Sensitivity analyses in Mendelian randomization studies

Dr Nicole Warrington
Institute for Molecular Bioscience, The University of Queensland
(Some slides adapted from Prof David Evans and Dr Daniel Hwang)

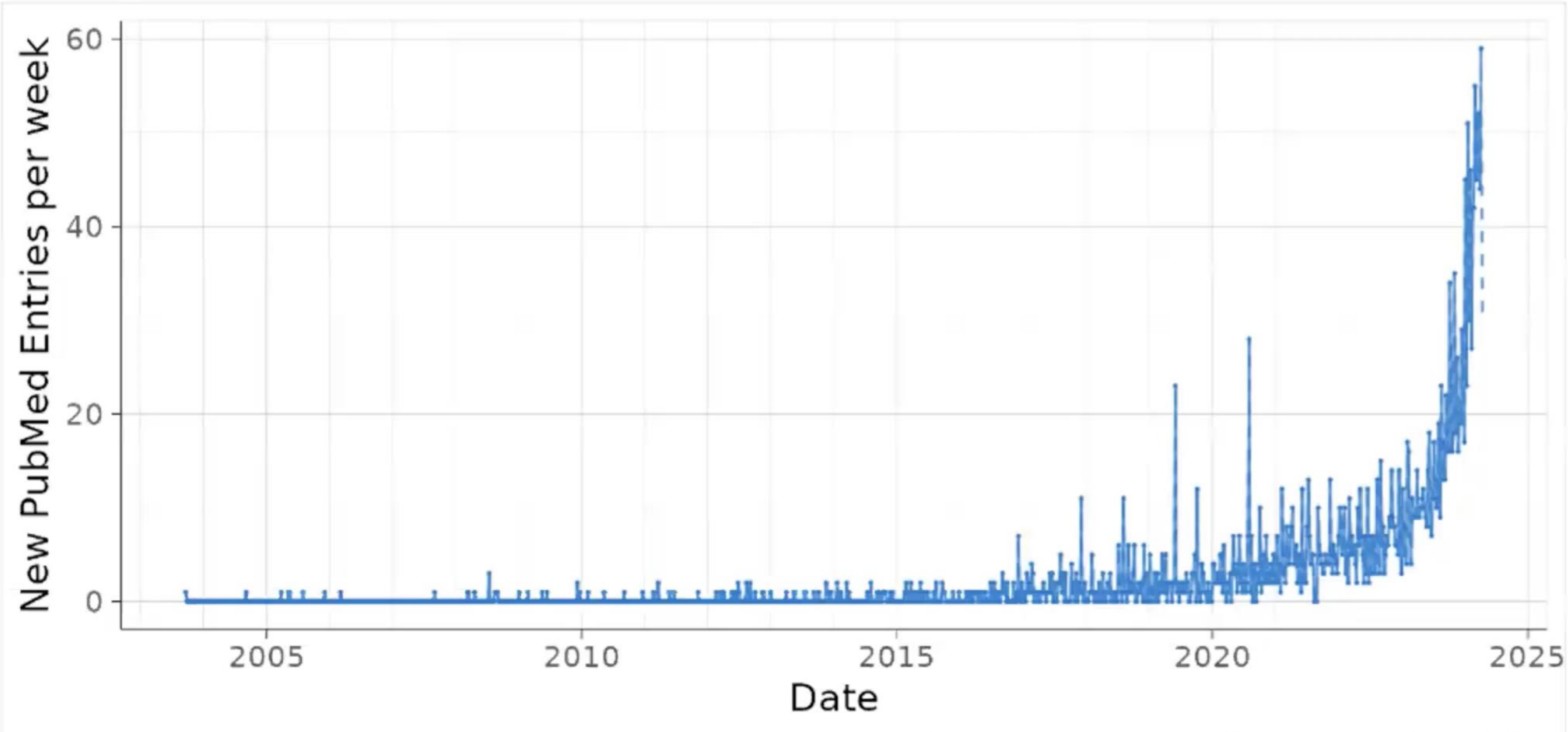
Acknowledgment of **Country**

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.

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



Up to 2024-04-13

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

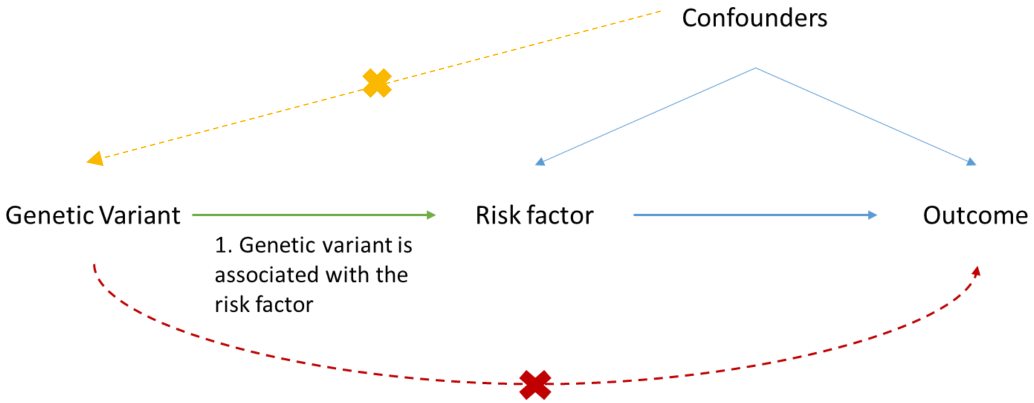
(3) Exclusion restriction: SNP ONLY associated outcome through the exposure

- 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

Recap

Assumption 1: Relevance assumption



Typically, SNPs which pass genome-wide significance ($P<5x10^{-8}$) and have been replicated in independent samples are used as IV's

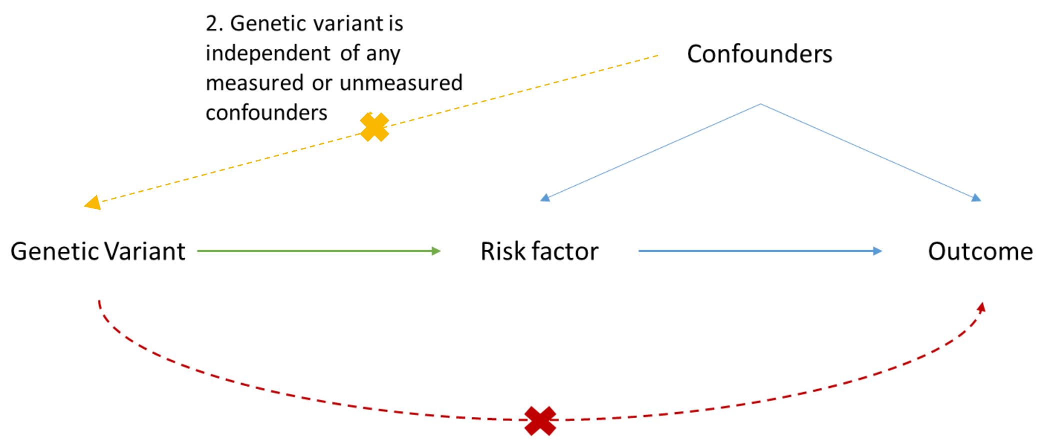
- Weak instruments:
  - Loss of power
  - Bias due to violations of the other assumptions will be amplified
  - Bias towards outcome-risk factor association in one-sample MR or towards the null in two-sample MR – precision is also underestimated.
- Weak instruments can be detected using an F-statistic in one-sample MR (F-statistic > 10)

$$F_{stat} = \frac{R^2 * (N-1)}{(1-R^2)}$$

Burgess S et al. *Int J Epidemiol.* (2011) 40(3):755-764
Burgess S and Thompson SG. *Stat Med* (2011) 30(11):1312-1323

Recap

Assumption 2: Independence assumption

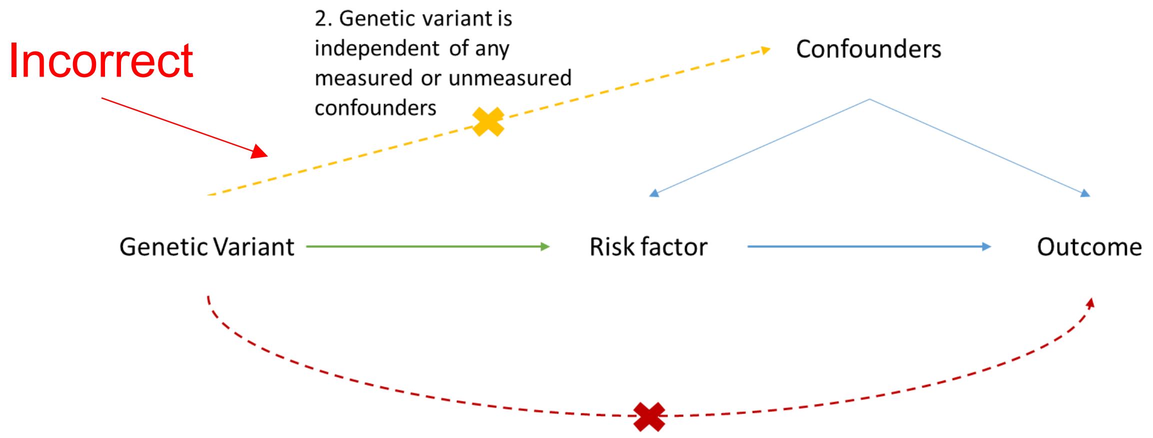


- Technically impossible to prove this assumption holds as we can't test for association with *unobserved* confounders (need to rely on good knowledge of the science)
- May be possible to disprove by checking that the genetic variant is independent of *measured* confounders of the exposure-outcome relationship
- Factors that could influence the genetic variants and outcome include population stratification or structure, intergenerational (dynastic) effects and assortative mating.

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

Recap

Assumption 2: Independence assumption

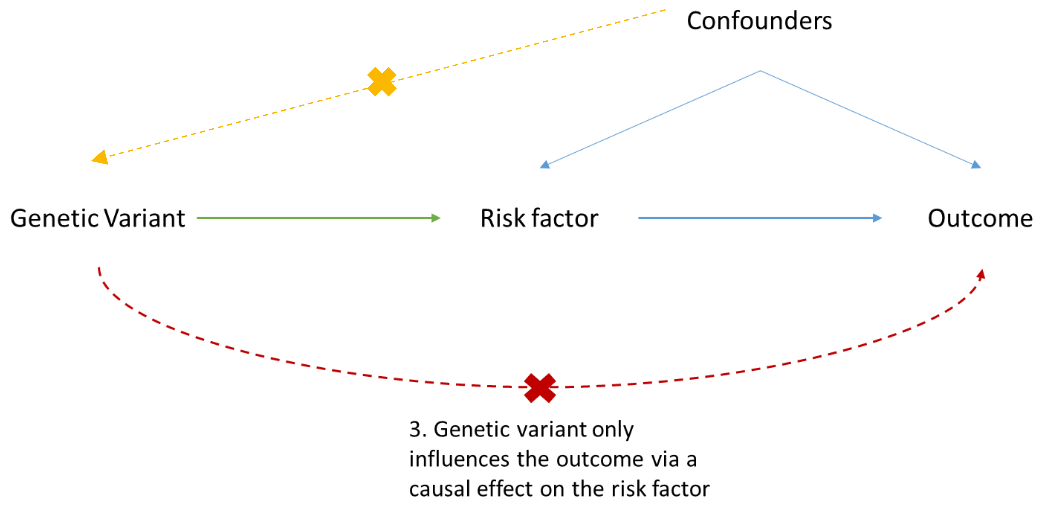


- Technically impossible to prove this assumption holds as we can’t test for association with *unobserved* confounders (need to rely on good knowledge of the science)
- May be possible to disprove by checking that the genetic variant is independent of *measured* confounders of the exposure-outcome relationship
- Factors that could influence the genetic variants and outcome include population stratification or structure, intergenerational (dynastic) effects and assortative mating.

