Supplementary Material for MR-CoPe: Causal Inference Without the Chaos — An End-to-End Reproducible Pipeline for Mendelian Randomisation Analysis

Guillermo Comesaña Cimadevila[1*] and Christian Pepler[2]

[1]Department of Life Sciences, University of Bath, Claverton Down, Bath BA2 7AY, UK

[2]School of Biosciences, Cardiff University, Cardiff CF10 3AT, UK

*Corresponding authors: gcc46@bath.ac.uk

Supplementary Material

Inverse-Variance Weighted (IVW) Estimator

Setup

Suppose we have L independent genetic variants (SNPs) that satisfy the conditions for valid instruments in Mendelian Randomisation (MR). For each SNP $l \in \{1, 2, ..., L\}$:

- $\hat{\beta}_{Xl}$ is the estimated association of SNP l with the exposure X.
- $\hat{\beta}_{Yl}$ is the estimated association of SNP l with the outcome Y.
- σ^2_{Yl} is the variance of $\hat{\beta}_{Yl}$.

The basic MR model for each SNP is:

$$\hat{\beta}_{Yl} = \beta_{\text{IVW}} \hat{\beta}_{Xl} + \varepsilon_l,$$

where β_{IVW} is the causal effect of X on Y we wish to estimate, and ε_l is an error term.

Assumptions

Assuming:

- No pleiotropy (no direct effect of SNPs on Y except through X).
- The error terms ε_l are independent and normally distributed: $\varepsilon_l \sim \mathcal{N}(0, \sigma^2_{Yl})$.

Weighted Least Squares Estimator

We aim to estimate β_{IVW} using **weighted least squares**, minimising the sum of squared residuals weighted by inverse variance:

$$Q(\beta) = \sum_{l=1}^{L} \frac{(\hat{\beta}_{Yl} - \beta \hat{\beta}_{Xl})^2}{\sigma^2_{Yl}}.$$

To minimize $Q(\beta)$, we take the derivative with respect to β and set it to zero:

$$\frac{dQ}{d\beta} = \sum_{l=1}^{L} \frac{-2\hat{\beta}_{Xl}(\hat{\beta}_{Yl} - \beta\hat{\beta}_{Xl})}{\sigma_{Yl}^2} = 0.$$

Rewriting:

$$\sum_{l=1}^{L} \frac{\hat{\beta}_{Xl}\hat{\beta}_{Yl}}{\sigma_{Yl}^2} - \beta \sum_{l=1}^{L} \frac{\hat{\beta}_{Xl}^2}{\sigma_{Yl}^2} = 0.$$

Solving for β gives the **IVW estimator**:

$$\hat{\beta}_{\text{IVW}} = \frac{\sum_{l=1}^{L} \frac{\hat{\beta}_{Xl}\hat{\beta}_{Yl}}{\sigma_{Yl}^2}}{\sum_{l=1}^{L} \frac{\hat{\beta}_{Xl}^2}{\sigma_{Yl}^2}}.$$

Interpretation

This estimator is a weighted average of the ratio estimates $\hat{\beta}_{Yl}/\hat{\beta}_{Xl}$, with weights proportional to $\hat{\beta}_{Xl}^2/\sigma_{Yl}^2$. This reflects greater confidence in instruments with stronger associations with the exposure and smaller standard errors.

Weighted Median Estimator (WME)

Setup

For each SNP $l \in \{1, 2, ..., L\}$, define the **ratio estimate**:

$$\hat{\beta}_l^{\text{ratio}} = \frac{\hat{\beta}_{Yl}}{\hat{\beta}_{Xl}}.$$

Let w_l be the inverse variance weight for SNP l:

$$w_l = \frac{1}{\sigma_{Yl}^2 / \hat{\beta}_{Xl}^2}.$$

The weighted median estimator is defined as the **median** of the ratio estimates $\hat{\beta}_l^{\text{ratio}}$, weighted by w_l. That is, it is the value $\tilde{\beta}$ such that:

$$\sum_{l:\hat{\beta}_l^{\text{ratio}} \leq \tilde{\beta}} w_l \geq \frac{1}{2} \sum_{l=1}^{L} w_l \quad \text{and} \quad \sum_{l:\hat{\beta}_l^{\text{ratio}} \geq \tilde{\beta}} w_l \geq \frac{1}{2} \sum_{l=1}^{L} w_l.$$

Assumptions

The WME is consistent for the true causal effect if:

- At least 50% of the total weight comes from valid instruments.
- The SNP-exposure estimates $\hat{\beta}_{Xl}$ are precise and non-zero.

Interpretation

Unlike IVW, the WME is robust to invalid instruments, as long as the majority (by weight) of instruments are valid. It reduces the influence of outlier ratio estimates due to pleiotropy or weak instruments.

MR-Egger Regression

Setup

MR-Egger regression extends the IVW model by allowing for an intercept term to account for directional pleiotropy:

$$\hat{\beta}_{Yl} = \alpha + \beta_{\text{Egger}}\hat{\beta}_{Xl} + \varepsilon_l.$$

Here:

- β_{Egger} estimates the causal effect.
- α captures the average directional pleiotropic effect.

Assumptions

MR-Egger relies on the **InSIDE** (Instrument Strength Independent of Direct Effect) assumption:

- The instrument strength ($\hat{\beta}_{Xl}$) is independent of the direct effect of the SNP on the outcome.
- Measurement error in $\hat{\beta}_{Xl}$ is negligible (no regression dilution bias).

Estimation

Parameters α and β_{Egger} are estimated using weighted linear regression, minimising:

$$Q(\alpha, \beta) = \sum_{l=1}^{L} \frac{(\hat{\beta}_{Yl} - \alpha - \beta\hat{\beta}_{Xl})^2}{\sigma_{Yl}^2}.$$

Interpretation

A non-zero intercept α suggests the presence of directional (unbalanced) pleiotropy. The slope β_{Egger} remains an unbiased estimate of the causal effect under the InSIDE assumption.