Mendelian Randomization

Daisy Crick

Data Agreement

To maximize your learning experience, we will be working with genuine human genetic data, during this module.

Access to this data requires agreement to the following in to comply with human genetic data ethics regulations

If you haven’t done so, please email <ctr-pdg-admin@imb.uq.edu.au> with your name and the below statement to confirm that you agree with the following:

“I agree that access to data is provided for educational purposes only and that I will not make any copy of the data outside the provided computing accounts.”

Learning materials

Instructions to access WiFi/desktop/server:

https://suave-pillow-de4.notion.site/Instruction-to-Computing-Resources-dcba658c9a584e6d80a443c5d64042d8?pvs=4

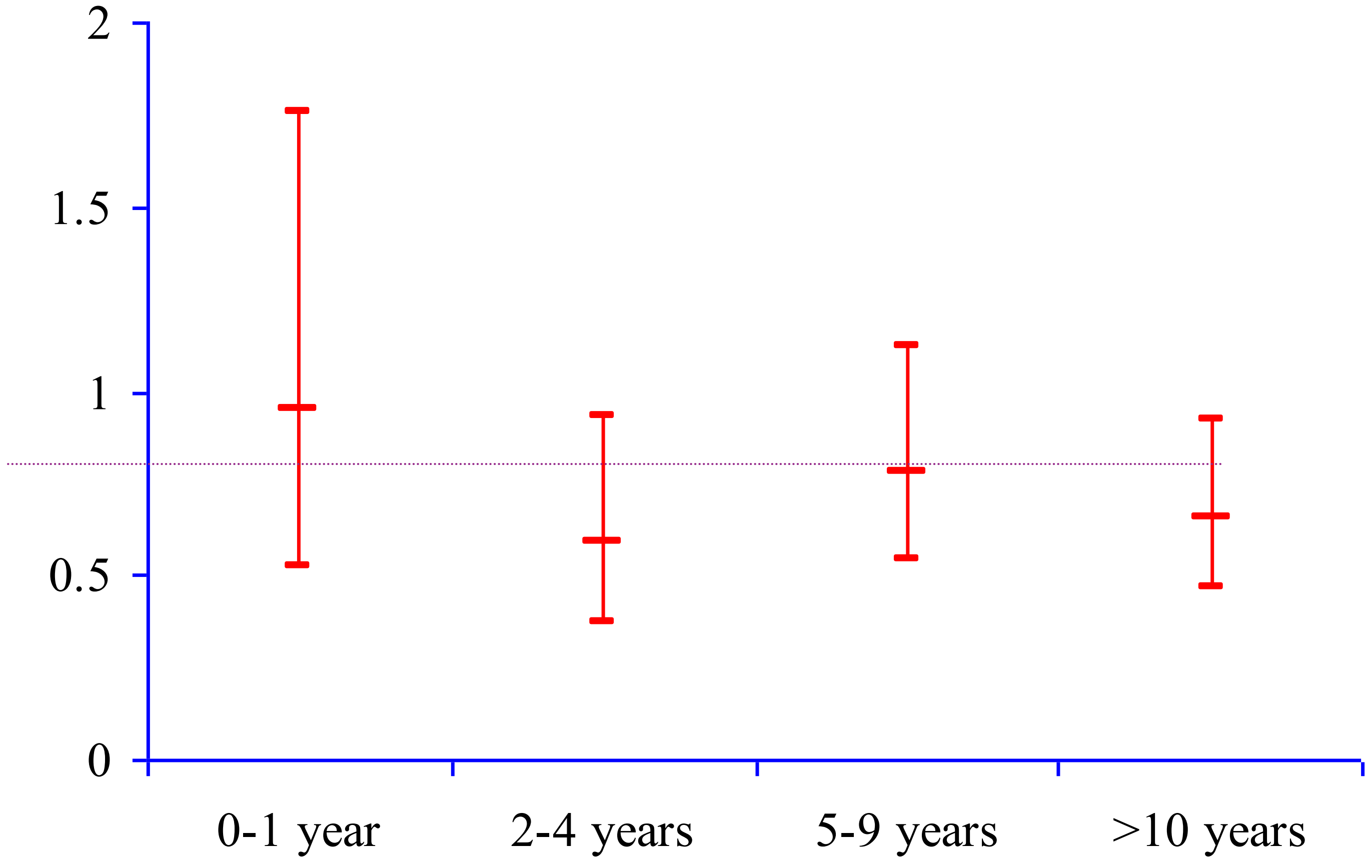
Slides and practical notes:

https://cnsgenomics.com/data/teaching/GNGWS25/module4

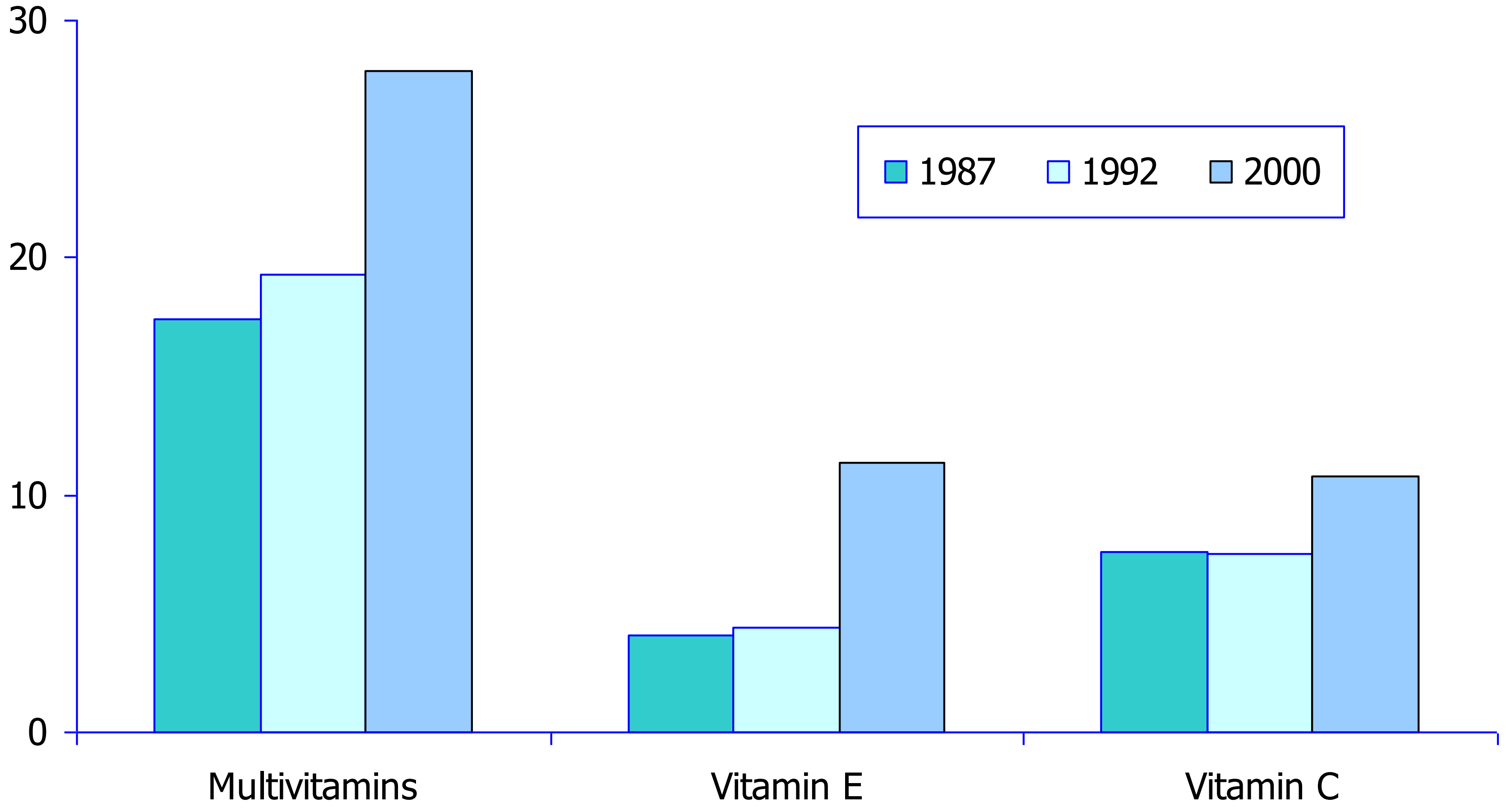
Learning Objectives

- Understand the issues of observational epidemiology.
- Understand how Mendelian randomization (MR) works, what its core assumptions and how to calculate causal effect estimates.
- Understand what directed acyclic graphs (DAGs) are and how they can be used to inform study design.
- Cover the basic limitations to Mendelian randomization.

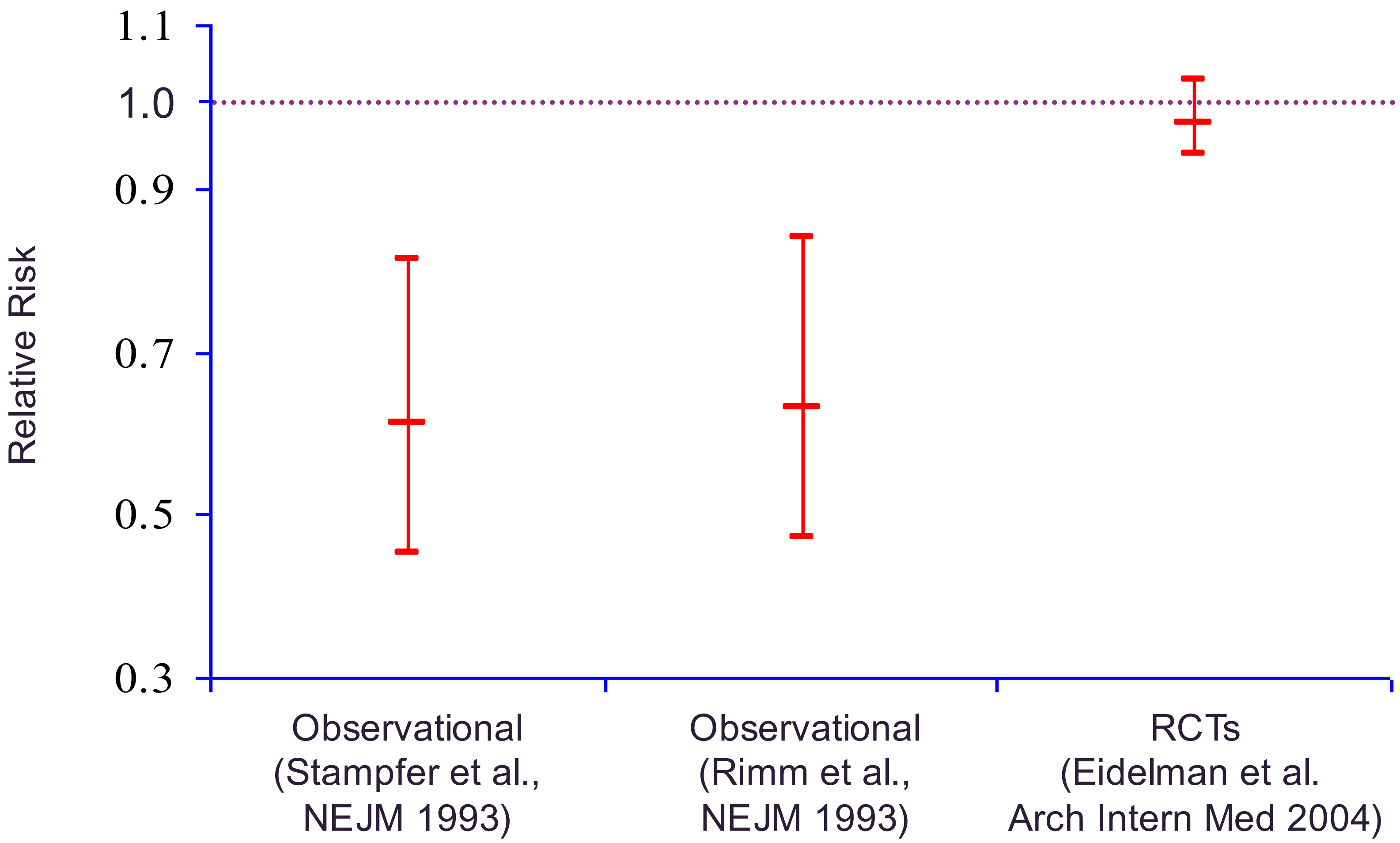
Vitamin E supplement use and risk of Coronary Heart Disease