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

Recap

Assumption 3: Exclusion restriction



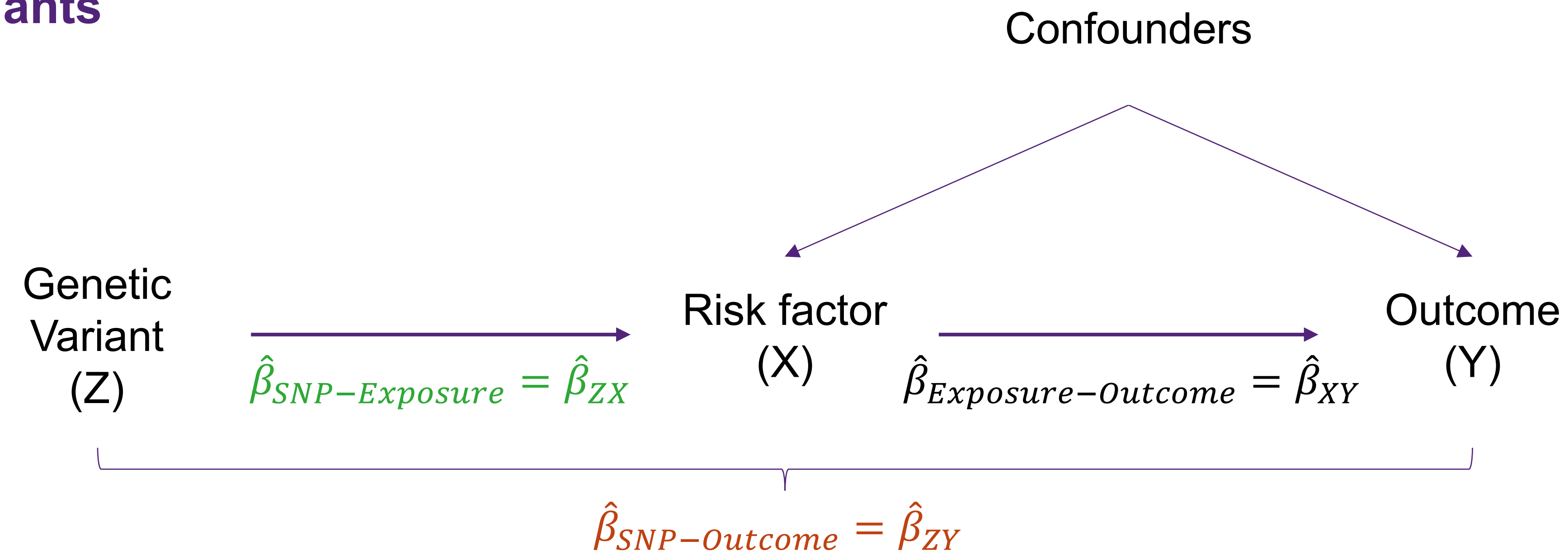
- Again, is difficult to prove this assumption holds
- Horizontal pleiotropy = SNP is associated with multiple traits independently of the exposure of interest
- Extensions to the basic MR design can be used to detect horizontal pleiotropy and estimate causal effect in its presence

Outline

- Inverse variance weighted MR
- Heterogeneity tests
- Multivariable MR
- MR Egger
- MR Weighted Median
- Steiger Filtering

Two-sample MR

Single variants



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)

Delta method to estimate SE of Wald ratio

$$Var(\hat{\beta}_{xy}) = Var\left(\frac{\hat{\beta}_{SNP-Outcome}}{\hat{\beta}_{SNP-Exposure}}\right)$$

$$\approx \frac{Var(\hat{\beta}_{SNP-Outcome})}{\hat{\beta}_{SNP-Exposure}^{\;2}} + \left(\frac{\hat{\beta}_{SNP-Outcome}^{\;2}}{\hat{\beta}_{SNP-Exposure}^{\;4}}\right) Var(\hat{\beta}_{SNP-Exposure}) - 2\left(\frac{\hat{\beta}_{SNP-Outcome}}{\hat{\beta}_{SNP-Exposure}^{\;3}}\right) Cov(\hat{\beta}_{SNP-Exposure}, \hat{\beta}_{SNP-Outcome})$$

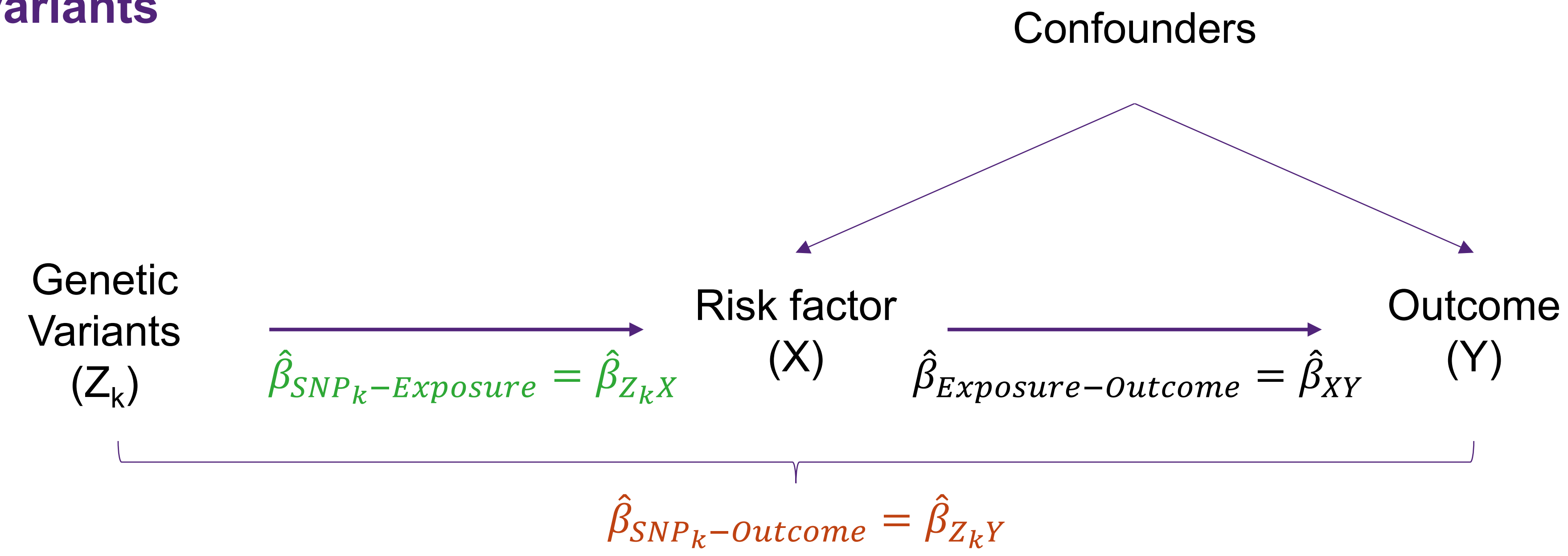
$$\approx \frac{Var(\hat{\beta}_{SNP-Outcome})}{\hat{\beta}_{SNP-Exposure}^{\;2}}$$

$$\text{SE}(\hat{\beta}_{XY}) = \hat{\sigma}_{XY} \approx \sqrt{\frac{Var(\hat{\beta}_{SNP-Outcome})}{\hat{\beta}_{SNP-Exposure}^{\;2}}}$$

$$\approx \frac{\sigma_{SNP-Outcome}}{\hat{\beta}_{SNP-Exposure}}$$

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

Two-sample MR

Multiple variants

Confounders

Genetic Variants (Z_k)

$\hat{\beta}_{SNP_k-Exposure} = \hat{\beta}_{Z_kX}$

Risk factor (X)

$\hat{\beta}_{Exposure-Outcome} = \hat{\beta}_{XY}$

Outcome (Y)

$\hat{\beta}_{SNP_k-Outcome} = \hat{\beta}_{Z_kY}$

Causal effect by Wald estimator:

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

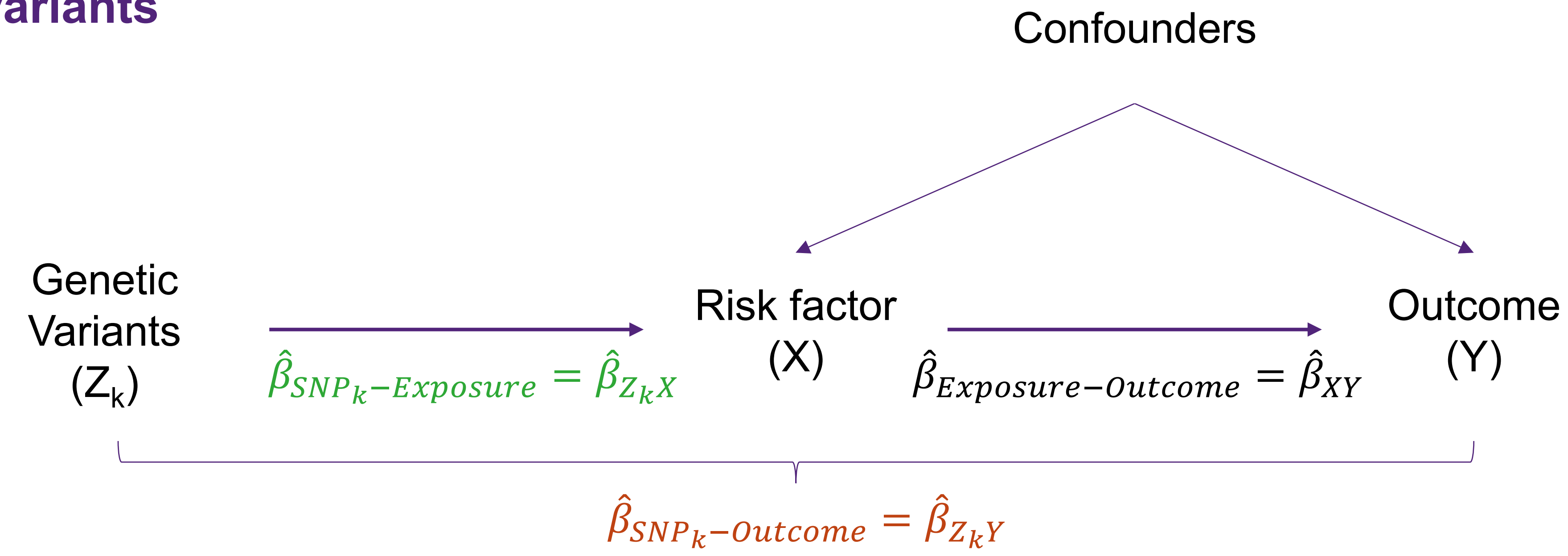
Inverse variance weighted (IVW) average causal effect:

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

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

Two-sample MR

Multiple variants



Causal effect by Wald estimator:

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

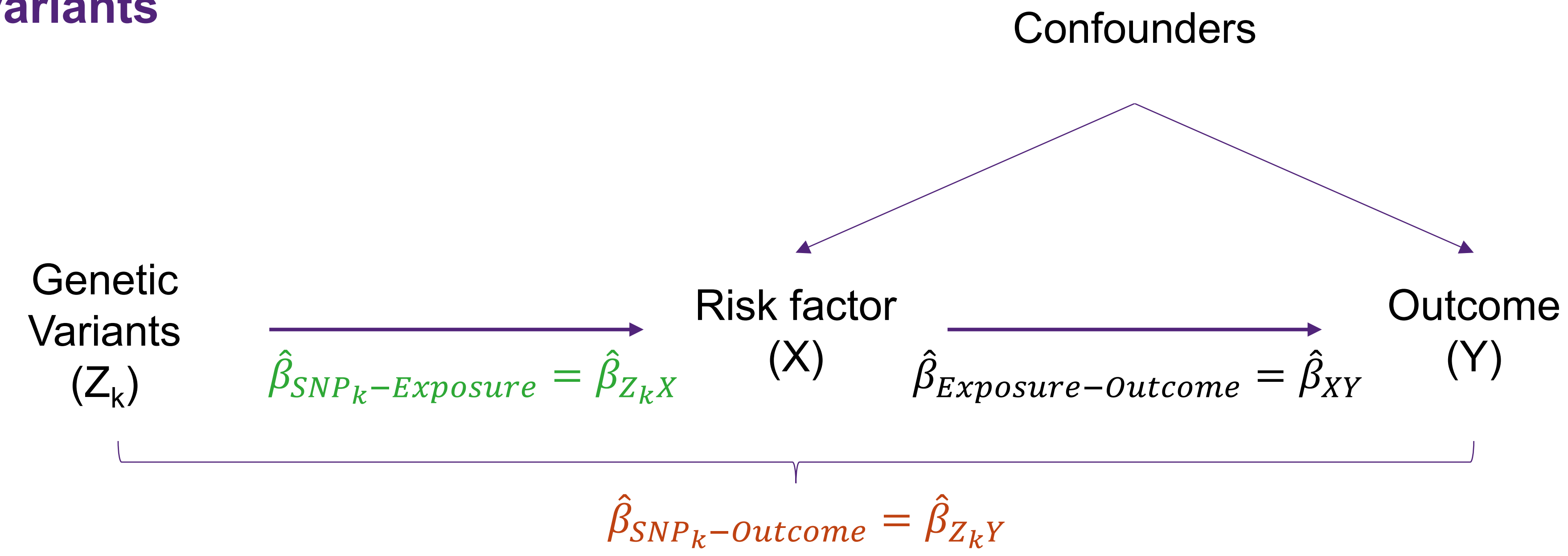
Inverse variance weighted (IVW) average causal effect:

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

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

Two-sample MR

Multiple variants



Causal effect by Wald estimator:

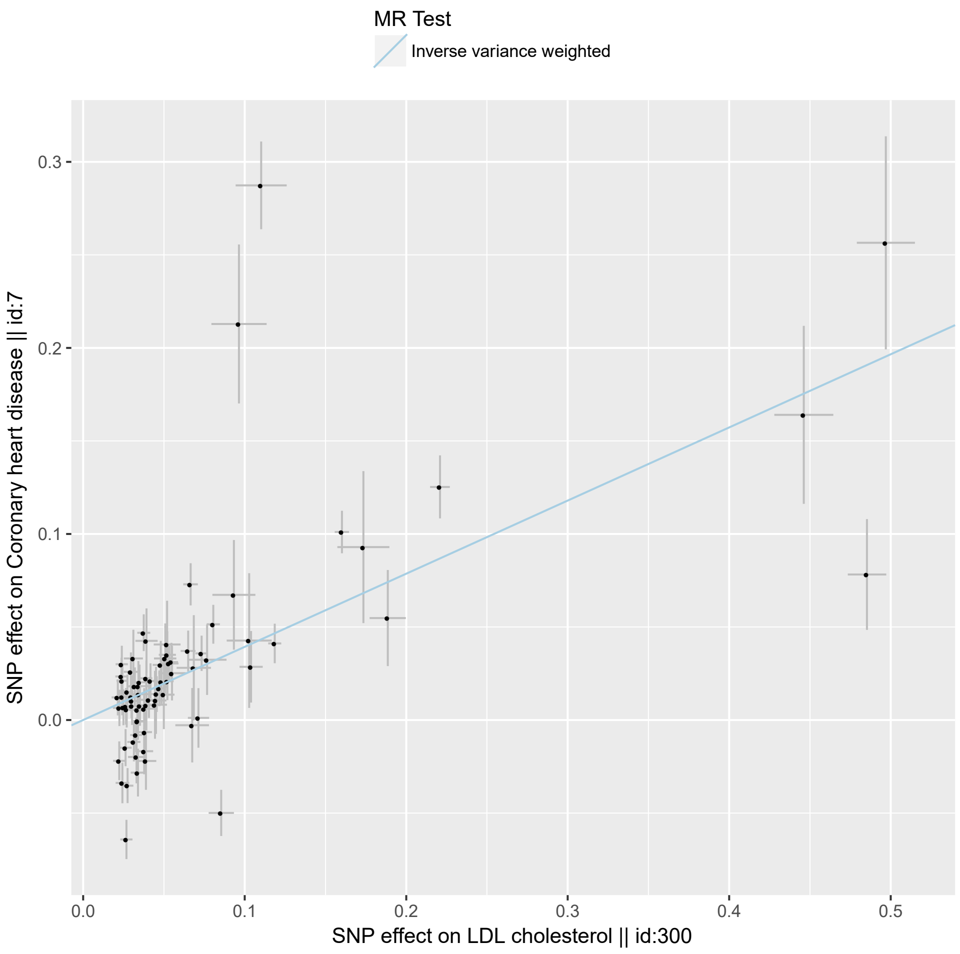
$$\hat{\beta}_{XY_k} = \frac{\hat{\beta}_{SNP_k-Outcome}}{\hat{\beta}_{SNP_k-Exposure}}$$

