Sensitivity analyses in Mendelian randomization studies

Dr Daniel Hwang

Institute for Molecular Bioscience, The University of Queensland

(Some slides adapted from Prof David Evans')

THE UNIVERSITY OF QUEENSLAND AUSTRALIA | IMB

Acknowledgment of **Country**

The University of Queensland (UQ) acknowledges the Traditional Owners and their custodianship of the lands on which we meet.

We pay our respects to their Ancestors and their descendants, who continue cultural and spiritual connections to Country.

We recognise their valuable contributions to Australian and global society.

Recap

- Mendelian randomization is a technique that uses genetically informative observational data to inform causality.

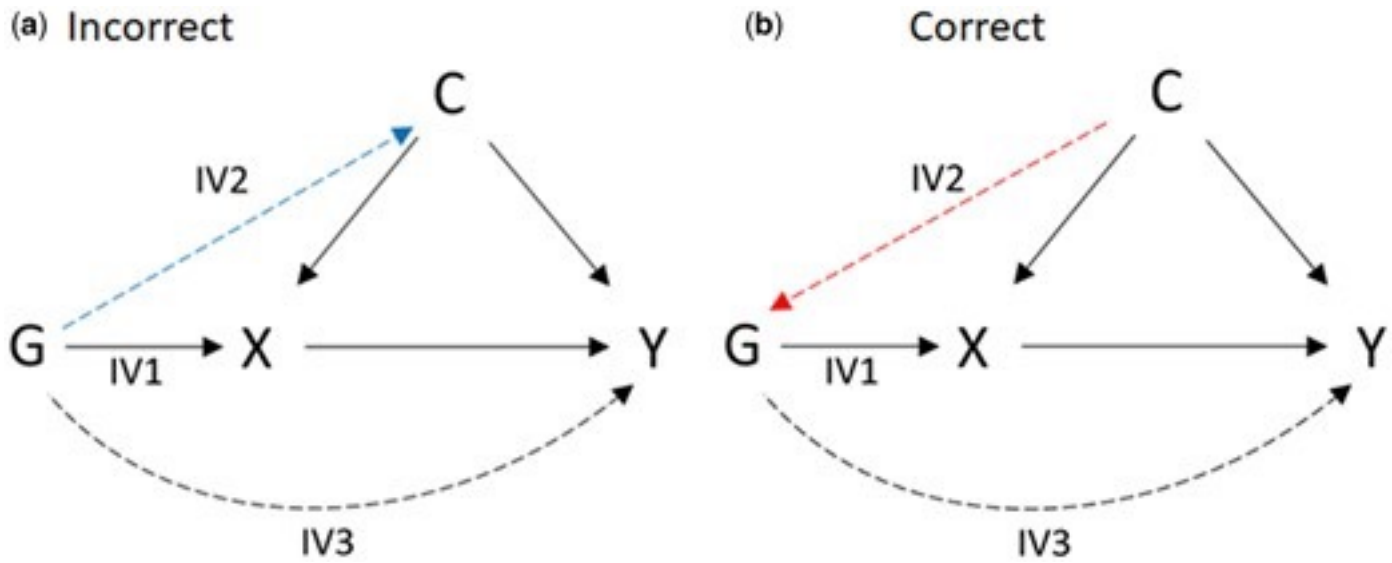
- Three core assumptions:
  (1) Relevance assumption: SNP is associated with the exposure
  (2) Independence assumption: SNP is NOT associated with confounding variables
    - **There are no confounders of the association between the instrumental variables (IVs) and the outcome.**
  (3) Exclusion restriction: SNP ONLY associated outcome through the exposure

- Pleiotropy: Genetic variant influences more than one trait

- One-sample MR is where the SNP, exposure and outcome are all available in the same study.
- Two-sample MR is where the SNP-exposure association is measured in one study and the SNP outcome association is measured in a second study.

Independence assumption (Second MR assumption)

There are no confounders of the association between the instrumental variables (IVs) and the outcome. As genetic variants are determined at conception it is not possible for them to be affected by confounders of exposure-outcome associations. When referring to the second MR assumption, factors that could influence the genetic variants and outcome include population stratification or structure, intergenerational (dynastic) effects and assortative mating. (MR Dictionary)



Carter & Anderson (2024) *International Journal of Epidemiology*

Delta method to estimate SE of Wald ratio

$BETA_{SNP\text{-}exposure} = X$

$BETA_{SNP\text{-}outcome} = Y$

$Var(BETA_{Wald\ ratio}) = Var(Y/X)$

$\approx Var(Y)/X^2 + (Y^2/X^4)*Var(X) - 2*(Y/X^3)*Cov(X,Y)$ <= This is based on the Delta method

$\approx Var(Y)/X^2$

$SE(BETA_{Wald\ ratio}) \approx \sqrt{Var(Y)/X^2}$

$= SE(Y)/X$

$= SE_{SNP\text{-}outcome}/BETA_{SNP\text{-}exposure}$

Outline

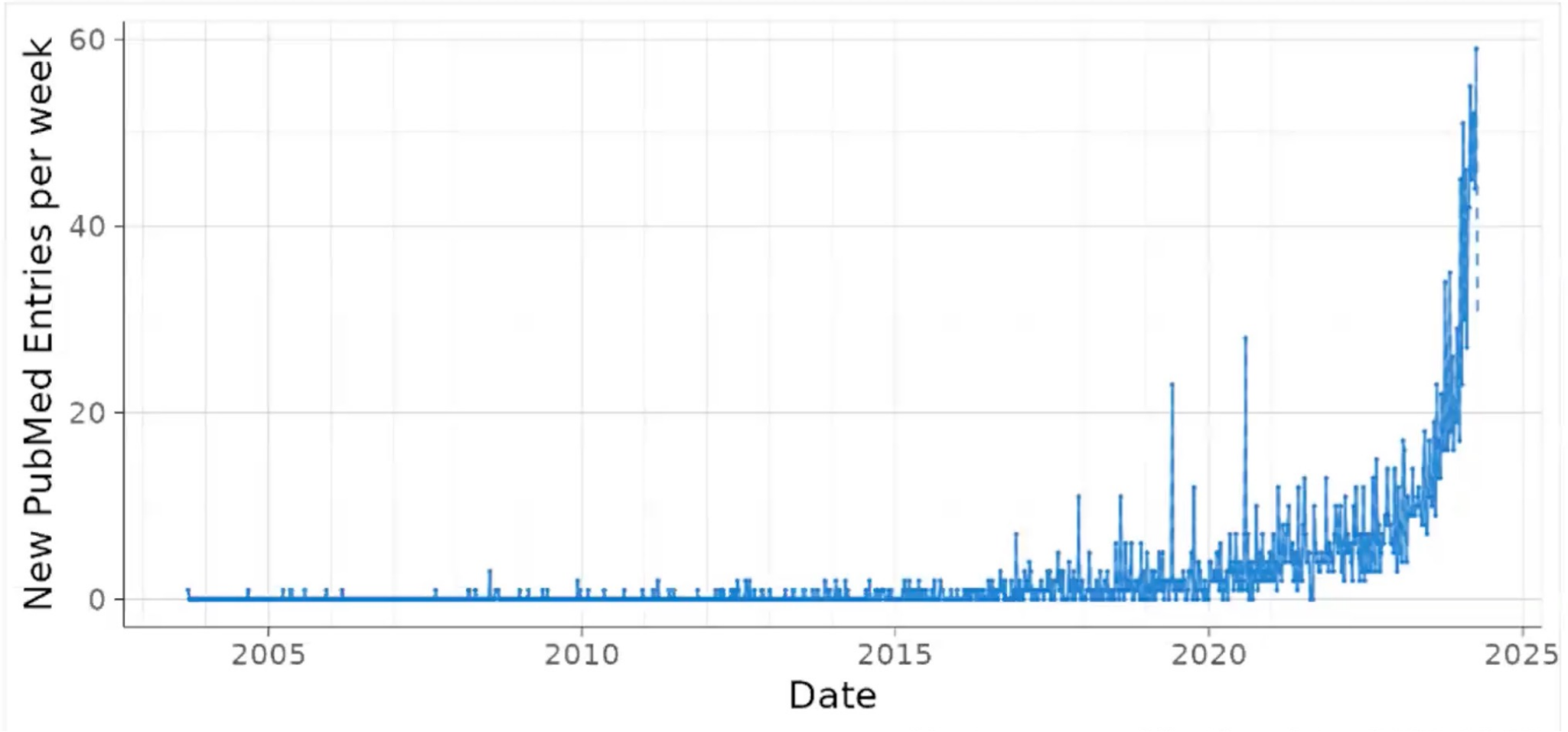
Set up a two-sample MR analysis using multiple genetic variants

Problems with pleiotropy and heterogeneity

15 minutes Break

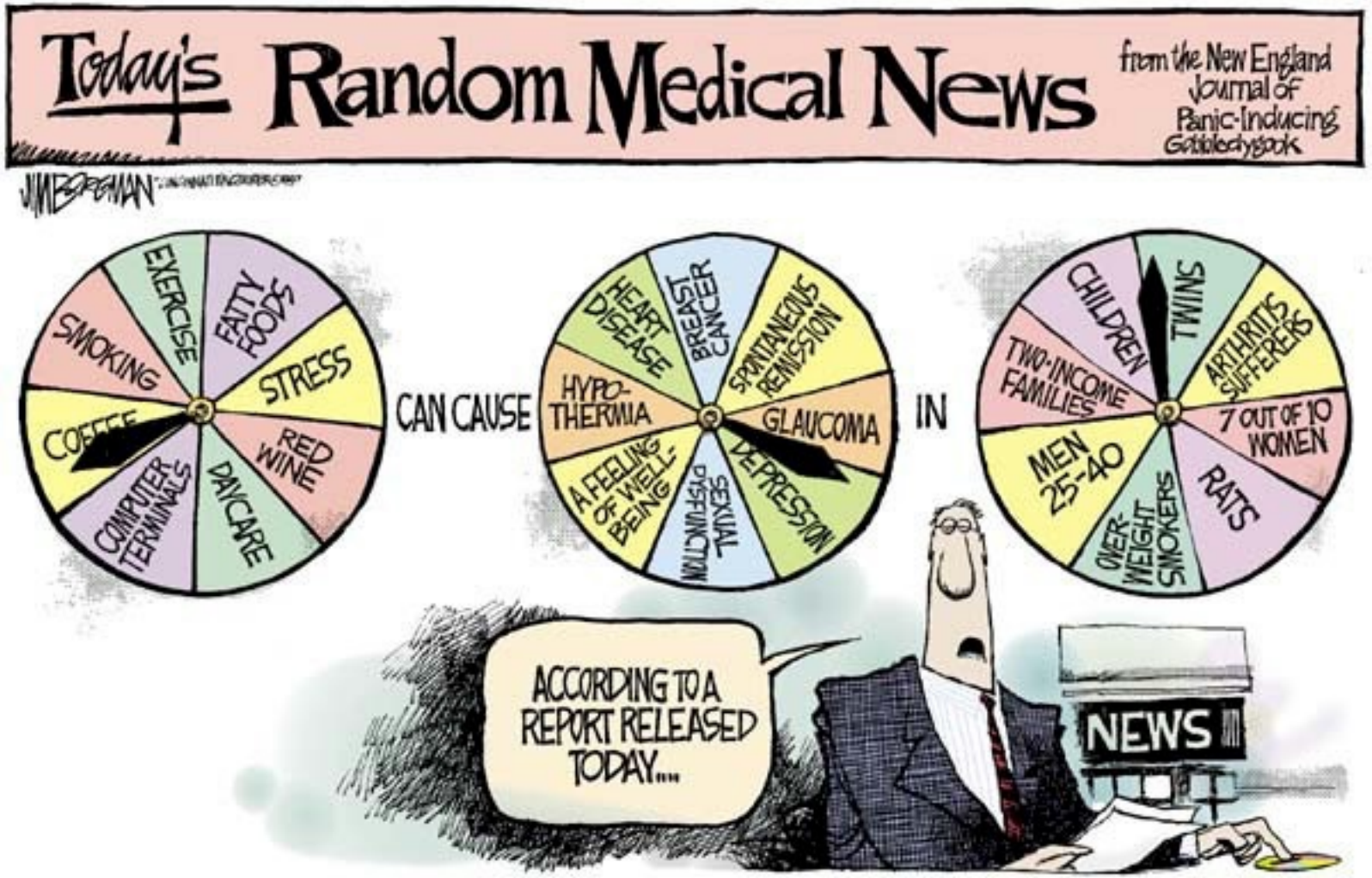
Methods for handling pleiotropy

PubMed search for Mendelian randomi[z/s]ation (title only)