Vitamin E supplement use and risk of Coronary Heart Disease



Vitamin E supplement use and risk of Coronary Heart Disease

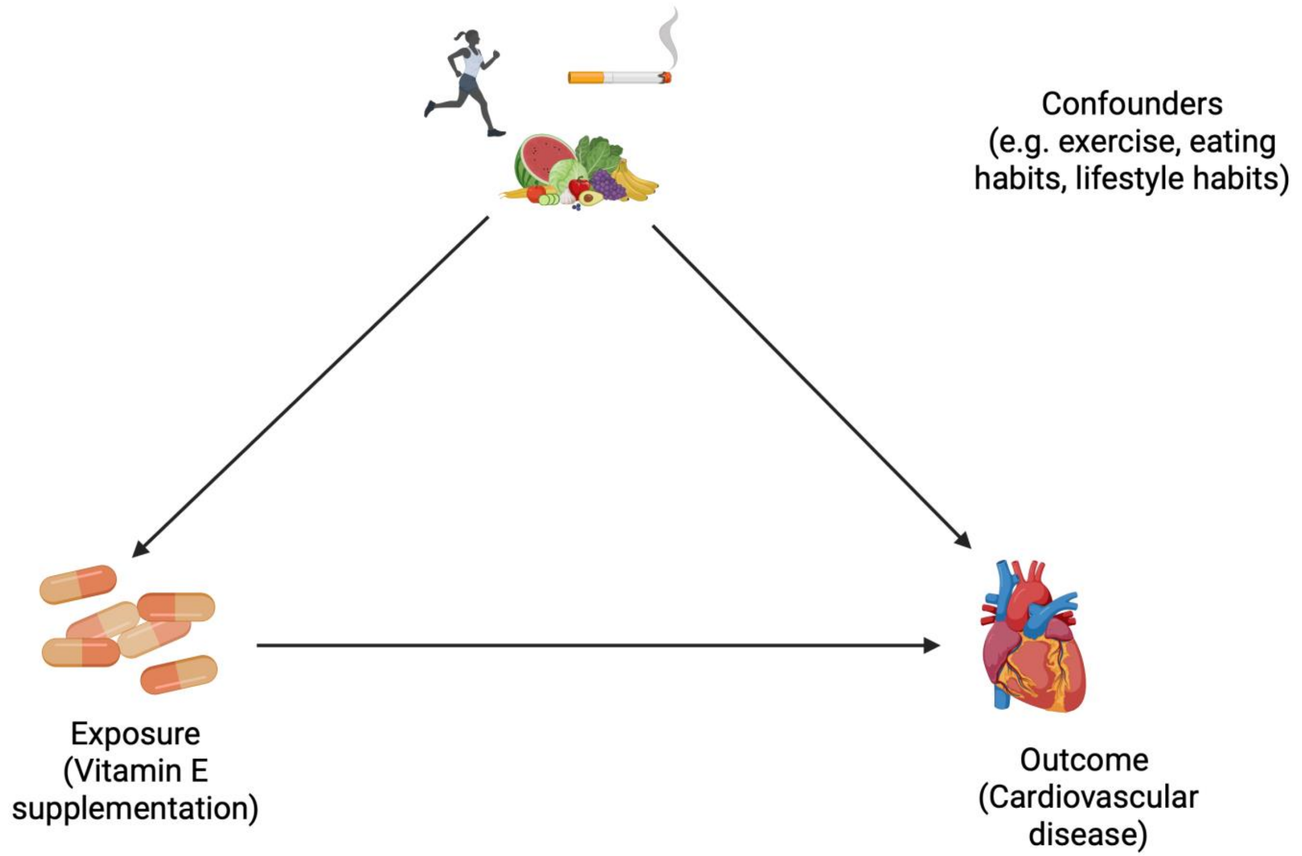


Inferring causality using observational data

- Results from observational studies can give the wrong answer.

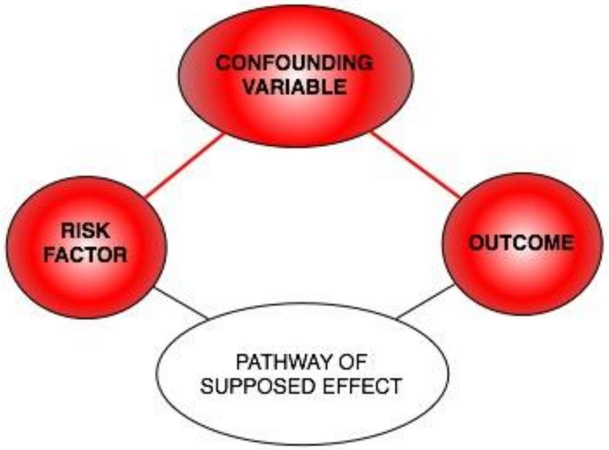
Inferring causality using observational data

- Results from observational studies can give the wrong answer.



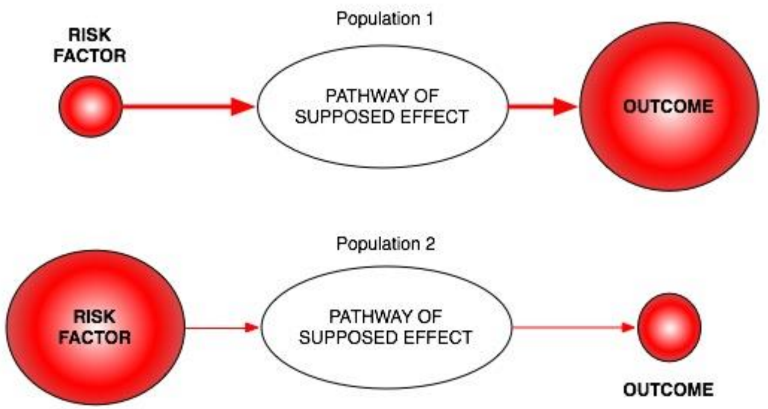
Classic limitations to observational science

Confounding



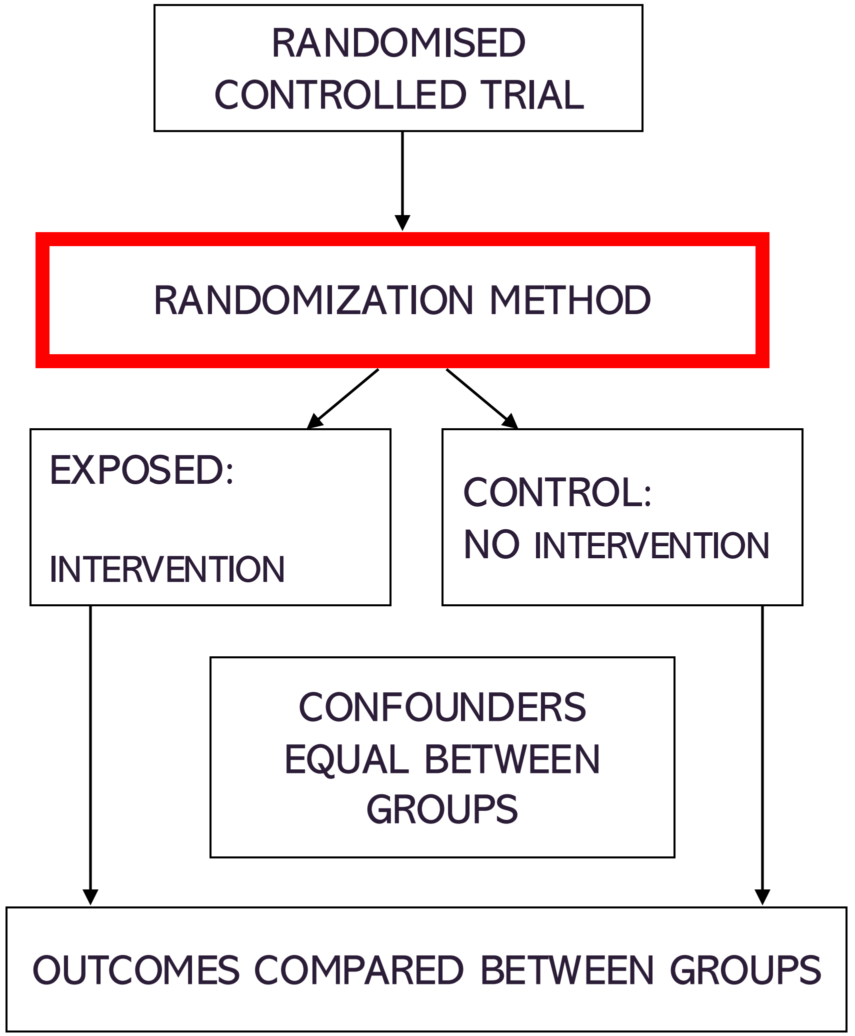
Reverse Causation

Bias



Randomised Control Trials (RCTs)

- The gold standard in inferring causality!



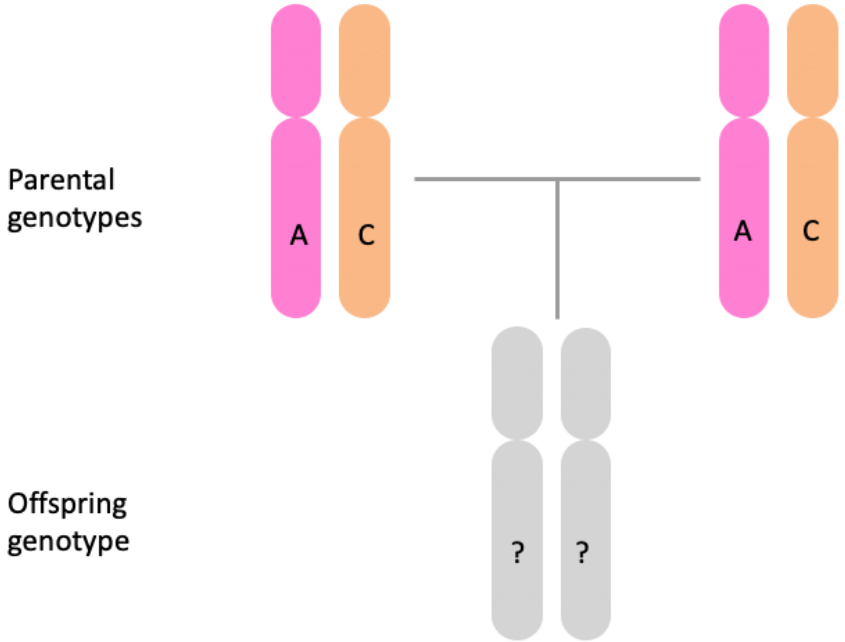
Mendelian randomization!

- A technique based on the idea that genetics can tell us about non-genetic factors and their effects on health and disease.
- MR uses genetic information as a proxy for non-genetic information.
- The modifiable exposure on the outcome will be the same whether the exposure is influenced by the environment or genetics.