Inverse variance weighted (IVW) average causal effect:

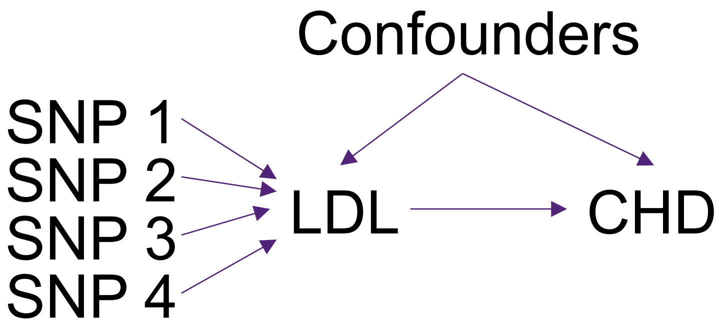
$$\hat{\beta}_{IVW} = \frac{\sum_{k=1}^{K} \hat{\beta}_{XY_k} w_k}{\sum_{k=1}^{K} w_k}$$

Where $w_k = \frac{1}{var(\hat{\beta}_{XY_k})} = \frac{1}{\hat{\sigma}_{XY_k}^2} = \frac{\hat{\beta}_{SNP-Exposure}^{\ 2}}{\sigma_{SNP-Outcome}^2}$ is the inverse variance of the causal effect estimated from the kth 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 $\frac{1}{\sigma_{Z_k Y}}$
- The slope is the estimate of the causal effect



Assumptions for two-sample MR

- Using summarized data for two-sample MR analyses is convenient when sharing individual level data is impractical

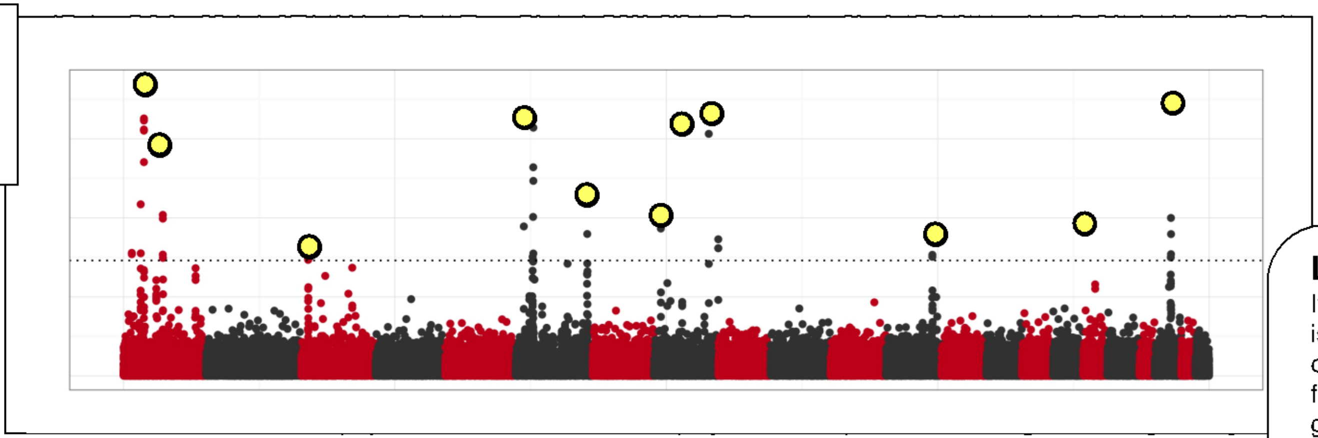
- If:
  - The K genetic variants are perfectly uncorrelated (not in LD) and do not interact
  - The two samples are homogenous (same underlying populations)
  - Constant causal effect at each level of the exposure

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

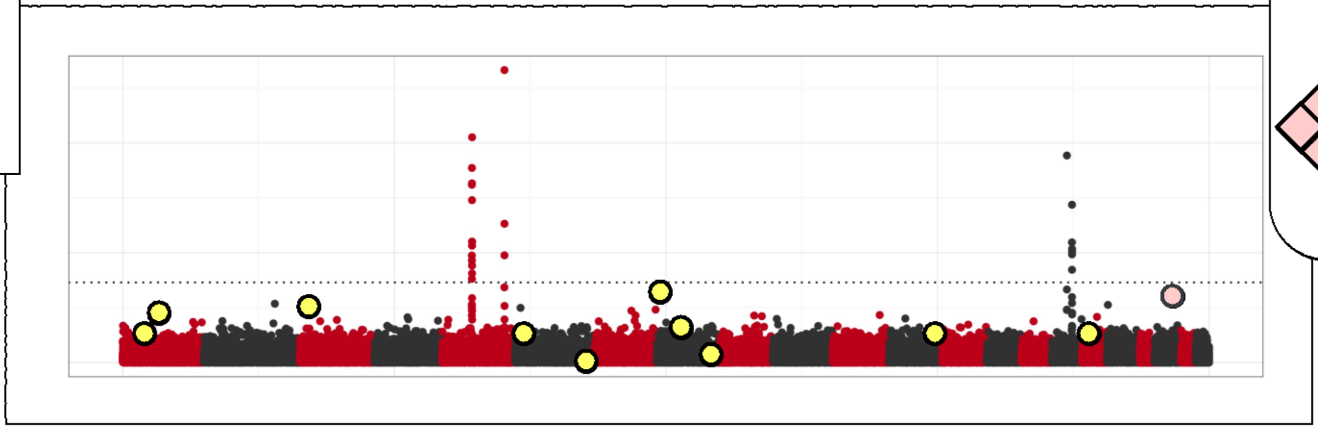
- Two-sample MR is still vulnerable to weak instrument bias
  - Bias towards the null effect, not the observational estimate
  - If approximate F-statistic ($\hat{\beta}^2_{Z_k X}/\sigma^2_{Z_k X}$) is greater than 10, then the expected dilution of $\hat{\beta}_{XY_k}$ towards zero is less than 10%

Performing MR with summary statistics

Obtain instruments from exposure GWAS

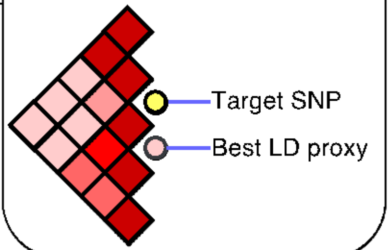


Extract SNP effects from outcome GWAS

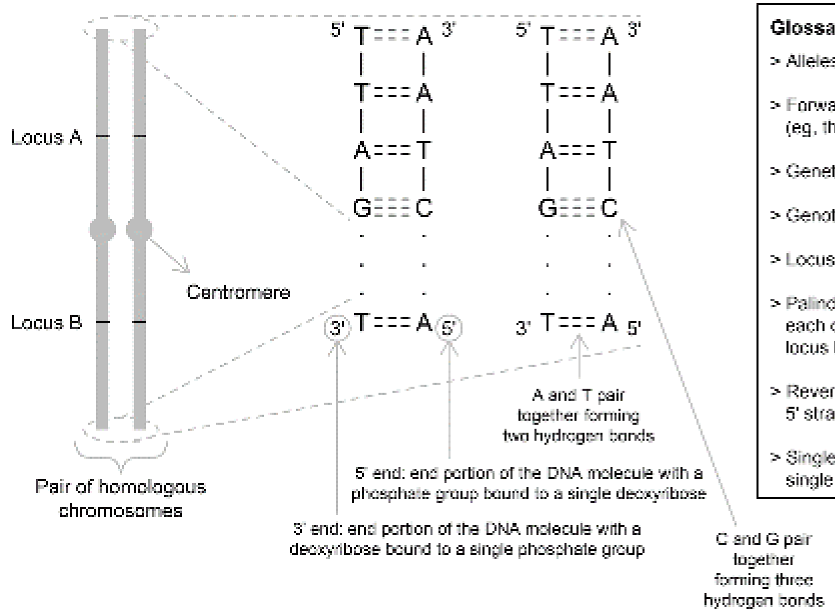


LD Proxies

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



The issue of strand



Glossary

> Alleles: variant forms that a locus may present.

> Forward or positive strand: the DNA strand that is read from the 5' to the 3' end (eg, the 5' TTAG...T 3' strand in the figure).

> Genetic variant: locus with more than one allele in a population.

> Genotype: combination of alleles that an individual presents at a given locus.

> Locus (plural loci): a specific location in a DNA sequence.

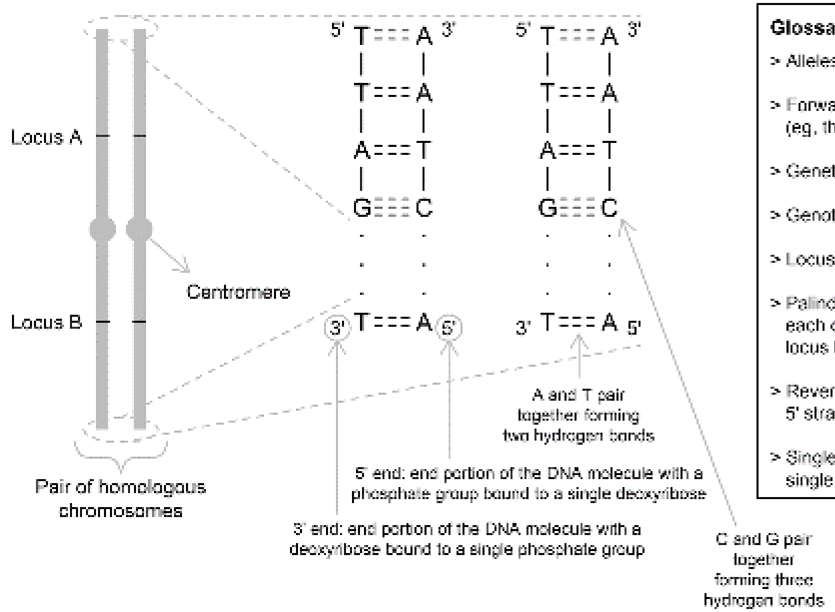
> Palindromic SNP: SNPs whose alleles correspond to nucleotides that pair with each other in a double-stranded DNA molecule. SNPs with A/T or G/C (as in locus B below) alleles are palindromic SNPs.

> Reverse or negative strand: the DNA strand that is read from the 3' to the 5' strand (eg, the 3' AATC...A 5' strand in the figure).

> Single nucleotide polymorphism (SNP): a type of genetic variant that involves single base pair changes.

Hartwig FP et al. *Int J Epidemiol.* (2016) 45(6):1717-1726

The issue of strand



Glossary

> Alleles: variant forms that a locus may present.

> Forward or positive strand: the DNA strand that is read from the 5' to the 3' end (eg, the 5' TTAG...T 3' strand in the figure).

> Genetic variant: locus with more than one allele in a population.

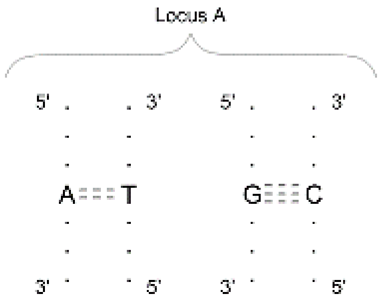
> Genotype: combination of alleles that an individual presents at a given locus.

> Locus (plural loci): a specific location in a DNA sequence.

> Palindromic SNP: SNPs whose alleles correspond to nucleotides that pair with each other in a double-stranded DNA molecule. SNPs with A/T or G/C (as in locus B below) alleles are palindromic SNPs.

> Reverse or negative strand: the DNA strand that is read from the 3' to the 5' strand (eg, the 3' AATC...A 5' strand in the figure).