Up to 2024-04-13

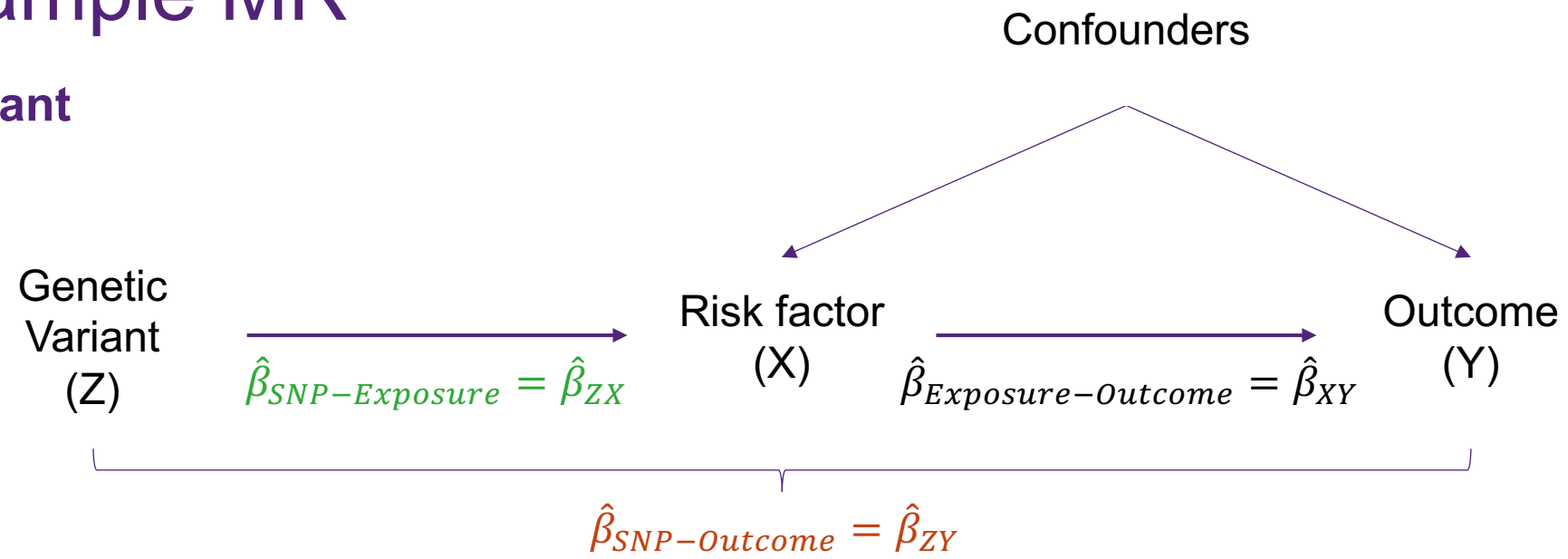
MR studies today have become yesterday's observational studies



Jim Borgman (1997)

Two-sample MR

Single variant



Causal effect ($\hat{\beta}_{XY}$) by Wald estimator: $\frac{\hat{\beta}_{SNP-Outcome}}{\hat{\beta}_{SNP-Exposure}}$

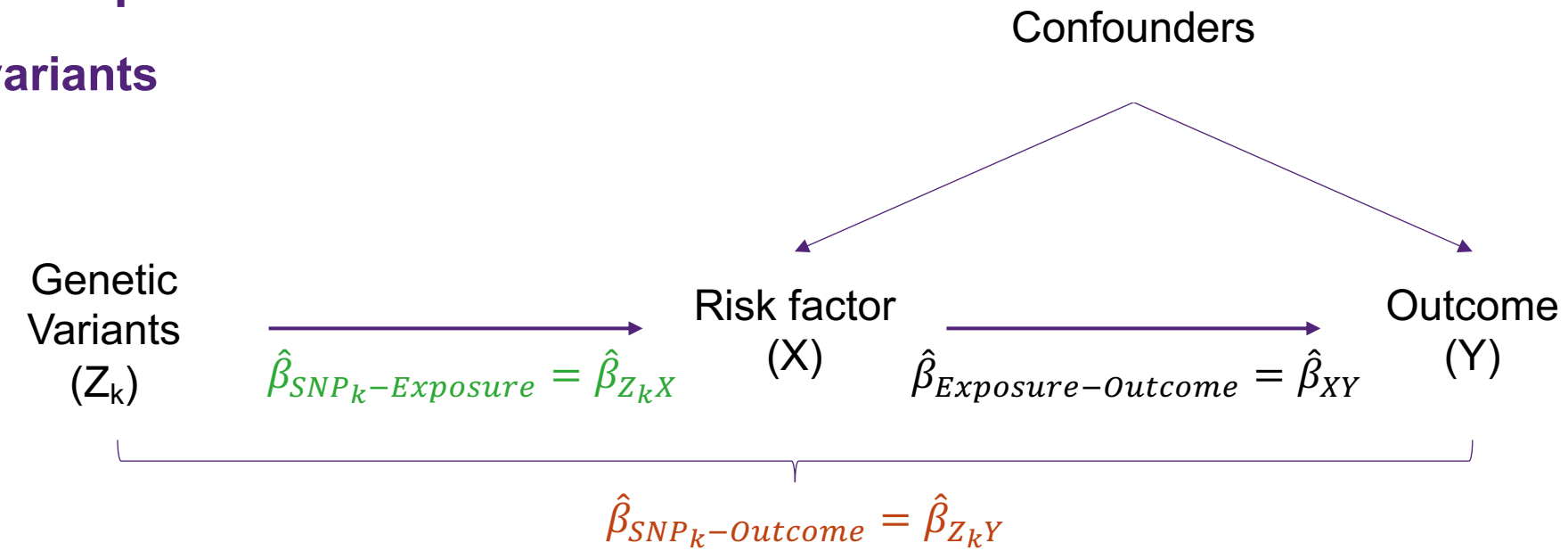
Standard error ($\hat{\sigma}_{XY}$) by Delta method: $\frac{\sigma_{SNP-Outcome}}{\hat{\beta}_{SNP-Exposure}}$

$\hat{\beta}_{SNP-Outcome} = \hat{\beta}_{SNP-Exposure} \times \hat{\beta}_{Exposure-Outcome}$

Can be estimated in different samples (e.g. two-sample MR)

Two-sample MR

Multiple variants



Causal effect by Wald estimator:

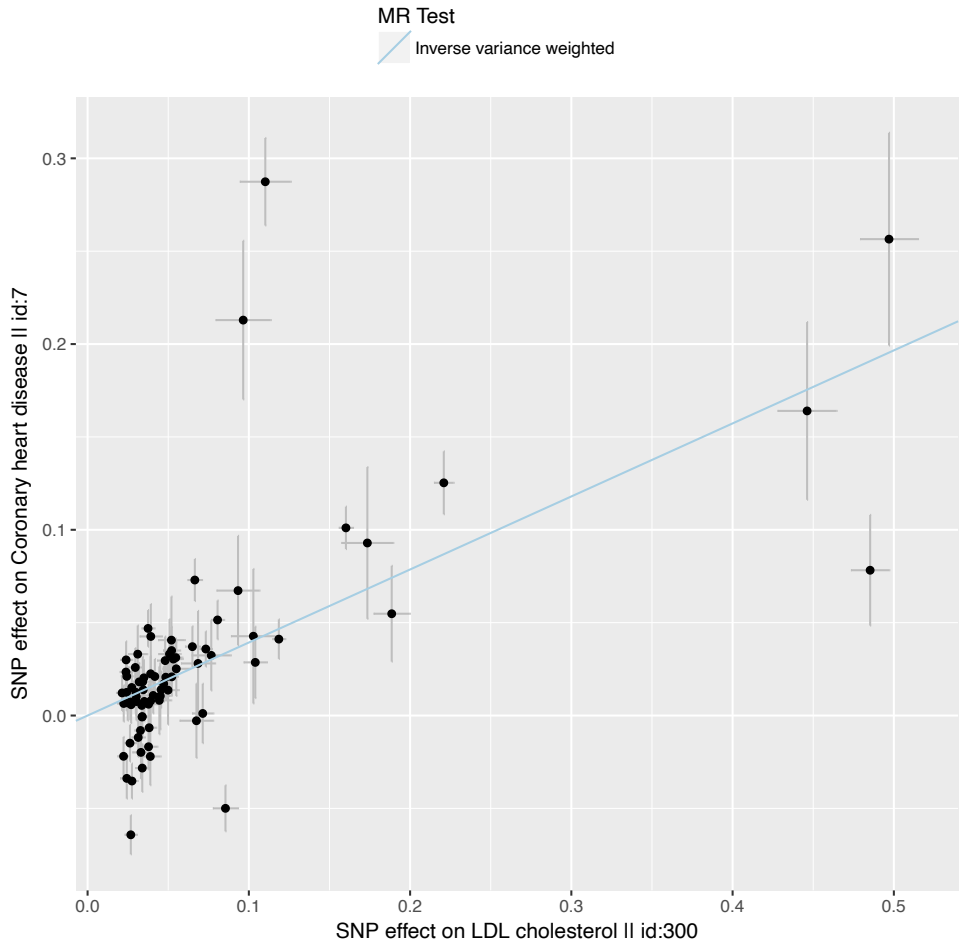
$$\hat{\beta}_{XY_k} = \frac{\hat{\beta}_{SNP_k-Outcome}}{\hat{\beta}_{SNP_k-Exposure}}$$

Inverse variance weighted (IVW) average causal effect:

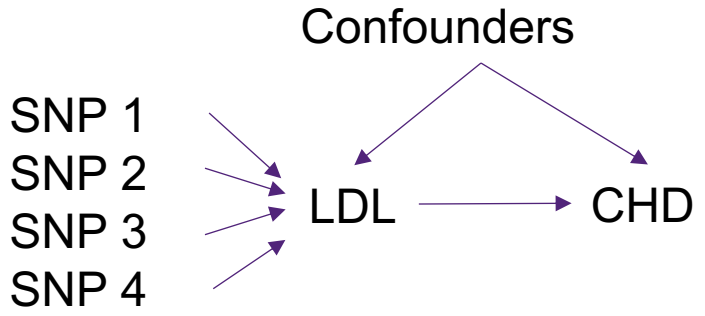
$$\hat{\beta}_{IVW} = \frac{\sum_{k=1}^{K} \hat{\beta}_{XY_k} w_k}{\sum_{k=1}^{K} w_k}$$

Where $w_k = \frac{1}{\boldsymbol{var}(\hat{\beta}_{XY_k})} = \frac{1}{\hat{\sigma}_{XY_k}^2}$ is the inverse variance of the causal effect estimated from the k^{th} genetic variant

Fixed effects IVW-MR and weighted linear regression



- IVW is equivalent to a weighted regression of SNP-outcome effects on SNP-exposure effects passing through the origin
- The weights are the inverse of the variance of the individual causal effect estimates, i.e., $\frac{1}{\boldsymbol{var}(\widehat{\beta}_{XY_k})}$
- The slope is the estimate of the causal effect



Assumptions for two-sample MR

If:

- The K genetic variants are perfectly uncorrelated (SNPs not in LD) and do not interact
- The two samples are homogenous (same underlying populations)
- No sample overlap (this could be relaxed if all IVs are “valid”)
- Constant causal effect at each level of the exposure

Then two-sample MR can consistently estimate the true causal effect.

