BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH

An Introduction to Mendelian Randomization

Guanghao Qi, PhD
Department of Biostatistics
University of Washington

Overview

- Basic theory of Mendelian randomization
- Robust methods for Mendelian randomization analysis
- Software and resources

Origin of Mendelian randomization

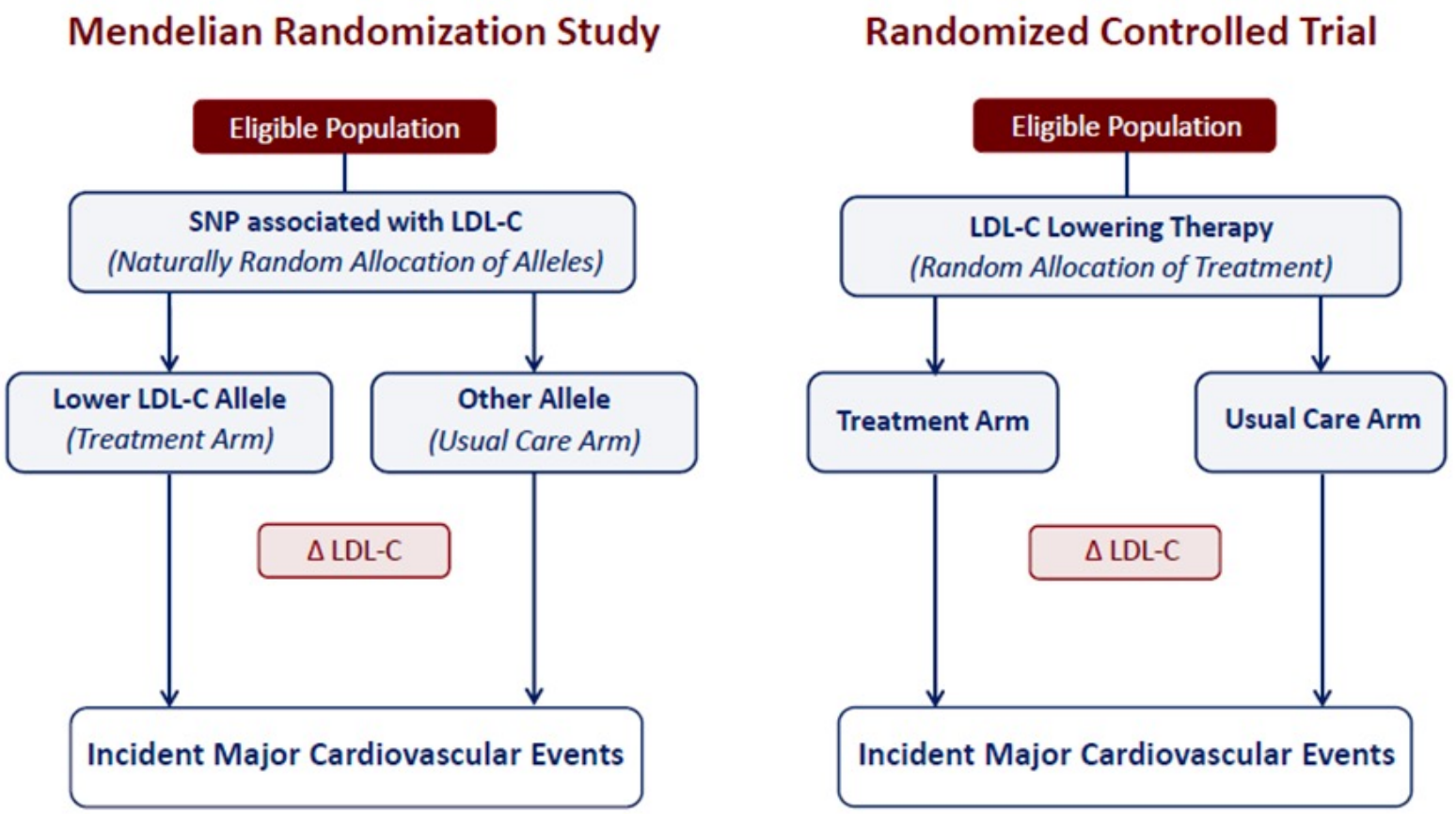
> "Genetics is indeed in a peculiarly favoured condition in that Providence has shielded the geneticist from many of the difficulties of a reliably controlled comparison. The different genotypes possible from the same mating have been beautifully randomised by the meiotic process. A more perfect control of conditions is scarcely possible, than that of different genotypes appearing in the same litter." — R.A. Fisher[5]

- Initially proposed in the 1980s
- Popularized by researchers in University of Bristol

Mendelian Randomization Analysis: Genetic Analog of Randomized Trials

This SNP is independent of most common confounders

Figure: Analogy Between a Mendelian Randomization Study and a Randomized Trial



Proof of Principle: Selenium and Prostate Cancer Risk

Table 1. Comparison of the effect of 114 ug/L selenium on overall prostate cancer, high-grade/advanced prostate cancer, and type 2 diabetes in SELECT and Mendelian randomization

Outcome	SELECT HR (95% CI)	Mendelian randomization OR (95% CI)
Overall prostate cancer	1.04 (0.91 to 1.19)	1.01 (0.89 to 1.13)
High-grade/advanced prostate cancer*	1.21 (0.97 to 1.52)	1.21 (0.98 to 1.49)
Type 2 diabetes	1.07 (0.87 to 1.18)	1.18 (0.97 to 1.43)

*"High-grade prostate cancer" pertains to SELECT results, and "advanced prostate cancer" pertains to Mendelian randomization results. CI = confidence interval; HR = hazard ratio; OR = odds ratio.

Yarmolinksky, JNCI, 2018

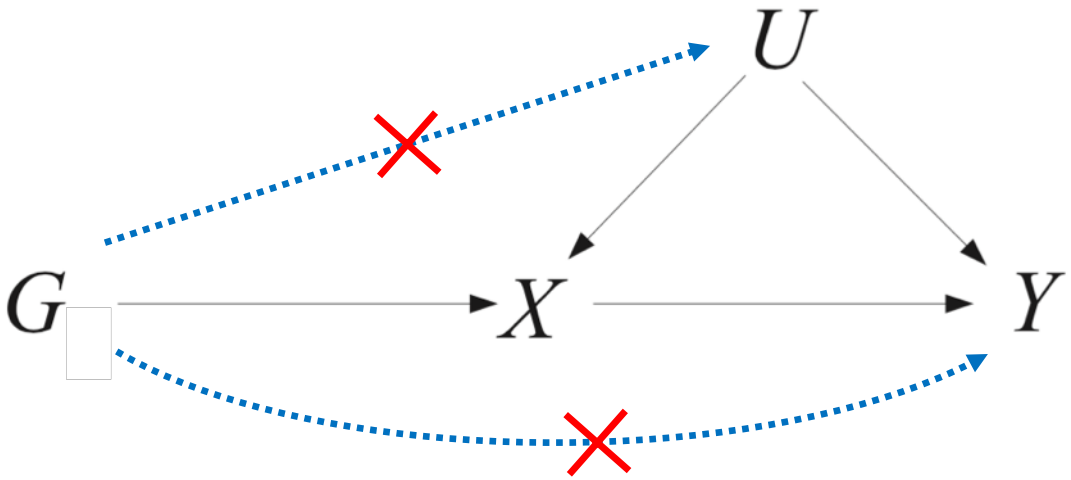
Evidence from MR analysis could be used to decide whether to launch costly prevention trials

