Mendelian Randomization Accounting for Horizontal and Correlated Pleiotropy Using Genome-Wide Summary Statistics
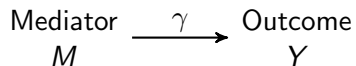
Jean Morrison, Nicholas Knoblauch, Joseph Marcus
Matthew Stephens and Xin He

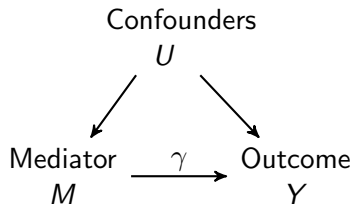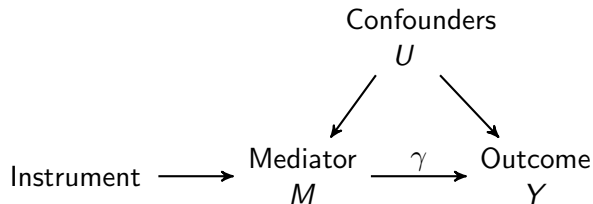University of Chicago, Department of Human Genetics

June 25, 2019

THE UNIVERSITY OF
CHICAGO

Genetic Variants as Instrumental Variables

Mediator $\xrightarrow{\quad\gamma\quad}$ Outcome
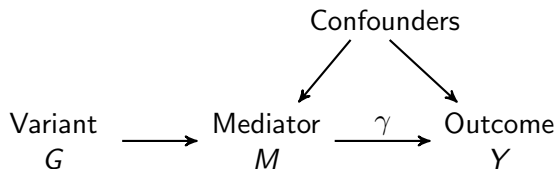M Y

Genetic Variants as Instrumental Variables

Genetic Variants as Instrumental Variables

Mendelian Randomization

Causal
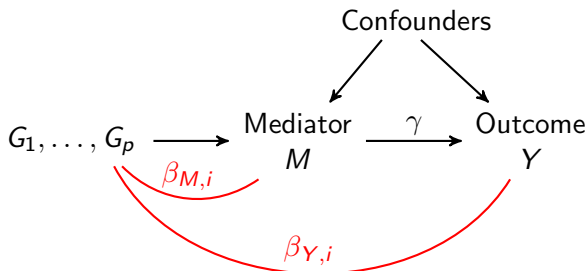Model

Mendelian Randomization

Causal Model

Confounders

Variant G → Mediator M — γ → Outcome Y

Statistical Associations

β_M

β_Y
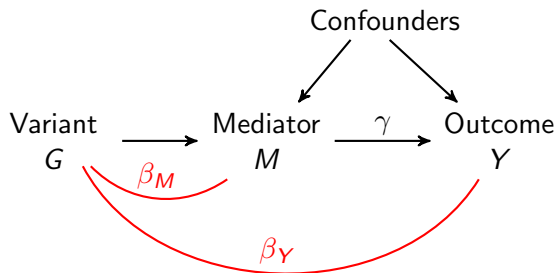
$$\beta_Y = \gamma \beta_M$$
$$\hat{\gamma} = \hat{\beta}_Y / \hat{\beta}_M$$
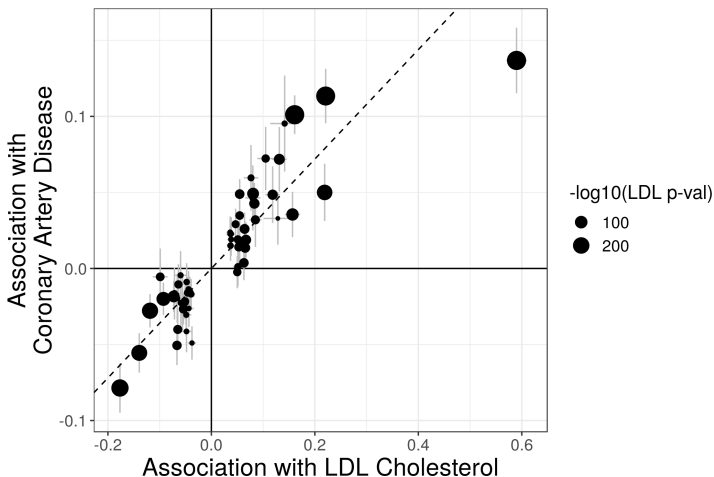
Mendelian Randomization