Two-sample MR is still vulnerable to weak instrument bias

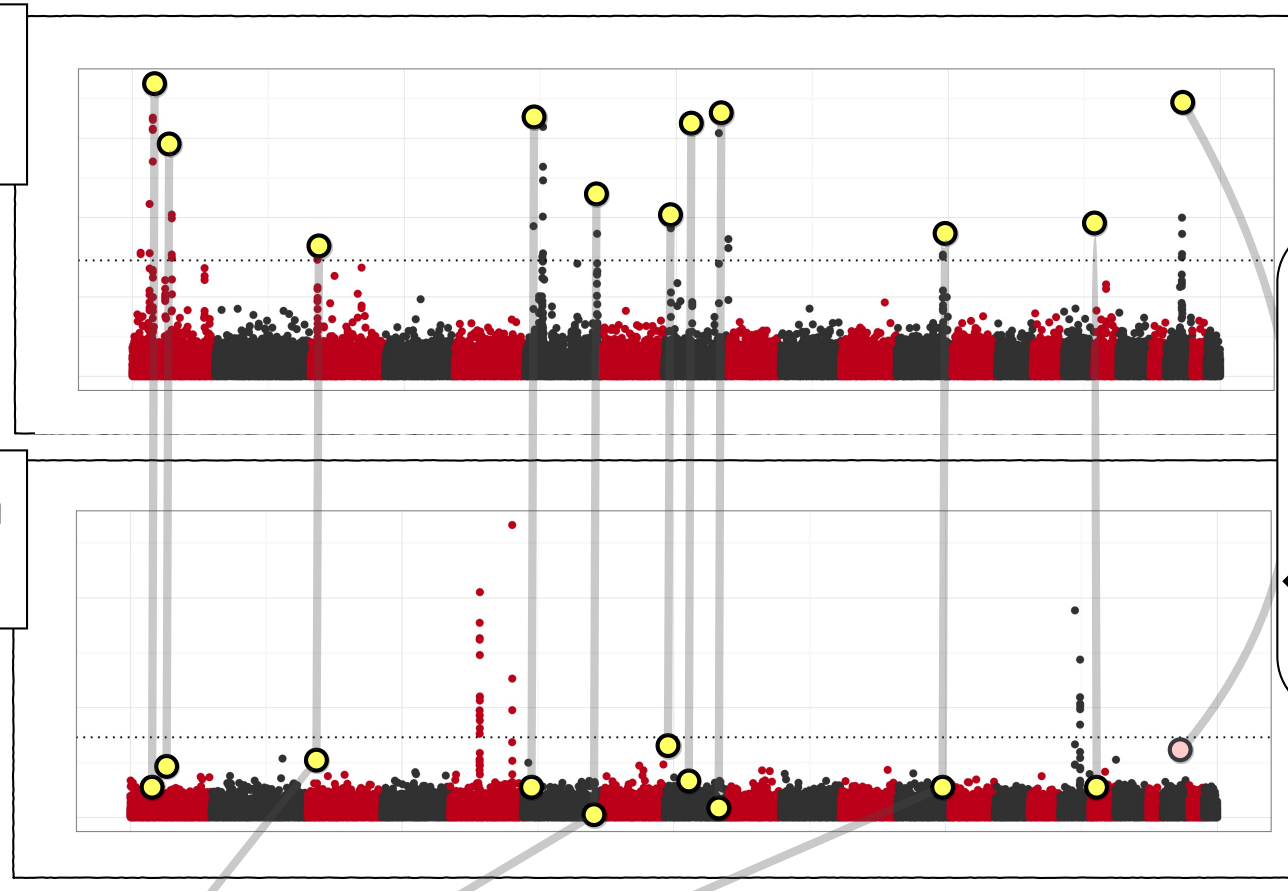
- Bias towards the null effect, not the observational estimate
- If approximate F-statistic ($\hat{\beta}^2_{Z_kX}/\sigma^2_{Z_kX}$) is greater than 10, then the expected dilution $\hat{\beta}_{XY_k}$ of towards zero is less than 10%

Performing two-sample MR with summary statistics

A convenient approach when sharing individual level data is impractical

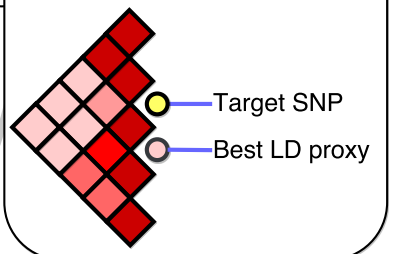
Obtain instruments from exposure GWAS

Extract SNP effects from outcome GWAS



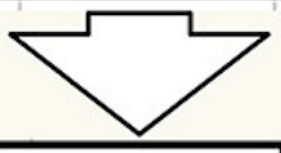
LD Proxies

If an exposure instrument is not available in the outcome GWAS then look for LD proxies in 1000 genomes



Harmonise exposure and outcome effects

	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**

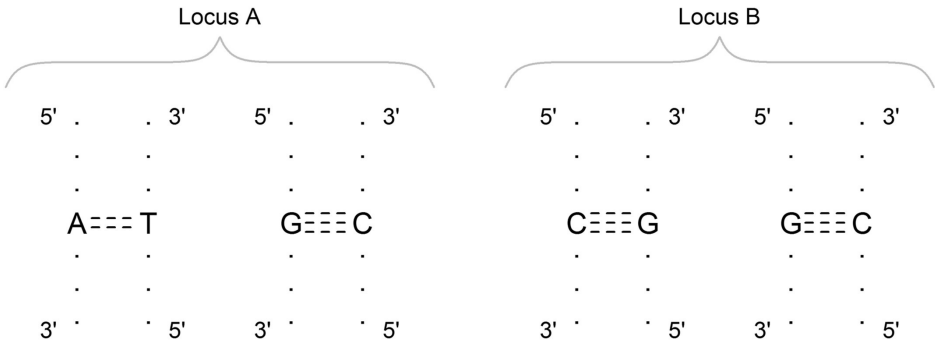


	Exposure GWAS				Outcome GWAS			
SNP	*Effect*	*Effect allele*	*Other allele*	*Effect allele frequency*	*Effect*	*Effect allele*	*Other allele*	*Effect allele frequency*
rs12345	0.132	A	G	0.28				
rs23456	-0.485	G	T	0.41				
rs34567	0.203	G	C	0.11				

The issue of strand (palindromic variant)

	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**

Palindromic



(Not palindromic) (Palindromic)

	Locus A	Locus B
Genotype of the forward strand (5' > 3')	A/G	C/G
Genotype of the reverse strand (3' > 5')	T/C	G/C

Exposure GWAS and outcome GWAS may be based on the genotypes of different strands. When there are palindromic SNPs, simply merging datasets based on effect alleles may result in the effect being the opposite.

Hartwig et al 2016

Harmonise exposure and outcome effects

	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**

Palindromic

	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**

Strand issue exercise (5 mins)

SNP	Study 1 alleles	Study 1 allele freq	Study 2 alleles	Study 2 allele freq	Verdict?
rs1	A/G	0.2	A/G	0.2	
rs2	G/T	0.3	T/G	0.72	
rs3	G/C	0.65	G/C	0.62	
rs4	A/T	0.49	A/T	0.5	
rs5	A/T	0.12	A/T	0.89	
rs6	A/G	0.4	A/T	0.4	

WHAT IS THE PROBLEM WITH MR?

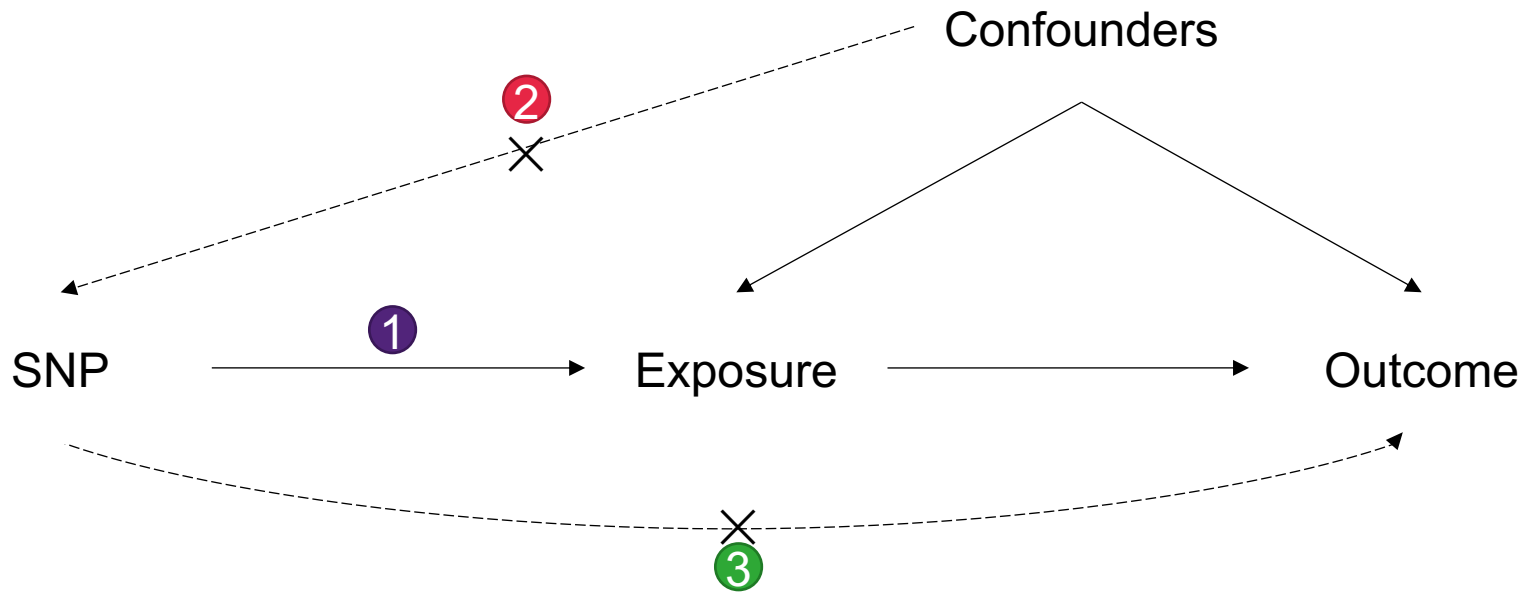
- MR uses genetic variants to test for causal relationships between phenotypic exposures and disease-related outcomes

- Due to the proliferation of GWAS, it is increasingly common for MR analyses to use large numbers of genetic variants

- Increased power but greater potential for **horizontal pleiotropy**

- Pleiotropic variants affect biological pathways other than the exposure under investigation and therefore can lead to biased causal estimates and false positives under the null

Three core MR assumptions



(1) Relevance assumption: SNP is associated with the exposure

(2) Independence assumption: SNP is NOT associated with confounding variables (population stratification, assortative mating, dynastic effects).