Mendelian Randomization is an Instrumental Variable Approach

Assumptions:

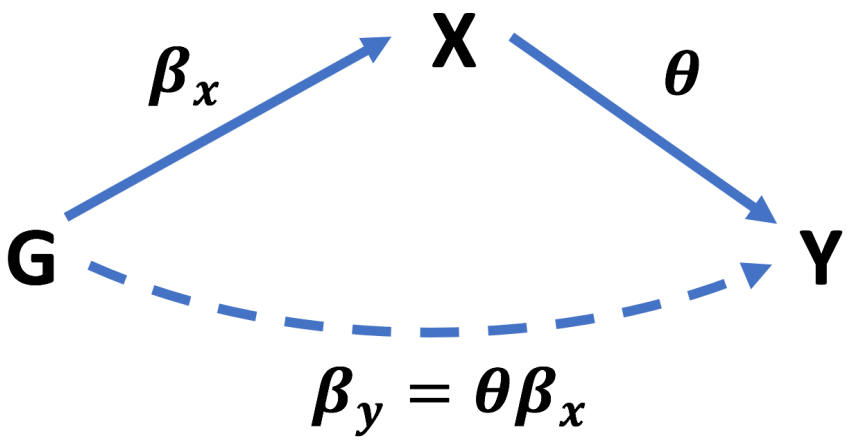
1. The SNP predicts the exposure X
2. The SNP is independent of the confounder U
3. The SNP has no effect on the outcome Y that is not mediated X

   (exclusion restriction; no **horizontal pleiotropy**)



Basic Mendelian Randomization

- Use genetic correlation across traits to estimate causal effect (assuming no horizontal pleiotropy)



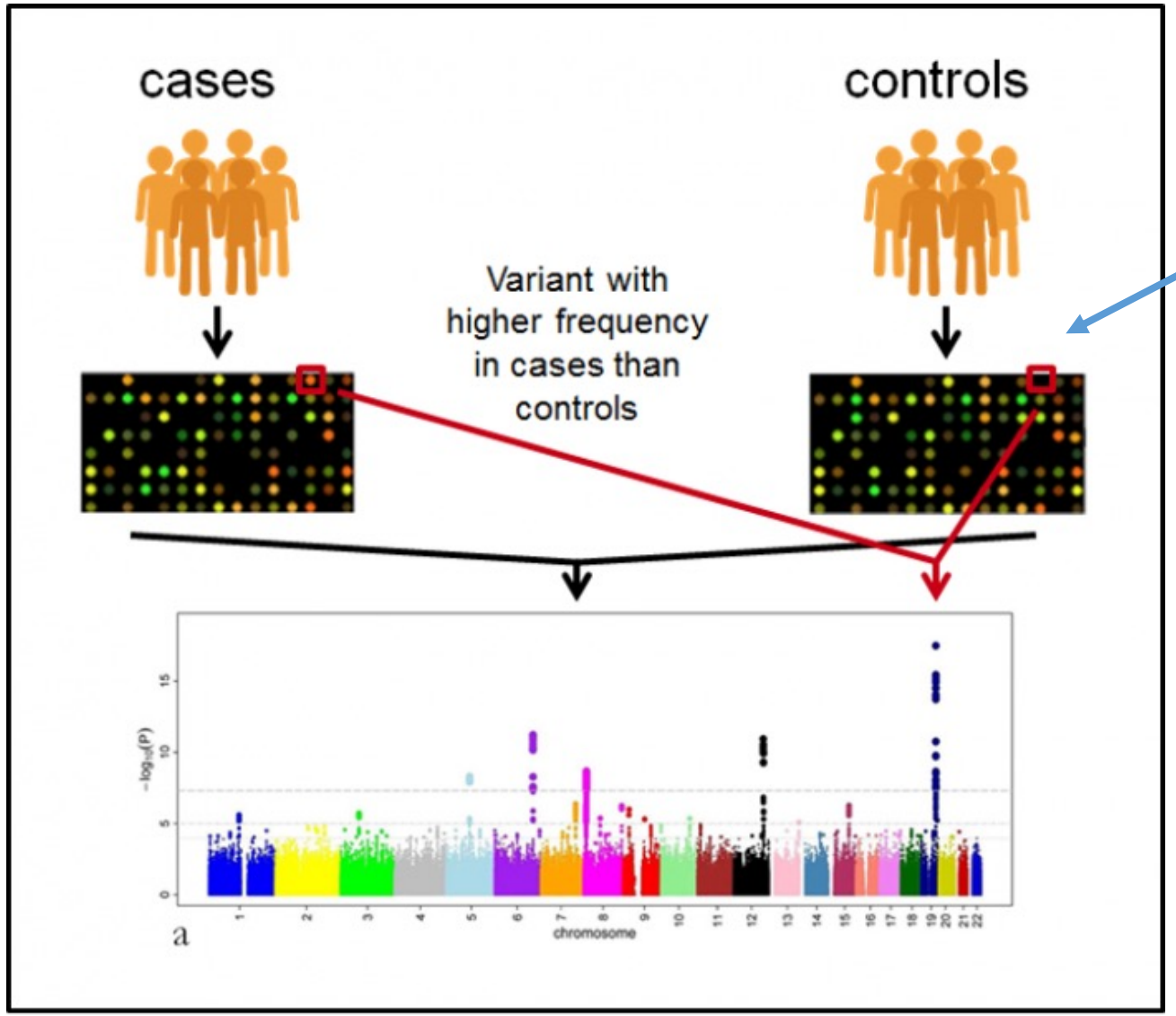
Assuming linear relationships

Single-marker Ratio Estimator: $\hat{\theta}=\frac{\hat{\beta}_y}{\hat{\beta}_x}$

