A Bayesian framework to address limitations of Mendelian Randomization

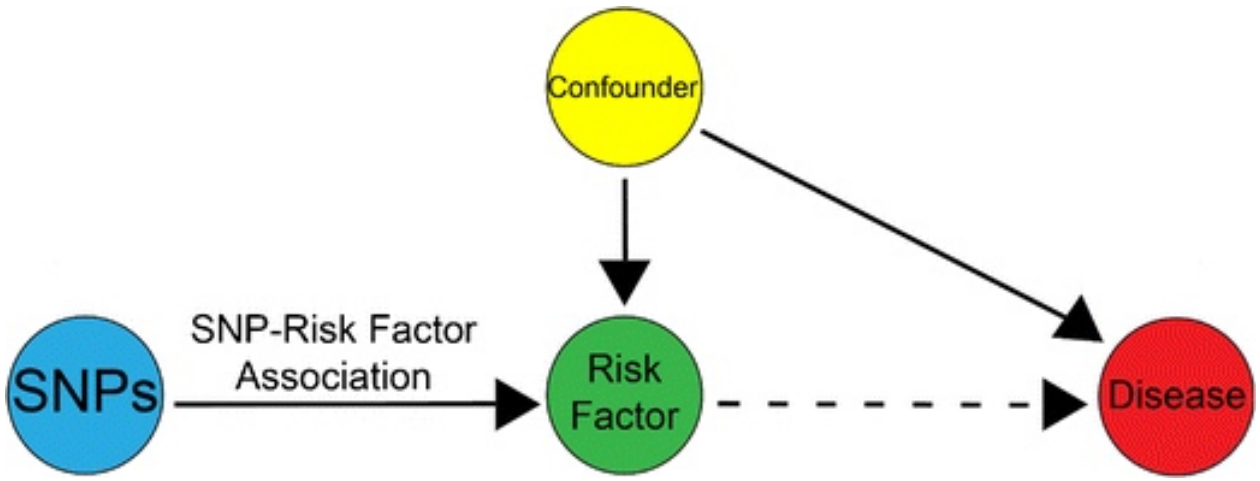
Xin He
Department of Human Genetics
University of Chicago

February 5, 2019

Inferring trait relations from GWAS

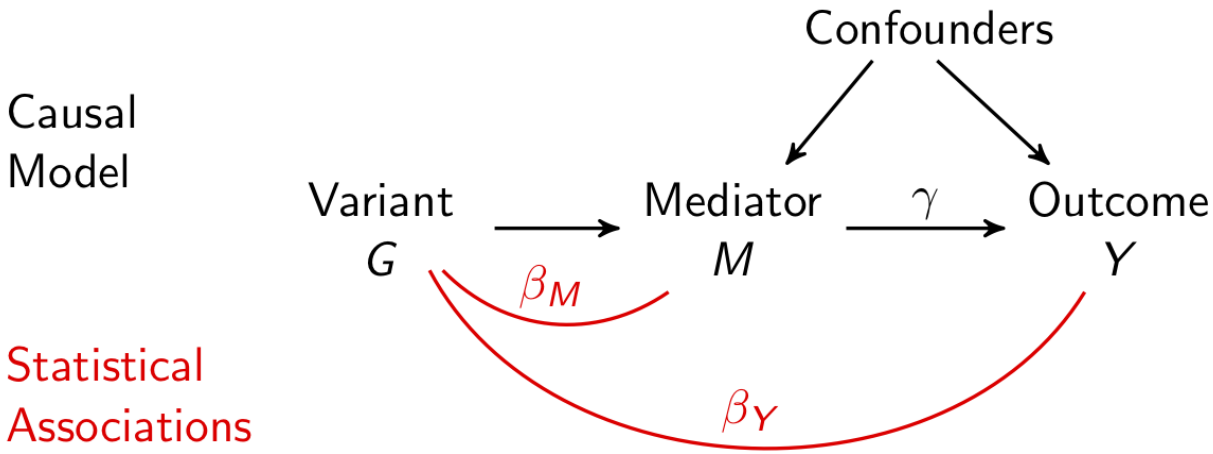
GWAS of thousands of phenotypes and even more molecular and cellular level traits have been performed. Can we use them to infer relationship among traits?

Mendelian Randomization (MR) is a general framework of causal inference



MR is similar to Randomized Clinical Trial (RCT): the SNP acts to randomize samples. We compare the disease risks in the two groups defined by the two alleles.

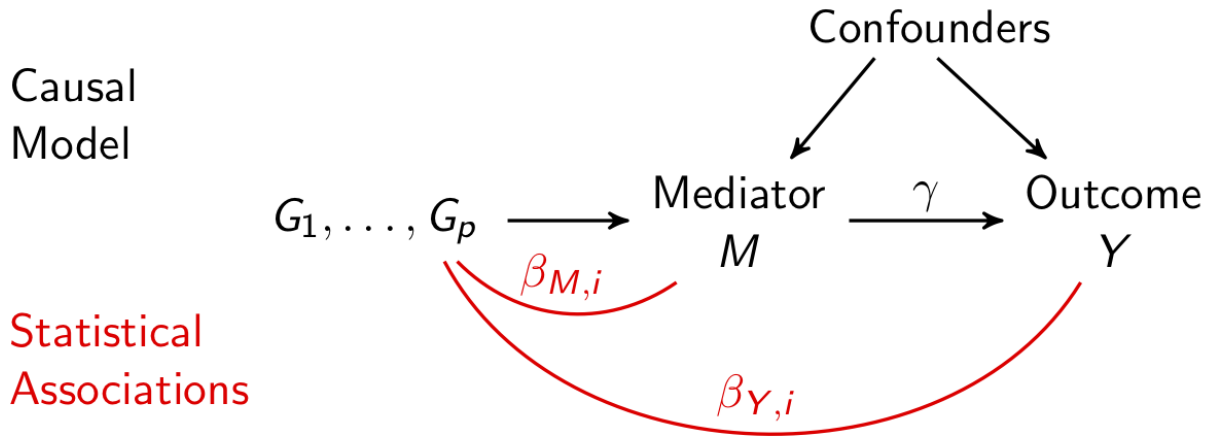
MR with single Instrument Variable (IV)