(3) Exclusion restriction: SNP ONLY associated outcome through the exposure

MR methods for handling horizontal pleiotropy

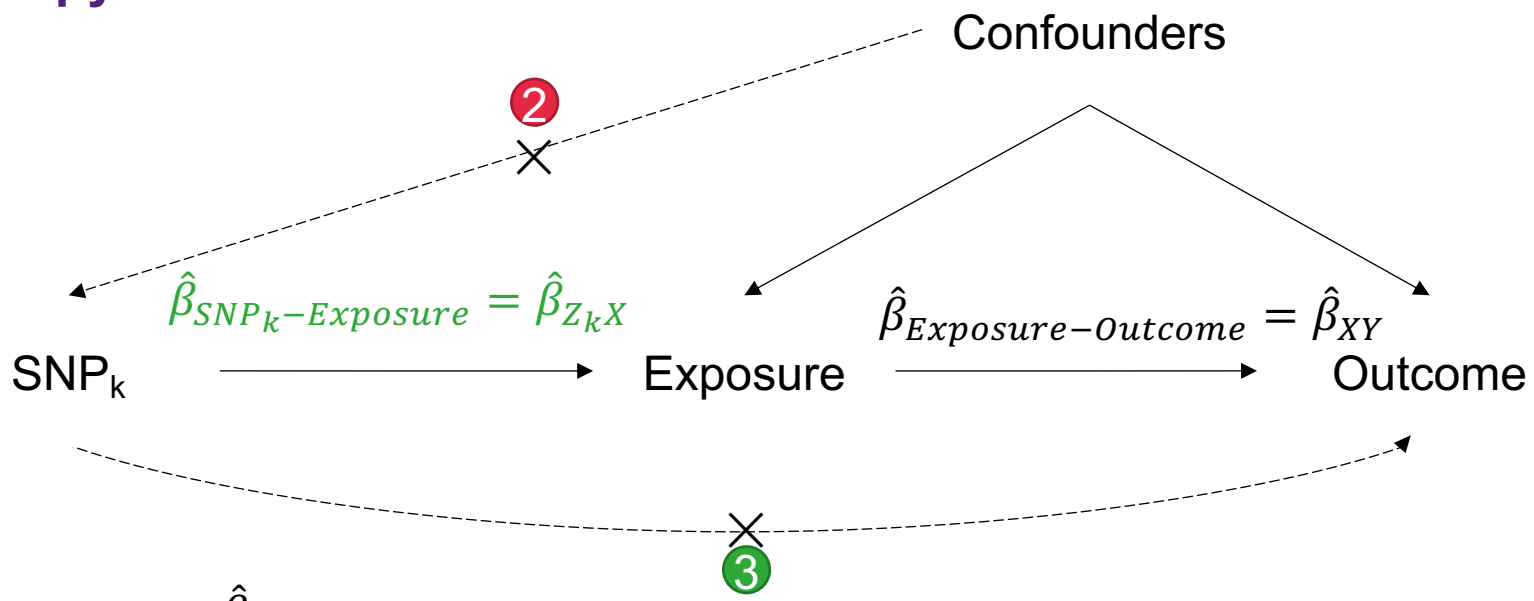
Table 1 | **List of MR estimation methods**

Category	Core IV assumption relaxed	Individual-level data	Summary data
'Basic' MR method	None	Wald ratio estimation, 2SLS regression analysis[a]	Wald ratio estimation, IVW[a,37]
Weak instrument robust methods	IV1; allows for weak instruments	LIML[26], allele score approaches[26]	MR RAPS[87], debiased IVW[187], MR GRAPPLE[88], NOME adjustment[188], two-sample AR[189]
Outlier/variant selection and removal	IV3; allows for balanced/sparse pleiotropy	Weighted median[190]	Weighted median[a,82]
Outlier/variant selection and removal	IV3; allows for (some) directional pleiotropy	sisVIVE[70], adaptive LASSO[71], weighted mode[190]	Weighted mode[a,83], MR LASSO[84], Steiger filtering[a,93], Welch-weighted Egger[94], contamination mixture[191], GSMR[79], MR-Clust[192], Bayesian MIMR[193], CIV[72]
Outlier/variant adjustment	IV3; allows for balanced pleiotropy	Limited approaches currently available	MR RAPS[87], MRCIP[194]
Outlier/variant adjustment	IV3; allows for (some) directional pleiotropy	Limited approaches currently available	MR TRYX[85], MR Robust[84], MR CAUSE[89], MR PRESSO[86], MR GRAPPLE[88], MRMix[195], MR-LDP[196], IMRP[197], regularization[198], MR-PATH (see preprint[199])
Estimation adjustment	IV3; allows for balanced pleiotropy	Limited approaches currently available	Debiased IVW[187]
Estimation adjustment	IV3; allows for (some) directional pleiotropy	Constrained IVs[72], multivariable MR[73]	MR Egger[90], multivariable MR[73,91], MR Link[200], hJAM[201], GIV[202], Bayesian network analysis[203], BMRE[204], BayesMR[205]
Environmental control adjustment	IV3; allows for (some) directional pleiotropy	MR GxE[75,76], MR GENIUS[77]	Limited approaches currently available

Sanderson et al. *Nat Rev Methods Primers* 2022

Two-sample MR

No direct pleiotropy



$\hat{\beta}_{SNP-Outcome} = \hat{\beta}_{SNP-Exposure} \times \hat{\beta}_{Exposure-Outcome}$

Causal effect by Wald estimator:

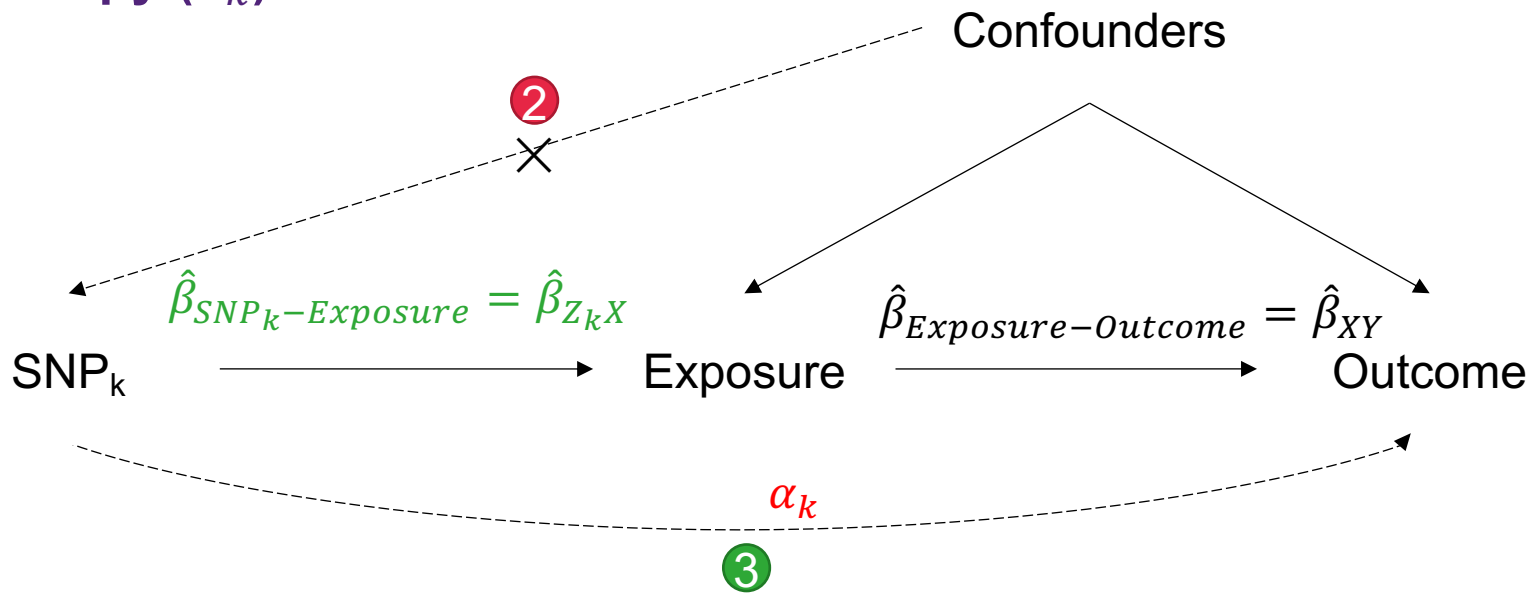
$$\hat{\beta}_{XY_k} = \frac{\hat{\beta}_{SNP_k-Outcome}}{\hat{\beta}_{SNP_k-Exposure}}$$

Inverse variance weighted (IVW) average causal effect:

$$\hat{\beta}_{IVW} = \frac{\sum_{k=1}^{K} \hat{\beta}_{XY_k} w_k}{\sum_{k=1}^{K} w_k}$$

Two-sample MR

With direct pleiotropy (α_k)



$$\hat{\beta}_{SNP-Outcome} = \hat{\beta}_{SNP-Exposure} \times \hat{\beta}_{Exposure-Outcome} + \alpha_k$$

Causal effect by Wald estimator:

$$\frac{\hat{\beta}_{SNP_k-Outcome}}{\hat{\beta}_{SNP_k-Exposure}} = \hat{\beta}_{Exposure-Outcome} + \frac{\alpha_k}{\hat{\beta}_{SNP_k-Exposure}}$$

Inverse variance weighted (IVW) average causal effect:

$$\frac{\sum_{k=1}^{K} \hat{\beta}_{XY_k} w_k}{\sum_{k=1}^{K} w_k} = \hat{\beta}_{Exposure-Outcome} + \text{Bias}\ (\alpha, \hat{\beta}_{SNP_k-Exposure})$$

Heterogeneity

We expect that each SNP represents an independent study, and each should give an unbiased (if imprecise) estimate of the causal effect of x on y.

Heterogeneity, where effect estimates are more different than expected due to standard errors, arises because at least some of the instruments are invalid.

Cochran's Q statistic (heterogeneity test):

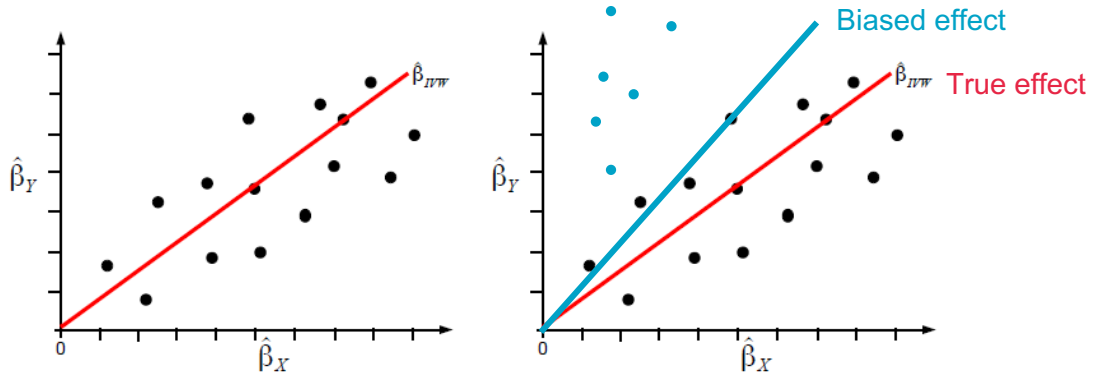
$$Q = \sum_{k=1}^{K} w_k \left(\hat{\beta}_{XY_k} - \hat{\beta}_{IVW}\right)^2$$

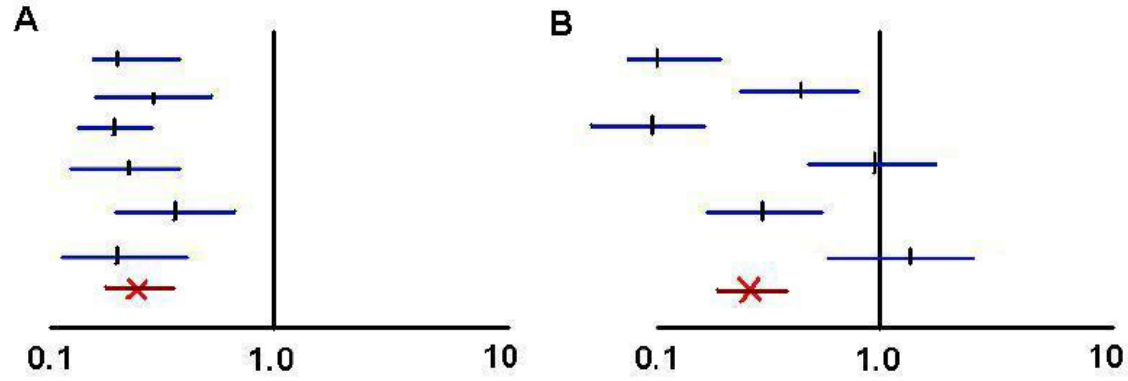
Where w_k is the weight (i.e. inversed variance) of the causal estimate at SNP *k.*

If MR model is correct, *Q* follows a χ^2 distribution with expected value *K*-1.

If *Q* is larger than *K*-1, then the estimates exhibit over-dispersion.

- SNPs are valid instruments
- SNPs associated with outcome via an independent pathway.