Polygenic Ratio Estimator (IVW): $\hat{\theta} = \sum_k w_k \hat{\theta}_k$

Burgess et al., Genetic Epidemiology, 2013

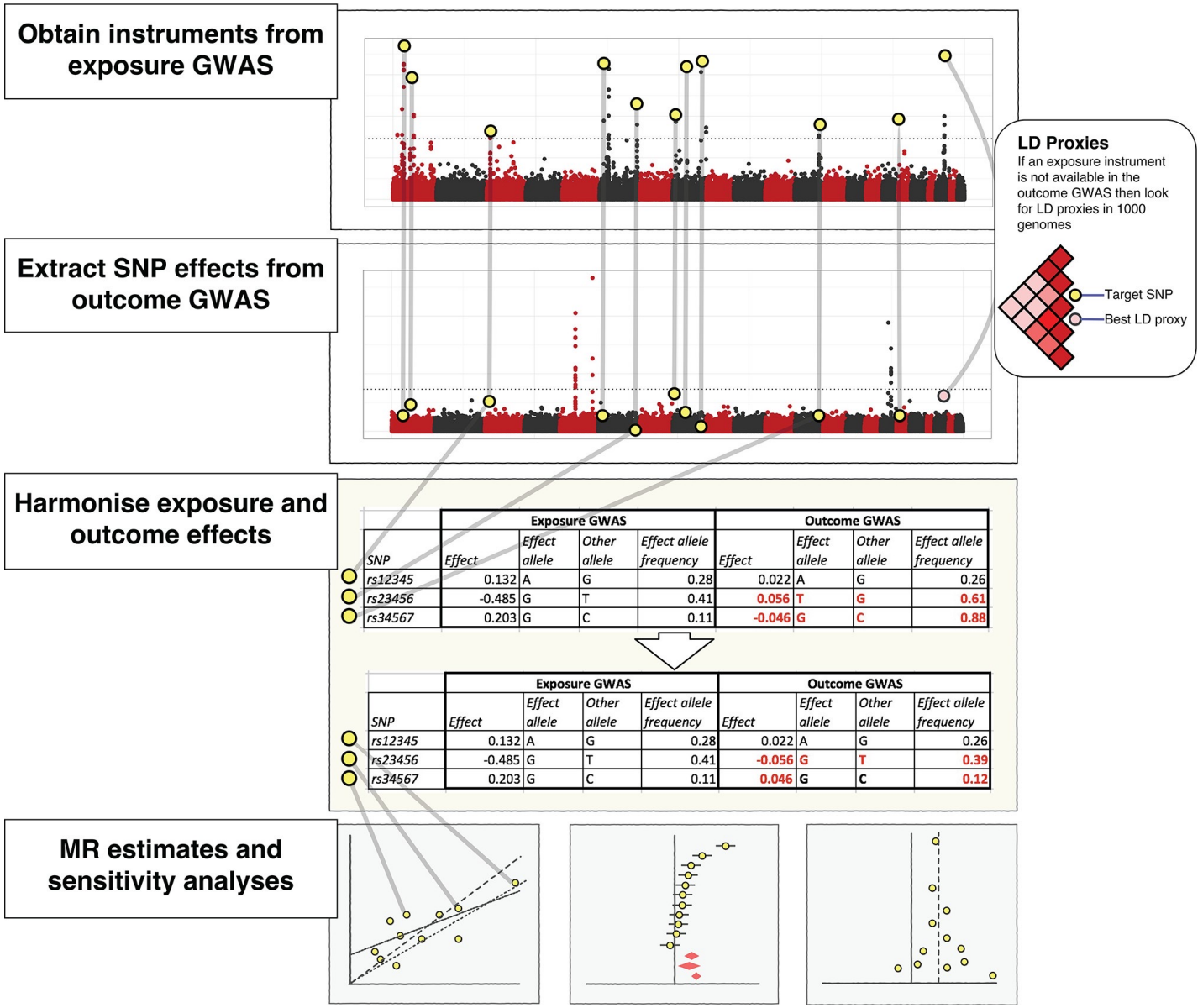
Genome-wide association studies (GWAS) provide the data for MR analysis



Single nucleotide polymorphism (SNP)

EMBL-EBI

From GWAS to Mendelian randomization (MR)

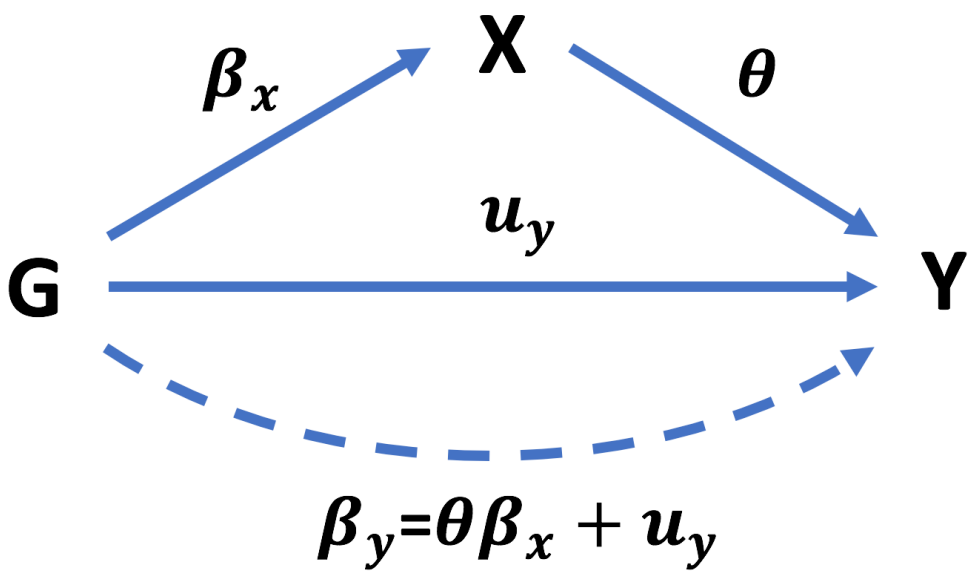


	Exposure GWAS				Outcome GWAS			
SNP	Effect	Effect allele	Other allele	Effect allele frequency	Effect	Effect allele	Other allele	Effect allele frequency
rs12345	0.132	A	G	0.28	0.022	A	G	0.26
rs23456	-0.485	G	T	0.41	0.056	T	G	0.61
rs34567	0.203	G	C	0.11	-0.046	G	C	0.88

	Exposure GWAS				Outcome GWAS			
SNP	Effect	Effect allele	Other allele	Effect allele frequency	Effect	Effect allele	Other allele	Effect allele frequency
rs12345	0.132	A	G	0.28	0.022	A	G	0.26
rs23456	-0.485	G	T	0.41	-0.056	G	T	0.39
rs34567	0.203	G	C	0.11	0.046	G	C	0.12