Estimator of causal effect γ:

$$\beta_Y = \gamma\beta_m \Rightarrow \hat{\gamma} = \frac{\hat{\beta_Y}}{\hat{\beta_M}}$$

MR with multiple IVs

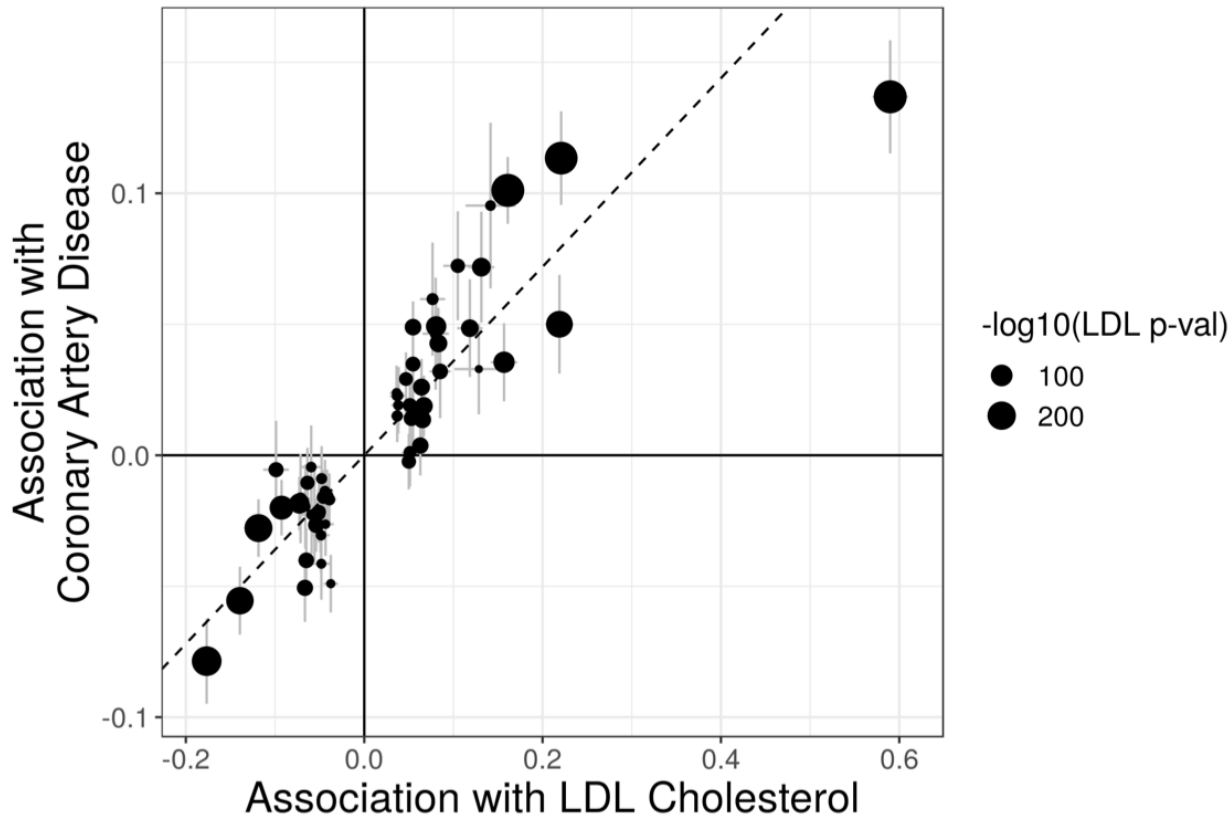


To estimate γ, regression of β_Y against β_M:

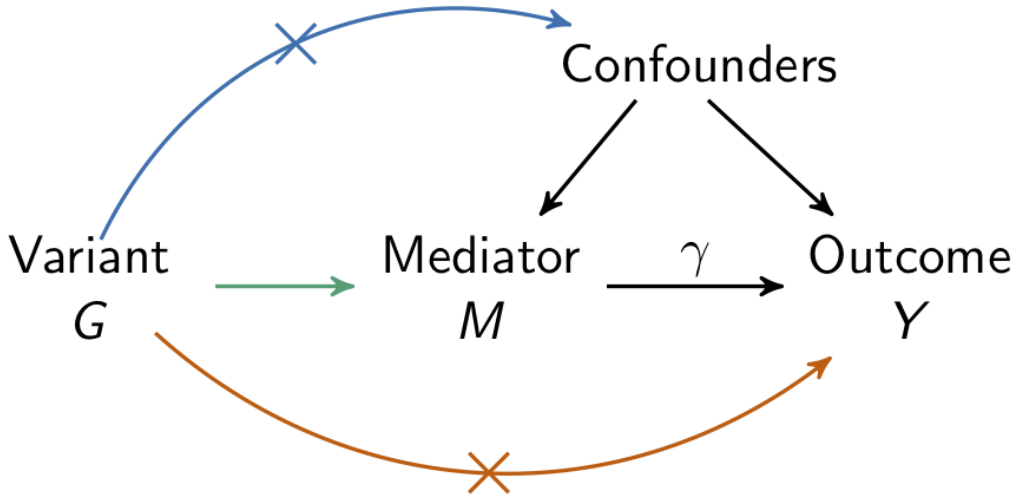
$$\hat{\beta}_{Y,i} = \gamma \hat{\beta}_{M,i}$$

Often the contribution of SNPs are weighted: IVW estimator.