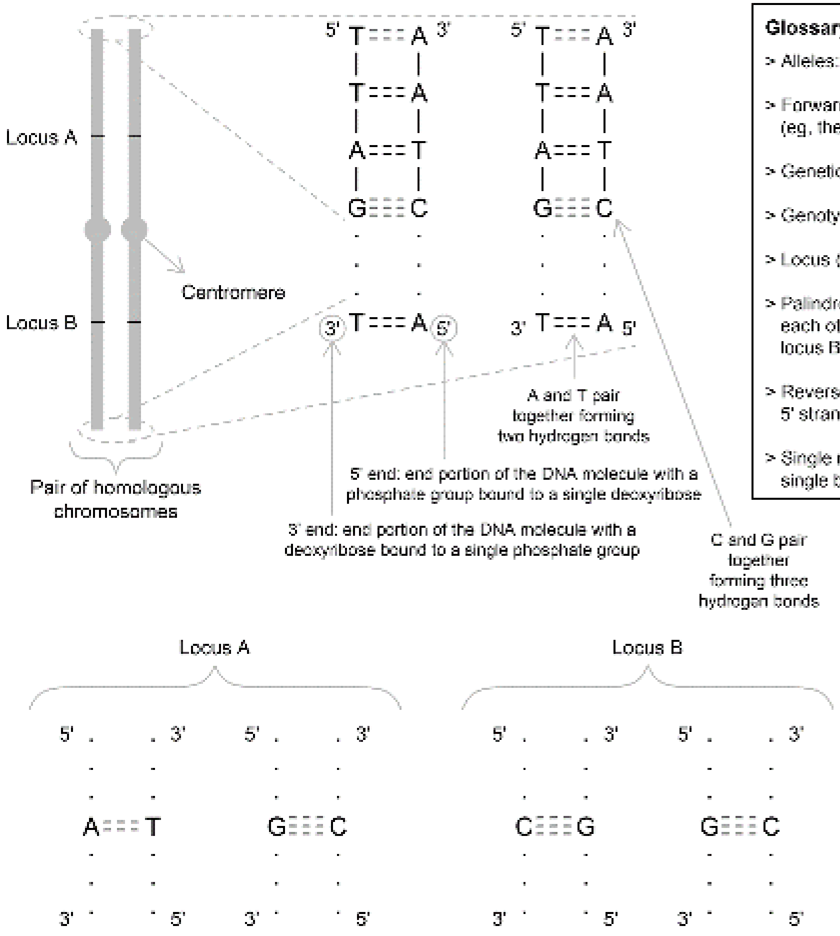
> Single nucleotide polymorphism (SNP): a type of genetic variant that involves single base pair changes.



	Locus A
Type of genetic variation	Single nucleoti polymorphism
Alleles (5' to 3')	A and G
Alleles (3' to 5')	T and C
Genotype (5' to 3')	AG
Genotype (3' to 5')	TC
Palindromic variant	No

Hartwig FP et al. *Int J Epidemiol.* (2016) 45(6):1717-1726

The issue of strand



Glossary

> Alleles: variant forms that a locus may present.

> Forward or positive strand: the DNA strand that is read from the 5' to the 3' end (eg, the 5' TTAG...T 3' strand in the figure).

> Genetic variant: locus with more than one allele in a population.

> Genotype: combination of alleles that an individual presents at a given locus.

> Locus (plural loci): a specific location in a DNA sequence.

> Palindromic SNP: SNPs whose alleles correspond to nucleotides that pair with each other in a double-stranded DNA molecule. SNPs with A/T or G/C (as in locus B below) alleles are palindromic SNPs.

> Reverse or negative strand: the DNA strand that is read from the 3' to the 5' strand (eg, the 3' AATC...A 5' strand in the figure).

> Single nucleotide polymorphism (SNP): a type of genetic variant that involves single base pair changes.

	Locus A	Locus B
Type of genetic variation	Single nucleotide polymorphism	Single nucleotide polymorphism
Alleles (5' to 3')	A and G	C and G
Alleles (3' to 5')	T and C	G and C
Genotype (5' to 3')	AG	CG
Genotype (3' to 5')	TC	GC
Palindromic variant	No	Yes

Hartwig FP et al. *Int J Epidemiol.* (2016) 45(6):1717-1726

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*

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	0.022	A	G	0.26

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	0.022	A	G	0.26
rs23456	-0.485	G	T	0.41	**-0.056**	**G**	**T**	**0.39**

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	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

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	

MR methods for handling horizontal pleiotropy

Many methods now exist!

Extensions to MR

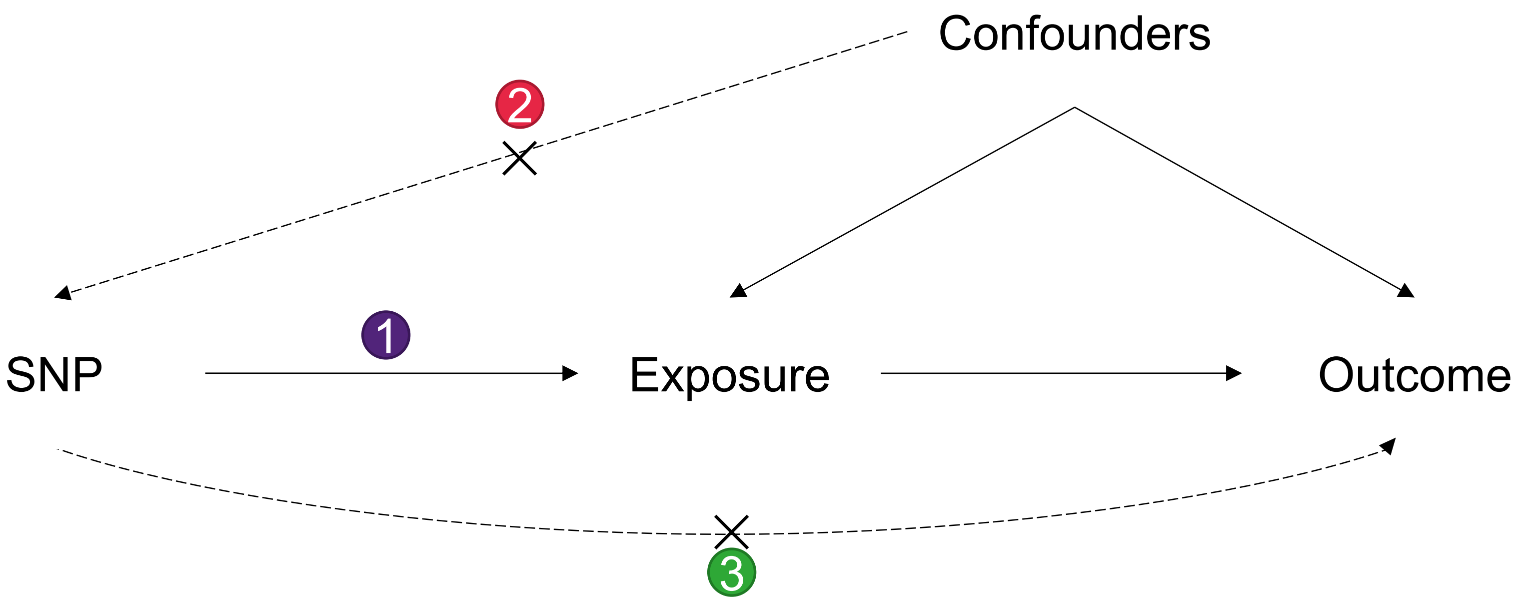
- MR uses genetic variants to test for causal relationships between phenotypic exposures and disease-related outcomes

- 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 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

Two-sample MR



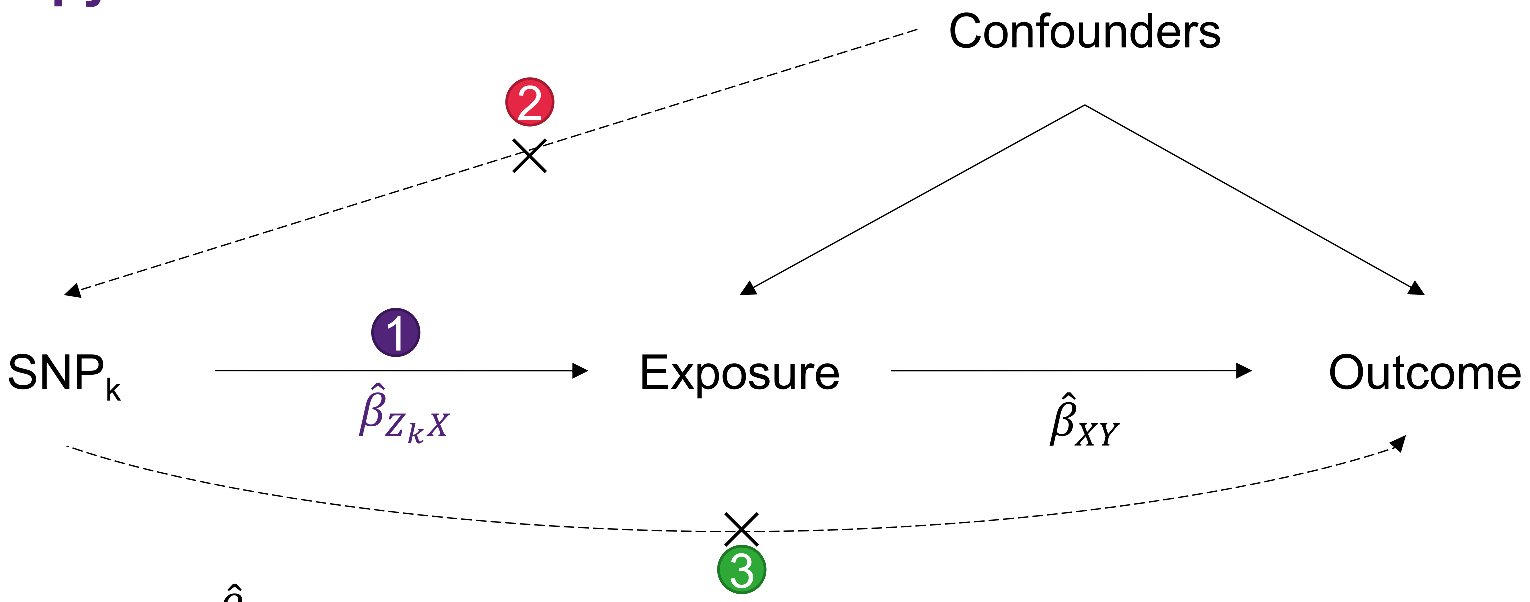
(1) Relevance assumption: SNP is associated with the exposure

(2) Independence assumption: SNP is NOT associated with confounding variables

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

Two-sample MR

No direct pleiotropy



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

Causal effect by Wald estimator:

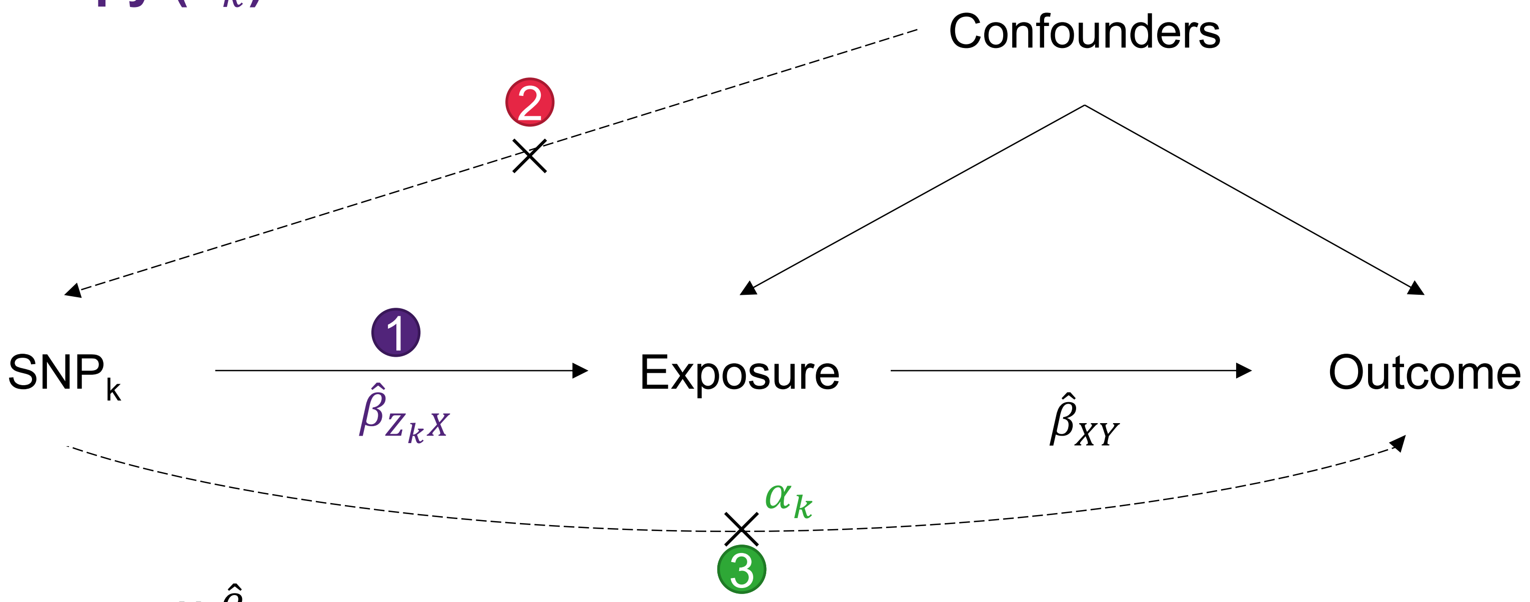
$$\hat{\beta}_{XY_k} = \frac{\hat{\beta}_{SNP_k-Outcome}}{\hat{\beta}_{SNP_k-Exposure}}$$

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)



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

Causal effect by Wald estimator:

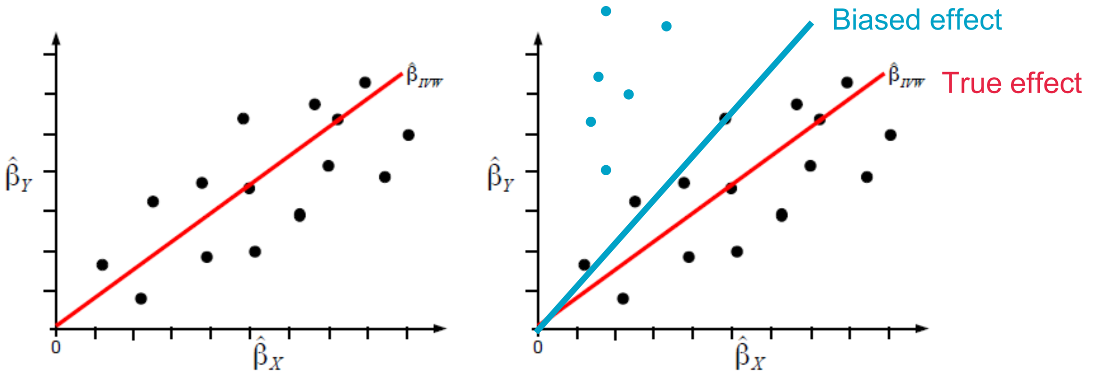
$$\frac{\hat{\beta}_{SNP_k-Outcome}}{\hat{\beta}_{SNP_k-Exposure}} = \hat{\beta}_{XY_k} + \frac{\alpha_k}{\hat{\beta}_{SNP_k-Exposure}}$$

Inverse variance weighted (IVW) average causal effect:

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

Heterogeneity

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



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, arises because at least some of the instruments are invalid.