Hemani et al, eLife 2018

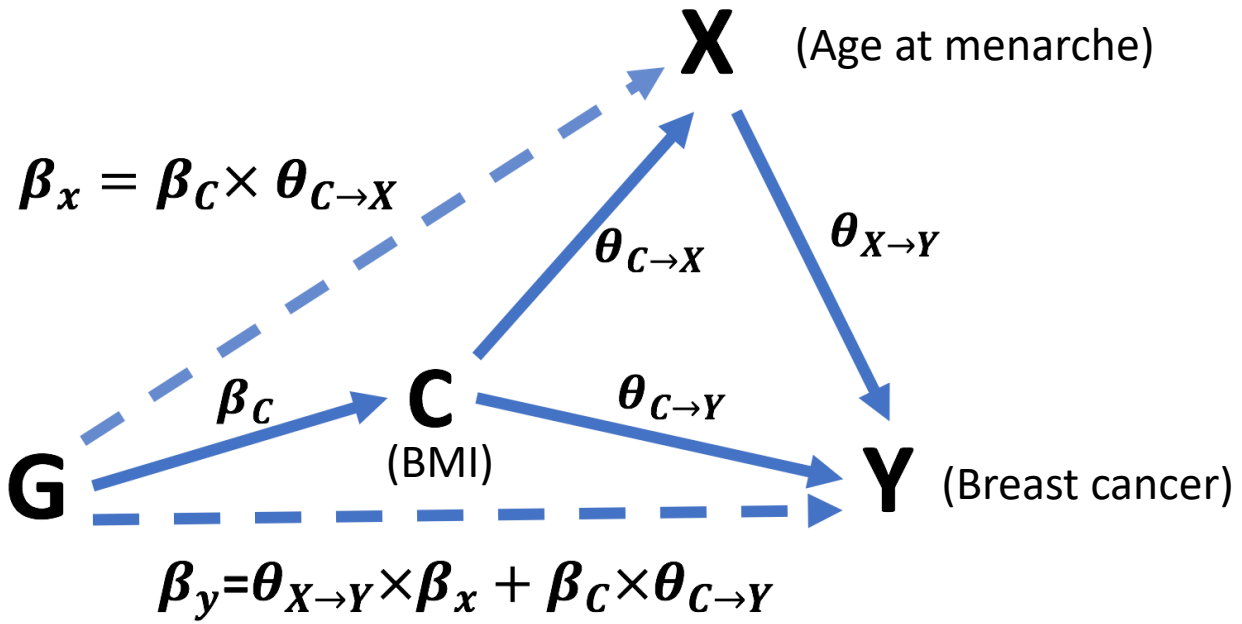
Horizontal Pleiotropy is Ubiquitous



But if direct (u_y) and indirect effects (β_x) are uncorrelated, genetic correlation can only arise due to causal effect (InSIDE Assumption)

Bowden et al., *International Journal of Epidemiology,* 2015

Violation of MR assumptions: age at menarche vs. risk of breast cancer



Genetically determined BMI (early BMI) confounding relationship between age at menarche and breast cancer

Burgess et al., Genetics, 2017

Robust MR methods to account for horizontal pleiotropy

Existing MR Methods

1. Location parameter of the distribution of ratio estimates
    - IVW (meta-analysis), Egger regression, median, mode
2. Robust loss functions
    - MR-RAPS (robust adjusted profile score), MR-cML (conditional maximum likelihood)
3. Outlier detection
    - MR-PRESSO
4. Direct modeling of pleiotropy
    - MRMix, contamination mixture, MR-CUE

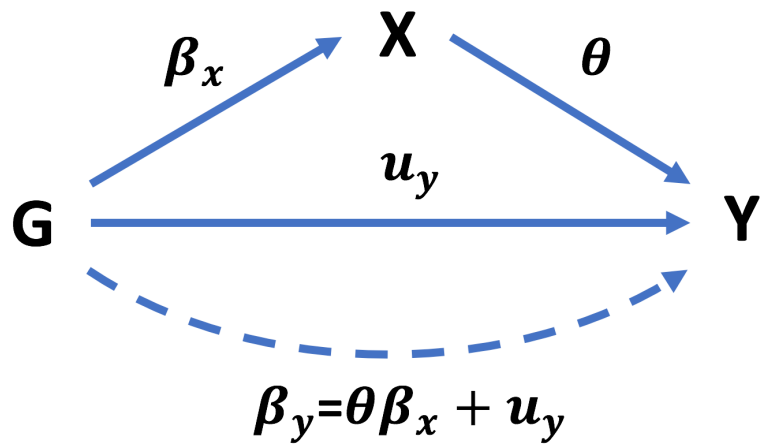
IVW: Burgess et al, Genetic Epidemiology, 2013
Egger: Bowden et al, International Journal of Epidemiology, 2015
Median: Bowden et al, Genetic Epidemiology, 2016
Mode: Hartwig et al, International Journal of Epidemiology, 2017
RAPS: Zhao et al, Annals of Statistics, 2020
cML: Xue et al, AJHG, 2021
PRESSO: Verbaneck et al, Nature Genetics, 2018
MRMix: Qi and Chatterjee, Nature Communications, 2019
Cont Mixture: Burgess et al, Nature Communications, 2020
CUE: Cheng et al, Nature Communications, 2022

IVW with random effects

- Assume the pleiotropic effect u_y are random error terms that is independent of β_x and centered at 0 (e.g. $u_y \sim N(0, \sigma^2)$)
  - Requires InSIDE assumption

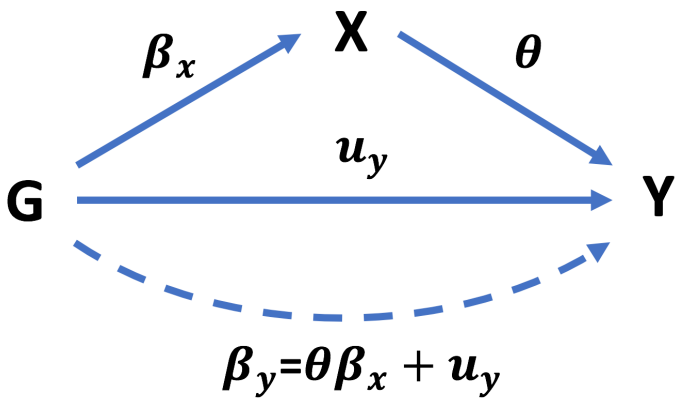
- Mitigate the pleiotropic effects by averaging across SNPs
  - Pleiotropic effects of different SNPs cancel with each other

- Estimator: $\hat{\theta} = \sum_k w_k{}' \hat{\theta}_k$
  - w_k' accounts for extra variation due to u_y



Egger regression: accounting for directional pleiotropy

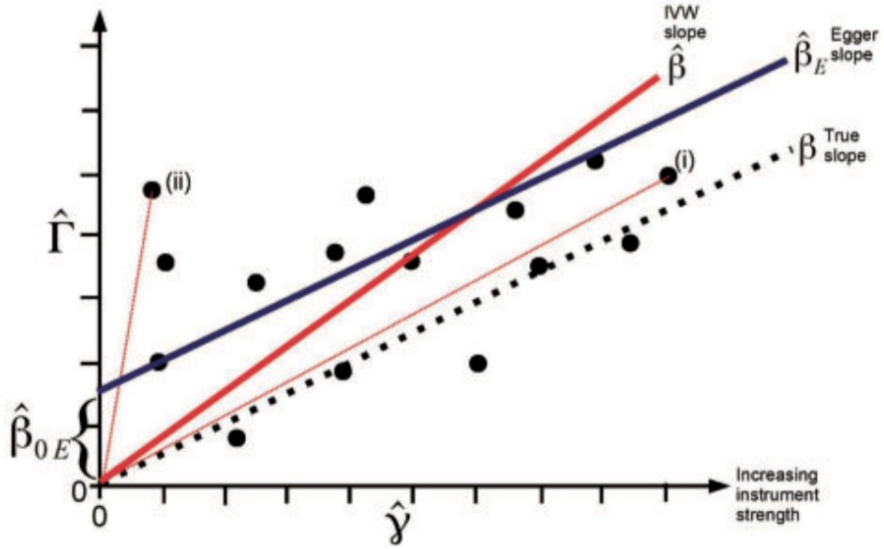
- Requires the InSIDE assumption $\beta_x \perp u_y$.



