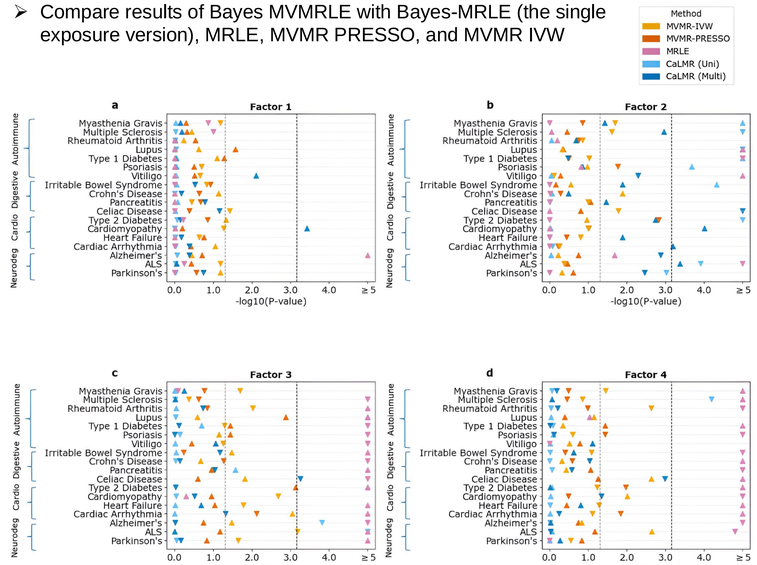
MR ON PSYCHIATRIC DISORDERS

We treated the 11 psychiatric disorders analyzed in stratified GenomicSEM (Grotzinger, 2022) as observed traits, and the four main factor groups they classified as the latent exposures in our models:

- Factor 1 – Compulsive (ANX, OCD, TS)
- Factor 2 – Psychotic (SCZ, BIP, ALCH)
- Factor 3 – Neurodevelopmental (TS, ALCH, ADHD, AUT, PTSD, MDD)
- Factor 4 – Internalizing (ALCH, PTSD, MDD, ANX)



➢ Compare results of Bayes MVMRLE with Bayes-MRLE (the single exposure version), MRLE, MVMR PRESSO, and MVMR IVW



CONCLUSIONS

- CaLMR outperforms the existing methods in terms of Type-I error control and power for detecting the direct effect of exposures in the presence of other correlated latent exposures.
- The R package “CaLMR” is available at https://github.com/yueuuy/CaLMR.