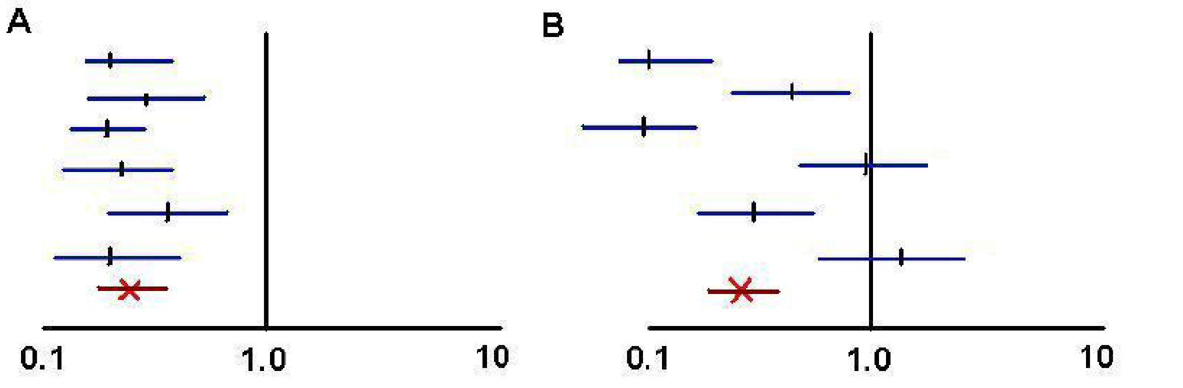
Cochran's Q statistic (heterogeneity test):

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

Where v_k is the 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 it's plausible that there are one or more genetic variants that have pleiotropic effects.



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

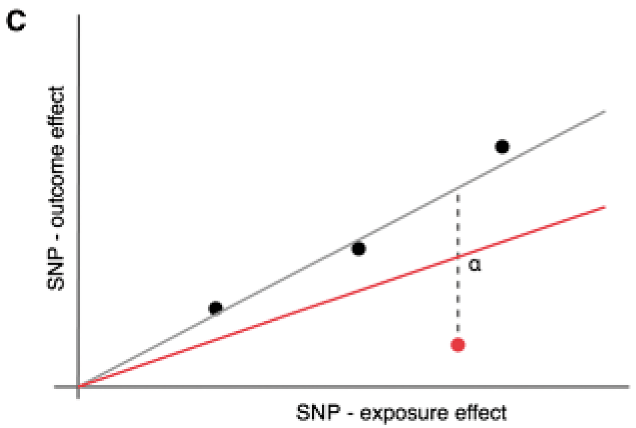
Bowden et al. Am J Epidemiol. (2018) 187(12):2681-2685

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

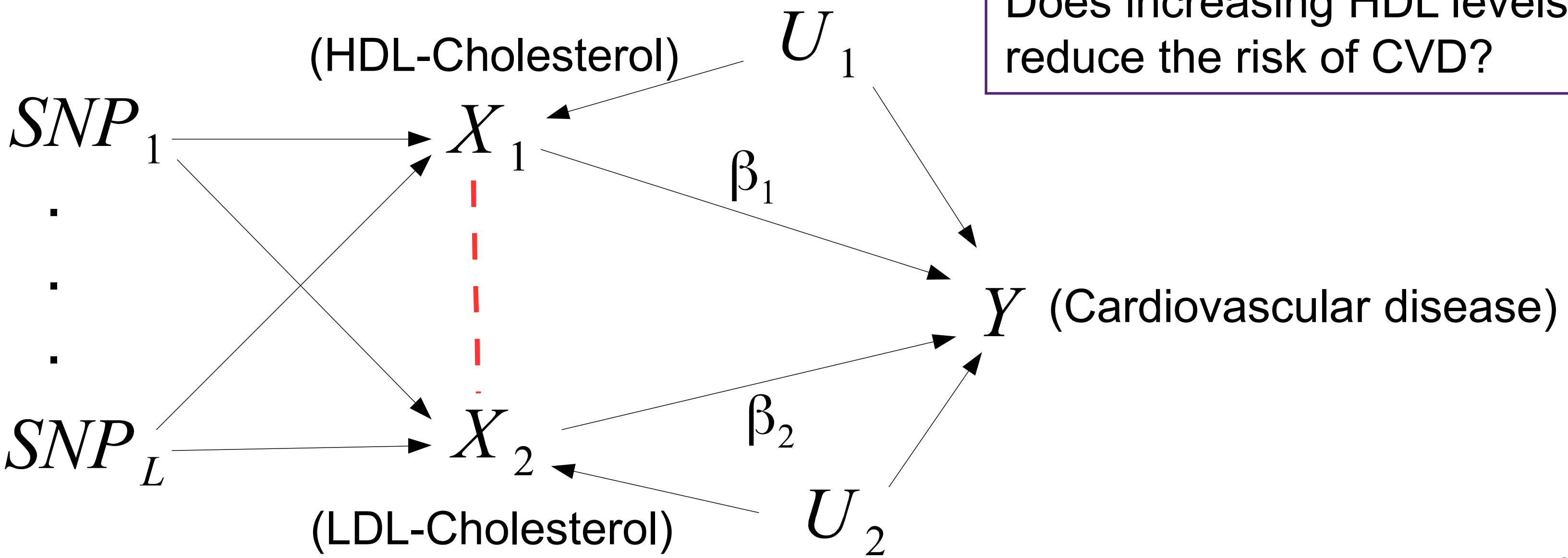
- However – beware of:
  - Cherry picking – removing outliers will artificially provide a more precise estimate
  - What if the outlier is the only valid instrument, and all the others are invalid?
    - E.g. cis-variants for gene expression, DNA methylation, protein levels.
    - CRP levels are best 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 estimate assumes all SNPs are valid instruments, and averages across them all

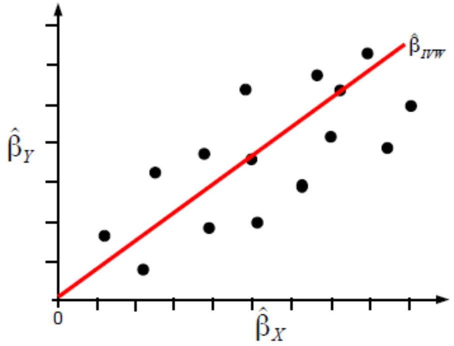
- Additive random effects estimate:
  - Estimate the between IV estimate of heterogeneity (denoted by τ^2), then calculate and update IVW estimate by replacing v_k with $v_k + \tau^2$
  - Point estimate and variance different from $\hat{\beta}_{IVW}$

- Multiplicative random effects model
  - Replace v_k with ϕv_k, where $\phi = \frac{Q}{K-1}$
  - Point estimate equals $\hat{\beta}_{IVW}$, but variance is inflated

- Additive random effects model popular in meta-analysis, but can perform poorly in the presence of pleiotropy

Heterogeneity and pleiotropy

- IVW assumes all variants are valid instrumental variables
  - Clear trend in estimates increasing with $\hat{\beta}_{Z_k X}$ from origin
  - Cochran's Q ≈ K – 1 (no heterogeneity)



- If there is an indication that these don't hold in the data, invalid "pleiotropic" variants could be the cause

Investigating heterogeneity and pleiotropy

- The IVW method assumes the underlying SNP-outcome model is

$$\hat{\beta}_{Yk} = \beta_{IVW}\beta_{Xk} + \varepsilon_{Yk} \qquad (\varepsilon_{Yk} \text{ independent of } \hat{\beta}_{Xk})$$

β_{Xk} replaced with $\hat{\beta}_{Xk}$ when fitting the model

- A more realistic model to account for heterogeneity might be:

$$\hat{\beta}_{Yk} = \alpha_k + \beta_{IVW}\beta_{Xk} + \varepsilon_{Yk}$$

Where α_k is the pleiotropic effect of variant *k*

- Can the IVW method still estimate the causal effect without bias even when all variants have pleiotropic effects? Yes, if:
  - α_k is independent of β_{Xk} across *K* SNPs (InSIDE assumption)
  - The mean value of α_k is zero
  - If satisfied, pleiotropy is said to be balanced

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

Investigating heterogeneity and pleiotropy

- The IVW method assumes the underlying SNP-outcome model is

$$\hat{\beta}_{Yk} = \beta_{IVW}\beta_{Xk} + \varepsilon_{Yk} \qquad (\varepsilon_{Yk} \text{ independent of } \hat{\beta}_{Xk})$$

β_{Xk} replaced with $\hat{\beta}_{Xk}$ when fitting the model

- A more realistic model to account for heterogeneity might be:

$$\hat{\beta}_{Yk} = \alpha_k + \beta_{IVW}\beta_{Xk} + \varepsilon_{Yk}$$

Where α_k is the pleiotropic effect of variant *k*

- Can the IVW method still estimate the causal effect without bias even when all variants have pleiotropic effects? Yes, if:
  - α_k is independent of β_{Xk} across *K* SNPs (InSIDE assumption)
  - The mean value of α_k is zero
  - If satisfied, pleiotropy is said to be balanced

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

Investigating heterogeneity and pleiotropy

- The IVW method assumes the underlying SNP-outcome model is

$$\hat{\beta}_{Yk} = \beta_{IVW}\beta_{Xk} + \varepsilon_{Yk} \qquad (\varepsilon_{Yk} \text{ independent of } \hat{\beta}_{Xk})$$

β_{Xk} replaced with $\hat{\beta}_{Xk}$ when fitting the model

- A more realistic model to account for heterogeneity might be:

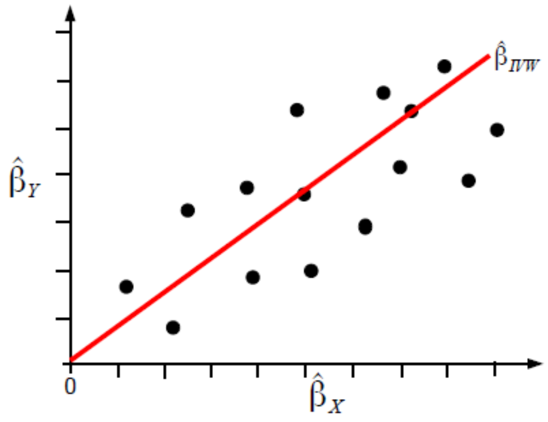
$$\hat{\beta}_{Yk} = \alpha_k + \beta_{IVW}\beta_{Xk} + \varepsilon_{Yk}$$

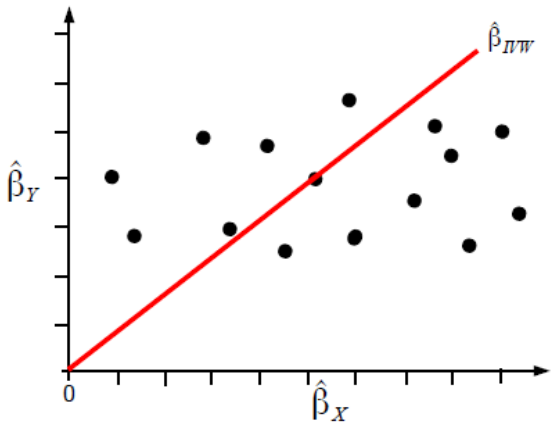
Where α_k is the pleiotropic effect of variant *k*

- Can the IVW method still estimate the causal effect without bias even when all variants have pleiotropic effects? Yes, if:
  - α_k is independent of β_{Xk} across *K* SNPs (InSIDE assumption)
  - The mean value of α_k is zero
  - If satisfied, pleiotropy is said to be balanced

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

Balanced or directional pleiotropy



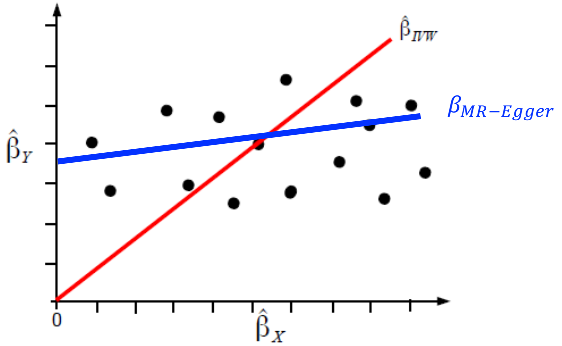


- Trend towards origin + heterogeneity
- Pleiotropy potentially causing heterogeneity -> IVW appears to be a good fit