Example of MR analysis: LDL → Coronary Artery Disease

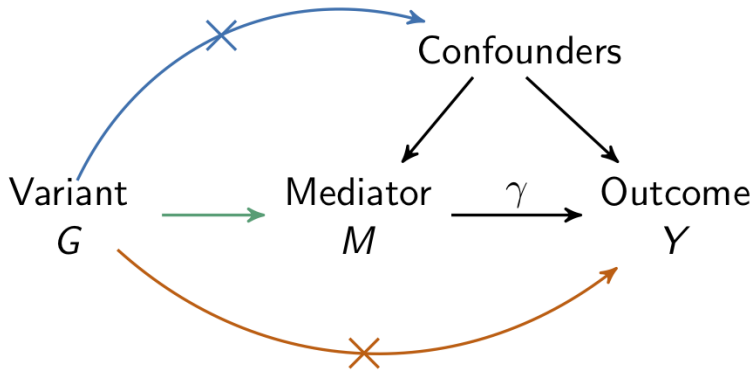


MR makes strong assumptions



- G causally affects M.
- G does not affect confounders.
- G does not affect Y through any other pathways.

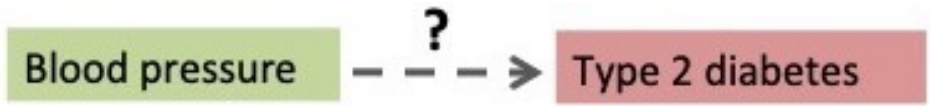
Problems of the three assumptions



- There is often uncertainty of identifying variants acting on M.
- A confounder may be a (heritable) biological process that acts on both M and Y.
- Horizontal pleiotropy can be common.

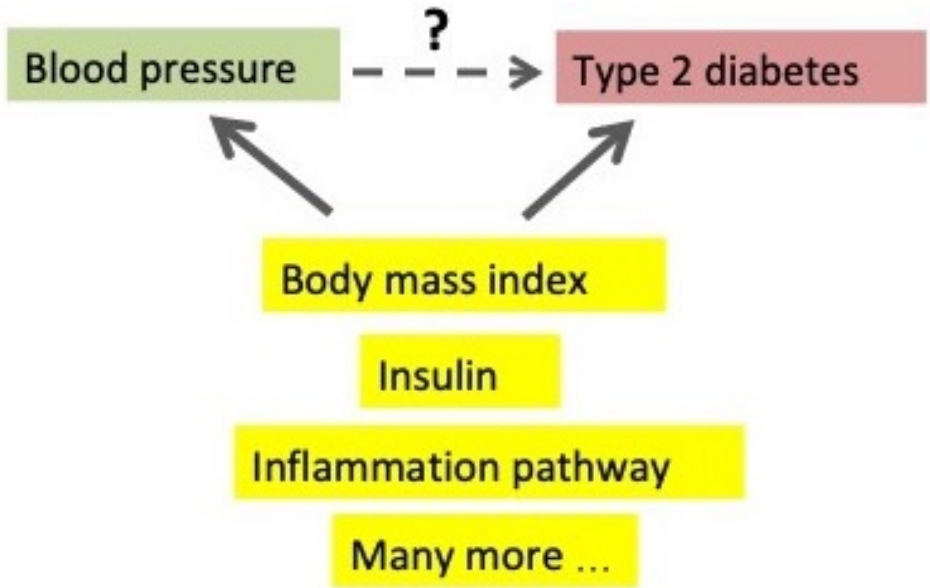
Heritable confounders lead to violation of MR assumption

Often MR researchers assume confounders are related to environmental exposure. However, confounders can be biological factors with genetic basis.



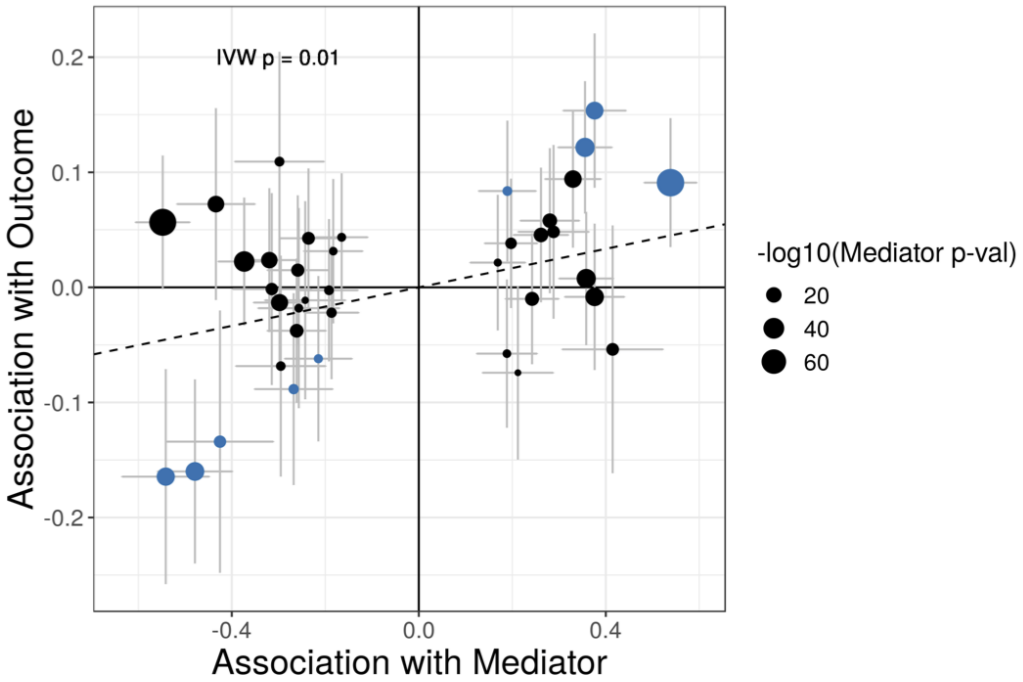
Heritable confounders lead to violation of MR assumption

Often MR researchers assume confounders are related to environmental exposure. However, confounders can be biological factors with genetic basis.