Causal Model

Statistical Associations

$$\beta_{Y,i} = \gamma \beta_{M,i} \qquad i = 1, \ldots, p$$

$$\hat{\gamma} = \frac{\sum_{i=1}^{p} \hat{\beta}_{Y,i} \hat{\beta}_{M,i} \sigma_Y^{-2}}{\sum_{i=1}^{p} \hat{\beta}_{M,i}^2 \sigma_Y^{-2}}$$
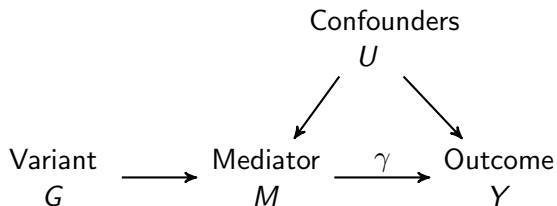
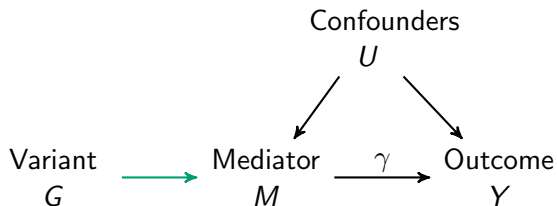# LDL Cholesterol and Coronary Artery Disease

Global Lipids Genetics Consortium (2013)
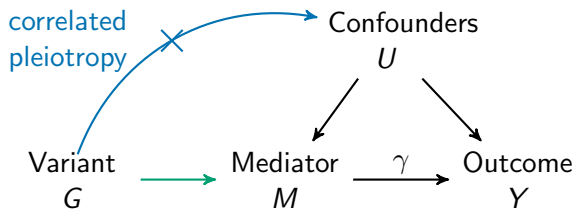van der Harst (2018)

Traditional MR Assumptions

Confounders
U

Variant
G

Mediator
M

γ

Outcome
Y

Traditional MR Assumptions
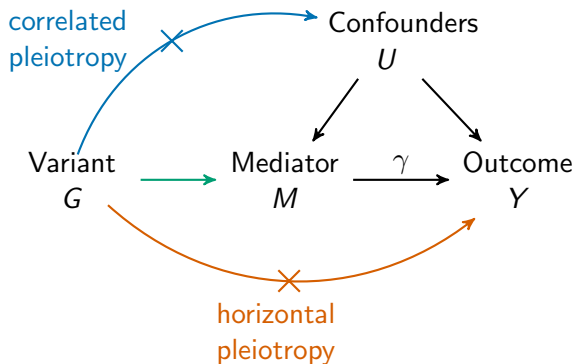
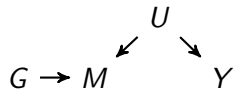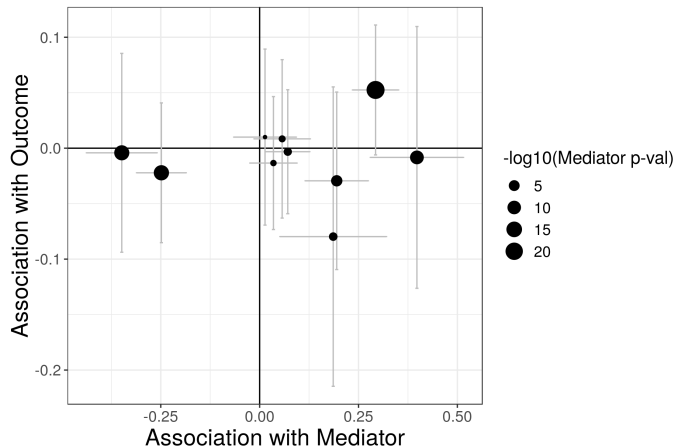

- G causally affects M.

Traditional MR Assumptions

- G causally affects M.
- G does not affect confounders.