Mendel's Laws of inheritance

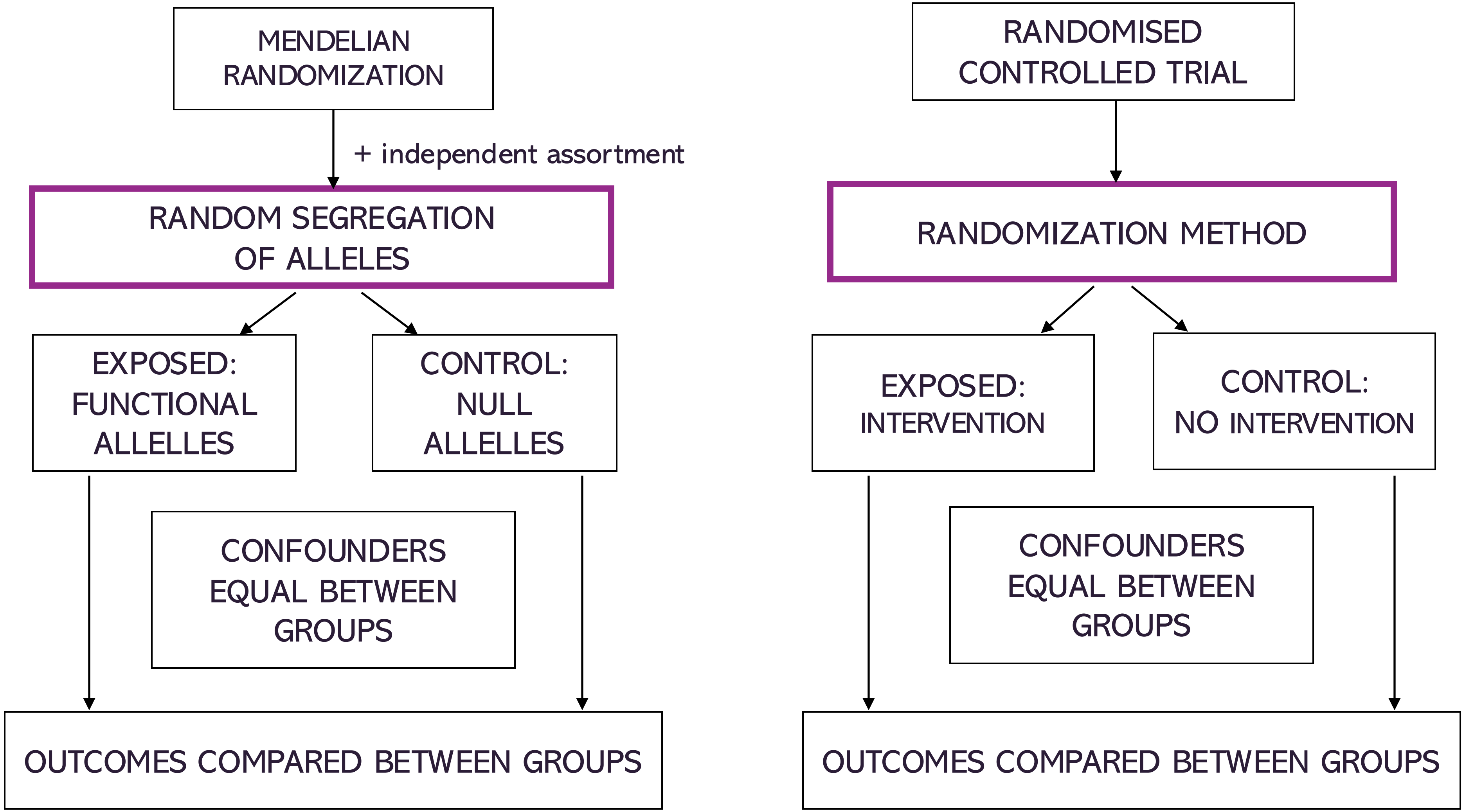
Gregor Mendel in 1862

1. **Segregation:** alleles separate at meiosis and a randomly selected allele is transmitted to offspring.

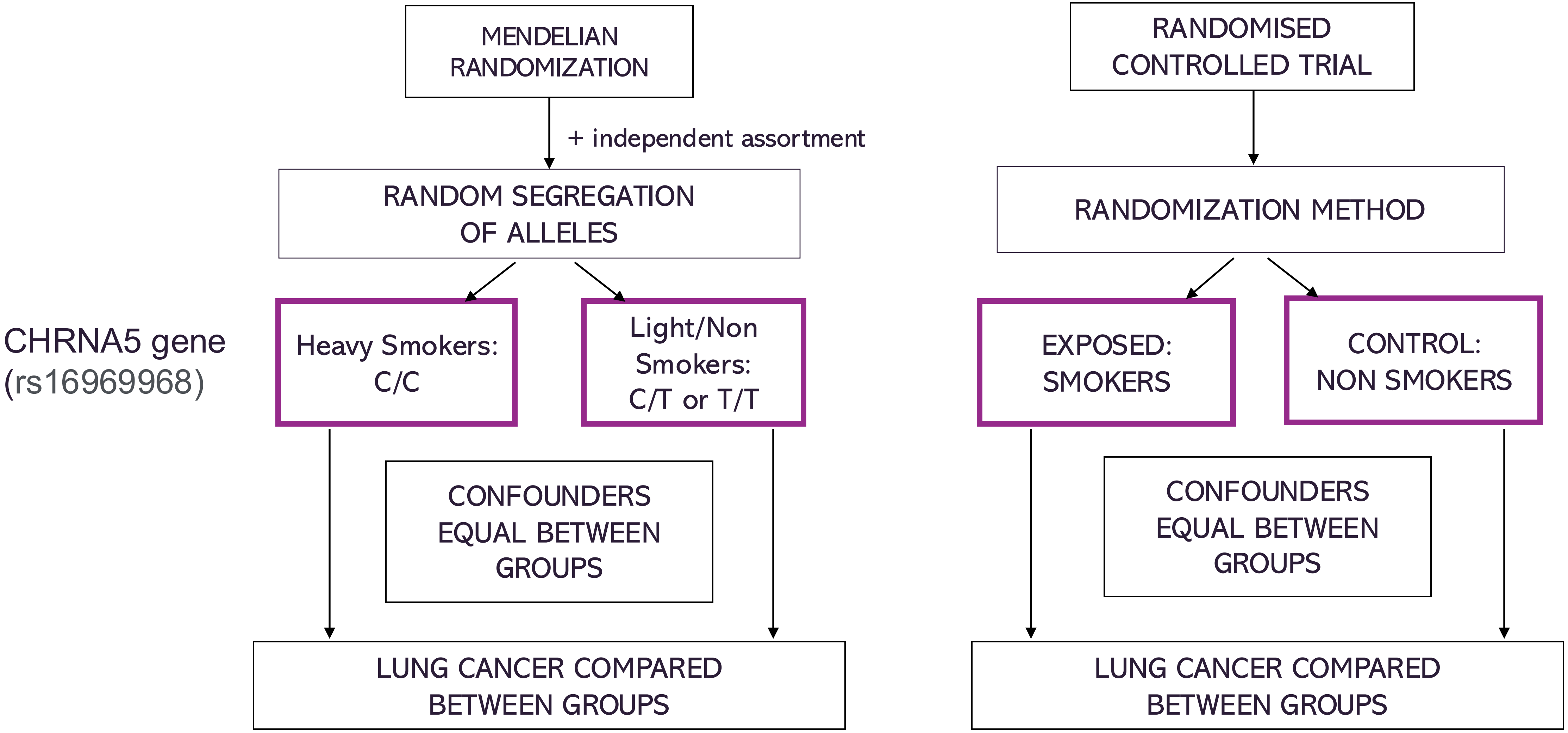


2. **Independent assortment:** alleles at different genetic loci (for different traits) are transmitted independently of one another.

Mendel’s Laws of inheritance



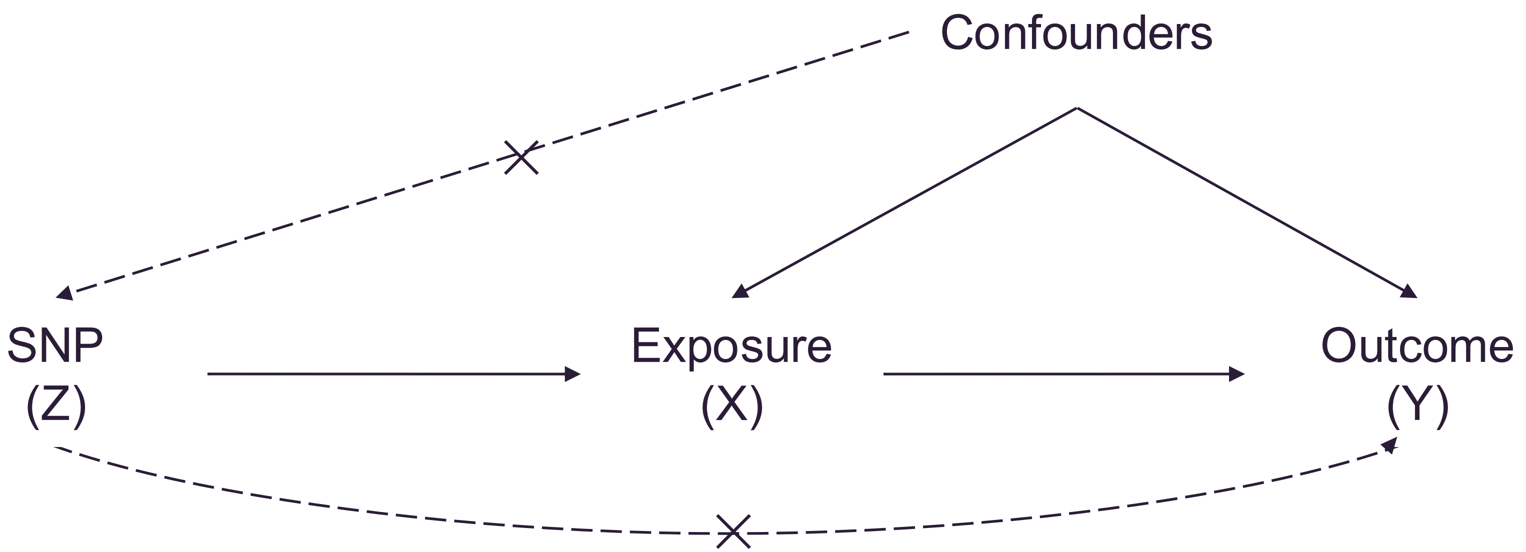
Mendel's Laws of inheritance



What is a DAG

- Directed Acyclic Graph.
- Systematic representation of causal relationships.
- Displays assumptions about the relationship between variables.
- Clarify study design.

What is a DAG



The DAG game

DAG Rules

- They have to be directed.
- They have to be acyclic.
- All common causes must be represented.
- Time flows from left to right.

DAG Rules

- They have to be directed.
- They have to be acyclic.
- All common causes must be represented.
- Time flows from left to right.

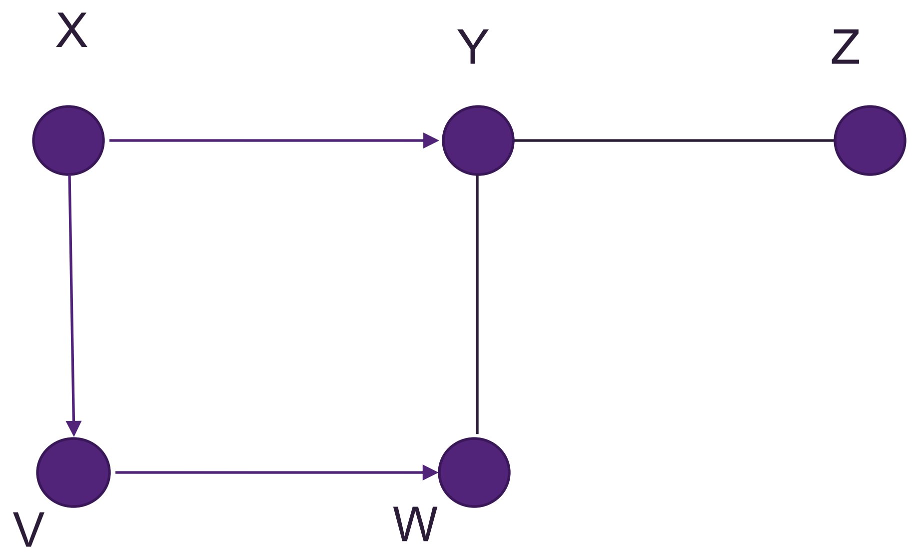
DAG Rules

- They have to be directed.
- They have to be acyclic.
- Common causes of two variables must be represented.
- Time flows from left to right.

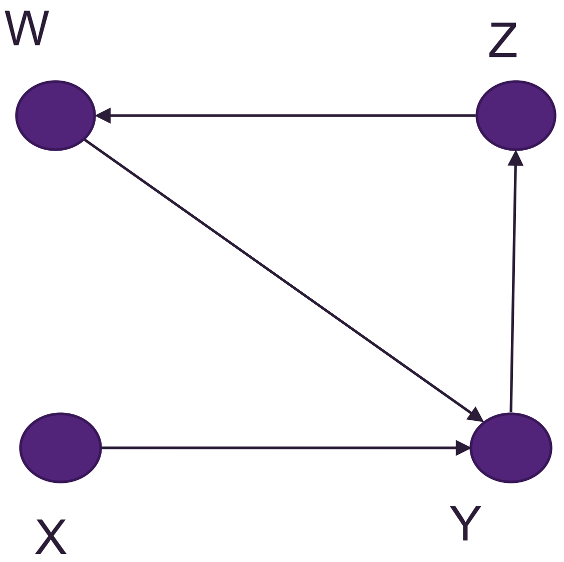
DAG Rules

- They have to be directed.
- They have to be acyclic.
- All common causes must be represented
- Time flows from left to right.