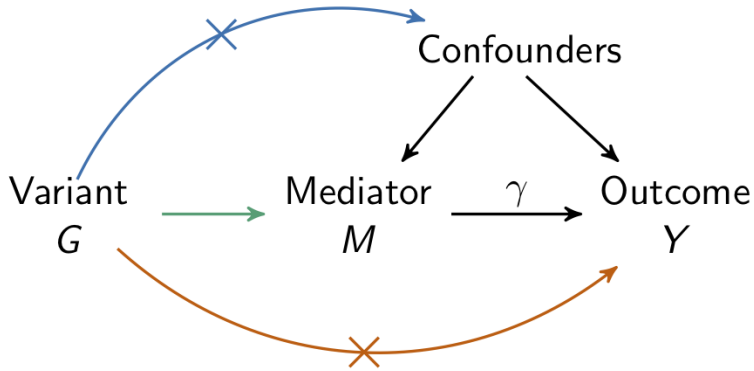
Variants of blood pressure are not valid IVs if they act on any of these processes.

A small number of invalid IVs driven by a confounder can lead to false findings by MR



Only 15% of shared SNPs can drive significant correlation in MR.

A unified model that addresses all three limitations



- Using genomewide summary statistics and model the uncertainty of IV: use a sparse prior for the true effects of a variant on mediator.
- Allow some variants of M to act on a shared factor (confounder).
- Introducing a sparse random effect of IV on Y not explained by the mediator or the shared factor.

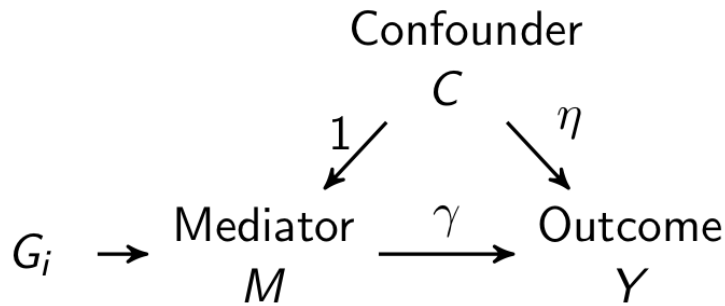
Modeling confounder in a non-causal model

Even when M does not causally act on Y, it is possible that some shared factor acts on both. We allow a small percent of variants of M to act on this shared factor.

Modeling shared factor as a mixture of genetic mechanisms

For generality, we allow both causal effect and a shared factor between M and Y:

G_i acts directly on M
probability $1 - q$

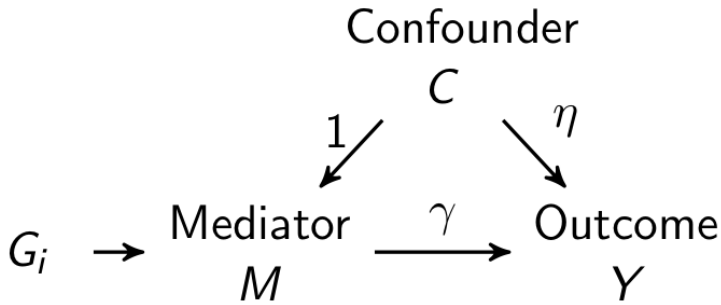


$$\beta_{Y,i} = \gamma \beta_{M,i}$$

Modeling shared factor as a mixture of genetic mechanisms

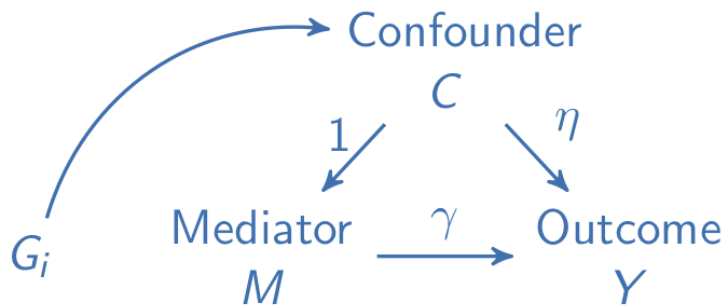
For generality, we allow both causal effect and a shared factor between M and Y:

G_i acts directly on M
probability $1 - q$



$$\beta_{Y,i} = \gamma\beta_{M,i}$$

G_i acts on M through confounder
probability q

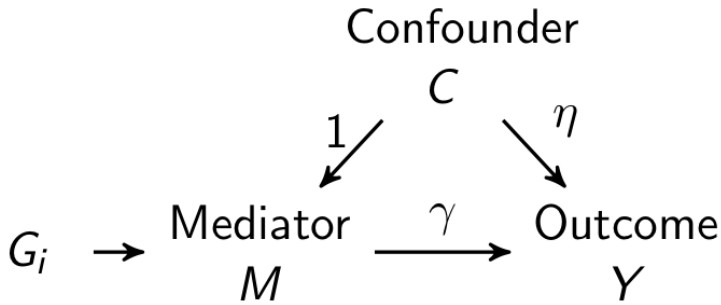


$$\beta_{Y,i} = (\gamma + \eta)\beta_{M,i}$$

Modeling shared factor as a mixture of genetic mechanisms

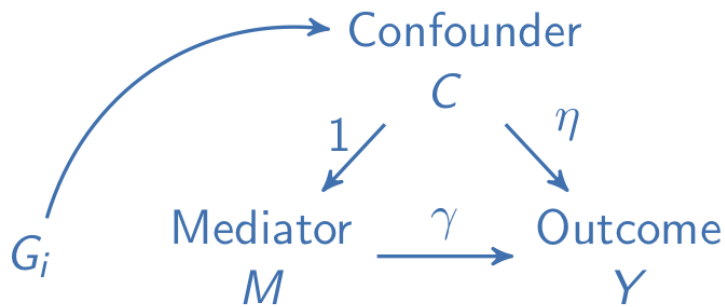
For generality, we allow both causal effect and a shared factor between M and Y:

G_i acts directly on M
probability $1 - q$



$$\beta_{Y,i} = \gamma\beta_{M,i}$$

G_i acts on M through confounder
probability q



$$\beta_{Y,i} = (\gamma + \eta)\beta_{M,i}$$

Whether variant i acts on the shared factor or not is denoted as Z_i, and $Z_i \sim \text{Bern}(q)$,

Modeling uncertainty of instruments

Our data $\hat{\beta}_{M,i}$ and $\hat{\beta}_{Y,i}$ are related to the true effects by:

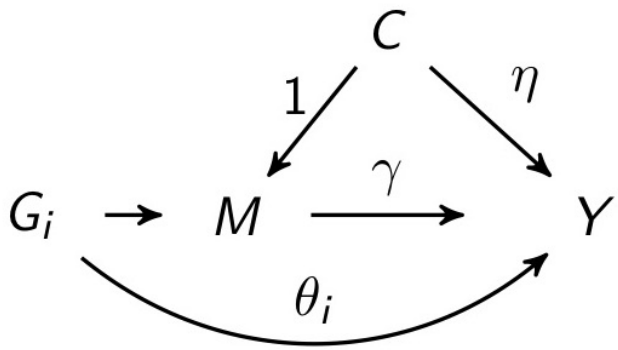
$$\hat{\beta}_{M,i} \sim N(\beta_{M,i}, s^2_{M,i}) \qquad \hat{\beta}_{Y,i} \sim N(\beta_{Y,i}, s^2_{Y,i})$$

The prior of true effect of variant i on M:

$$\beta_{M,i} \sim \pi_{M,0}\delta_0 + \sum_{k=1}^{K_M} \pi_{M,k} N(0, \sigma^2_{M,k})$$

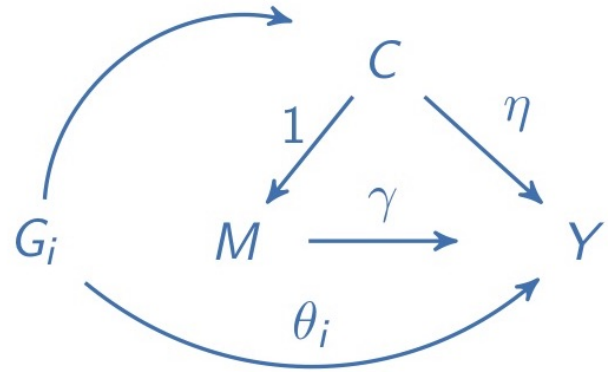
Modeling horizontal pleiotropy

G_i acts directly on M
With probability $1 - q$



$$\beta_{Y,i} = \gamma\beta_{M,i} + \theta_i$$

G_i acts on M through confounder
With probability q



$$\beta_{Y,i} = (\gamma + \eta)\beta_{M,i} + \theta_i$$

Horizontal pleiotropy: $\theta_i \sim \pi_{Y,0}\delta_0 + \sum_{k=1}^{K_Y} \pi_{Y,k} N(0, \sigma^2_{Y,k})$.

Testing causality by model comparison

Compare sharing model $\gamma = 0$ vs causal model $\gamma \neq 0$:

$$P(\hat{\beta}_M, \hat{\beta}_Y | M) = \int P(\hat{\beta}_M, \hat{\beta}_Y | q, \gamma, \eta) P(q, \gamma, \eta | M) dq d\gamma d\eta$$

Testing causality by model comparison

Compare sharing model $\gamma = 0$ vs causal model $\gamma \neq 0$:

$$P(\hat{\beta}_M, \hat{\beta}_Y | M) = \int P(\hat{\beta}_M, \hat{\beta}_Y | q, \gamma, \eta) P(q, \gamma, \eta | M) dq d\gamma d\eta$$

At $\gamma = 0$, the model is reduced to:

$$\beta_{Y,i} = Z_i \eta \beta_{M,i} + \theta_i \qquad Z_i \sim \text{Bern}(q)$$

A proportion q of variants of M show correlated effects in Y. The prior of q is chosen to be small (much smaller than 1).

Testing causality by model comparison

Compare sharing model $\gamma = 0$ vs causal model $\gamma \neq 0$:

$$P(\hat{\beta}_M, \hat{\beta}_Y | M) = \int P(\hat{\beta}_M, \hat{\beta}_Y | q, \gamma, \eta) P(q, \gamma, \eta | M) dq d\gamma d\eta$$

At $\gamma = 0$, the model is reduced to:

$$\beta_{Y,i} = Z_i \eta \beta_{M,i} + \theta_i \qquad Z_i \sim \text{Bern}(q)$$

A proportion q of variants of M show correlated effects in Y. The prior of q is chosen to be small (much smaller than 1).

At $\gamma \neq 0$, the model is approximately (ignoring shared factor):

$$\beta_{Y,i} \approx \gamma \beta_{M,i} + \theta_i$$

Testing causality by model comparison

Compare sharing model $\gamma = 0$ vs causal model $\gamma \neq 0$:

$$P(\hat{\beta}_M, \hat{\beta}_Y | M) = \int P(\hat{\beta}_M, \hat{\beta}_Y | q, \gamma, \eta) P(q, \gamma, \eta | M) dq d\gamma d\eta$$

At $\gamma = 0$, the model is reduced to:

$$\beta_{Y,i} = Z_i \eta \beta_{M,i} + \theta_i \qquad Z_i \sim \text{Bern}(q)$$

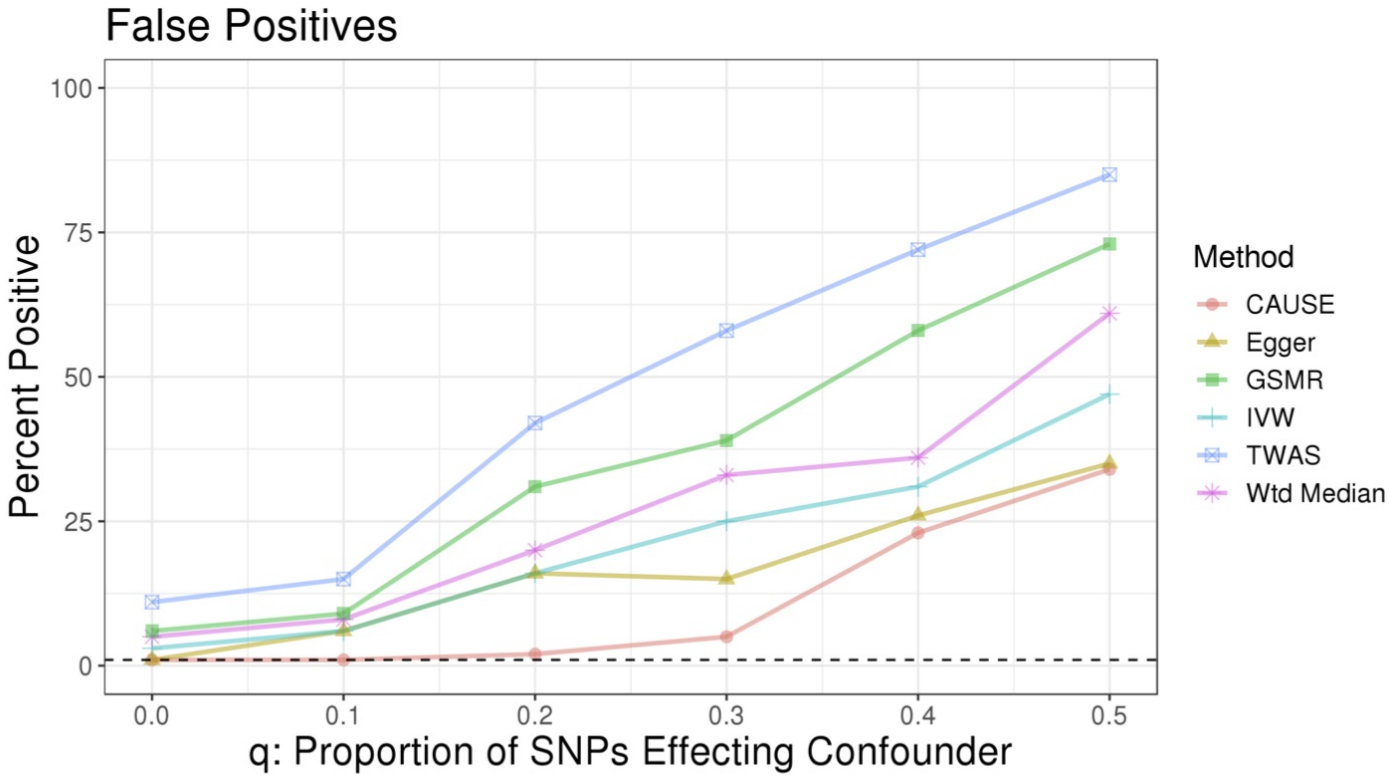
A proportion q of variants of M show correlated effects in Y. The prior of q is chosen to be small (much smaller than 1).

At $\gamma \neq 0$, the model is approximately (ignoring shared factor):

$$\beta_{Y,i} \approx \gamma \beta_{M,i} + \theta_i$$

MR-CAUSE (**C**ausal **A**nalysis **U**sing **S**ummary **E**ffects)

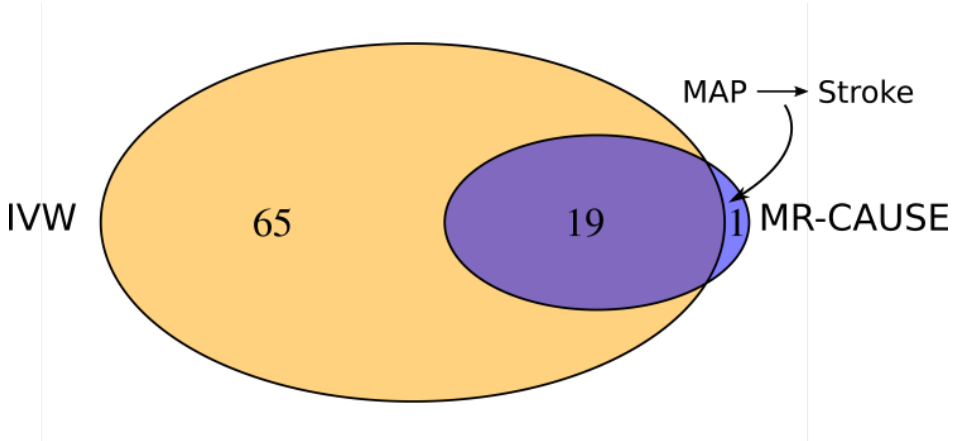
MR-CAUSE reduces false positives in simulation



The power of MR-CAUSE is comparable to existing methods.

Application of MR-CAUSE

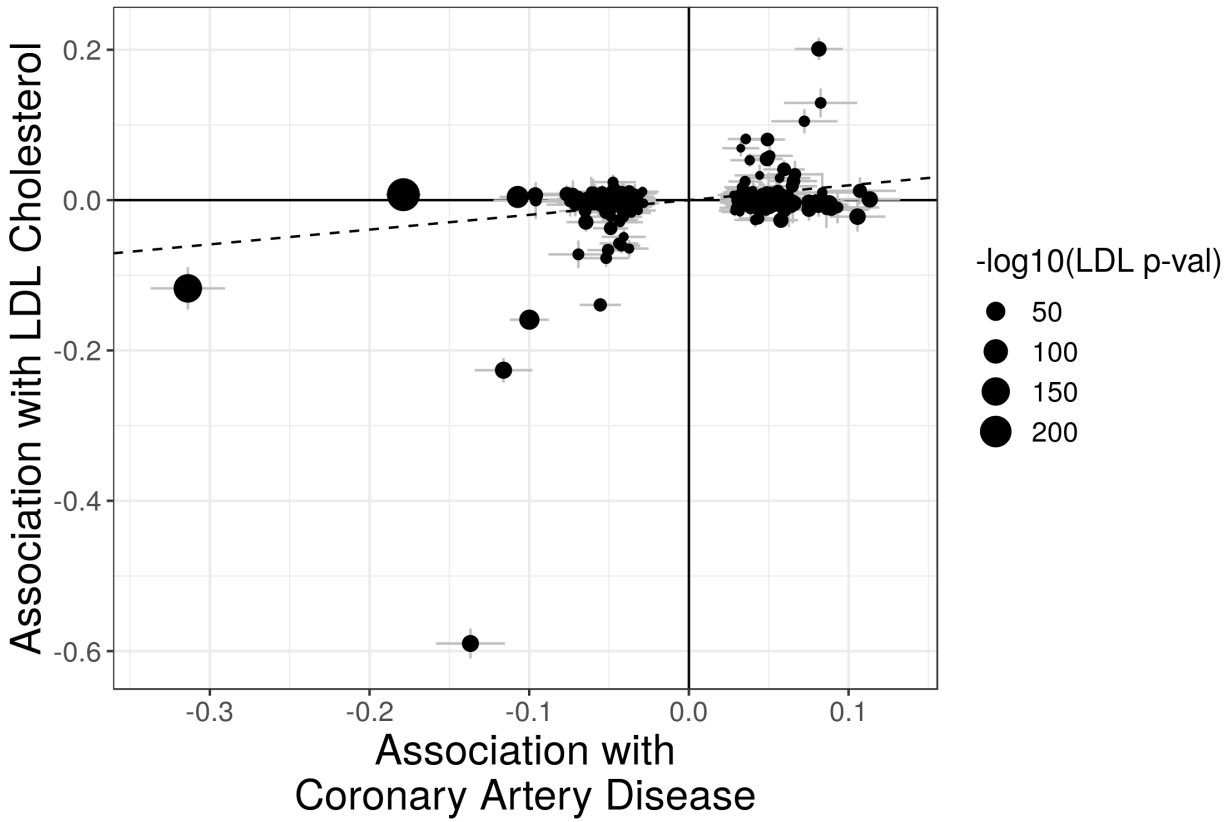
20 GWAS traits with summary statistics: comparison of IVW and MR-CAUSE at FDR < 0.05.