Traditional MR Assumptions

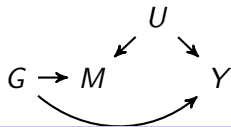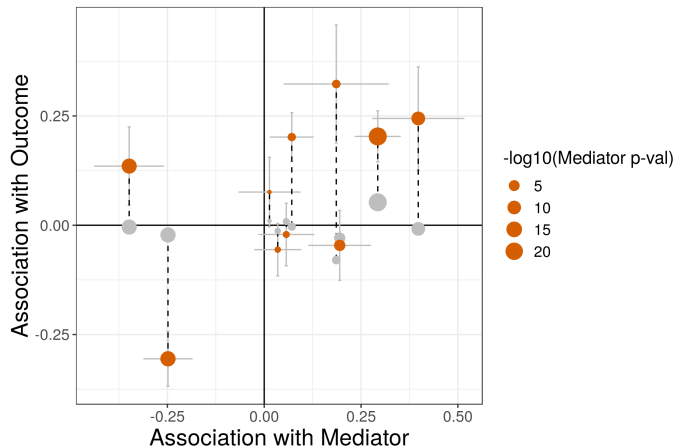- G causally affects M.
- G does not affect confounders.
- G does not affect Y through any other pathways.

No Cauasl Effect; No Pleiotropy

No Causal Effect; Horizontal Pleiotropy

No Causal Effect; Correlated Pleiotropy
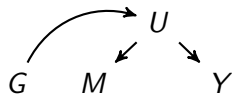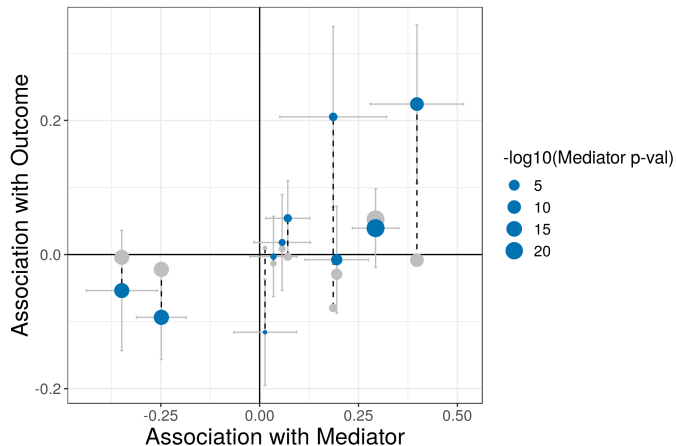
Correlated Pleiotropy Leads to False Positives
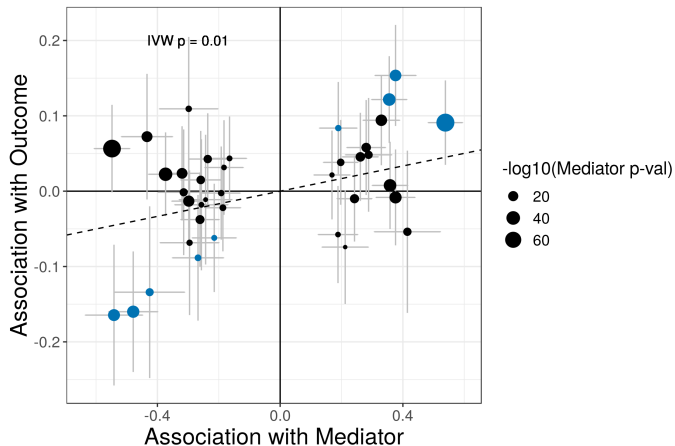
Correlated Pleiotropy Arises from Shared Biological Pathways

Example: HDL cholesterol and coronary artery disease

- Clinical trials do not show evidence of a causal effect.
- Simple MR analysis suggests a protective effect.
- If variants associated with LDL or triglycerides are excluded, there is no effect (Voight 2012).