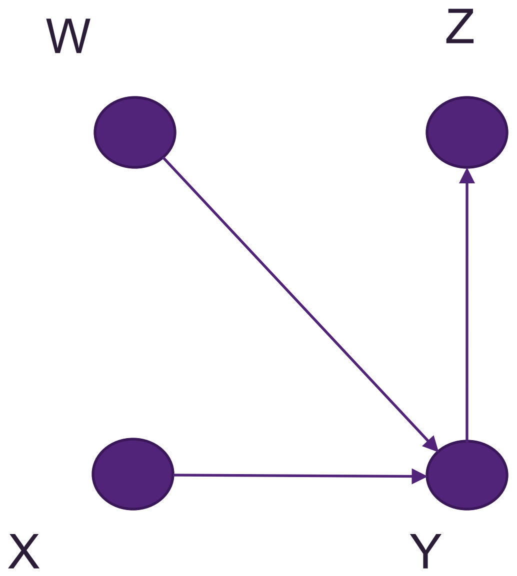
Is it a DAG?



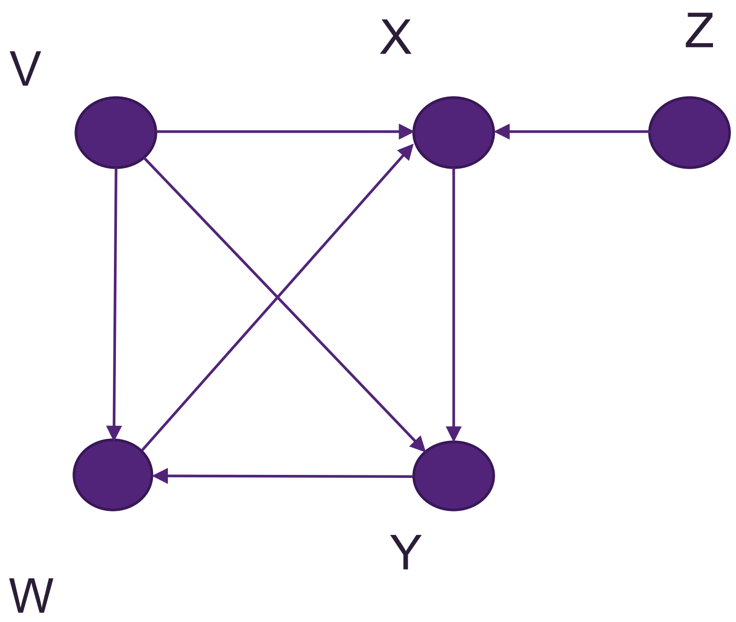
IS IT A DAG?



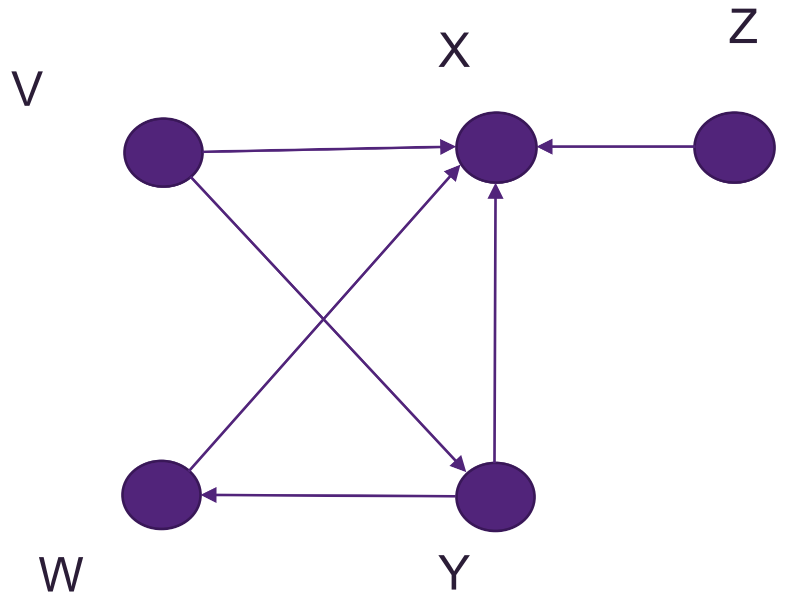
IS IT A DAG?



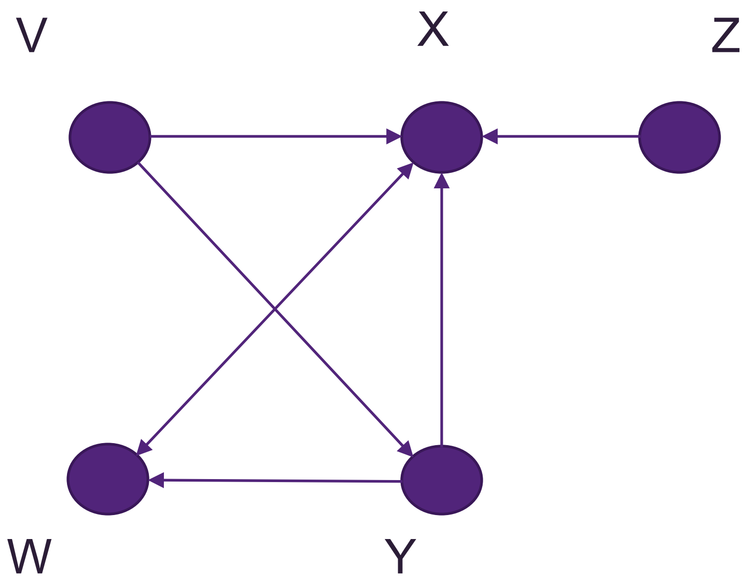
IS IT A DAG?



IS IT A DAG?

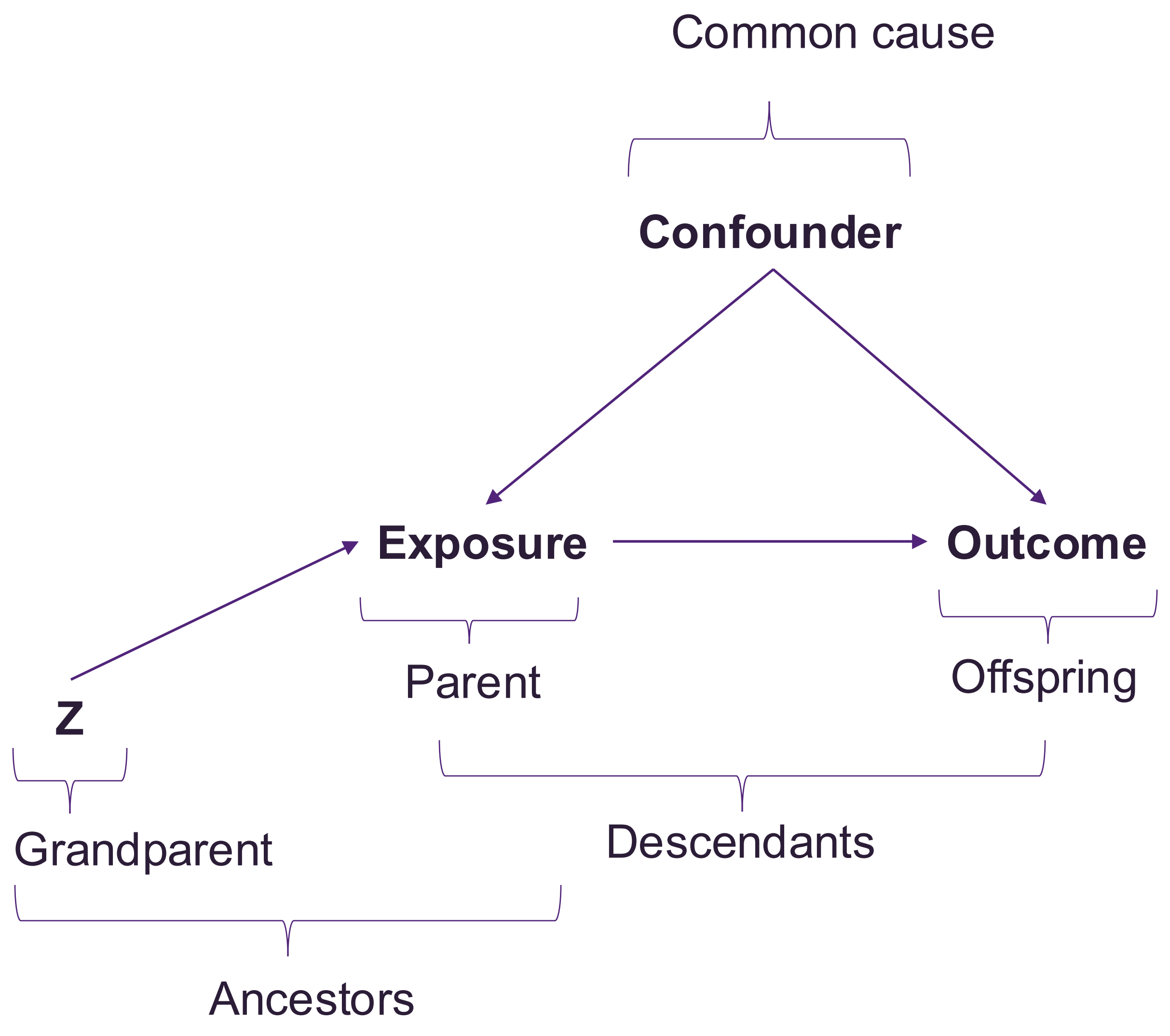


IS IT A DAG?



Glossary

- **Parent:** a direct cause of a particular variable.
- **Ancestor:** a direct cause or indirect cause of a particular variable.
- **Child:** The direct effect of a particular variable.
- **Descendant:** a direct effect or indirect effect of a particular variable.
- **Common cause:** A variable that is an ancestor of two other variables.