- Egger regression: fit linear model with intercept

$$\hat{\beta}_y = \theta_0 + \theta\hat{\beta}_x$$

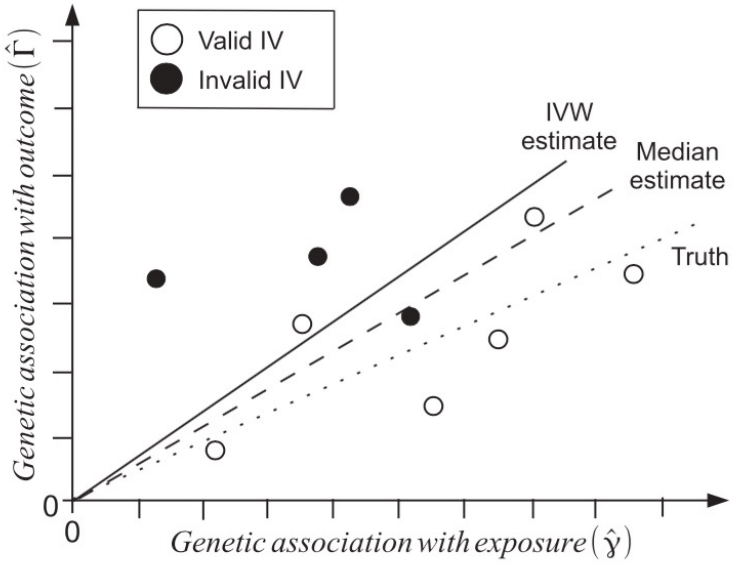
- Large standard error.



Bowden et al, IJE 2015

Median method

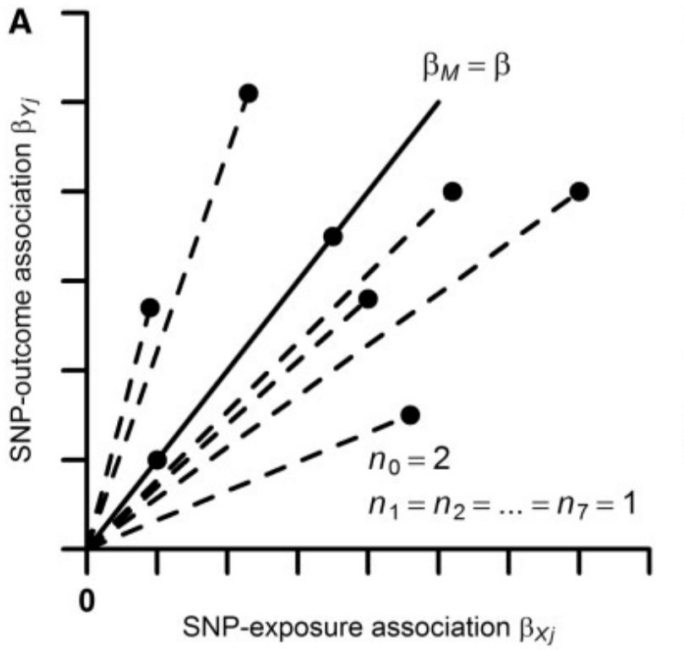
- Median of ratio estimates of multiple SNPs: $\hat{\theta} = median(\hat{\theta}_k)$
- Requires >50% of instruments to be valid.
- Variation: weighted median, penalized weighted median



Bowden et al, Genetic Epidemiology 2016

Mode-based estimator

- Mode of estimated distribution of $\hat{\theta}_k$
- Requires ZEro Modal Pleiotropy Assumption (ZEMPA): the most common value of the pleiotropic effect is 0
- Variation: weighted mode



2 SNPs: $u_y = 0$
1 SNP: $u_y = 0.5$
1 SNP: $u_y = 0.3$
1 SNP: $u_y = 0.1$

Hartwig, Davey Smith and Bowden, IJE 2017

MR by robust loss functions

- MR-RAPS: robust adjusted profile score

- MR-cML: maximum likelihood with L0 regularization

$$\min_{\theta, b_{Xi}, r_i, 1 \le i \le m} -L\left(\theta, \{b_{Xi}, r_i\}; \left\{\widehat{\beta}_{Xi}, \widehat{\beta}_{Yi}, \widehat{\sigma}^2_{Xi}, \widehat{\sigma}^2_{Yi}\right\}\right) \text{subject to}$$

