Mendelian Randomization Accounting for Correlated and Uncorrelated Pleiotropy Using Genome Wide Summary Statistics

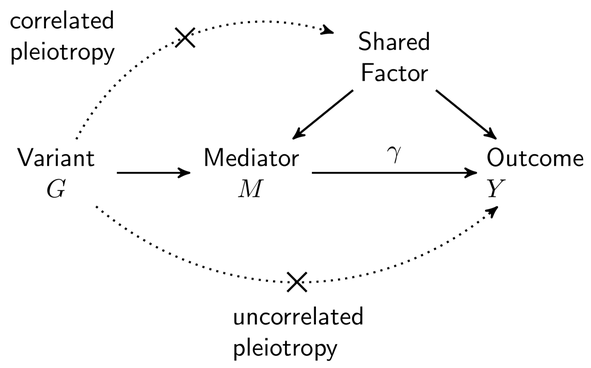
Jean Morrison, Nicholas Knoblauch, Joseph Marcus, Matthew Stephens and Xin He
University of Chicago, Department of Human Genetics

Mendelian Randomization

Mendelian randomization (MR) uses genetic variants as instrumental variables to estimate the causal effect of a mediator on an outcome.

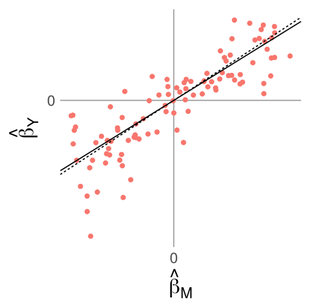
Most MR methods rely on strong assumptions:

1. Variants included must be causally associated with M.
2. Variants do not exhibit horizontal pleiotropy: Effects on Y that are not mediated by M.
   - Correlated pleiotropy occurs when G affects a factor affecting both M and Y.
   - Uncorrelated pleiotropy occurs when G affects Y through separate mechanisms.



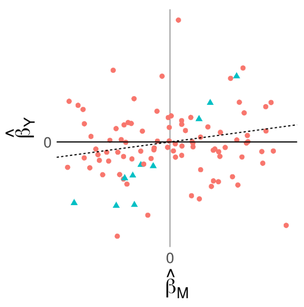
Correlated Pleiotropy Leads to False Positives

If MR assumptions hold, γ can be estimated by regressing variant associations with M on variant associations with Y.



Simulated effect estimates with a true causal effect of M on Y. The regression slope estimates the causal effect.

Even a small number of variants associated with a shared heritable factor can create a false positive.



Simulated effect estimates with no causal effect. 10% of variants exhibit correlated pleiotropy resulting in an IVW estimate significantly different from zero.

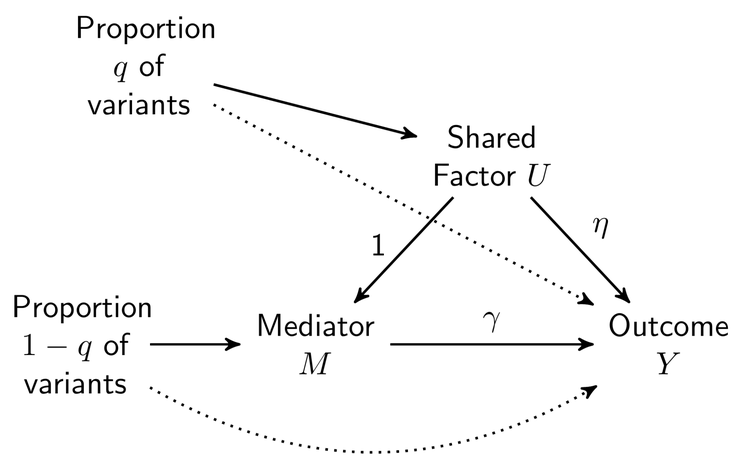
Detecting Causal Signatures by Modeling an Unobserved Shared Factor
CAUSE: Causal Analysis Using Summary Statistics

Model Features and Assumptions:

- Some variants affect the mediator through an unobserved heritable shared factor U.
- The proportion of variants acting through U is small.
- All variants may have additional mean zero pleiotropic effects on Y.
- Effects are assumed sparse and all variants are included.

Parameters:

- q: Proportion of variants that act through the shared factor
- γ: Causal effect of M on Y
- η: Shared factor effect of U on Y



Estimation

- The model defines a likelihood for summary statistics.
- Effect prior distributions are estimated empirically.
- Posterior distributions for q, γ, and η are obtained from the prior and likelihood.

Testing:

- Fit two nested models:
  - **Sharing Model:** γ is fixed at zero
  - **Causal Model:** γ can be non-zero
- Compare fit of posterior distributions using ELPD.
- If the Causal Model is a significantly better fit, the data are consistent with a causal effect.
- Otherwise, the data can be adequately explained with only a shared factor.

Examples

LDL Cholesterol → Coronary Artery Disease

P-value comparing causal and sharing models: 6.1×10^{-11}
Posterior medians and 95% credible intervals:

Model	γ	η	q
Sharing		0.40 (0.33, 0.47)	0.79 (0.66, 0.89)
Causal	0.36 (0.30, 0.42)	-0.01 (-0.67, 0.57)	0.03 (0.00, 0.25)

The data are consistent with a causal effect of LDL cholesterol on risk of CAD.