- Trend away from origin + heterogeneity
- Pleiotropy potentially causing heterogeneity and bias
  - IVW does not appear to be good fit
  - Zero-intercept condition unreasonable

MR-Egger regression: Central concept

We could therefore regress the SNP-outcome associations on the SNP-exposure associations, but allow for a non-zero intercept in the regression

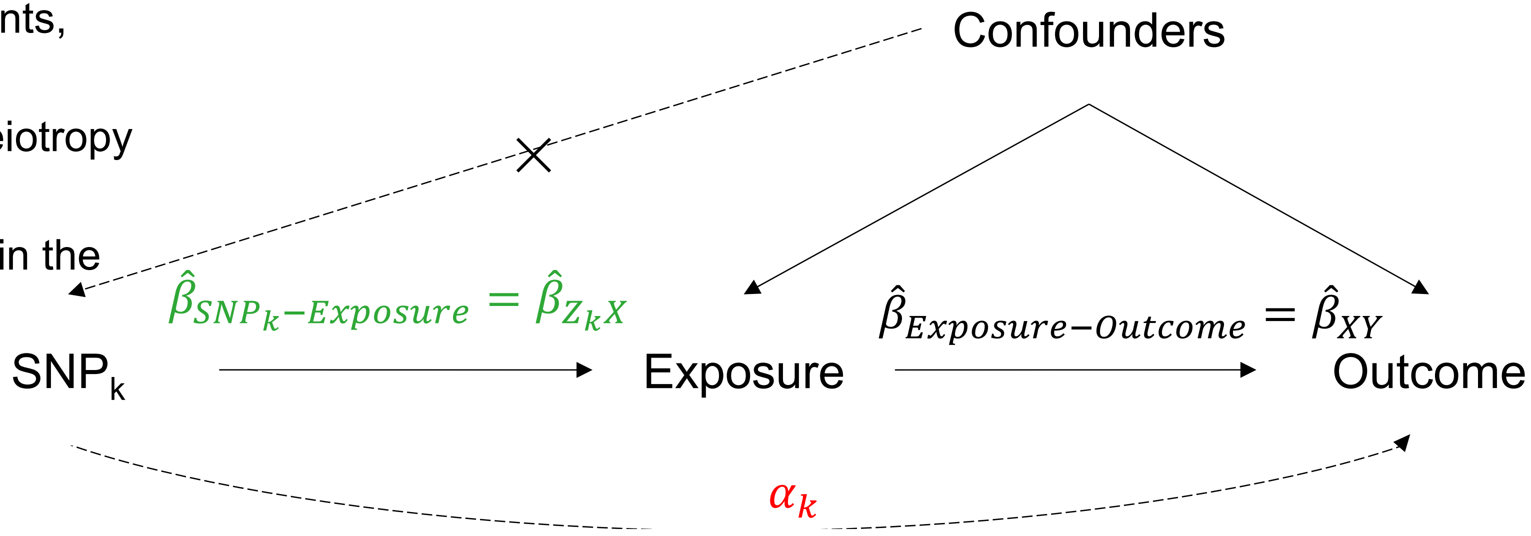


MR-Egger allows for a **non-zero intercept** in the regression.

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)

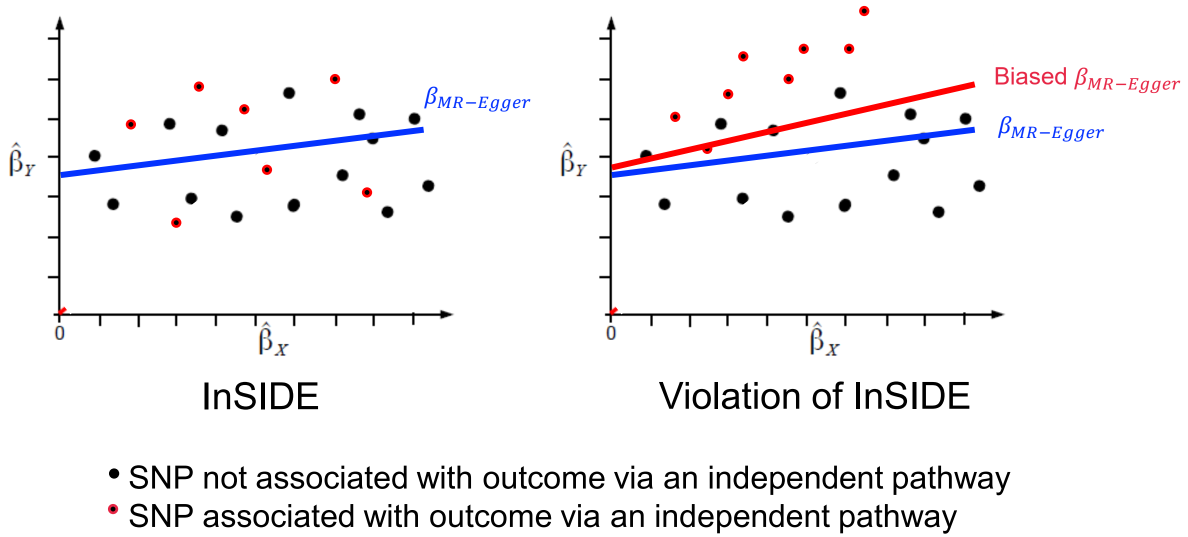
MR-Egger lacks power.



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

MR-Egger regression

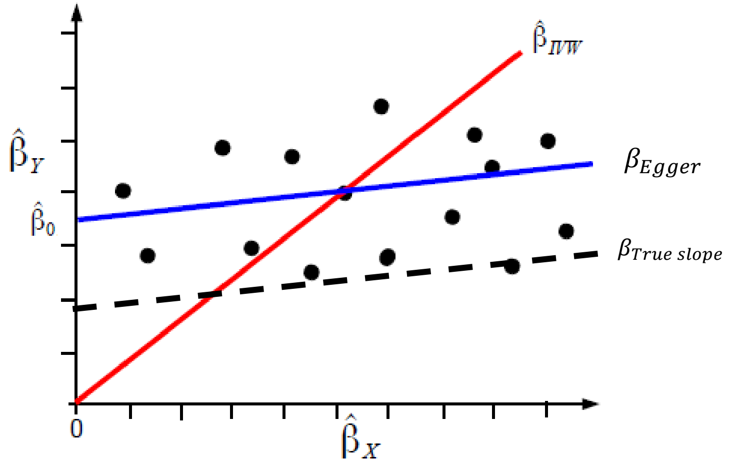
MR-Egger regression replies on the InSIDE (INstrument Strength Independent of Direct Effect) assumption, which states that the pleiotropic effects of SNPs must be independent of their strength as instruments.



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

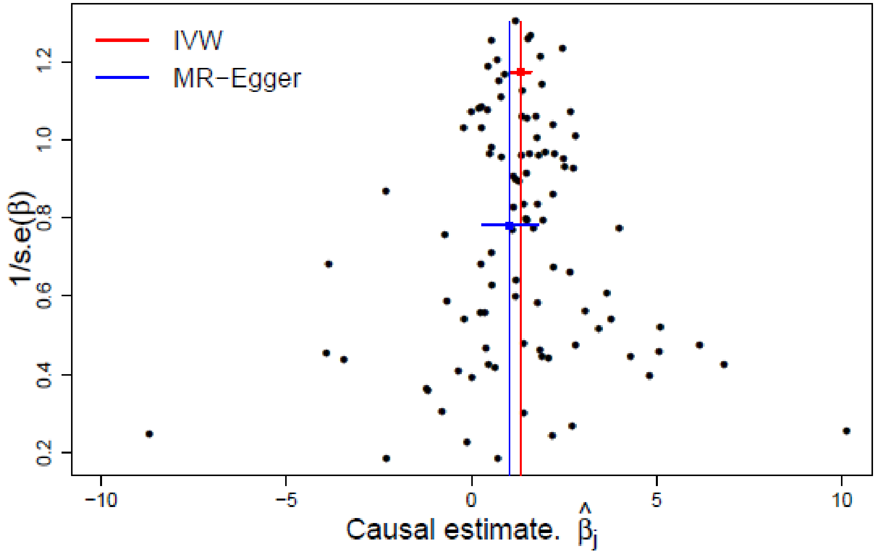
MR-Egger model: $\hat{\beta}_{Yk} = \beta_0 + \underbrace{\beta_{Egger}}_{\text{Slope}} \hat{\beta}_{Xk} + \varepsilon_{Yk}$

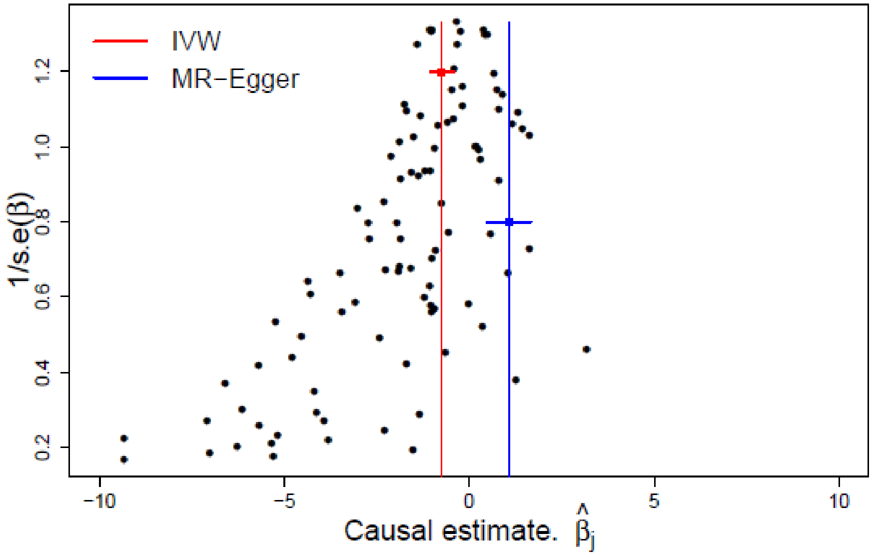
- β_0 is the intercept term. β_0 can be interpreted as the average pleiotropic effect across all genetic variants. A non-zero β_0 indicates directional pleiotropy.
- β_{Egger} is the causal estimate adjusted for directional pleiotropy



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

Funnel plot: balanced versus directional pleiotropy





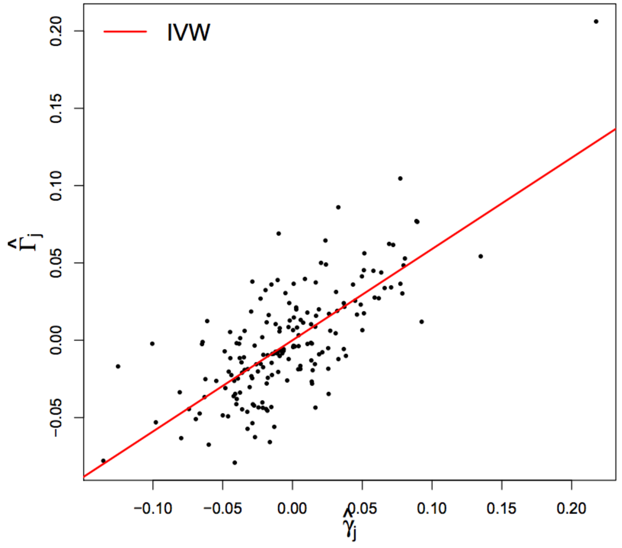
Funnel is symmetric -> pleiotropy appears to be balanced so IVW is okay

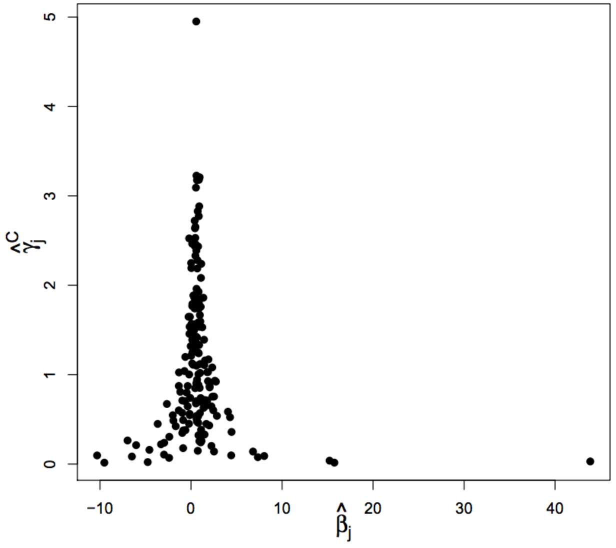
Funnel is asymmetric -> pleiotropy appears to be directional so IVW is not okay

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)