Modeling Correlated Pleiotropy
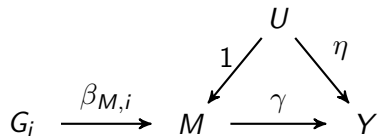
G_i acts directly on M
probability $1 - q$

Modeling Correlated Pleiotropy

G_i acts directly on M
probability $1 - q$

G_i acts on M through U
probability q

Modeling Correlated Pleiotropy

G_i acts directly on M
probability $1 - q$
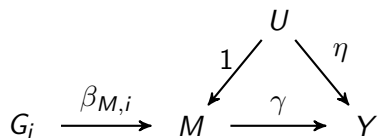
G_i acts on M through U
probability q
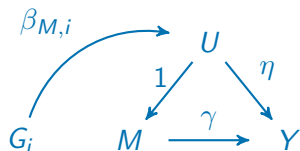
$$\beta_{Y,i} = (1 - Z_i)\gamma\beta_{M,i} + Z_i(\gamma + \eta)\beta_{M,i}$$
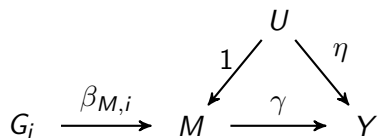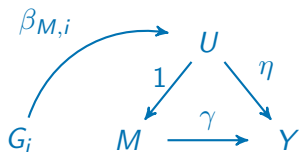
Modeling Horizontal Pleiotropy

G_i acts directly on M
probability $1 - q$

G_i acts on M through U
probability q
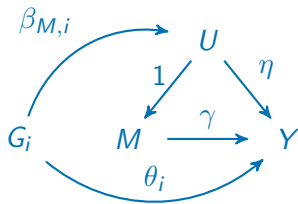
$$\beta_{Y,i} = (1 - Z_i)\gamma\beta_{M,i} + Z_i(\gamma + \eta)\beta_{M,i} + \theta_i$$
$$= \underbrace{\gamma\beta_{M,i}}_{\text{causal effect}} + \underbrace{Z_i\eta\beta_{M,i}}_{\substack{\text{correlated} \\ \text{pleiotropy}}} + \underbrace{\theta_i}_{\substack{\text{horizontal} \\ \text{pleiotropy}}}$$

Estimating Posterior Distributions from Summary Statistics

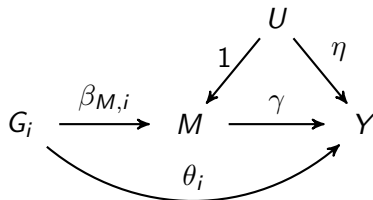▶ $(\hat{\beta}_{M,i},\ s_{M,i})$, $(\hat{\beta}_{Y,i},\ s_{Y,i})$ marginal effect estimates from GWAS

$$\begin{pmatrix}\hat{\beta}_{M,i}\\\hat{\beta}_{Y,i}\end{pmatrix} \sim N_2\left(\begin{pmatrix}\beta_{M,i}\\\beta_{Y,i}\end{pmatrix}, \begin{pmatrix}\hat{s}_{M,i}^2 & \rho\hat{s}_{M,i}\hat{s}_{Y,i}\\\rho\hat{s}_{M,j}\hat{s}_{Y,i} & \hat{s}_{Y,i}^2\end{pmatrix}\right)$$

▶ Estimate ρ and a joint empirical prior for $\beta_{M,i}$ and θ_i (unimodal centered at zero)