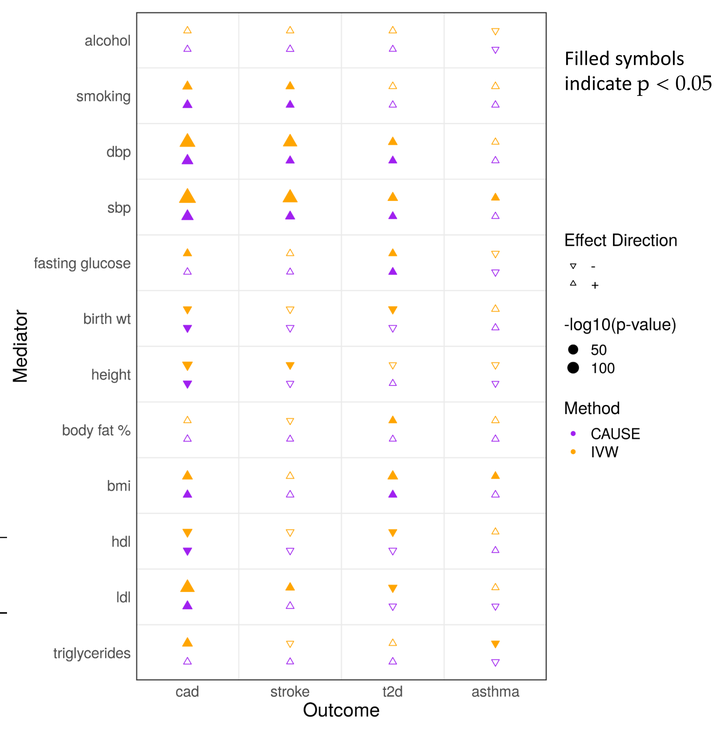
Birth Weight → Type 2 Diabetes

P-value comparing causal and sharing models: 0.07
Posterior medians and 95% credible intervals:

Model	γ	η	q
Sharing		-0.98 (-3.13, 0.00)	0.12 (0.01, 0.39)
Causal	-0.27 (-0.47, -0.08)	-0.03 (-2.30, 2.10)	0.04 (0.00, 0.26)

The data are not consistent with a causal effect but there is genetic sharing that leads to significant genetic correlation.

Anthropometric and Metabolic GWAS Traits

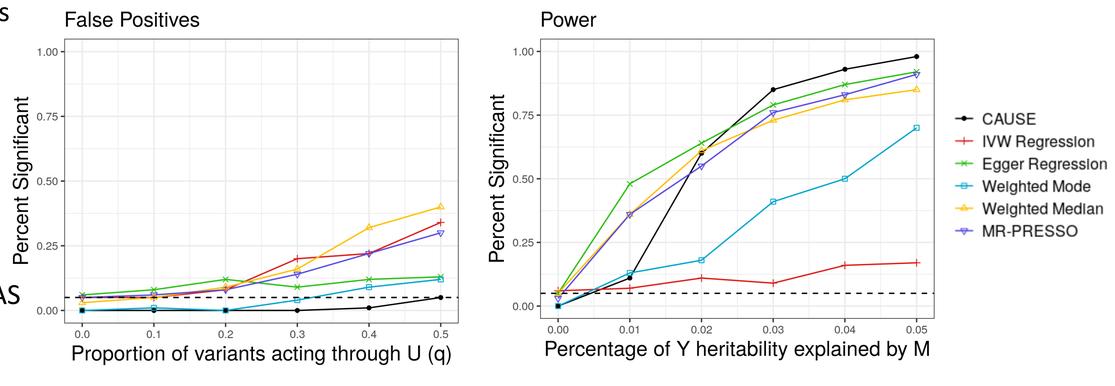


- CAUSE rejects the sharing model for plausibly causal trait pairs: LDL → CAD, BMI → CAD, Blood pressure → Stroke
- IVW identifies several pairs that are likely false positives: BMI → Asthma, Triglycerides → Asthma
- Both methods falsely identify HDL → CAD. These traits have a very high proportion of shared variants that is beyond what CAUSE can accommodate.

Learn More, Use CAUSE!
Pre-Print: BioRxiv 682237
R package: github.com/jean997/cause
Website: jean997.github.io/cause/
Contact: jeanm@uchicago.edu

CAUSE is Robust to Confounding While Retaining Power in Simulations

- Summary statistics simulated with LD patterns estimated from 1K Genomes.
- With small amounts of correlated pleiotropy, CAUSE controls the false positive rate below the nominal level.
- CAUSE has a better false positive rate than other methods for all levels of correlated pleiotropy.
- CAUSE has the best advantage when the GWAS for trait M has low power (results shown).
- Check the website for full simulation results!



References:
MR Review, IVW: Burgess, S., Dudbridge, F., & Thompson, S. G. (2016). *Stat Med*, *35*(11), 1880–1906.
ELPD: Vehtari, A., Gelman, A. & Gabry, *J Stat Comput* (2017) 27: 1413
Egger Regression: Bowden, J., Smith, G. D., & Burgess, S. (2015). *Int J Epidemiol*, (June), 512–525.
MR-PRESSO: Verbanck, M., Chen, C., Neale, B., & Do, R. (2018). *Nat Gen*, *50*(May).
Weighted Median: Bowden, J., Fabiola Del Greco, M., Minelli, C., Smith, G. D., Sheehan, N. A., & Thompson, J. R. (2016). *Int J Epidemiol*, *45*(6), 1961–1974.
Weighted Mode: Hartwig, F. P., Smith, G. D., & Bowden, J. (2017). *Int J Epidemiol*, *46*(6), 1985–1998.