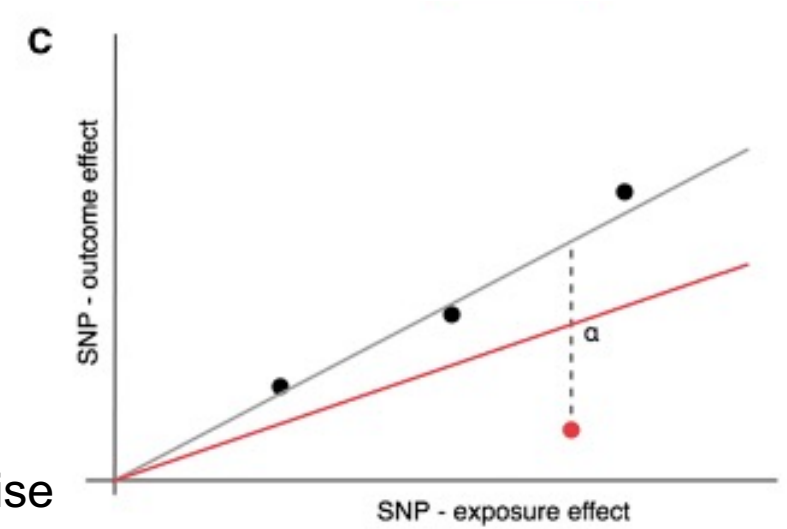
N=6 instruments
(A): No heterogeneity; all variants estimating the same quantity: $Q \approx 5$
(B): Heterogeneity; variants estimating different quantities: $Q >> 5$

Bowden et al. Am J Epidemiol. (2018) 187(12):2681-2685

Accounting for heterogeneity

Option 1: Remove outliers

- Some SNPs might contribute to the majority of the heterogeneity.
- If we assume these are the invalid instruments, then the IVW estimate excluding them should be less biased.



However – beware of:

- Cherry picking – remove outliers will artificially provide a more precise estimate
- What if the outlier is the only valid instrument, and all the others are invalid?

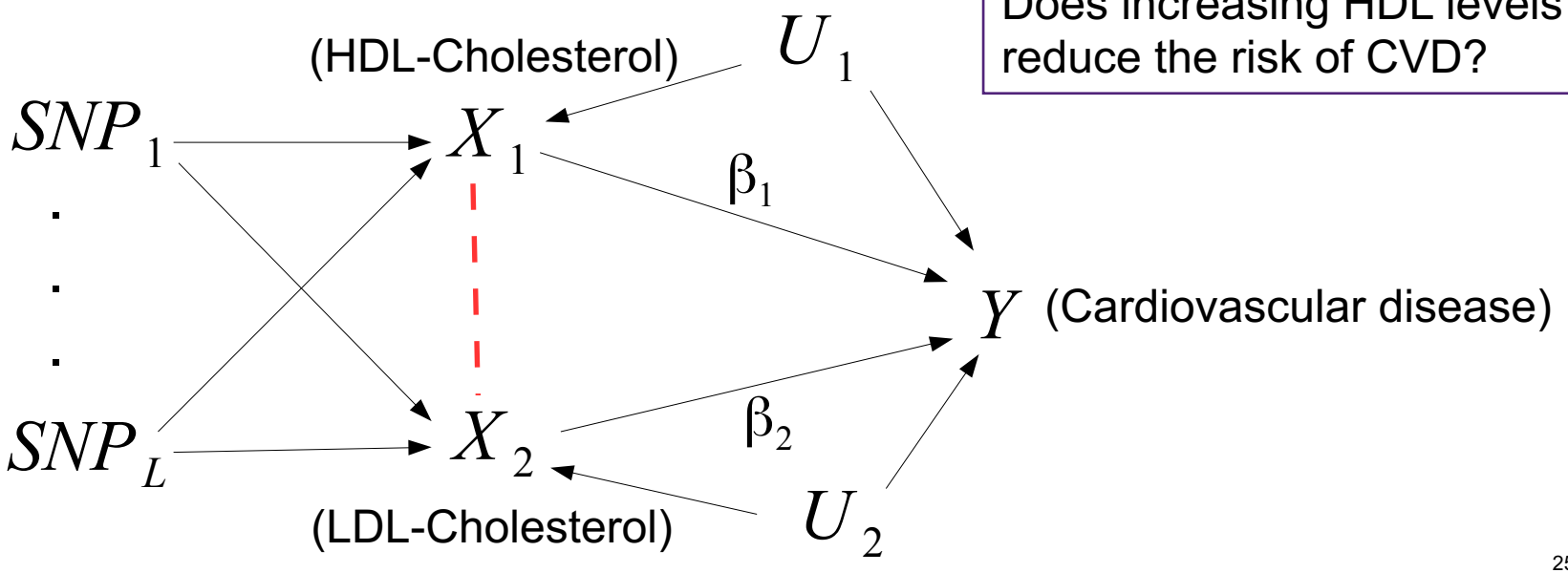
- E.g. cis-variants for gene expression, DNA methylation, and protein levels. CRP levels are best instrumented by variants within the *CRP* gene region. Most other variants that come up in CRP GWAS are upstream effects related to inflammation.

Accounting for heterogeneity

Option 2: Multivariable MR

- We are testing for whether X_1 has an influence on Y
- We know that some instruments for X_1 also have influences on X_2
- This opens up the possibility of horizontal pleiotropy biasing our estimate
- What is the X_1-Y association adjusting for X_2?

Does increasing HDL levels reduce the risk of CVD?

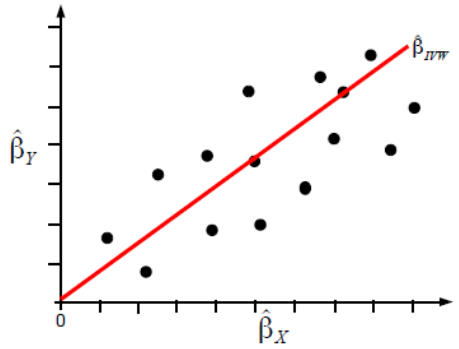


Accounting for heterogeneity

Option 3: Fit a model that is robust to some model of horizontal pleiotropy

IVW fixed effects estimator assumes all SNPs are valid instruments and averages across them all.

- Clear trend in estimates increasing with $\hat{\beta}_{Z_k X}$ from origin
- Cochran's $Q \approx K - 1$ (no heterogeneity)

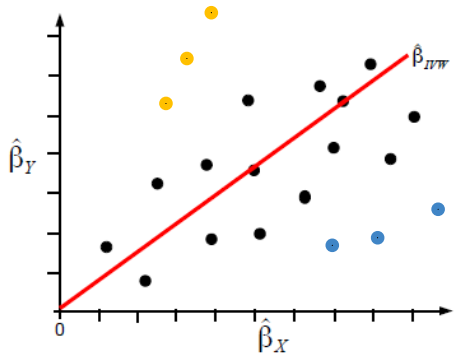


Accounting for heterogeneity

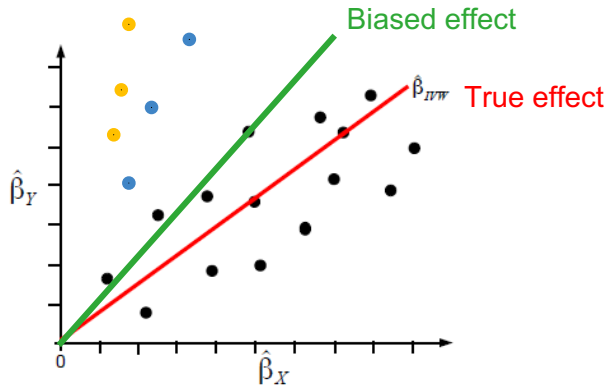
Option 3: Fit a model that is robust to some model of horizontal pleiotropy

IVW random effects estimator allows all SNPs to be invalid due to pleiotropy as long as the pleiotropy is balanced.

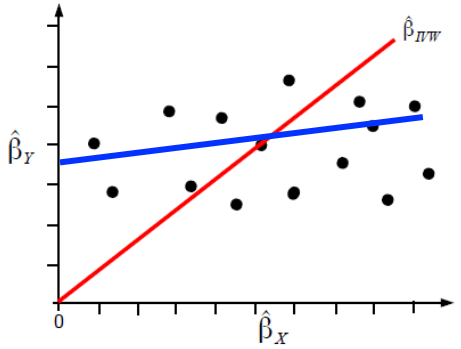
- The standard error of the causal estimate increases with the degree of heterogeneity.



Balanced



Imbalanced
(zero-intercept)



Imbalanced

We could therefore regress the SNP-outcome associations on the SNP-exposure associations, but allow for a **non-zero intercept** in the regression
This is the principal behind **MR-Egger regression**.

- SNPs are valid instruments
- SNP associated with outcome via confounder.
- SNP associated with outcome via an independent pathway.

Trend away from origin + heterogeneity

- Zero-intercept condition unreasonable
- IVW does not appear to be a good fit

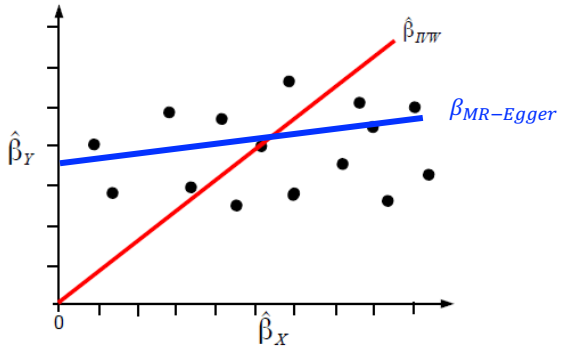
Pleiotropy potentially causes heterogeneity and bias

Break time

by Canva AI image generator

MR-Egger regression: Central concept

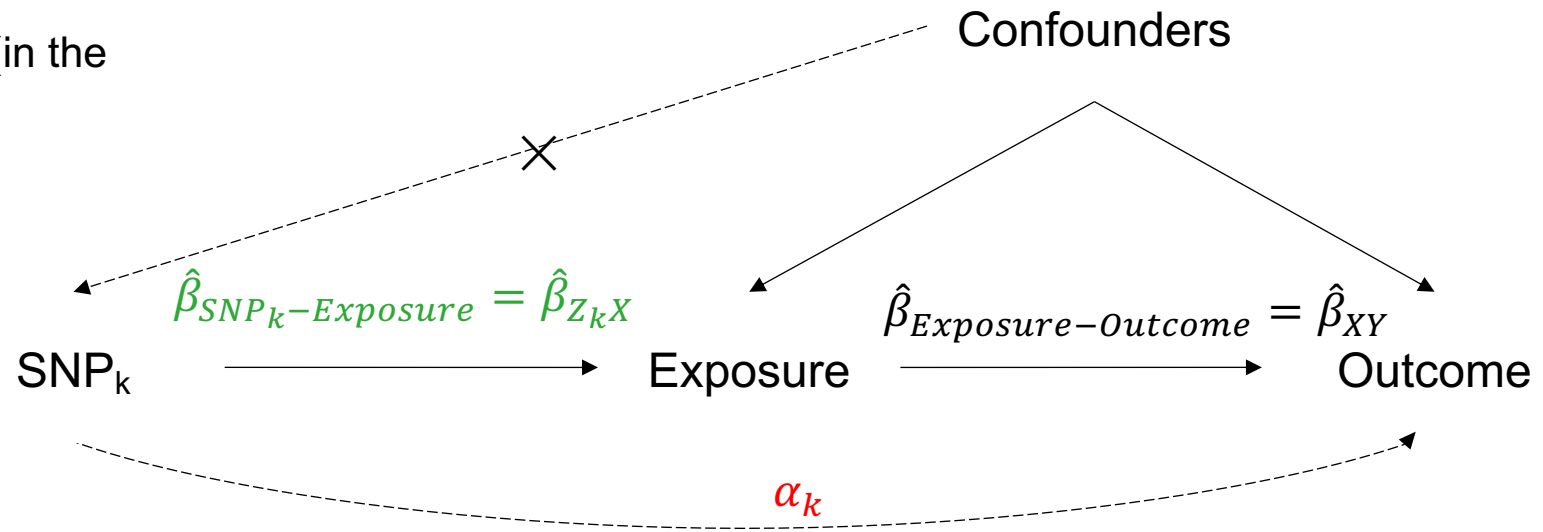
MR-Egger allows for a **non-zero intercept** in the regression.