Many pairs detected by IVW are likely false: e.g. height $\rightarrow$ 11 traits including CAD, lipids, kidney function.

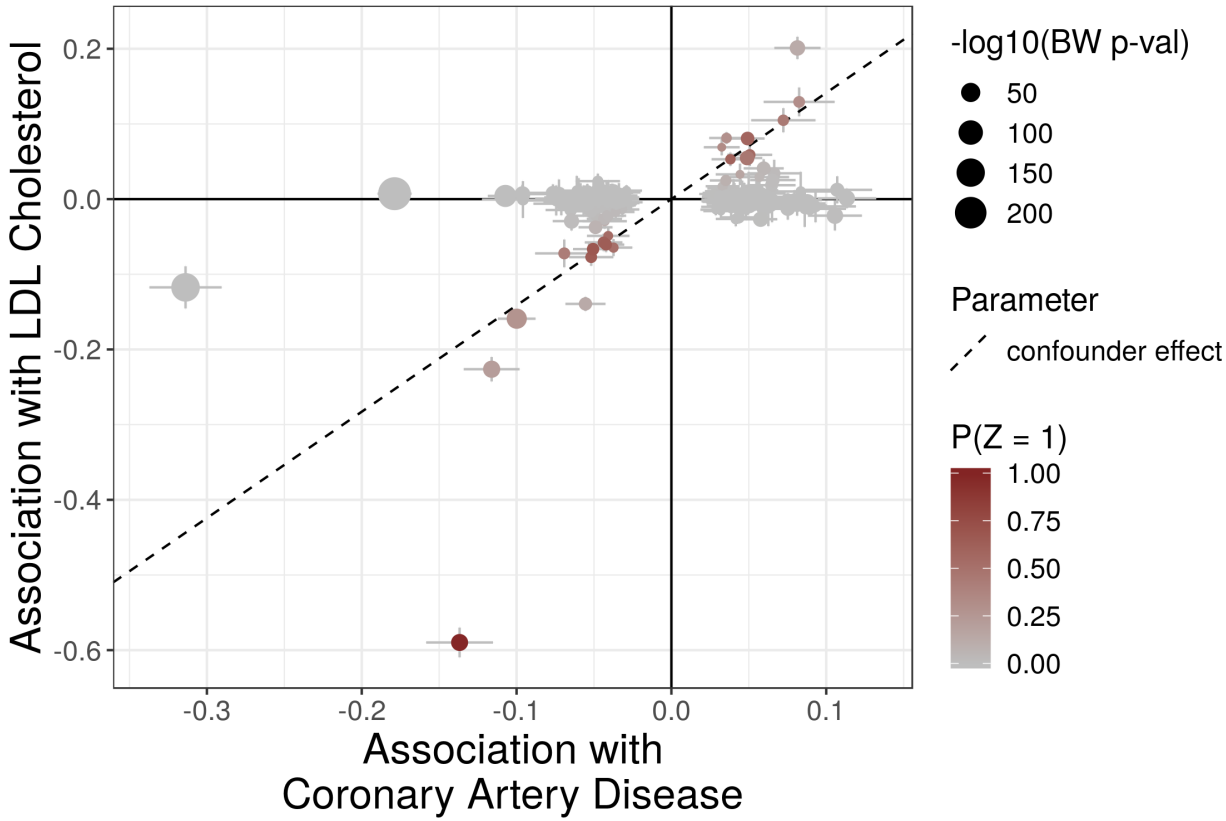
Testing CAD $\rightarrow$ LDL cholesterol

We know that LDL $\rightarrow$ CAD risk, so the opposite direction should NOT be causal.