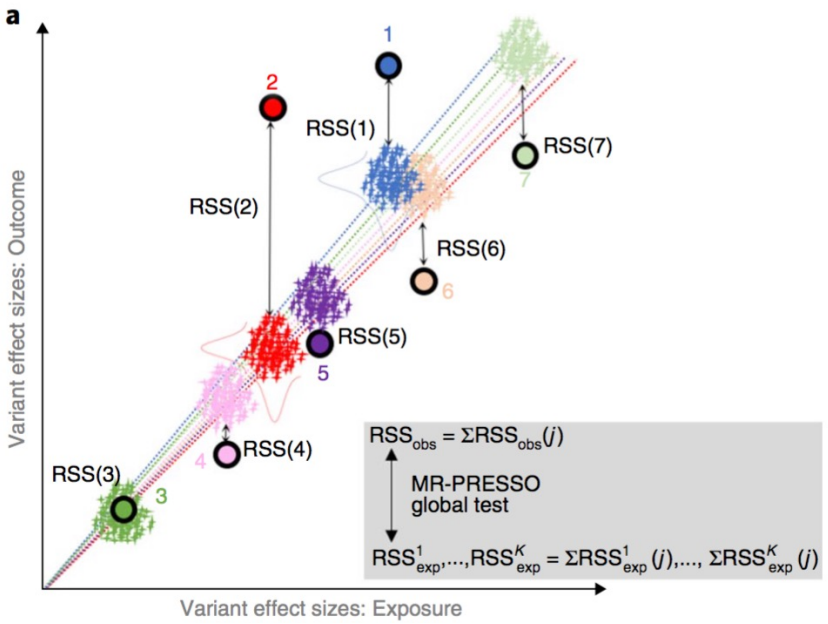
$$\sum_{i=1}^{m} I(r_i \neq 0) = K.$$

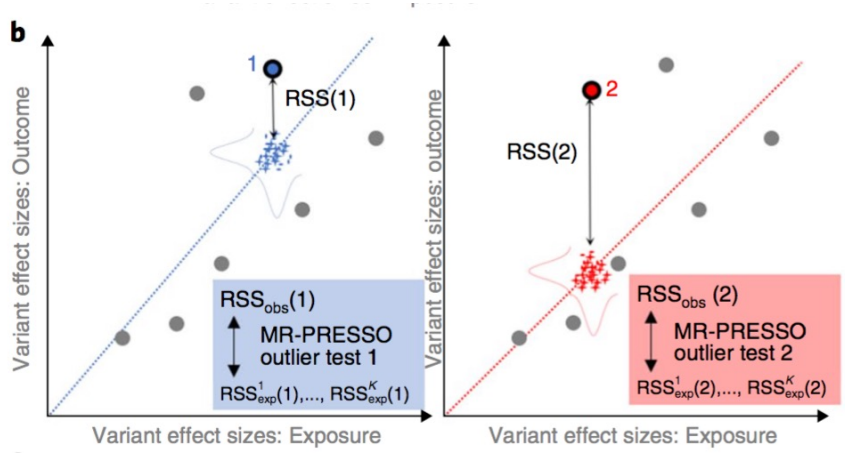
Zhao et al, 2020
Xue et al, 2021

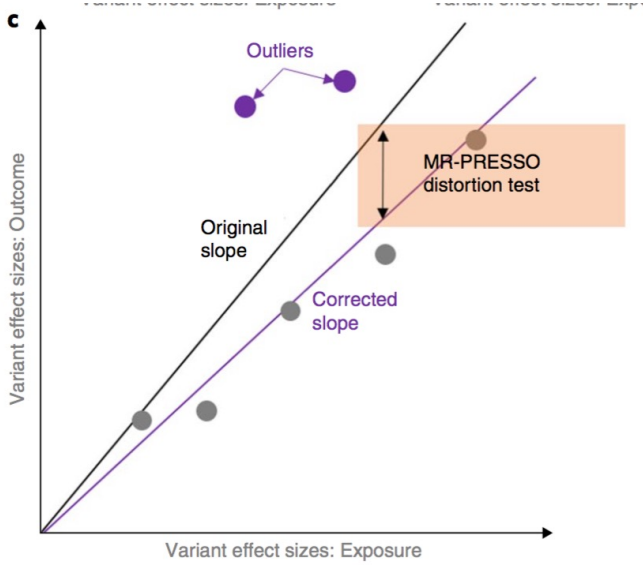
MR-PRESSO

Leave-one-out analysis

$$\mathrm{RSS}_{\mathrm{obs}} = \sum_j \mathrm{RSS}_{\mathrm{obs}}(j) = \sum_j (\hat{\Gamma}_j - \hat{\beta}_{-j}\hat{\gamma}_j)^2$$







Verbaneck et al, Nat Genet 2018

MRMix: Mendelian randomization using mixture models (motivation)

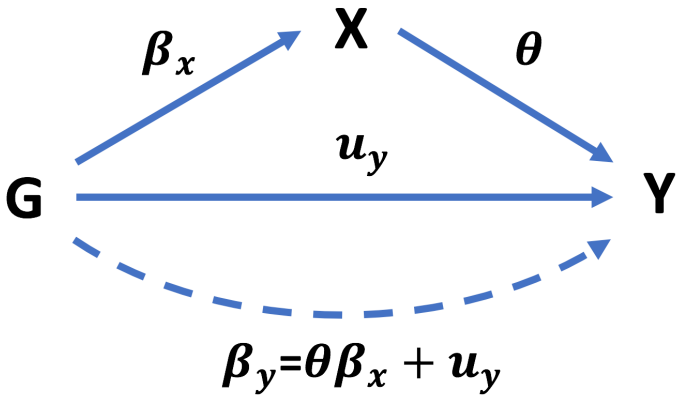
- Only across valid instruments, the proportionality assumption is satisfied,

$$\beta_y = \theta \beta_x$$

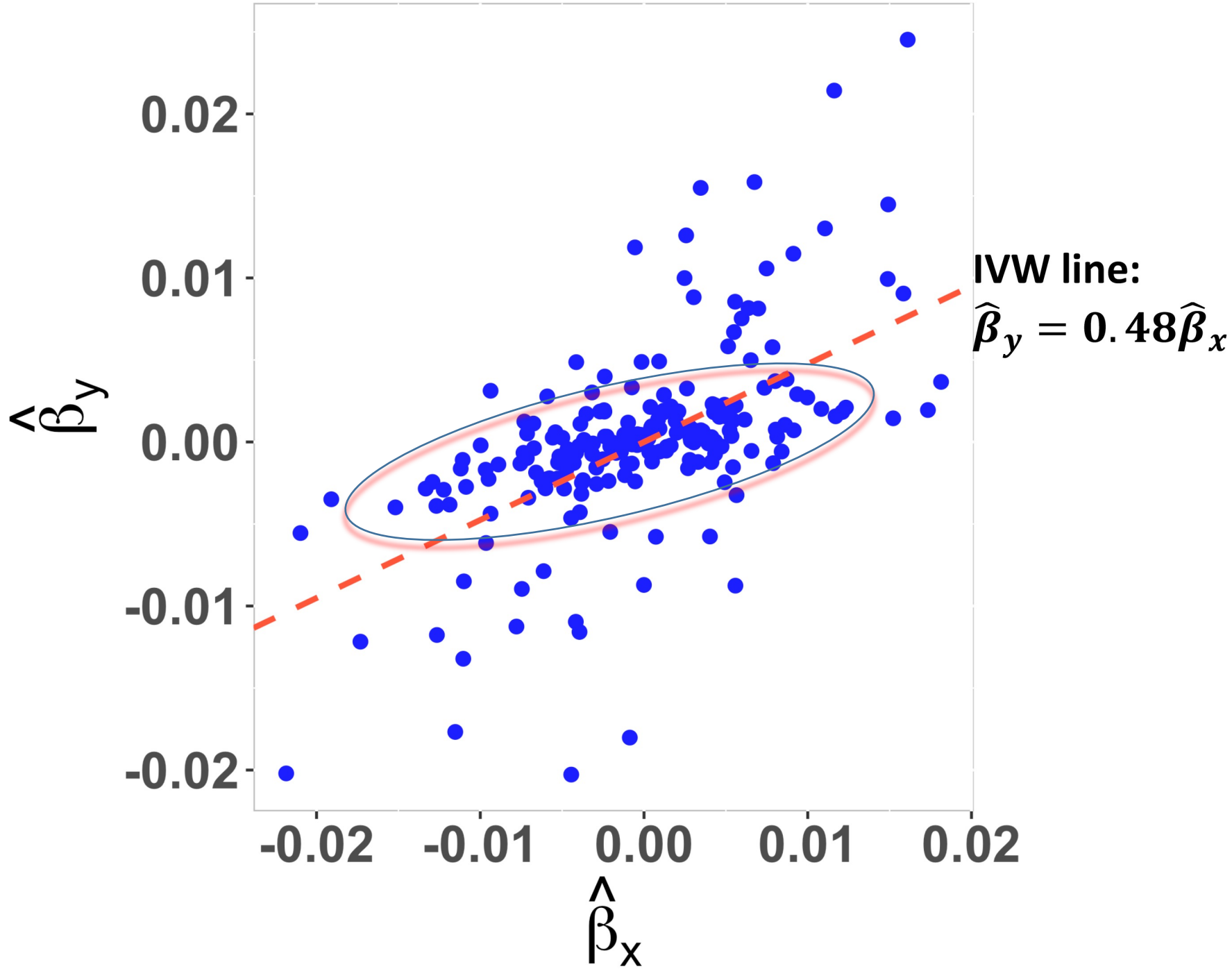
- For other potential instruments, there will "extra" variation around the proportional relationship

$$\beta_y = \theta \beta_x + u_y$$

- Use multitude of instruments available to distinguish this pattern and hence estimate causal effect



Qi and Chatterjee, Nat Commun 2019


0.02
0.01
0.00
-0.01
-0.02
$\hat{\beta}_y$
-0.02
-0.01
0.00
0.01
0.02
$\hat{\beta}_x$
IVW line:
$\hat{\beta}_y = 0.48\hat{\beta}_x$

MR analysis using mixture model for effect size distribution (MRMix)