When multiple SNPs are used as instruments, MR-Egger can:

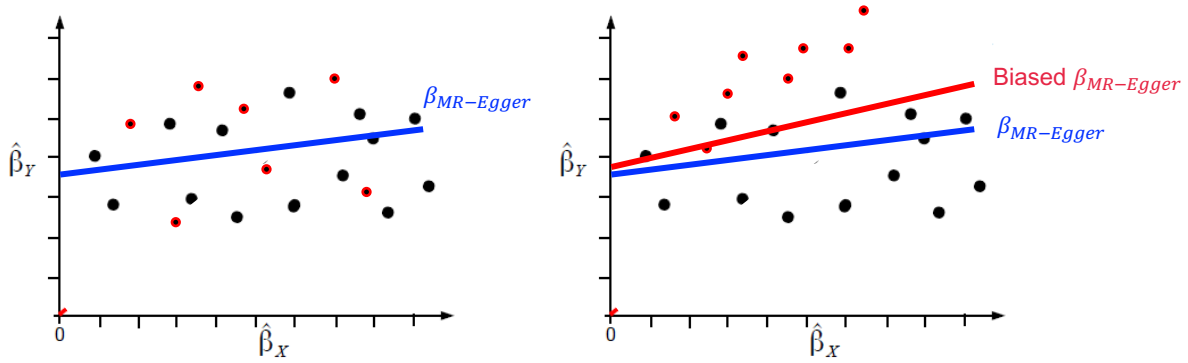
- Identify the presence of “directional” pleiotropy (biasing the causal estimate in IVW)
- Provide a less biased causal estimate (in the presence of pleiotropy)

MR-Egger lacks power.



MR-Egger regression

MR-Egger regression replies on the InSIDE (INstrument Strength Independent of Direct Effect) assumption, which states that the pleiotropic effects of SNPs must be independent of their strength as instruments.



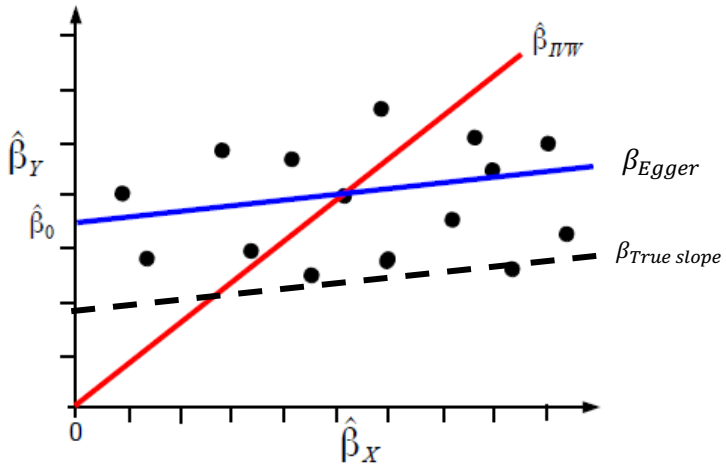
InSIDE | Violation of InSIDE

- SNP not associated with outcome via an independent pathway
- SNP associated with outcome via an independent pathway

IVW model: $\hat{\beta}_{Yk} = \underbrace{\beta_{IVW}}_{\text{Slope}} \hat{\beta}_{Xk} + + \varepsilon_{Yk}$

MR-Egger model: $\hat{\beta}_{Yk} = \beta_0 + \underbrace{\beta_{Egger}}_{\text{Slope}} \hat{\beta}_{Xk} + \varepsilon_{Yk}$

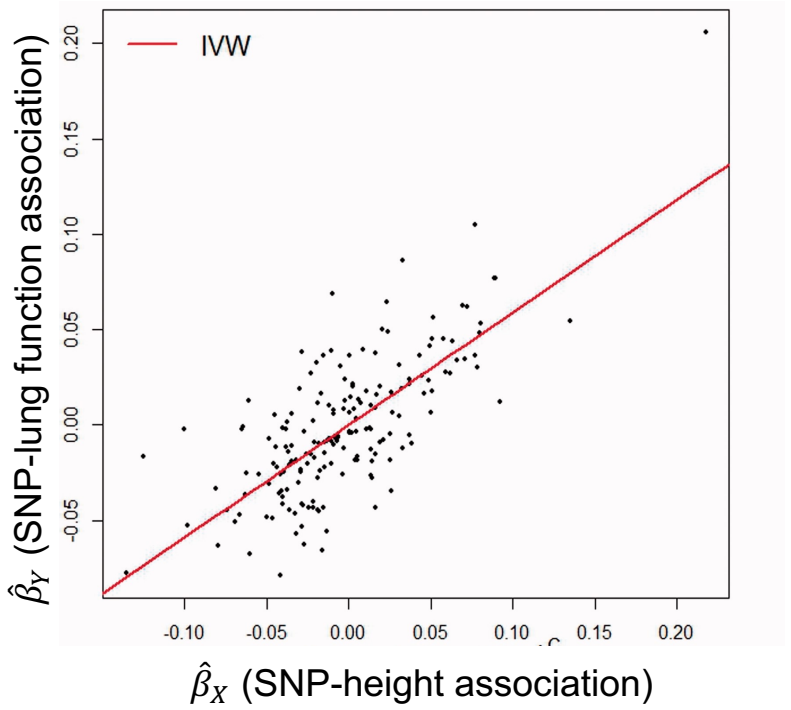
- β_0 is the intercept term. β_0 can be interpreted as the average pleiotropic effect across all genetic variants. A non-zero β_{0E} indicates directional pleiotropy.
- β_{Egger} is the causal estimate adjusted for directional pleiotropy



Bowden et al. Int J Epidemiol. (2015) 44(2):512-25

MR-Egger regression

Example: height and lung function



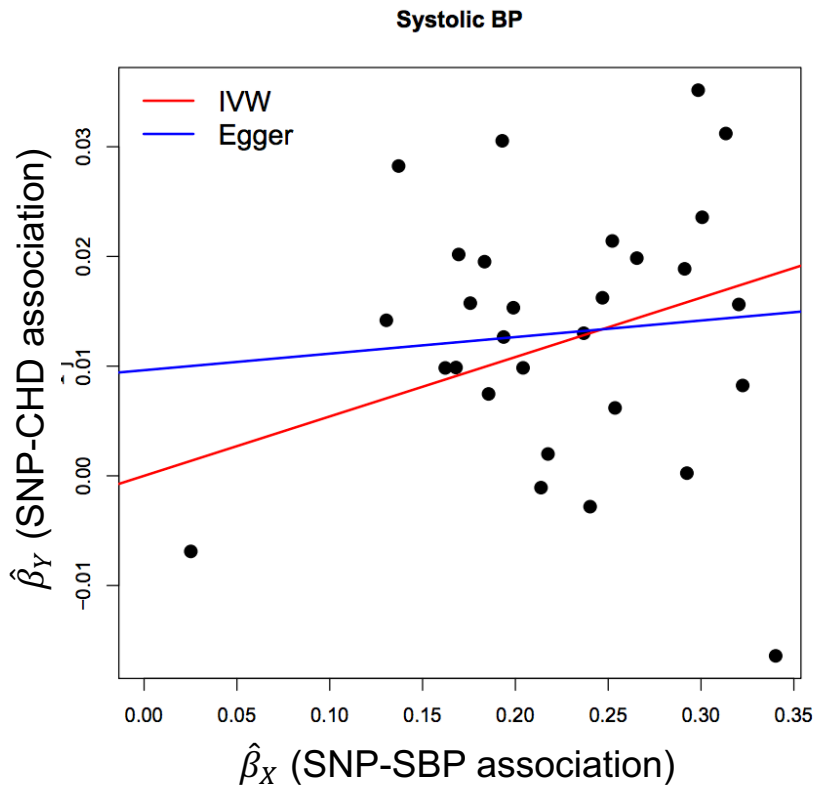
IVW = 0.59 (95% CI: 0.50, 0.67)

MR-Egger = 0.58 (95% CI: 0.50, 0.67);
intercept = -0.001 (p = 0.5)

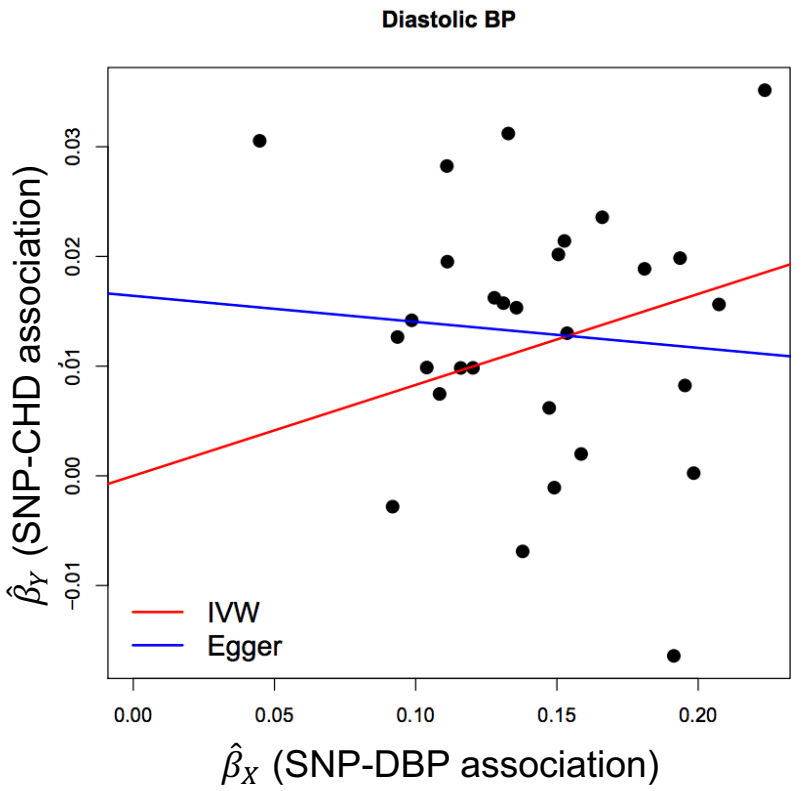
Bowden et al. Int J Epidemiol. (2015) 44(2):512-25

MR-Egger regression

Example: BP and Coronary Heart Disease



IVW = 0.054 logOR/mmHg; p = 4×10^{-6}
Egger = 0.015 logOR/mmHg; p = 0.6



IVW = 0.083 logOR/mmHg; p = 1×10^{-5}
Egger = -0.024 logOR/mmHg; p = 0.7

Median based methods (Median Estimator)

Order causal estimates (Wald ratio) and take the median.

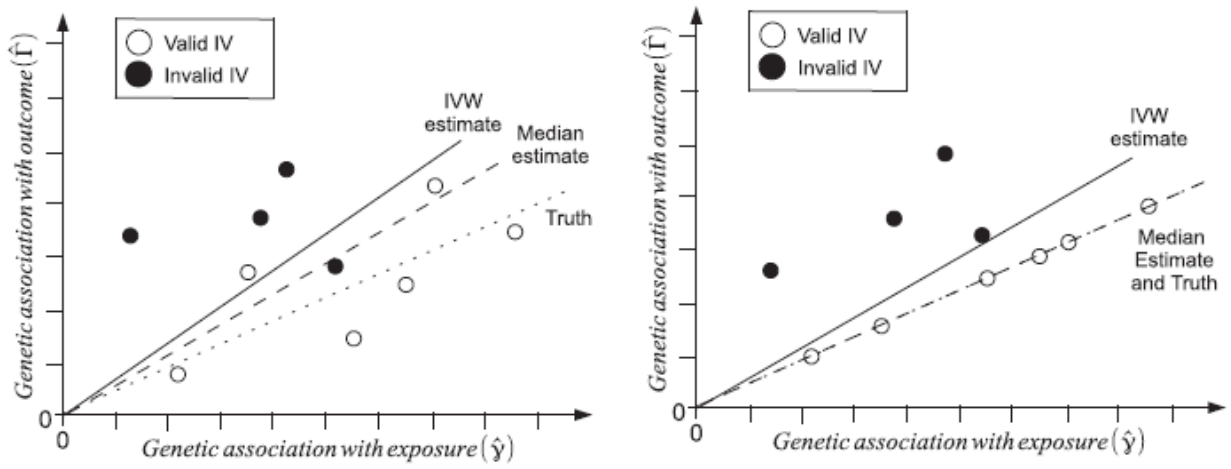


Figure 2. Fictional example of a Mendelian randomization analysis with 10 genetic variants—six valid instrumental variables (hollow circles) and four invalid instrumental variables (solid circles) for finite sample size (left) and infinite sample size (right) showing IVW (solid line) and simple median (dashed line) estimates compared with the true causal effect (dotted line). The ratio estimate for each genetic variant is the gradient of the line connecting the relevant datapoint for that variant to the origin; the simple median estimate is the median of these ratio estimates.