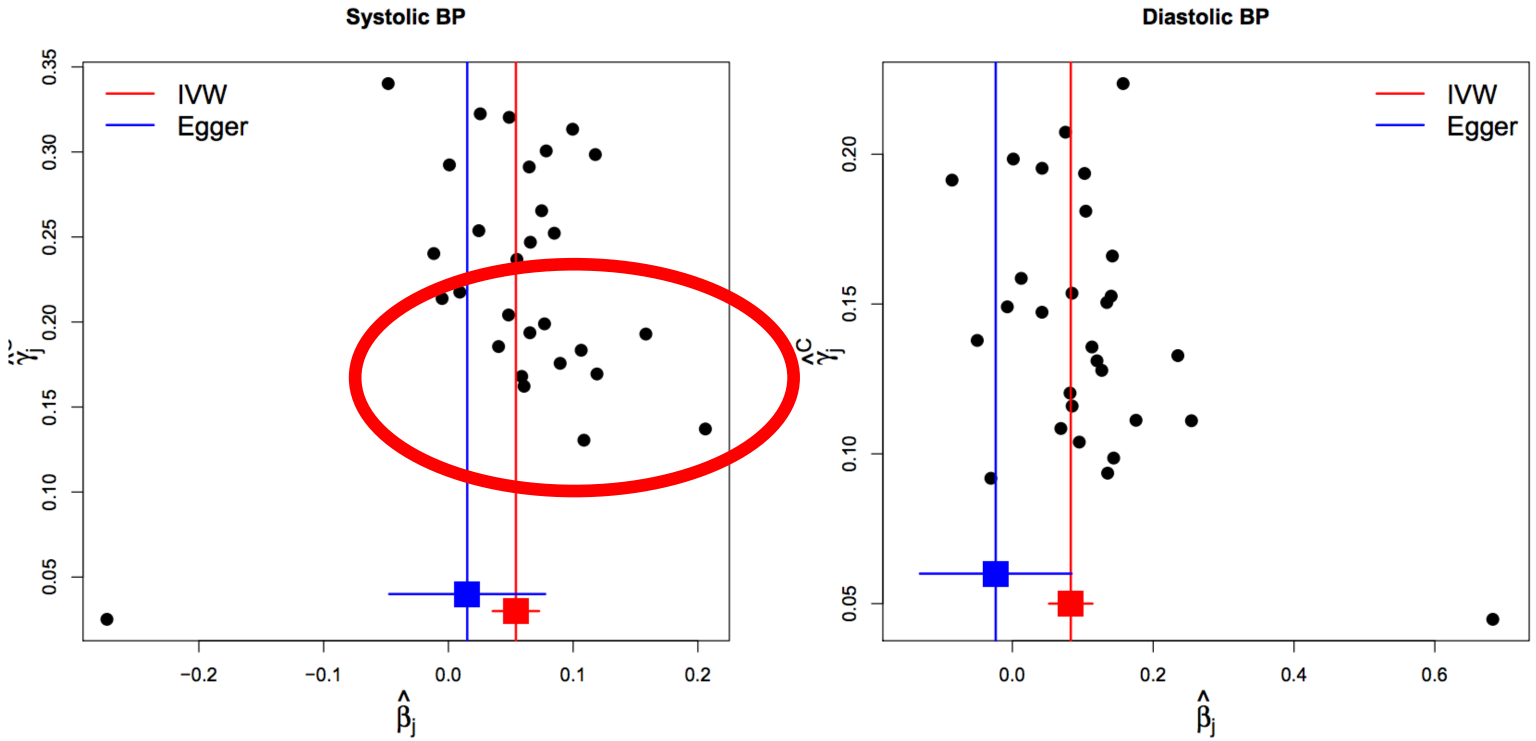
MR-Egger = 0.58 (95% CI: 0.50, 0.67);

intercept = -0.001 p=0.5

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

MR-Egger regression

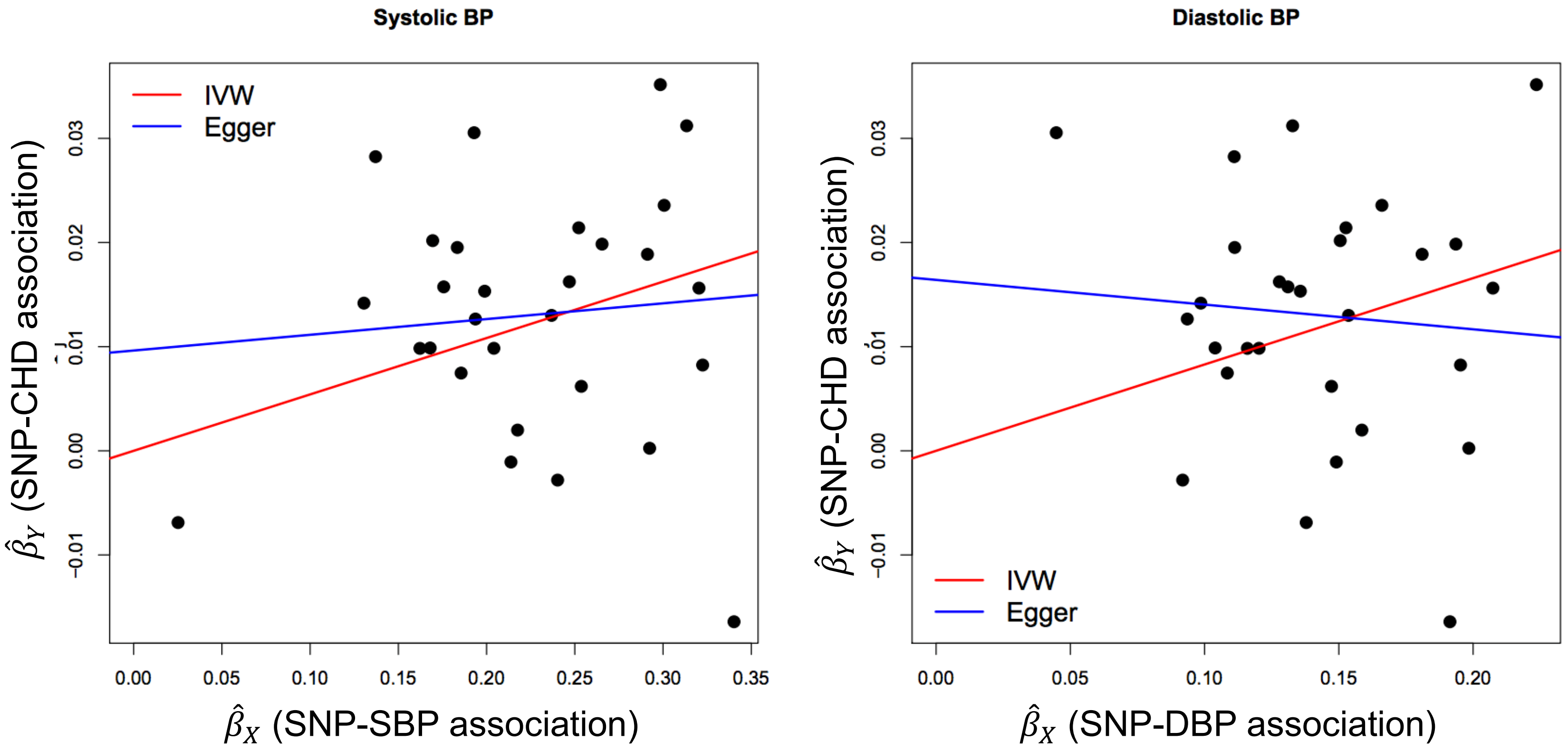
Example: BP and Coronary Heart Disease



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 $p=4\times10^{-6}$
Egger =0.015 logOR/mmHg p=0.6

IVW= 0.083 logOR/mmHg $p=1\times10^{-5}$
Egger =-0.024 logOR/mmHg p=0.7

Median based methods (Median Estimator)

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

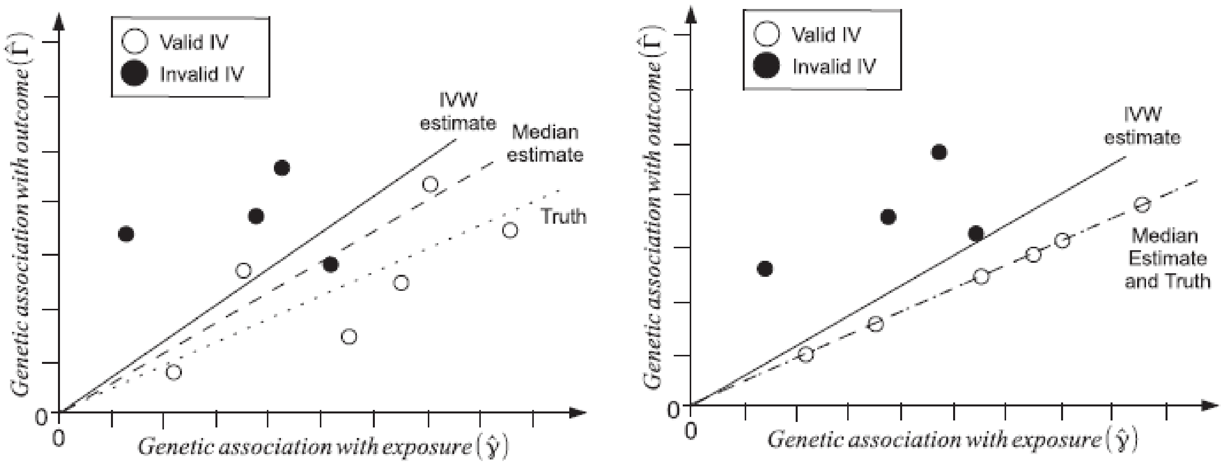


Figure 2. Fictional example of a Mendelian randomization analysis with 10 genetic variants—six valid instrumental variables (hollow circles) and four invalid instrumental variables (solid circles) for finite sample size (left) and infinite sample size (right) showing IVW (solid line) and simple median (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 estimation

- Simple median estimator:
  - Odd number of genetic variants: middle ratio estimate
  - Even number of genetic variants: median is interpolated between the two middle estimates $\left(\frac{1}{2}(\hat{\beta}_k + \hat{\beta}_{k+1})\right)$
  - 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

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 estimation

- 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: inverse of the variance of the ratio estimate: $w'_k = \frac{\hat{\beta}^2_{Z_k X}}{\sigma^2_{Z_k Y}}$

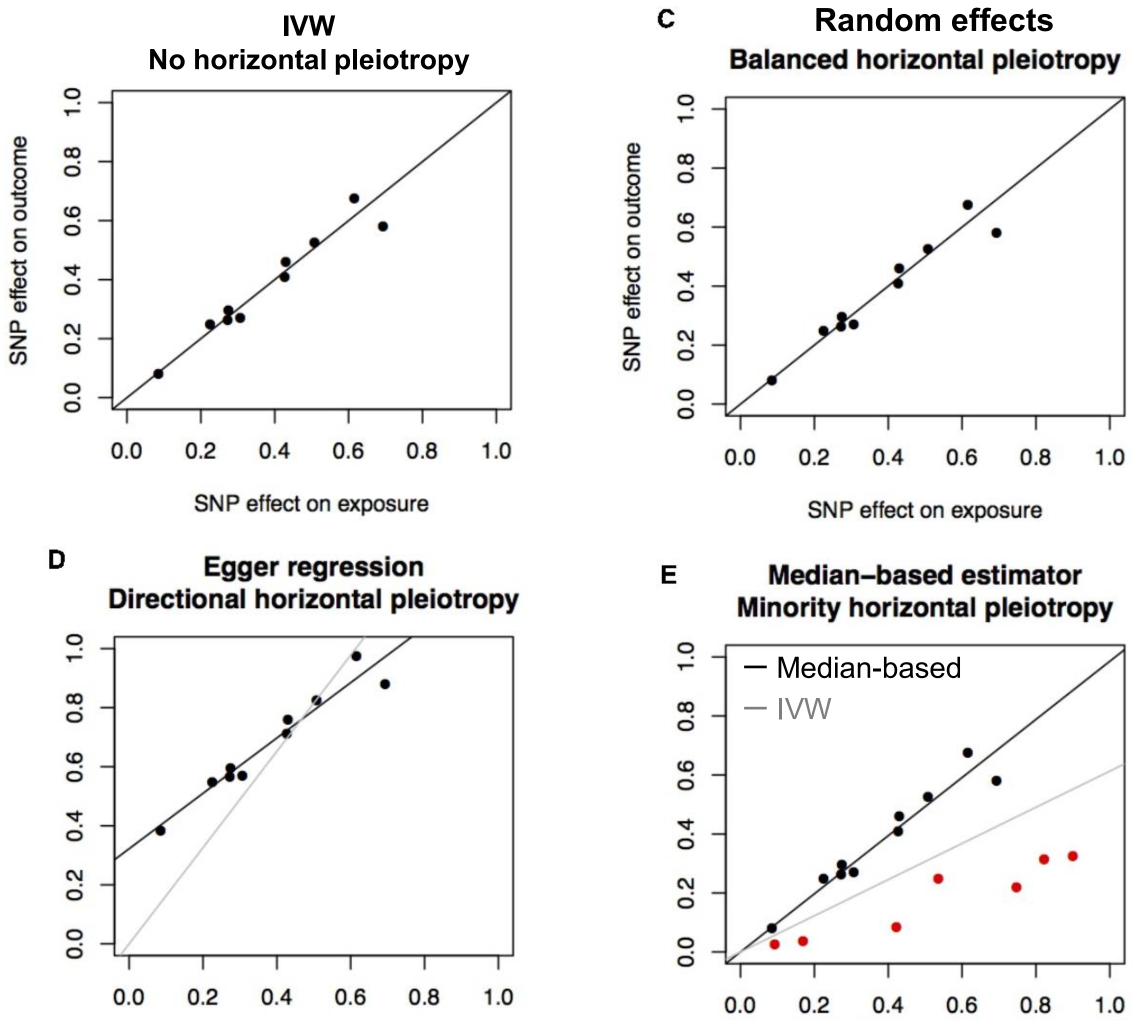
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$	$\hat{\beta}_{10}$
Simple median										
Weight $(1/v_k)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$
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