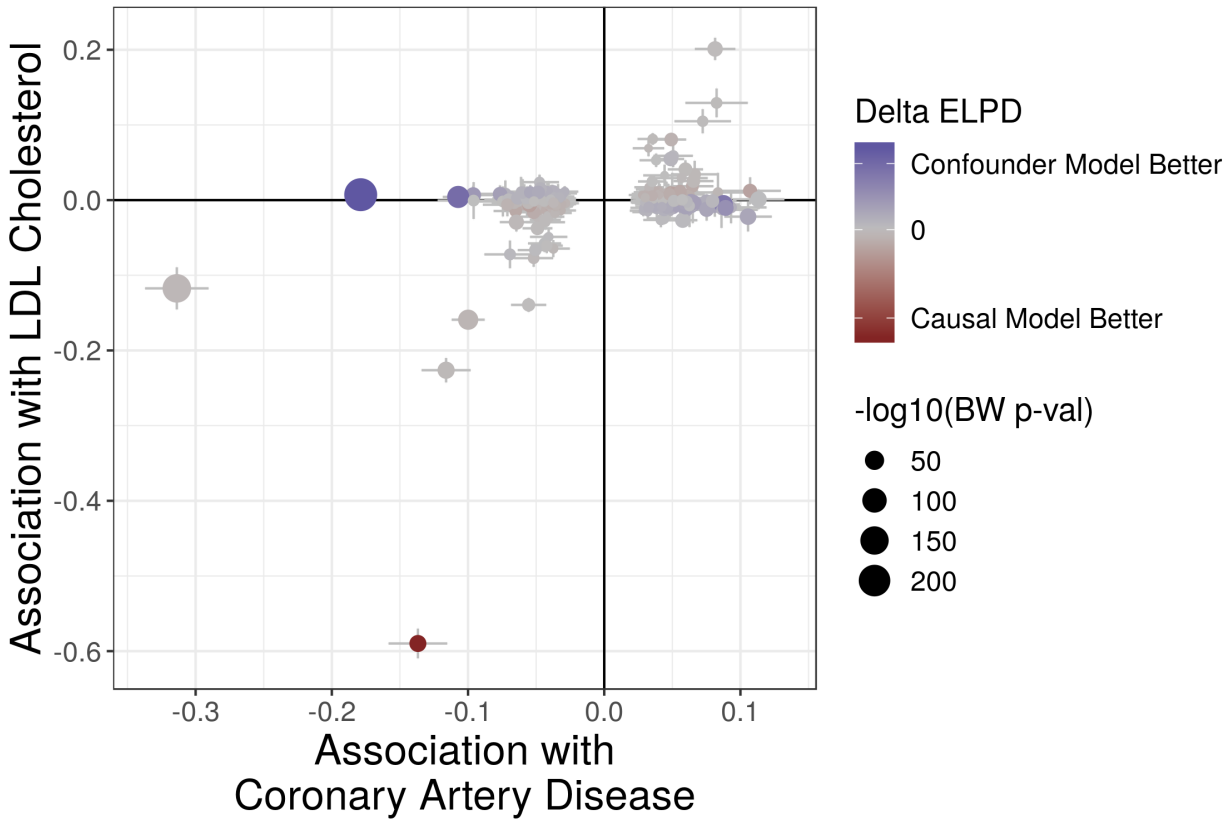
IVW p value $= 6 \times 10^{-6}$. MR-CAUSE: $p = 0.11$.

A small number of loci drive correlation of effect sizes



Exploration of the variants acting on the shared factor may provide insights on common biological pathways between traits.

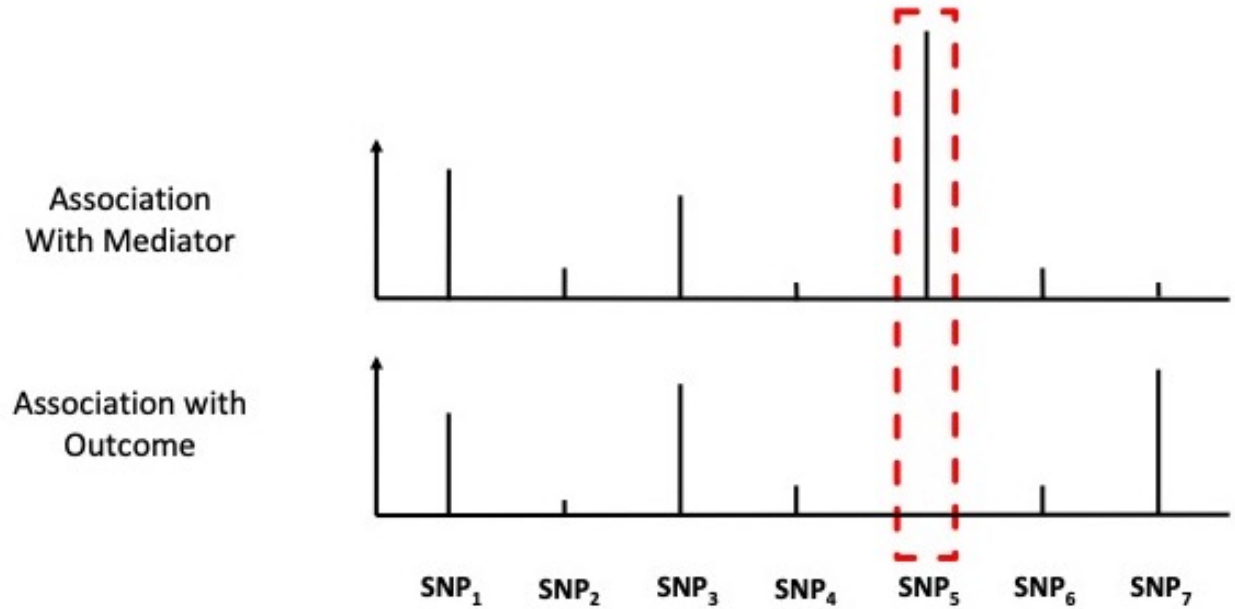
MR-CAUSE ranks SNPs by their contributions to model comparison



A small number of SNPs allow us to reject the causal model.

How to prove causality?

Proving causality is hard, but rejecting it is "easy":
We can reject if there is a variant with large effect on M, but not on Y (assuming GWAS of Y is sufficiently powered).



Summary

- Current MR methods are prone to false findings due to heritable confounders.
- MR-CAUSE explicitly models the shared factors, and address other limitations of current MR.
- MR results can provide information about which variants drive genetic sharing, and useful diagnosis for understanding the results.

Acknowledgments

Jean Morrison

Matthew Stephens

Nicholas Knoblauch

Joe Marcus

R Package: https://github.com/jean997/cause

THE UNIVERSITY OF CHICAGO

Different signatures of causal vs. non-causal relations