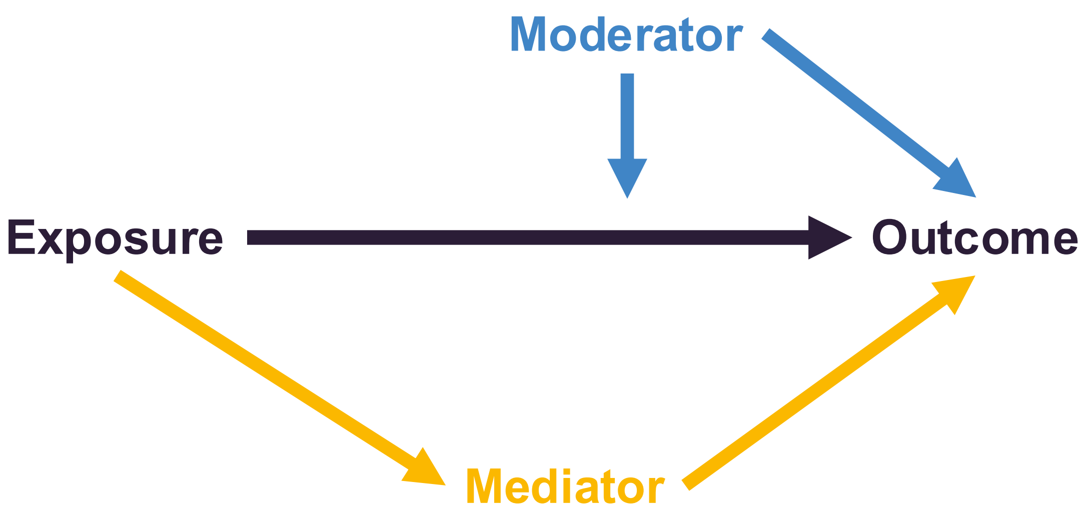
How to construct a DAG

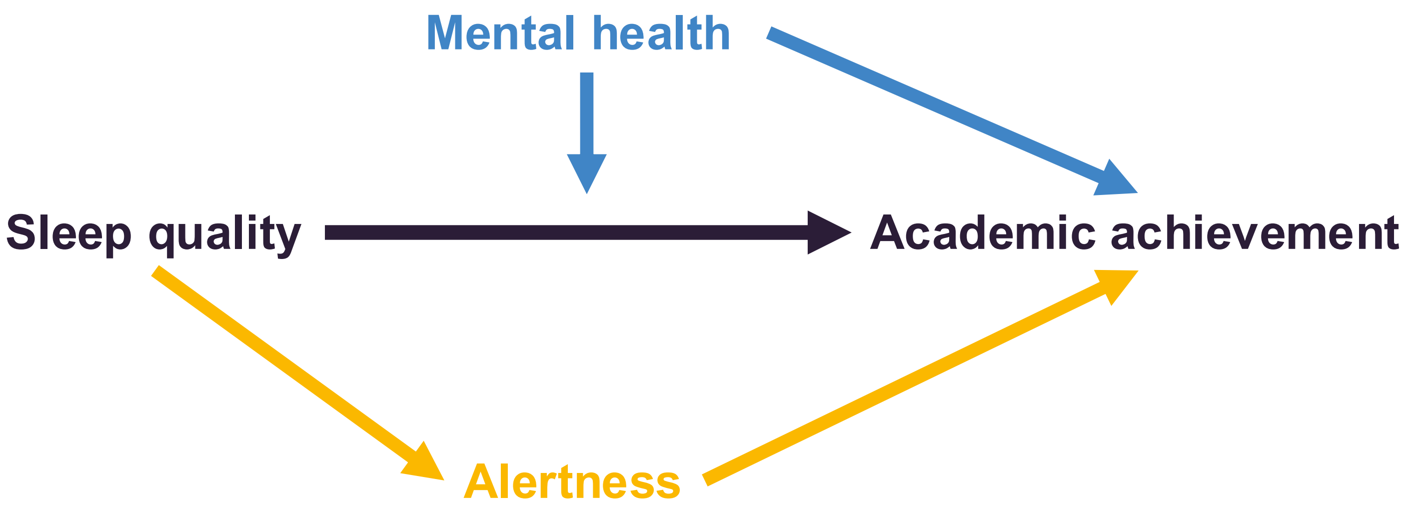
- Start with the exposure/treatment and the outcome/endpoint.

Exposure $\longrightarrow$ **Outcome**

How to construct a DAG

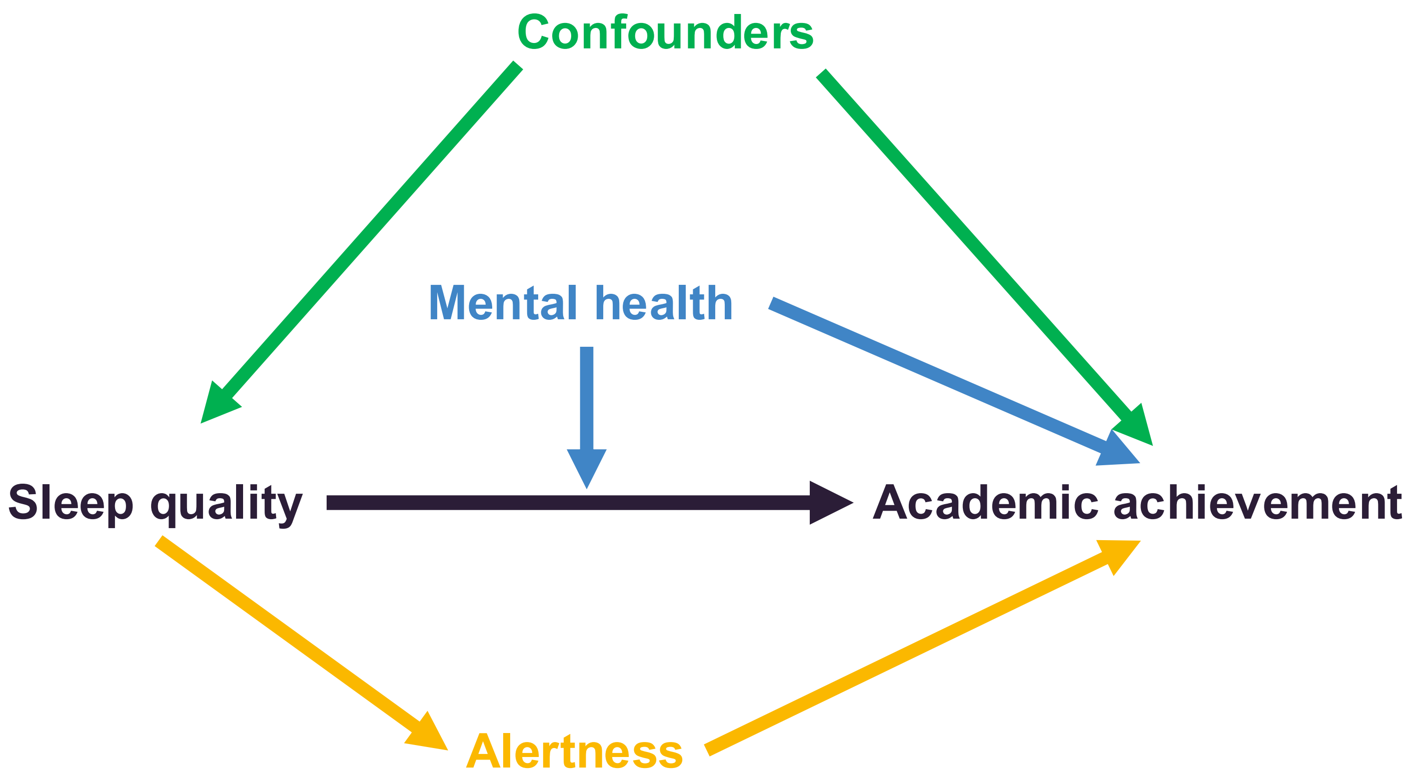
- Start with the exposure/treatment and the outcome/endpoint.
- Consider variables embedded in the question (e.g. mediators/moderators).

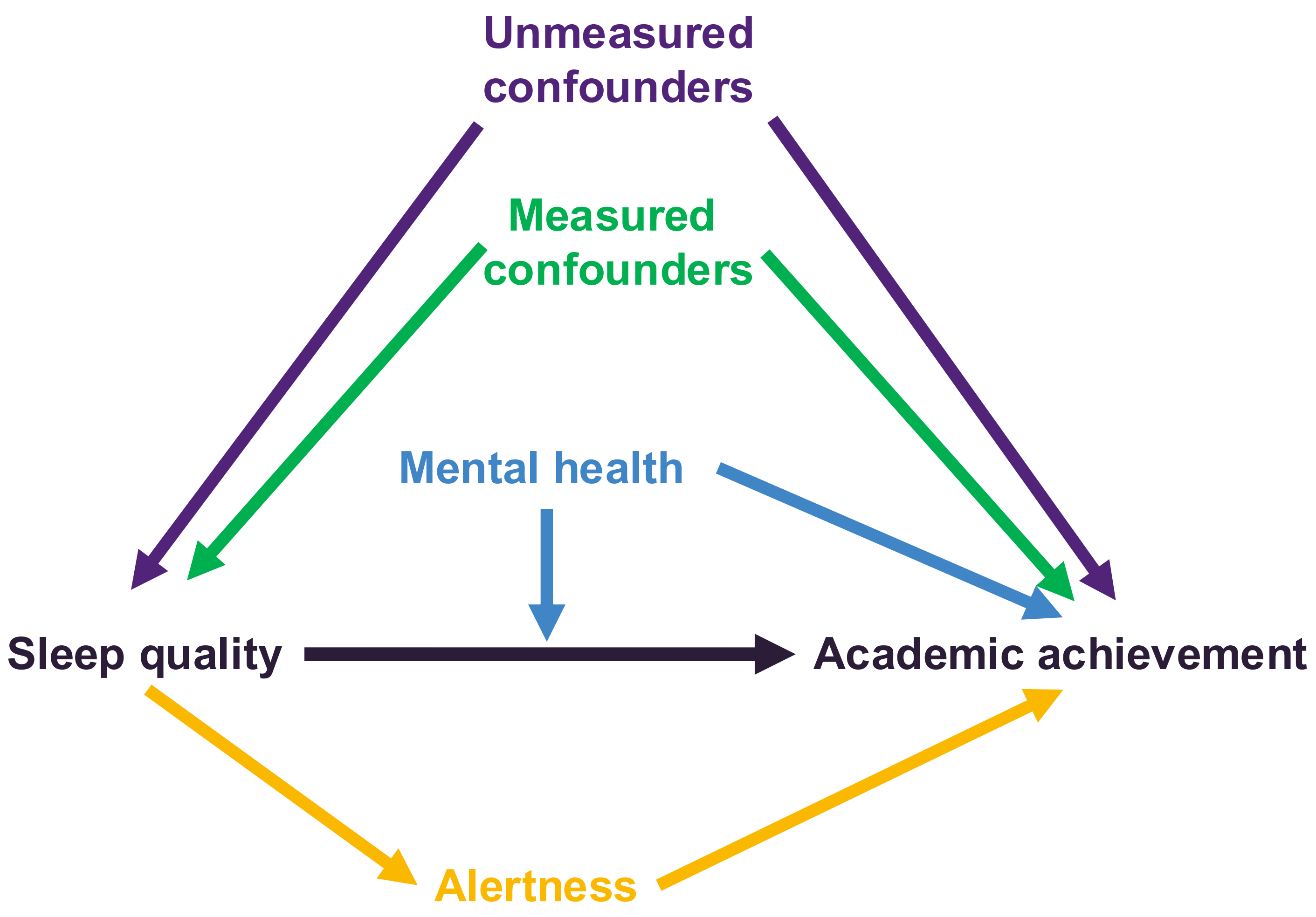

Moderator
Exposure
Outcome
Mediator


Mental health
Sleep quality
Academic achievement
Alertness

How to construct a DAG

- Start with the exposure/treatment and the outcome/endpoint.
- Consider variables embedded in the question (e.g. mediators/moderators).
- Consider confounding variables and add to the DAG.


Confounders
Mental health
Sleep quality
Academic achievement
Alertness


Unmeasured confounders
Measured confounders
Mental health
Sleep quality
Academic achievement
Alertness

How to construct a DAG

Must be included	Not required
All common causes of any 2 variables (confounders)	Variables that cause Y but not A (moderators)
Unmeasured and unmeasurable common causes (use U notation)	
Selection variables (i.e. inclusion criteria)	

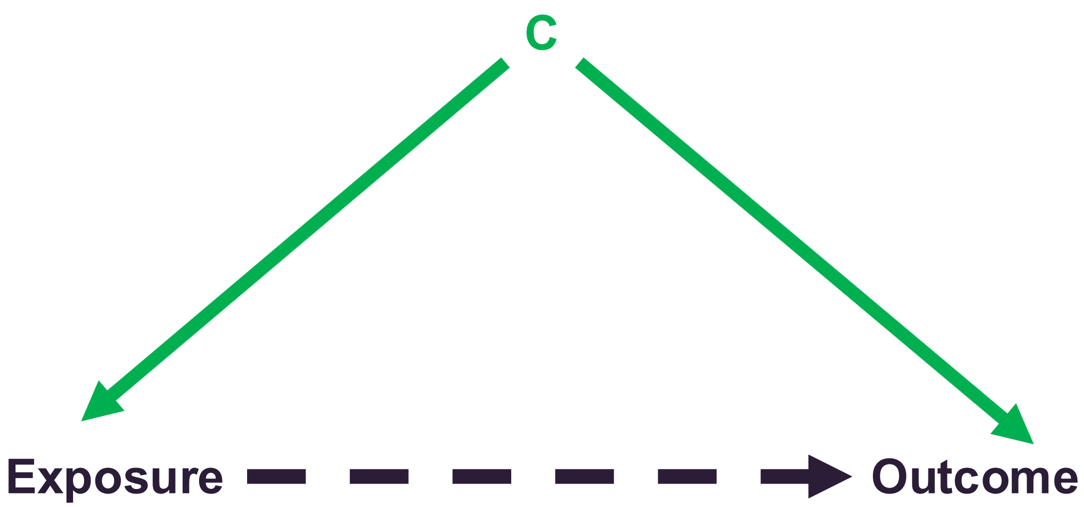
Remember:

- Assumptions must be made.
- There are often more than 1 appropriate DAG
- Alternative DAGs can make excellent sensitivity analyses.