▶ $q \sim \text{Beta}(1,10)$, $\gamma, \eta \sim N(0, \sigma_{\gamma\eta})$

▶ Estimate posteriors for γ, η, and q using an adaptive grid approximation

Model Comparison

- **Partial Sharing Model:** $\gamma = 0$
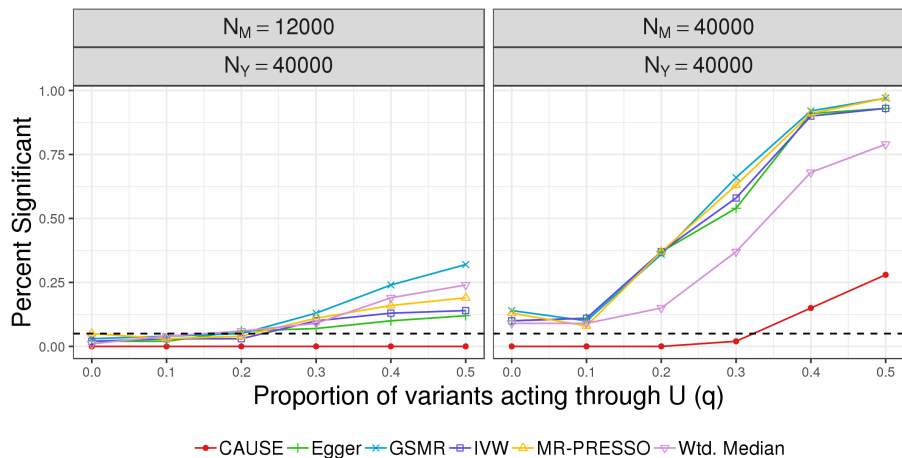- **Causal model:** γ is a free parameter

Model Comparison

- **Partial Sharing Model:** $\gamma = 0$
- **Causal model:** γ is a free parameter
- Compare model fits using ELPD [Vehtari et al (2016)]

Model Comparison

- **Partial Sharing Model:** $\gamma = 0$
- **Causal model:** γ is a free parameter
- Compare model fits using ELPD [Vehtari et al (2016)]
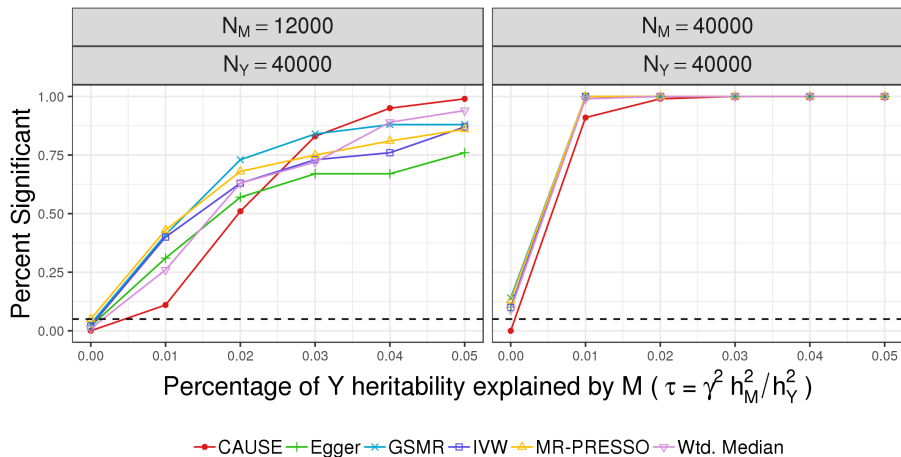- If posteriors estimated under the causal model predict the data significantly better than posteriors estimated under the sharing model we say that the **data are consistent with a causal effect**.

Simulated Data: False Positives

# Simulated Data: Power

Percent Significant vs. Percentage of Y heritability explained by M ($\tau = \gamma^2 h_M^2 / h_Y^2$)

Panels: $N_M = 12000$, $N_Y = 40000$ and $N_M = 40000$, $N_Y = 40000$

Legend: CAUSE, Egger, GSMR, IVW, MR-PRESSO, Wtd. Median

Pairs of Complex Traits

- We used CAUSE to evaluate all pairs of 20 GWAS traits.
- CAUSE identifies fewer pairs as causal than IVW regression, avoiding likely false positives like CAD -¿ LDL.
- CAUSE is able to detect an effect of blood pressure on stroke that is missed by IVW due to low power in BP GWAS.
- Many pairs of traits have evidence of substantial correlated pleiotropy.