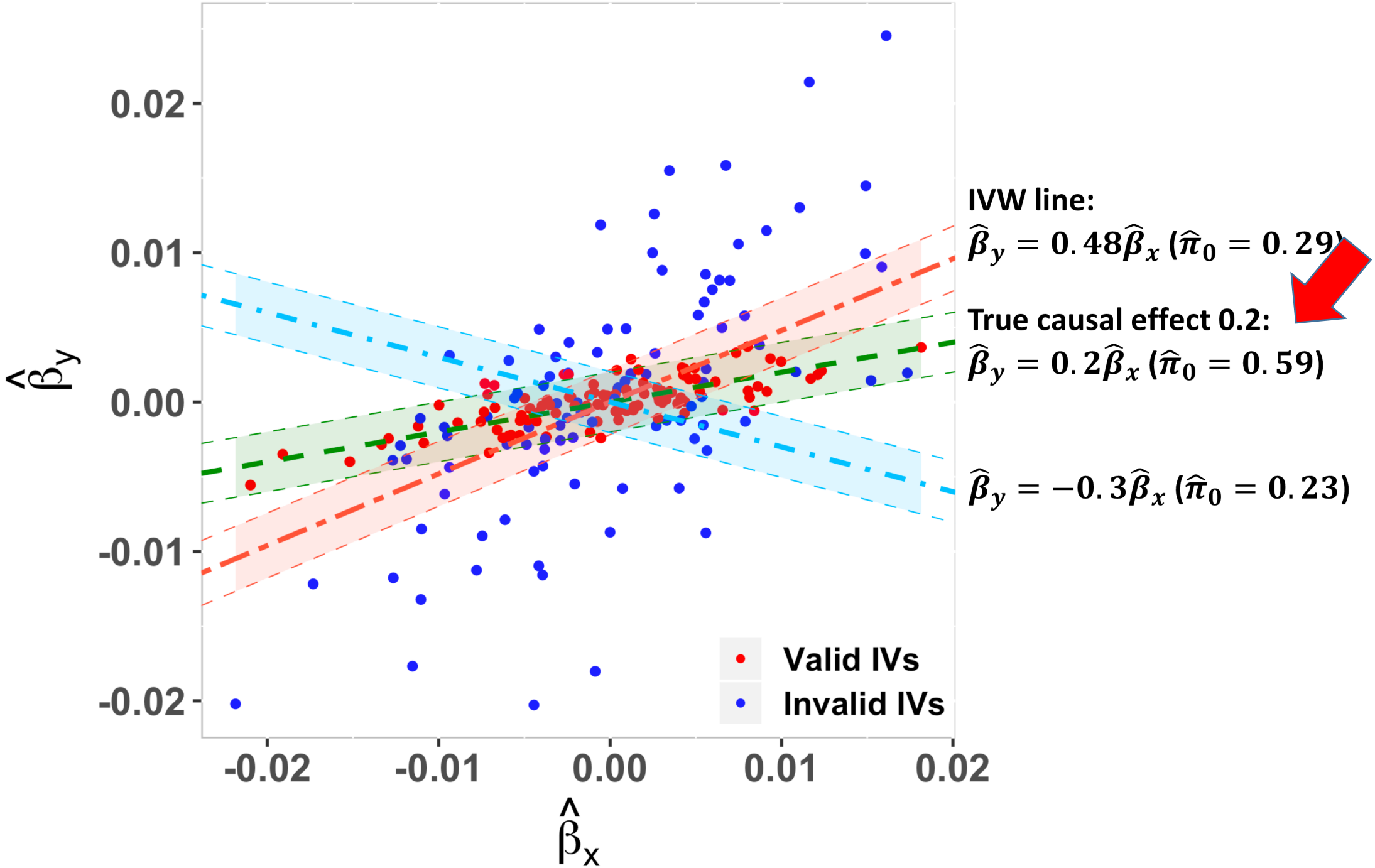
- Use the representation (Qi and Chatterjee, Nat Commun 2019)

$$\hat{\beta}_y\text{-}\theta\ \hat{\beta}_x \sim \pi N(0, \sigma_0^2\ ) + (1-\pi)\mathrm{N}(0,\ \sigma^2)$$

Normal distribution with known variance corresponding to valid instruments

→ Exists only when θ represents true causal effect

- Search for θ that maximizes the probability concentration of the null component


0.02
0.01
0.00
-0.01
-0.02
$\hat{\beta}_y$
-0.02
-0.01
0.00
0.01
0.02
$\hat{\beta}_x$
Valid IVs
Invalid IVs
IVW line:
$\hat{\beta}_y = 0.48\hat{\beta}_x$ ($\hat{\pi}_0 = 0.29$)
True causal effect 0.2:
$\hat{\beta}_y = 0.2\hat{\beta}_x$ ($\hat{\pi}_0 = 0.59$)
$\hat{\beta}_y = -0.3\hat{\beta}_x$ ($\hat{\pi}_0 = 0.23$)

Testing for pleiotropy and sensitivity analysis

- Cochran's Q (IVW) and Rucker's Q (Egger)[1]

$$Q = \sum_{j=1}^{L} \left(\frac{\hat{\Gamma}_j}{\sigma_{Yj}} - \frac{\hat{\beta}_{IVW}\gamma_j}{\sigma_{Yj}} \right)^2 = \sum_{j=1}^{L} w_j(\hat{\beta}_j - \hat{\beta}_{IVW})^2$$

- MR-Egger intercept: test for directional pleiotropy[2]

$$\hat{\beta}_y = \theta_0 + \theta\hat{\beta}_x$$

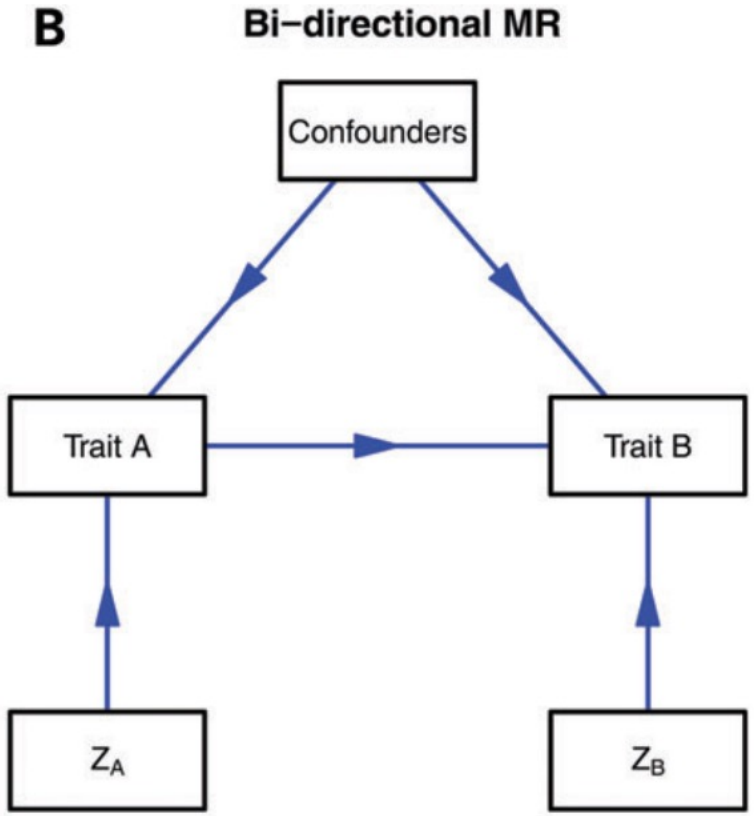
- MR-PRESSO[3]: leave-one-out

[1]Bowden et al, Stat Med 2017
[2]Bowden et al, IJE 2015
[3]Verbaneck et al, Nat Genet 2018