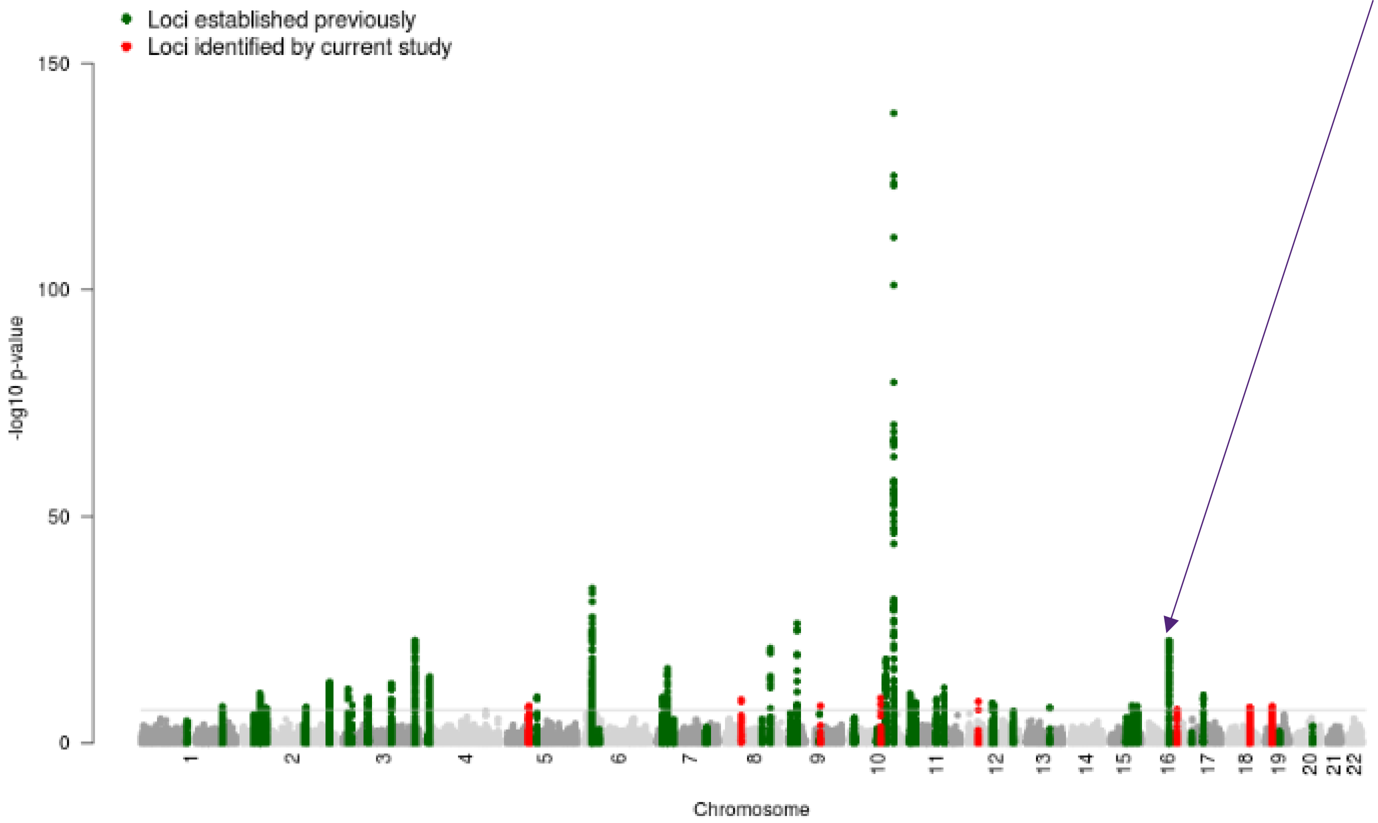


● SNPs associated with outcome via an independent pathway.

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

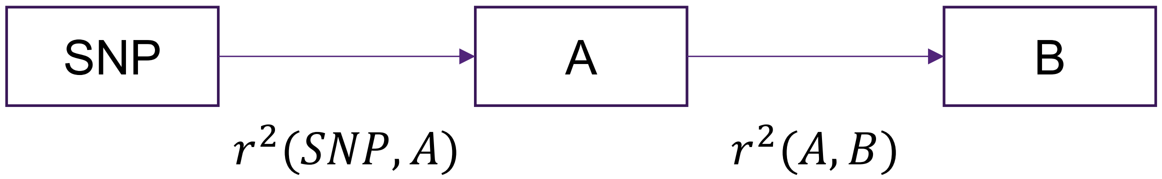
T2D ⟵ BMI ⟵ *FTO*

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

$$r^2(SNP, B) = r^2(SNP, A) \times \underbrace{r^2(A, B)}_{\text{This term is <1}}$$

- 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

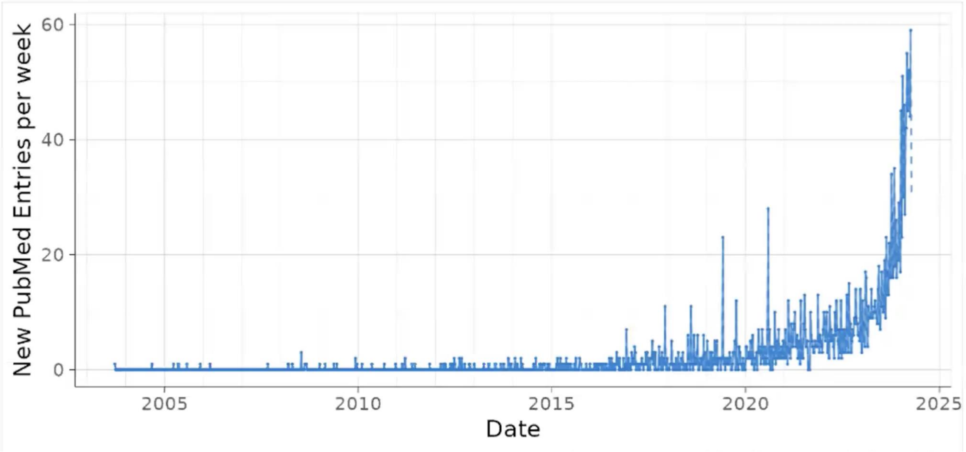
Links
Browse source code
Report a bug

License
Full license
MIT + file LICENSE

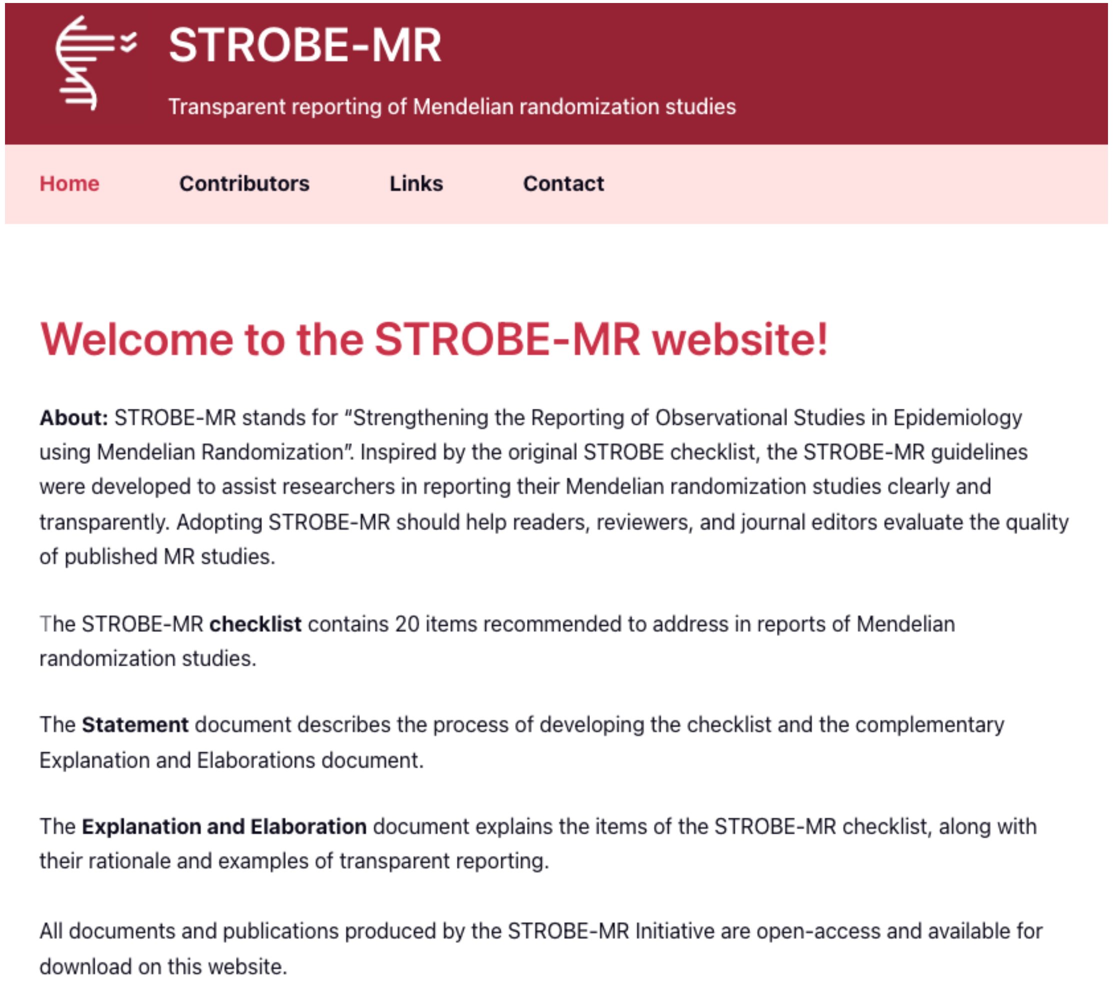
Citation
Citing TwoSampleMR

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

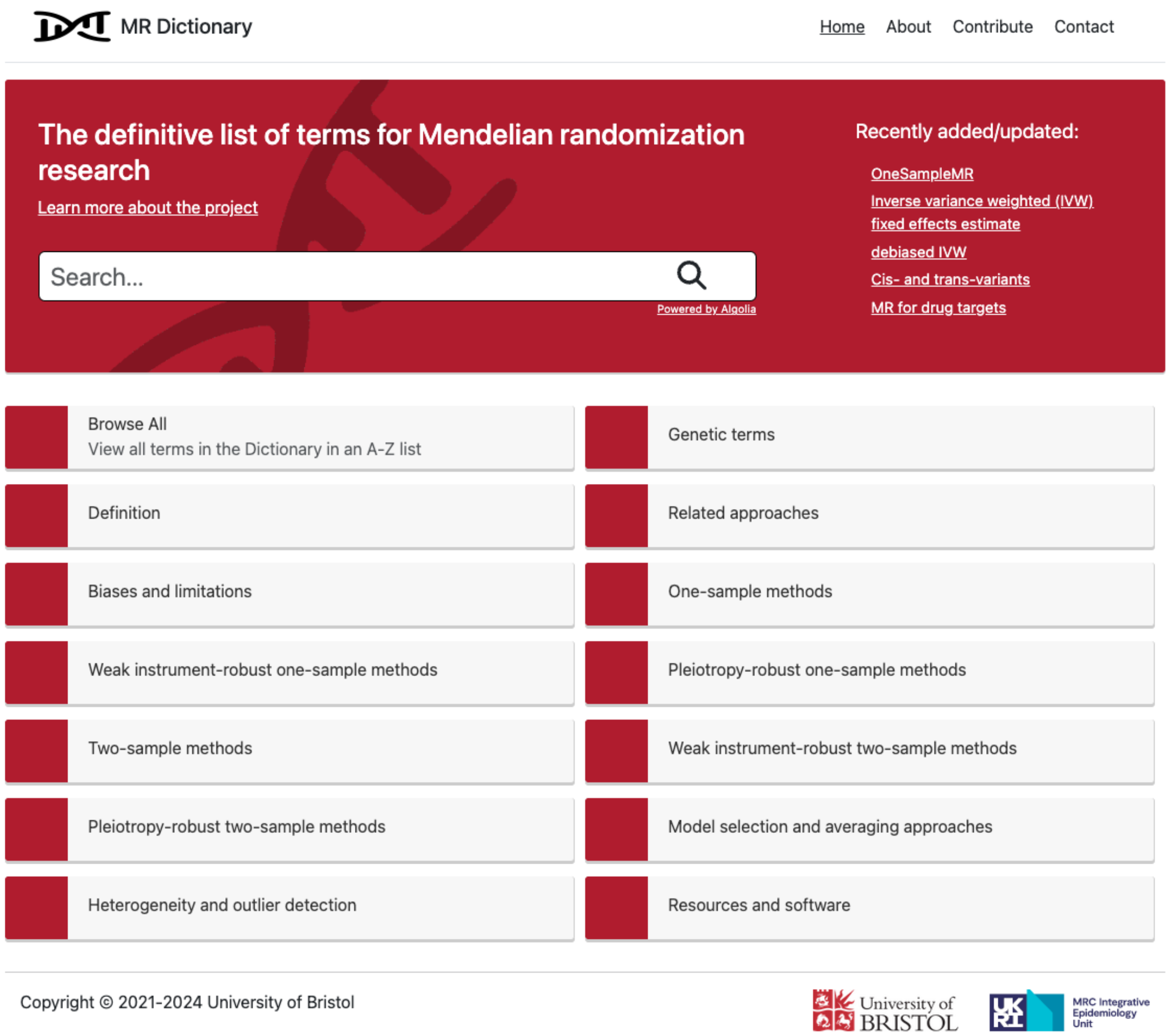
Dev status
R-CMD-check passing
lifecycle experimental
DOI 10.5281/zenodo.10684540
codecov 36%
r-universe 0.6.4



STROBE-MR



MR Dictionary



The University of Queensland
Australia

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 Epideml 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.