Assumption: >50% of the instrumental variables are valid.

No restrictions need to be placed on the invalid IVs:

- InSIDE assumption not required
- Violations of #2 and #3 MR assumptions are allowed

Bowden et al. Genet Epidemiol. (2016) 40(4):304-314

Median based methods

Simple median estimator

- Odd number of IVs: middle ratio estimate
- Even number of IVs: median is the average of the two middle estimates $\left(\frac{1}{2}\left(\hat{\beta}_k + \hat{\beta}_{k+1}\right)\right)$
  - Inefficient when the precision of individual variants varies considerably

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$	$\hat{\beta}_{10}$
Simple median										
Weight $(1/v_k)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$
Percentile (p_k)	5	15	25	35	45	55	65	75	85	95
Weighting 1										
Weight $(1/v_k)$	$\frac{1}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{4}{30}$	$\frac{5}{30}$	$\frac{5}{30}$	$\frac{4}{30}$	$\frac{3}{30}$	$\frac{2}{30}$	$\frac{1}{30}$
Percentile	1.67	6.67	15.00	26.67	41.67	58.33	73.33	85.00	93.33	98.33
Weighting 2										
Weight $(1/v_k)$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
Percentile (p_k)	2.78	9.72	27.78	52.78	70.83	81.94	88.89	93.06	95.83	98.61

$$\hat{\beta}_{Simple\ median} = \frac{\hat{\beta}_5 + \hat{\beta}_6}{2}$$

Bowden et al. Genet Epidemiol. (2016) 40(4):304-314

Median based methods

Weighted median estimator

- Weighted median estimator takes into account the differing precisions
- Weighted median: $\hat{\beta}_{WM} = \hat{\beta}_3 + (\hat{\beta}_4 - \hat{\beta}_3) \times \frac{50-27.78}{52.78-27.78}$
- Suggested weights: inversed variance of the ratio estimate: $w'_k = \frac{1}{\boldsymbol{var}(\hat{\beta}_{XY_k})}$

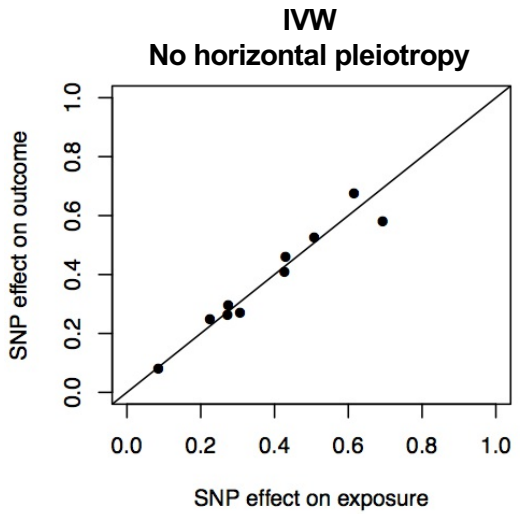
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$	$\hat{\beta}_{10}$
Simple median										
Weight $(1/v_k)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$
Percentile (p_k)	5	15	25	35	45	55	65	75	85	95
Weighting 1										
Weight $(1/v_k)$	$\frac{1}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{4}{30}$	$\frac{5}{30}$	$\frac{5}{30}$	$\frac{4}{30}$	$\frac{3}{30}$	$\frac{2}{30}$	$\frac{1}{30}$
Percentile	1.67	6.67	15.00	26.67	41.67	58.33	73.33	85.00	93.33	98.33
Weighting 2										
Weight $(1/v_k)$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
Percentile (p_k)	2.78	9.72	27.78	52.78	70.83	81.94	88.89	93.06	95.83	98.61

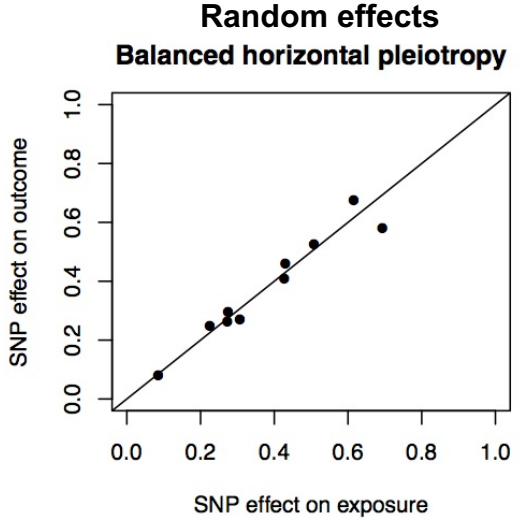
$$\hat{\beta}_{WM} = \hat{\beta}_5 + (\hat{\beta}_6 - \hat{\beta}_5) \times \frac{50 - 41.67}{58.33 - 41.67}$$

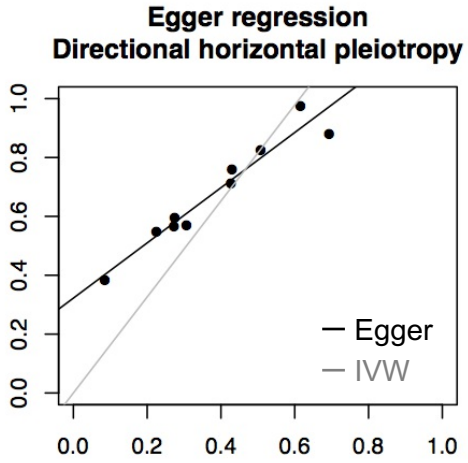
$$\hat{\beta}_{WM} = \hat{\beta}_3 + (\hat{\beta}_4 - \hat{\beta}_3) \times \frac{50 - 27.78}{52.78 - 27.78}$$

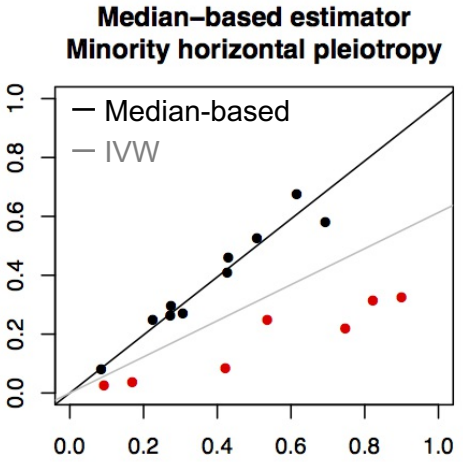
Bowden et al. Genet Epidemiol. (2016) 40(4):304-314

Summary of robust estimators







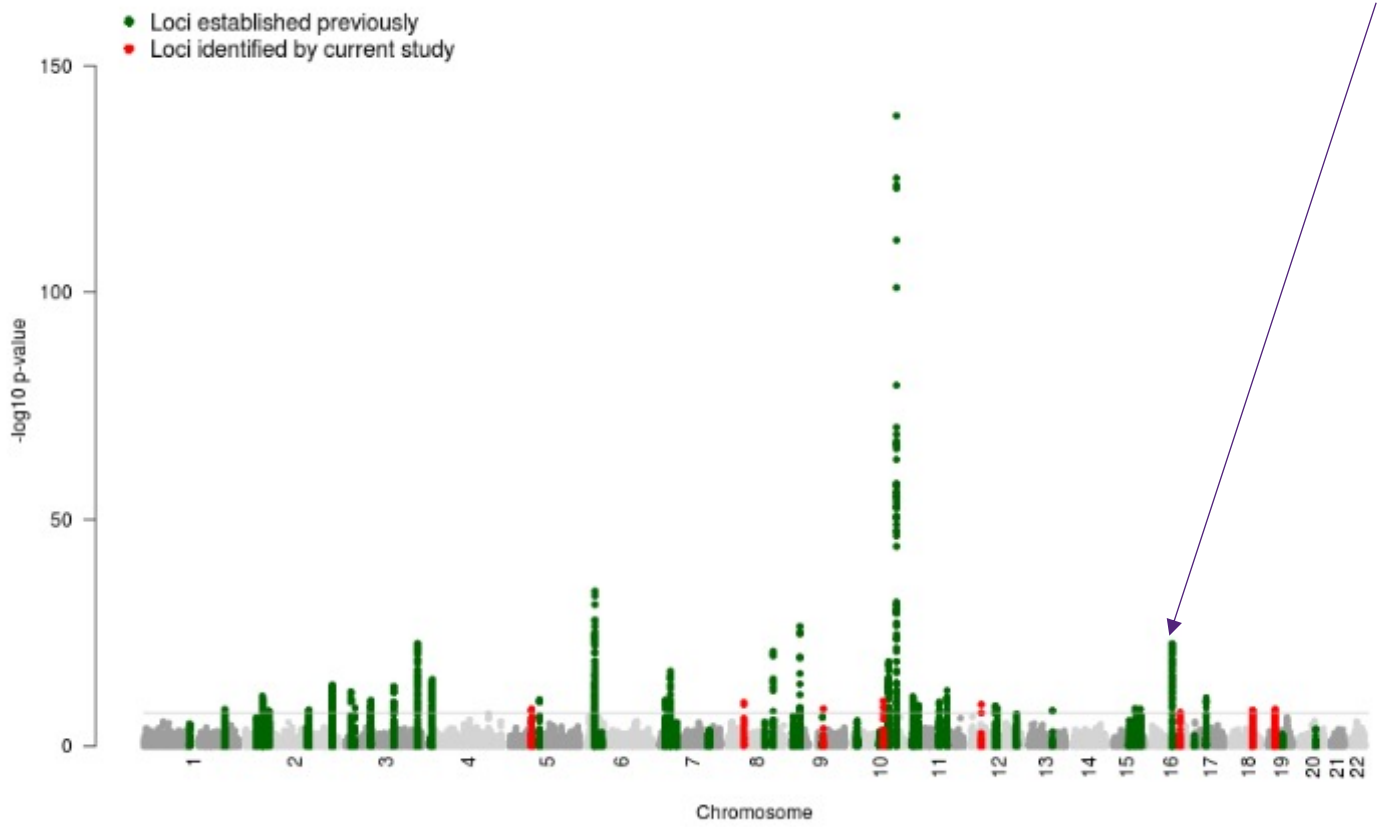


● SNPs associated with outcome via an independent pathway.

Hemani et al. eLife (2018)

Reverse causal instruments

Problem: MR of type 2 diabetes on BMI



GWAS of T2D reveals *FTO* variant
- Famously associated with BMI
- A reverse causal instrument?

FTO → T2D → BMI

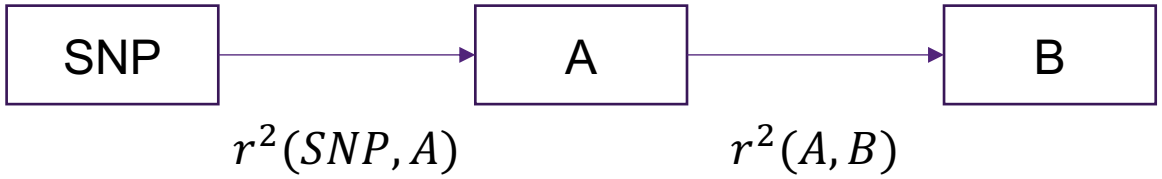
T2D ← BMI ← *FTO*

FTO may be associated with T2D through its effect on BMI
>> Not a good instrument for T2D

Can we avoid including reverse-causal SNPs as instruments?