Bidirectional MR

1. Use the instruments for trait A (Z_A) to estimate the effect of trait A on trait B.

2. Use the instruments for trait B to estimate the effect of trait B on trait A.



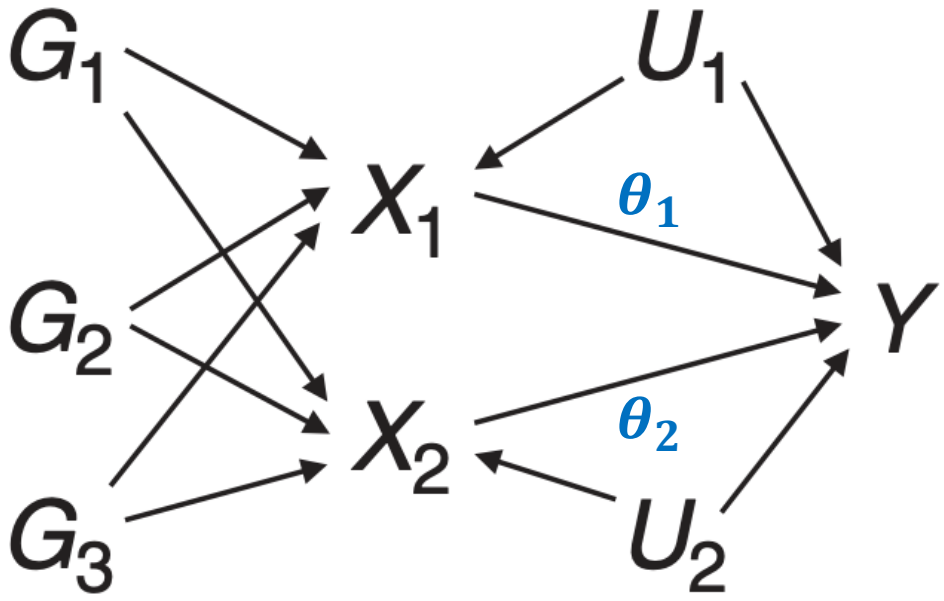
Davey Smith 2014

Multivariable MR

- Univariate MR analysis will be biased by horizontal pleiotropy
- Multivariable MR fits a multivariable linear regression

$$\hat{\beta}_y = \theta_1 \hat{\beta}_{x_1} + \theta_2 \hat{\beta}_{x_2} + \epsilon$$

- Simultaneously estimate θ_1 and θ_2

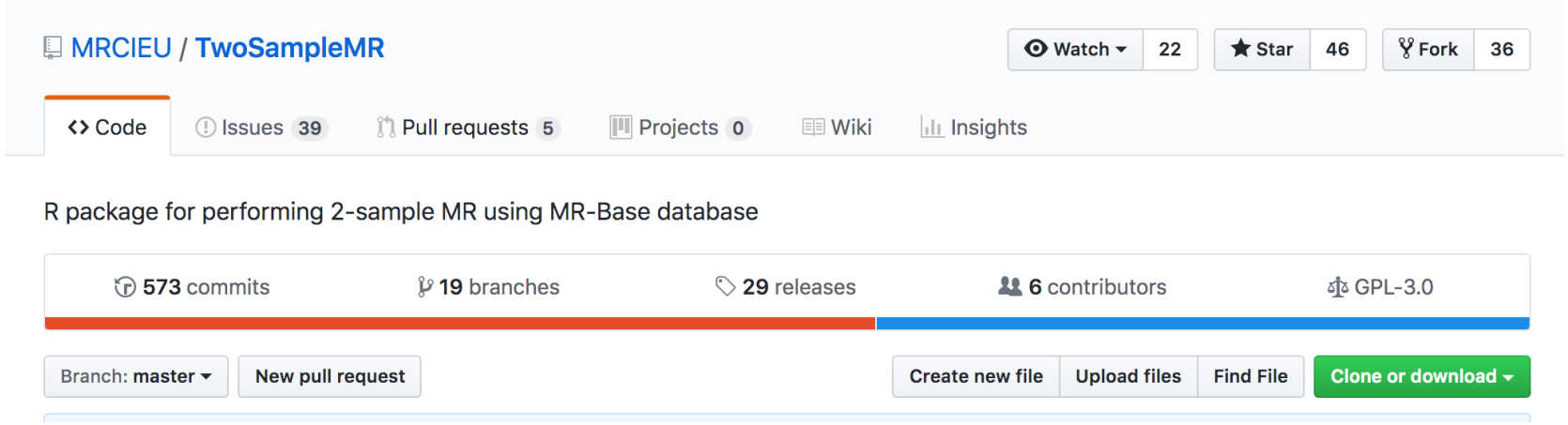


Other MR methods

- Debiased IVW: https://onlinelibrary.wiley.com/doi/full/10.1111/biom.13732
- MR-Horse and MVMR-Horse: https://www.biorxiv.org/content/10.1101/2023.05.30.542988v1
- CAUSE: https://www.nature.com/articles/s41588-020-0631-4
- …

Software and resources

R packages for MR analysis



MendelianRandomization: Mendelian Randomization Package

Encodes several methods for performing Mendelian randomization analyses with summarized data. Summarized data on genetic assoc obtaining causal estimates using instrumental variable methods.

Version:	0.4.1
Depends:	R (≥ 3.0.1), methods
Imports:	knitr, rmarkdown, plotly (≥ 3.6.0), ggplot2 (≥ 1.0.1), robustbase (≥ 0.92-6), Matrix (≥ 1.2), iterpc (≥ 0.3), rjson
Published:	2019-04-09
Author:	Olena Yavorska James Staley
Maintainer:	Stephen Burgess <sb452 at medschl.cam.ac.uk>
License:	GPL-2 | GPL-3
NeedsCompilation:	no
Materials:	NEWS
In views:	MetaAnalysis
CRAN checks:	MendelianRandomization results

MR-Base

2-sample Mendelian Randomisation

Home | MR-Base web app | R package | PheWAS | Publications

MR-base is a database and analytical platform for Mendelian randomization being developed by the MRC Integrative Epidemiology Unit at the University of Bristol.

You can either use the web application or our TwoSampleMR R package.

Launch MR-Base webapp | R package | *beta!*

Note - by clicking the "Launch MR-Base webapp" button you consent to the use of a cookie which enables us to ensure you have consented to the terms and conditions of data access. Information about how to control or delete cookies can be found at www.aboutcookies.org

MR-Base paper published

The MR-Base paper has now been published in eLife. See the publications page for details.

Telomeres paper published

Our paper reporting Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases has been published in Jama Oncology. See the publications page to access supporting data.

Further resources

- George Davey Smith explaining Mendelian randomization: https://www.youtube.com/watch?v=rjMwcTttKoQ&t=286s
- A review paper: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5711966/
- TwoSampleMR package: https://github.com/MRCIEU/TwoSampleMR
- MendelianRandomization package: https://cran.r-project.org/web/packages/MendelianRandomization/index.html
- MR-Base: http://www.mrbase.org/