Glossary

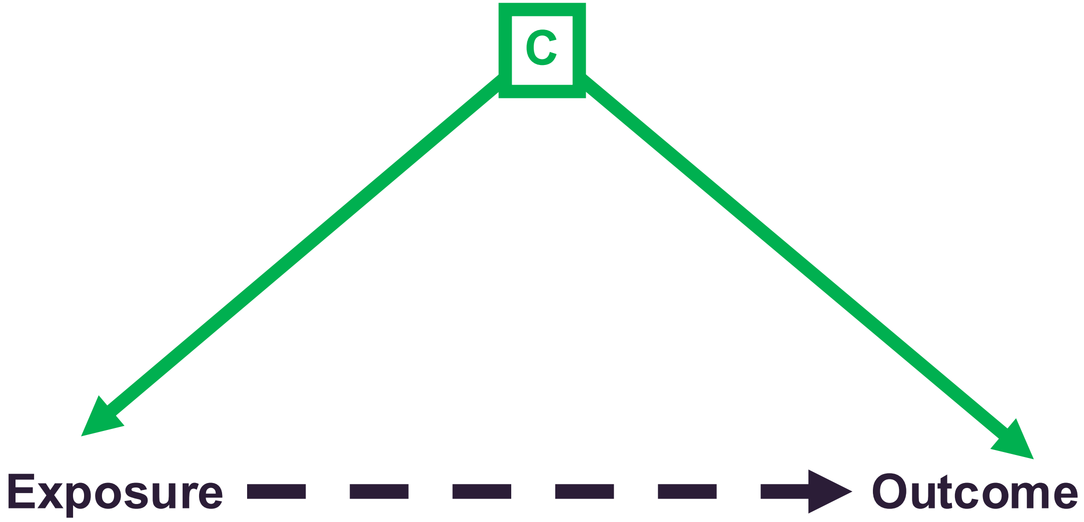
- **Back door path:** A connection between X and Y that does not follow the path of the arrows.
- **Collider:** A variable that is a descendant of two other variable. The term **collider** is used because the arrows "collide" at the descendant node.
- **Conditioning:** Conditioning on a variable means using either sample restriction, stratification, adjustment to examine the association of X and Y.

Back door path



- **Back door path:** A connection between X and Y that does not follow the path of the arrows.

Back door path



- **Back door path:** A connection between X and Y that does not follow the path of the arrows.

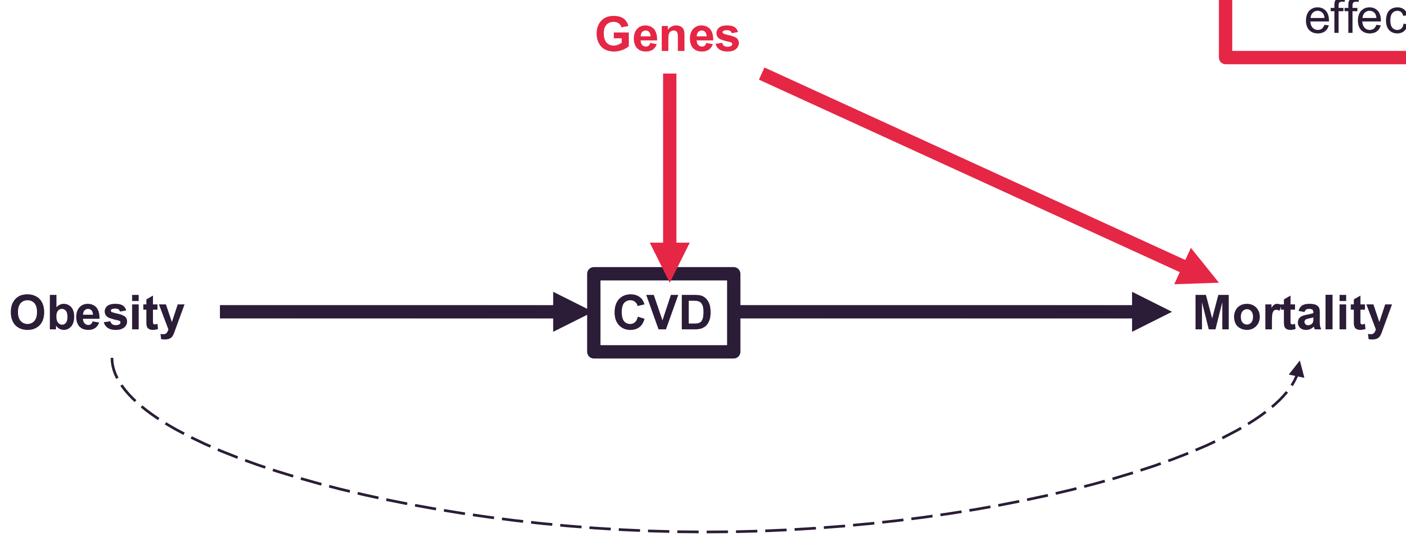
Collider

- **Collider:** A descendant of two other variables (where two arrows collide).
- **Collider Bias:** A phenomenon involving conditioning on common effects.

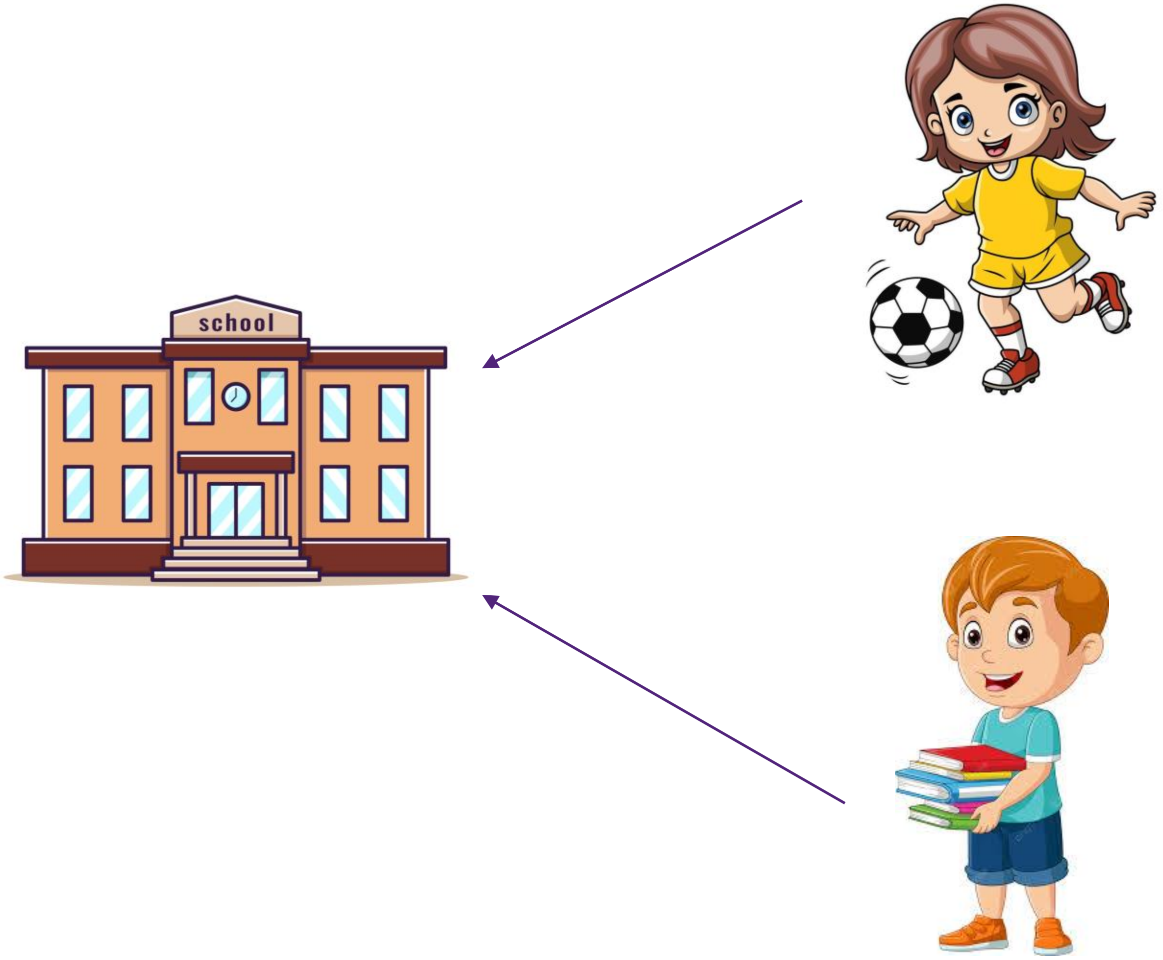
Obesity $\longrightarrow$ **Mortality**

Collider

- **Collider:** A descendant of two other variables (where two arrows collide).
- **Collider Bias:** A phenomenon involving conditioning on common effects.



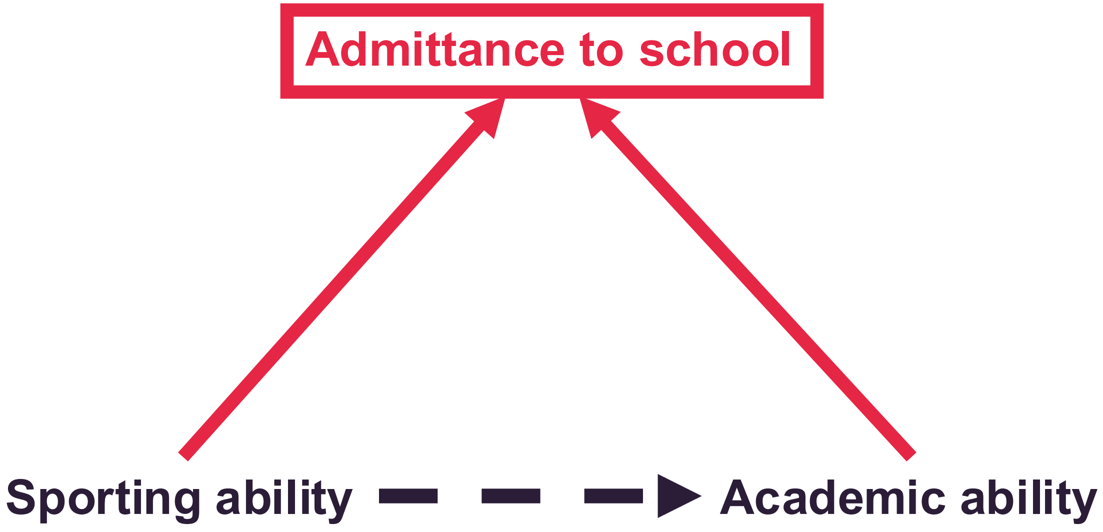
Collider



Collider

=

Collider



Sporting ability and admittance to the school are dependent

Academic ability and admittance to the school are dependent

Sporting ability and academic ability are independent

BUT

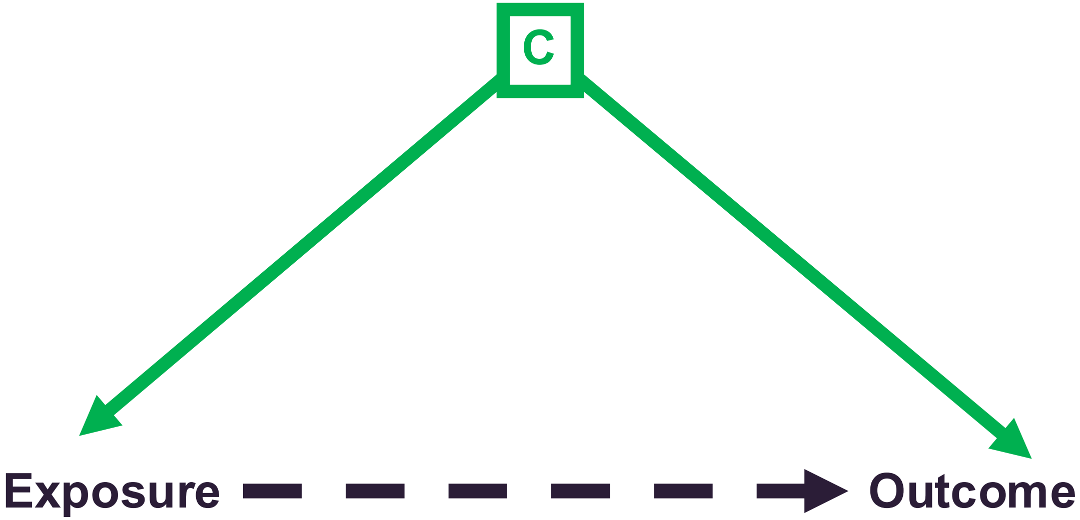
Sporting ability and academic ability are dependen. Conditional on the school!