Steiger filtering test

- If SNP causes A and A causes B
- The effect of SNP on A should be larger than the effect of SNP on B



Expect that

$$r^2(SNP, B) = r^2(SNP, A) \times r^2(A, B)$$

This term is <1

- Steiger test used to evaluate if r^2(SNP,A) > r^2(SNP,B)
- If this is not satisfied, infer that this instrument is not influencing the exposure primarily.

Ideal instruments are genetic variants with a known biological function related to the exposure

Bitter taste receptors (*TAS2Rs*) → Bitter taste perception (+/-) → Bitter beverage intake

(Ong et al. 2018 *Sci Rep*)

Caffeine metabolism genes (*AHR/CYP1A2*) → Maternal coffee intake (×) → Children's birth weight/ Perinatal outcomes

(Brito Nunes et al. 2023 *Int J Epidemiol*)

Aldehyde dehydrogenase (*ALDH2*) → Alcohol intake (×) → Breast/Ovarian cancer

(Ong et al. 2020 *Int J Cancer*)

Lactase (*LCT*) → Milk intake (*) → 18 diseases

(Yuan et al. 2022 *BMC Med*)

TwoSampleMR R Package

TwoSampleMR 0.6.4 Guide Functions Changelog Search for Source

Mendelian randomization with GWAS summary data

A package for performing Mendelian randomization using GWAS summary data. It uses the IEU GWAS database to obtain data automatically, and a wide range of methods to run the analysis. You can use the MR-Base web app to try out a limited range of the functionality in this package, but for any serious work we strongly recommend using this R package.

January 2020 major update

We have made substantial changes to the package, database and reference panels. For full details of the changes, please visit https://mrcieu.github.io/TwoSampleMR/articles/gwas2020.html

Installation

Users running Windows and macOS, to install the latest version of TwoSampleMR please install from our MRC IEU r-universe

```
install.packages("TwoSampleMR", repos = c("https://mrcieu.r-universe.dev", "https://cloud.r-project.o
```

Users running Linux or WebR please see the following instructions.

To update the package run the same command again.

Installing from source

```
install.packages("remotes")
remotes::install_github("MRCIEU/TwoSampleMR")
```

To update the package just run the `remotes::install_github("MRCIEU/TwoSampleMR")` command again.

Docker

A docker image containing R with the TwoSampleMR package pre-installed is available here: https://hub.docker.com/r/mrcieu/twosamplemr

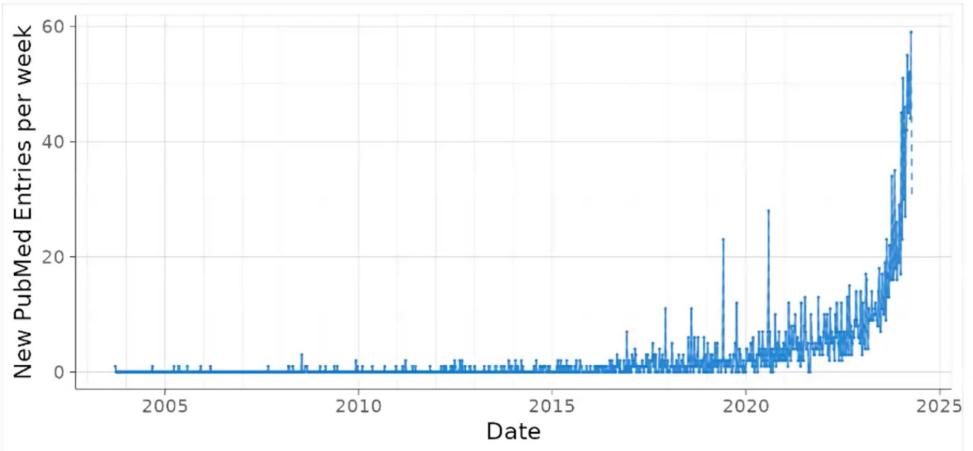
Links
Browse source code
Report a bug

License
Full license
MIT + file LICENSE

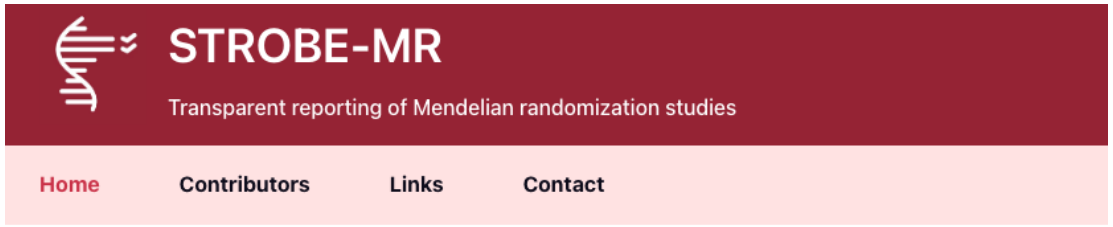
Citation
Citing TwoSampleMR

Developers
Gibran Hemani
Author, maintainer
Philip Haycock
Author
Jie Zheng
Author
Tom Gaunt
Author
Ben Elsworth
Author
Tom Palmer
Author

Dev status
R-CMD-check passing
lifecycle experimental
DOI 10.5281/zenodo.10684540
codecov 36%
r-universe 0.6.4



STROBE-MR



Welcome to the STROBE-MR website!

About: STROBE-MR stands for "Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization". Inspired by the original STROBE checklist, the STROBE-MR guidelines were developed to assist researchers in reporting their Mendelian randomization studies clearly and transparently. Adopting STROBE-MR should help readers, reviewers, and journal editors evaluate the quality of published MR studies.

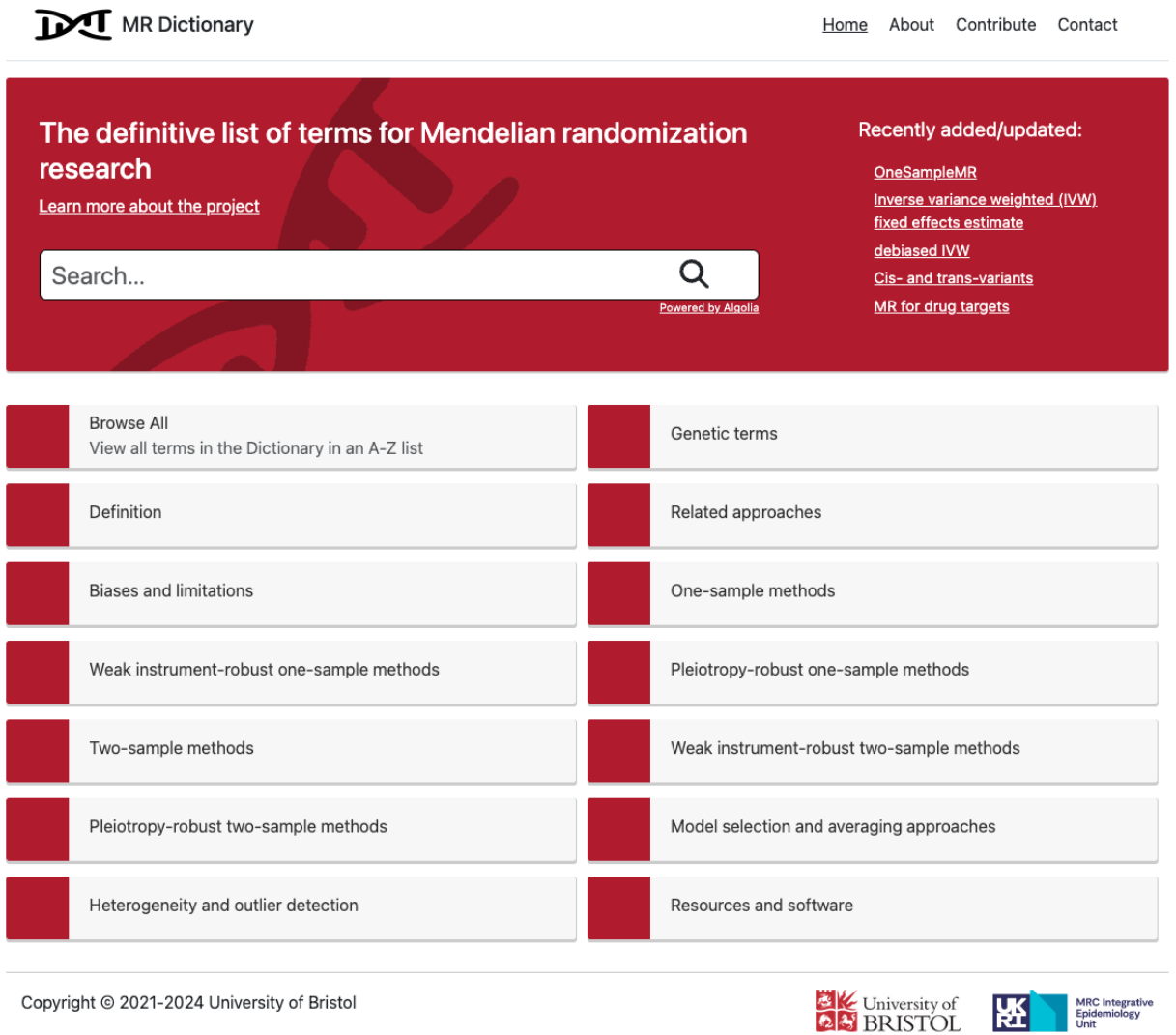
The STROBE-MR **checklist** contains 20 items recommended to address in reports of Mendelian randomization studies.

The **Statement** document describes the process of developing the checklist and the complementary Explanation and Elaborations document.

The **Explanation and Elaboration** document explains the items of the STROBE-MR checklist, along with their rationale and examples of transparent reporting.

All documents and publications produced by the STROBE-MR Initiative are open-access and available for download on this website.

MR Dictionary



Summary

- MR uses natural randomization to mimic an RCT
- It is useful, data is abundant, but it is not a panacea for causal inference
- Often valuable for proving that a hypothesized association is not causal
- Horizontal pleiotropy is one of the main threats to the validity of MR studies
  - Multiple methods developed to detect and adjust for horizontal pleiotropy
- Crucial to perform sensitivity analyses and obtain metrics regarding the likely reliability of the MR estimates
- Consistency of results across methods is key to reliable causal inference

Additional References

Bowden et al. **Detecting individual and global horizontal pleiotropy in Mendelian randomization: a job for the humble heterogeneity statistic?** *American Journal of Epidemiology 2018*, kwy185

Bowden et al. **Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator** Genet Epidemol 40(4), 304-14

Bowden et al. **Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption** Int J Epidemiol 46(6), 1985-1998

Bowden et al. **Mendelian randomization with invalid instruments: effect estimation and bias through Egger regression.** Int J Epidemiol, 44(2), 512-25

Burgess et al. **Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects.** Am J Epidemiol, 181(4), 251-60

Hemani et al. **Evaluating the potential role of pleiotropy in Mendelian randomization studies** Human Molecular Genetics, Volume 27, Issue R2, 1 August 2018, Pages R195–R208

Hemani et al. **Orienting the causal relationship between imprecisely measured traits using GWAS summary data** PLoS Genet 2017 13(11): e1007081

Burgess S, et al. **Bias due to participant overlap in two-sample Mendelian randomization**. Genet Epidemiol. 2016; 40(7): 597-608.

VanderWeele TJ et al. **Methodological challenges in Mendelian randomization**. Epidemiology. 2014; 25(3): 427-35.

Hartwig FP, et al. **Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique.** Int J Epidemiol. 2016; 45(6): 1717-1726.

Greco M FD, et al. **Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome**. Stat Med. 2015; 34(21): 2926-40.

Burgess S, et al. **Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology.** 2017; 28(1): 30-42.

Burgess S, Thompson SG. **Interpreting findings from Mendelian randomization using the MR-Egger method.** Eur J Epidemiol. 2017; 32(5): 377-389.

Bowden J, et al. **Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic.** Int J Epidemiol. 2016; 45(6): 1961-1974.

Hartwig FP, et al. **Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption.** Int J Epidemiol. 2017; 46(6): 1985-1998.