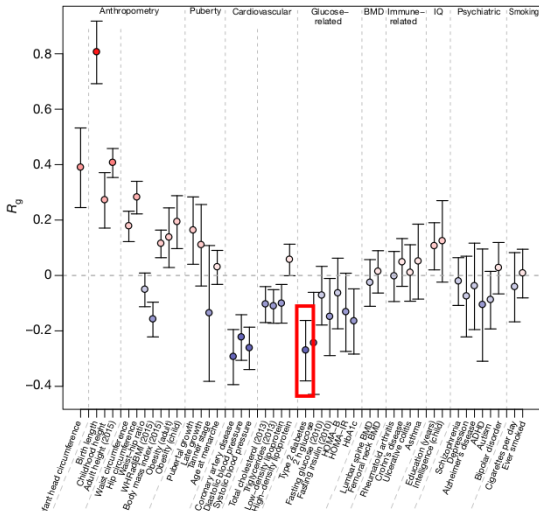# Does higher birth weight protect against metabolic disease?

Genetic Correlations with BW
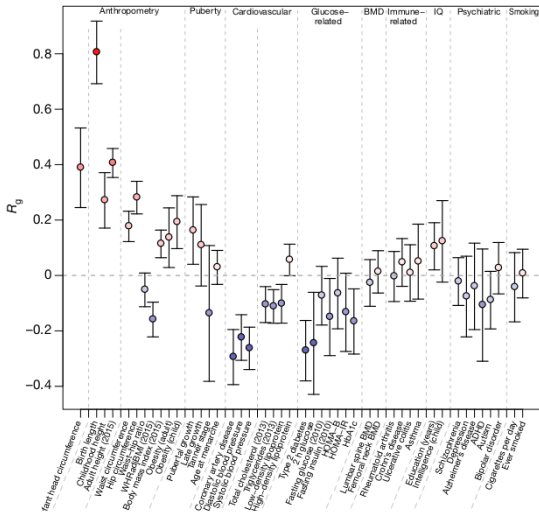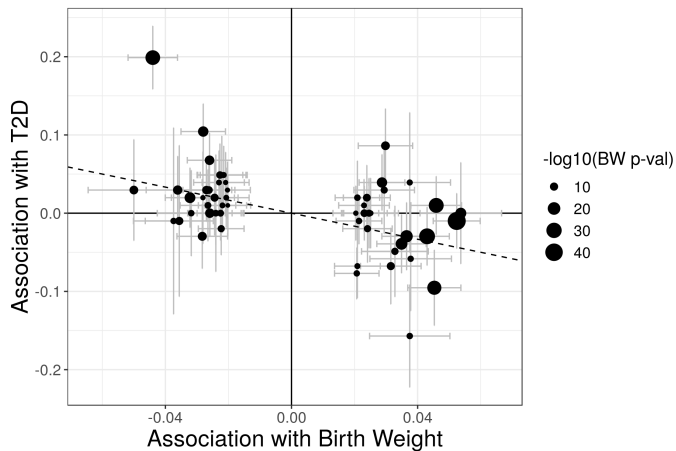
Horikoshi et al (2016)

Does higher birth weight protect against metabolic disease?
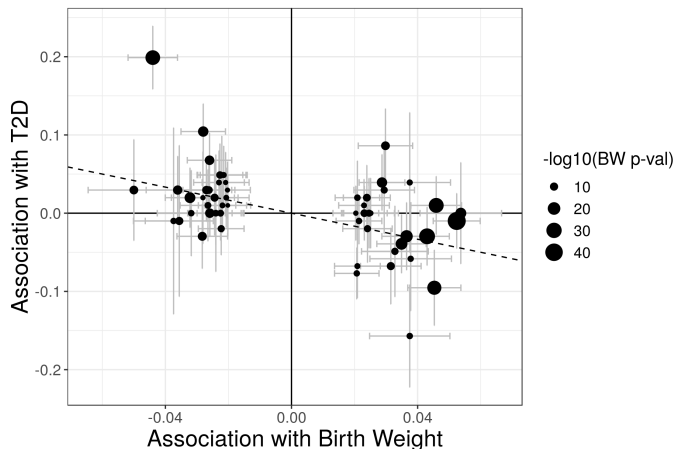
Genetic Correlations with BW

Horikoshi et al (2016)