Conditioning

Draw a box around the conditioned variables.

1. Conditioning on a variable in an open backdoor path removes the non-causal association (controls for confounding).
2. Conditioning on a collider opens the path that the collider was blocking.
3. Conditioning on a variable in the causal pathway (mediator) removes part of the causal effect.

Conditioning

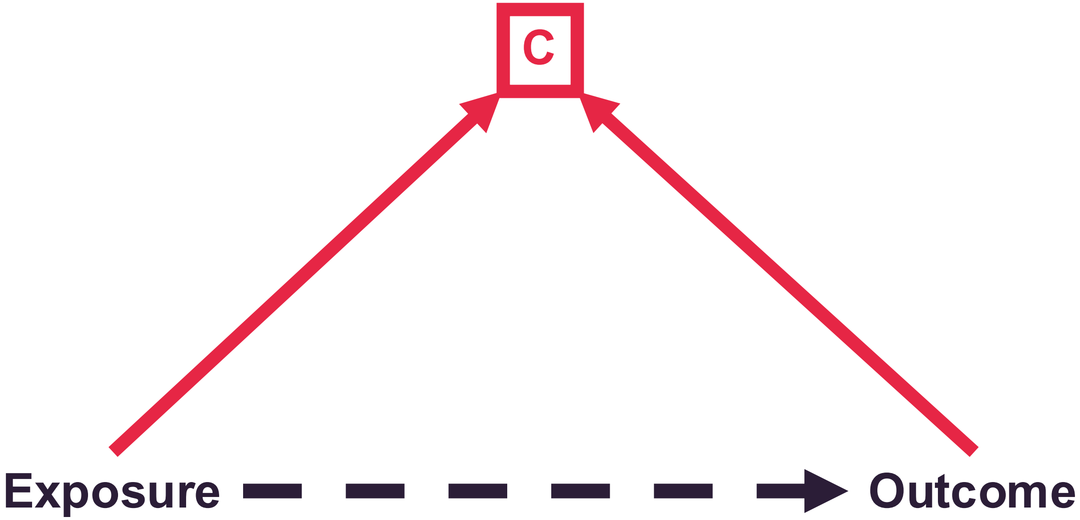


Conditioning

Draw a box around the conditioned variables.

1. Conditioning on a variable in an open backdoor path removes the non-causal association (controls for confounding).
2. Conditioning on a collider opens the path that the collider was blocking.
3. Conditioning on a variable in the causal pathway (mediator) removes part of the causal effect.

Conditioning

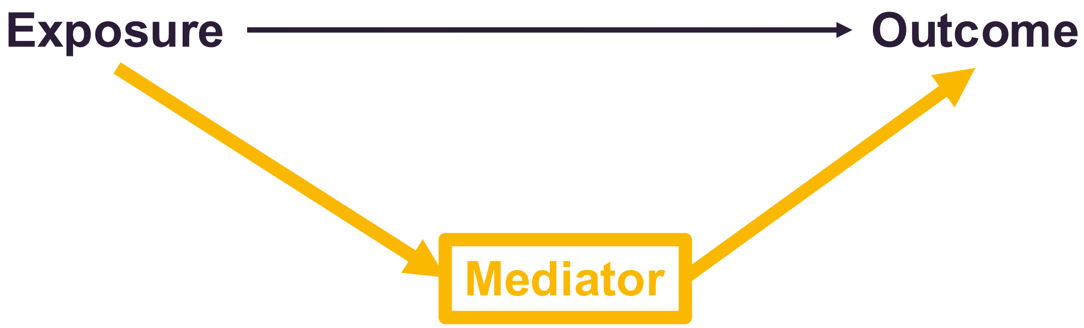


Conditioning

Draw a box around the conditioned variables.

1. Conditioning on a variable in an open backdoor path removes the non-causal association (controls for confounding).
2. Conditioning on a collider opens the path that the collider was blocking.
3. Conditioning on a variable in the causal pathway (mediator) removes part of the causal effect.

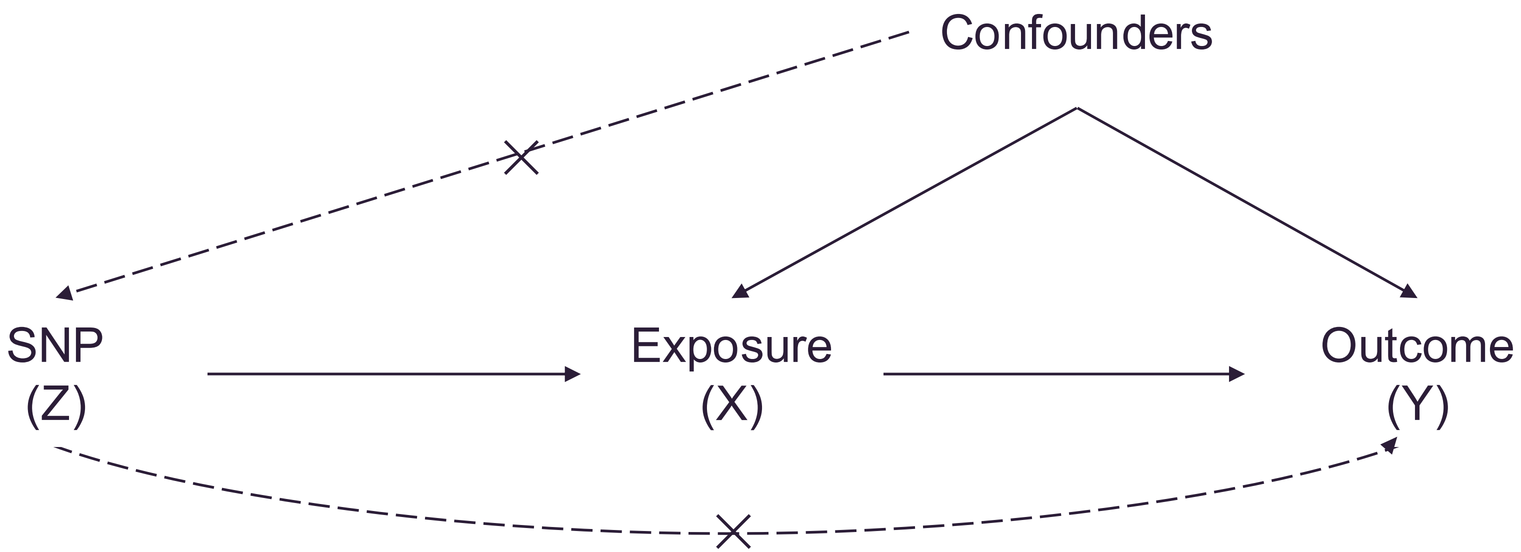
Conditioning



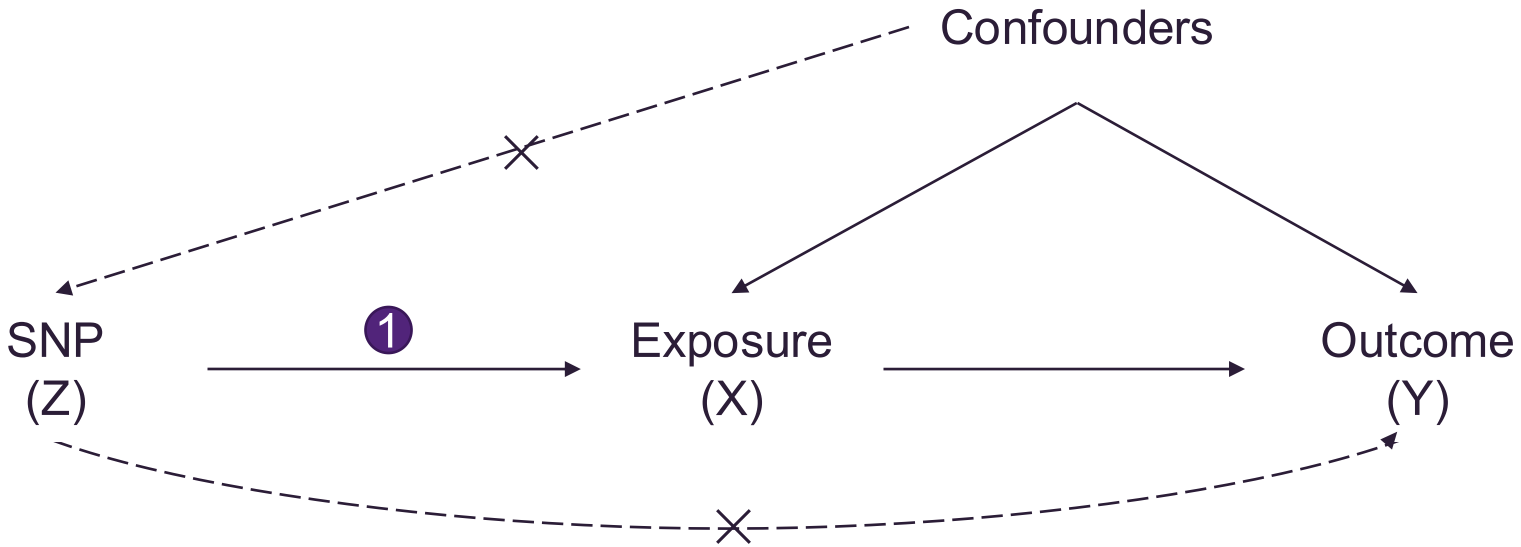
DAG elements

Element	Description
	Boxed elements indicate that the variable is conditioned on.
	An arrow with a solid line indicates direct association between two variables.
	An arrow with a dashed line indicates indirect association between two variables
C	Confounders.
U	Unmeasured confounders