Birth Weight and T2D Risk

Horikoshi et al (2016); Morris et al (2012)
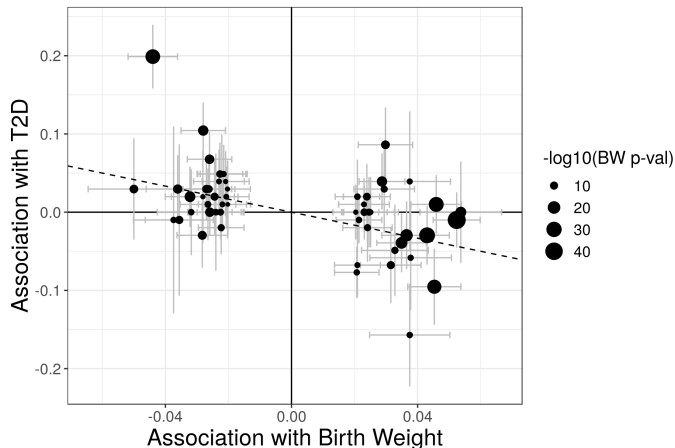
Birth Weight and T2D Risk

IVW *p*-value: $7.0 \cdot 10^{-5}$ MR-PRESSO *p*-value: $2.18 \cdot 10^{-5}$

GSMR *p*-value: $3.8 \cdot 10^{-6}$

Horikoshi et al (2016); Morris et al (2012)

Birth Weight and T2D Risk

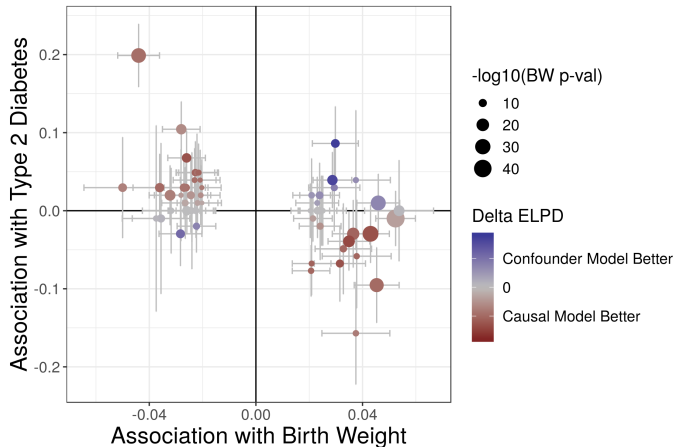IVW *p*-value: $7.0 \cdot 10^{-5}$ MR-PRESSO *p*-value: $2.18 \cdot 10^{-5}$
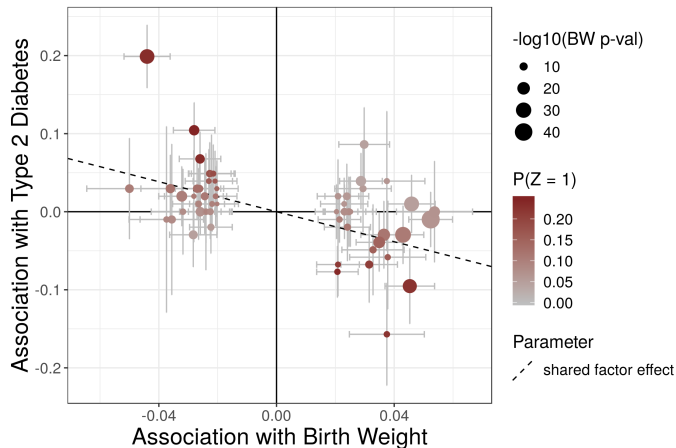GSMR *p*-value: $3.8 \cdot 10^{-6}$ CAUSE *p*-value: 0.066

Horikoshi et al (2016); Morris et al (2012)

Model Comparison: Variant Contribution to Test Statistic

Sharing Model Posteriors

Summary

- CAUSE is robust to both correlated and horizontal pleiotropy.
- Correlated pleiotropy may be a common source of MR false positives.
- Parameter estimates provide information about the amount of sharing and which variants affect the shared factor.

Acknowledgments

Xin He   Matthew Stephens

He Lab:
Nicholas Knoblauch

Stephens Lab:
Joseph Marcus

Website: https://jean997.github.io/cause/

THE UNIVERSITY OF
CHICAGO