Assumptions underlying MR

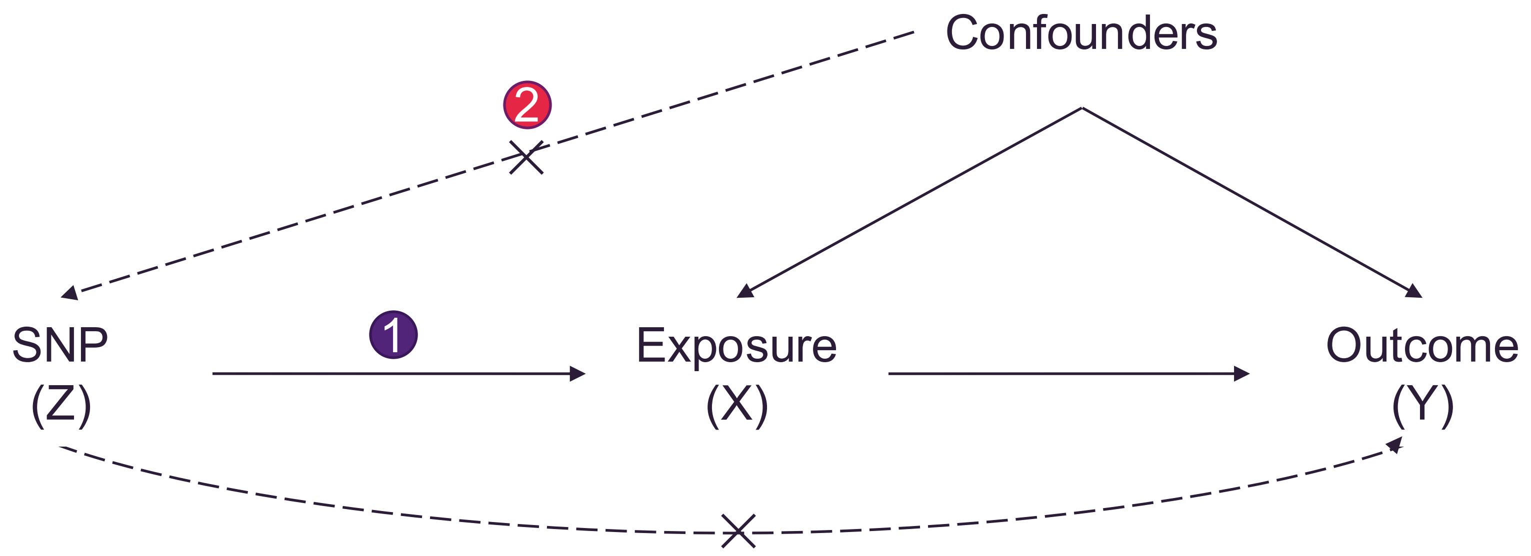


Assumptions underlying MR



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

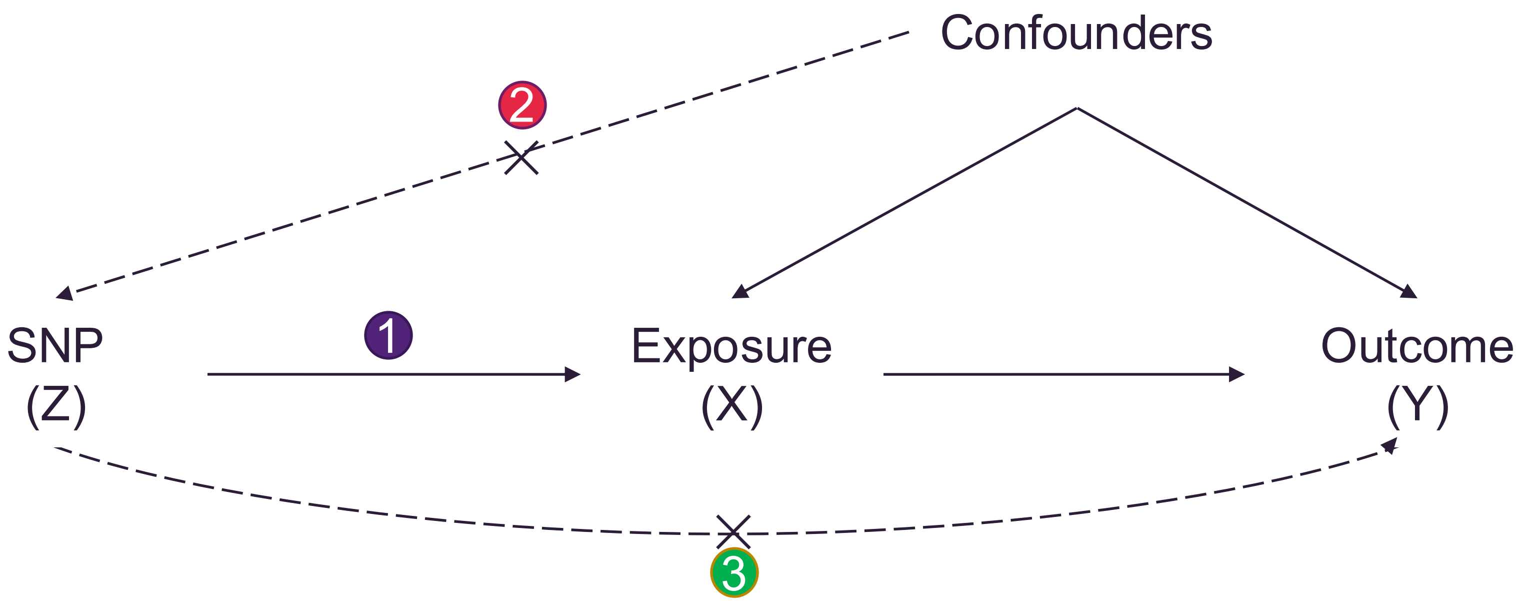
Assumptions underlying MR



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

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

Assumptions underlying MR

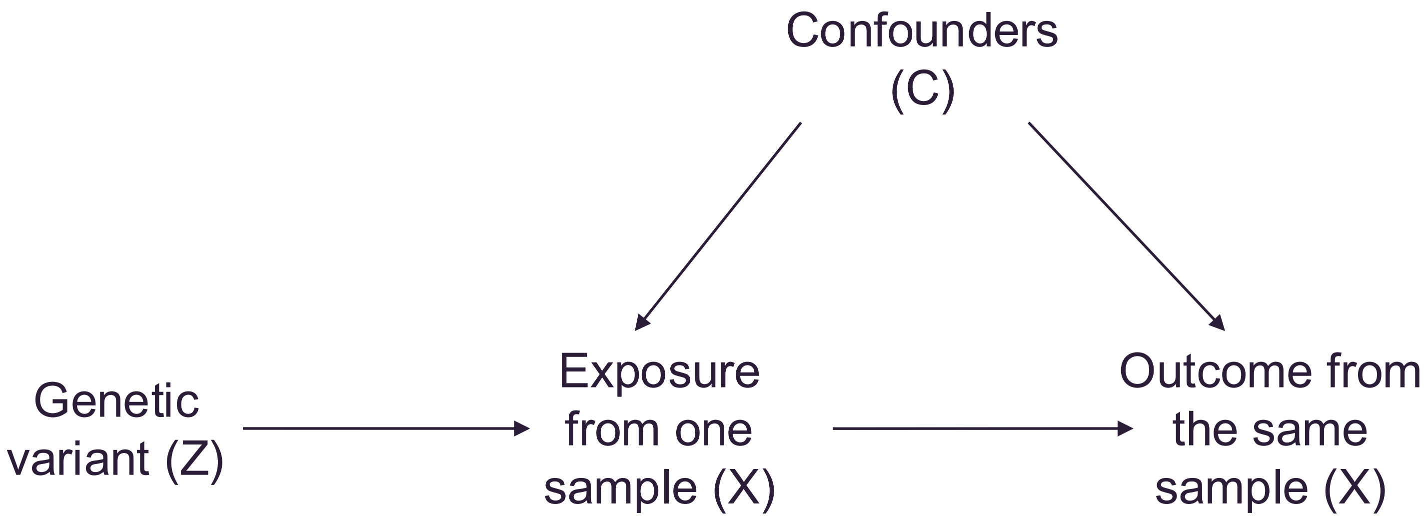


(1) Relevance assumption: SNP is associated with the exposure

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

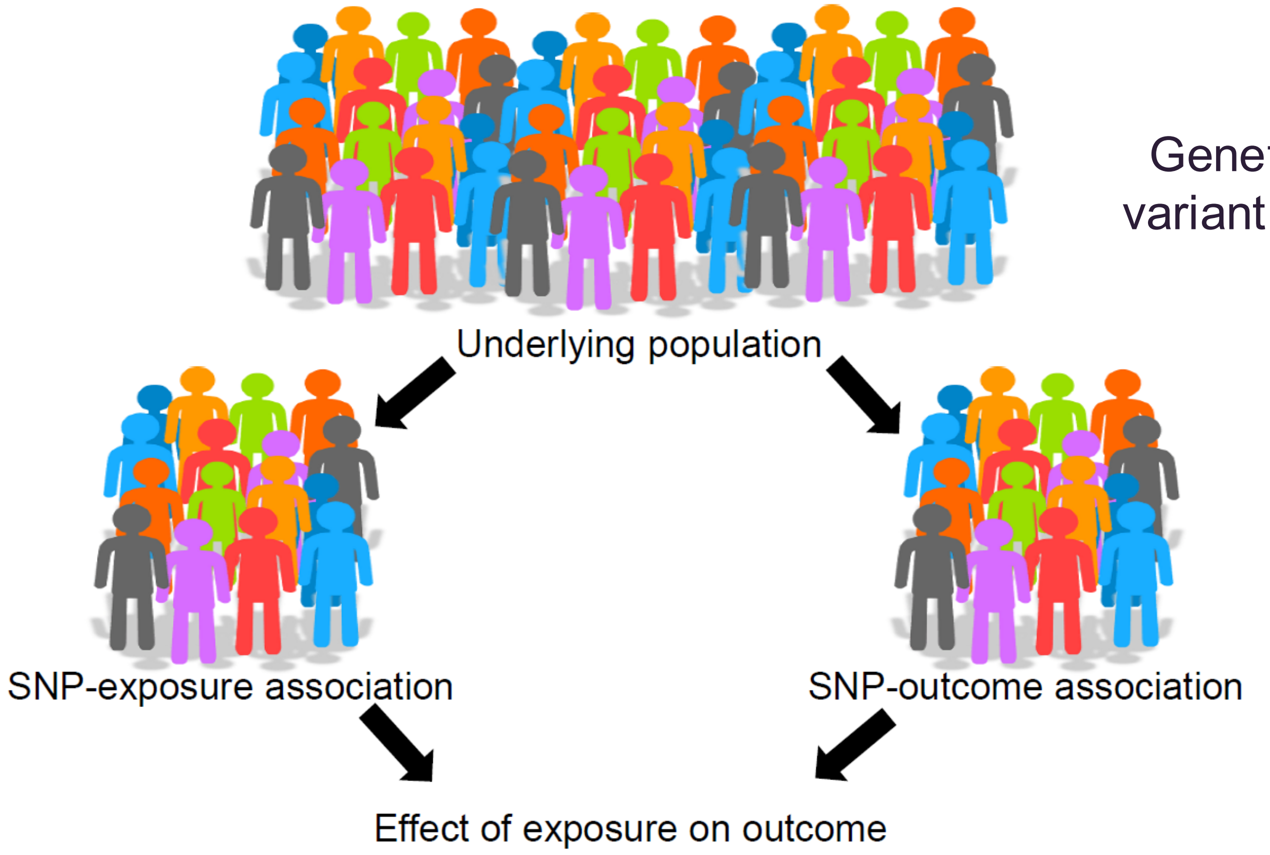
(3) Exclusion restriction: SNP ONLY associated outcome through the exposure

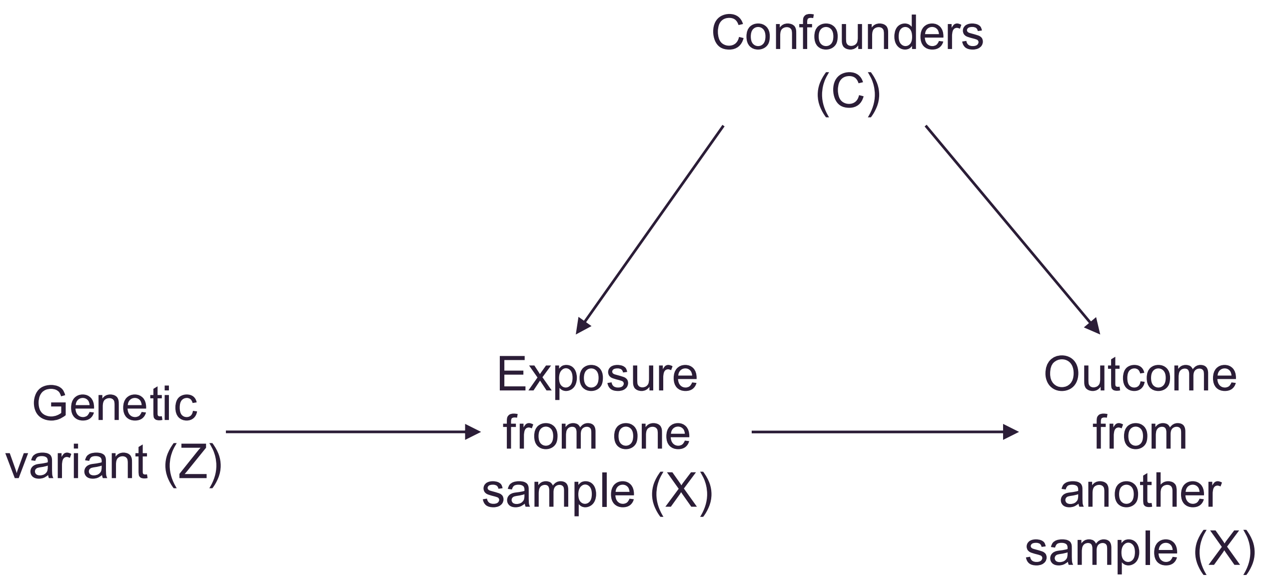
One-Sample MR



Genotypes, exposure and outcome are available on individuals from the same sample.

Two sample MR





SNP-exposure and SNP-outcome association estimates from **two independent samples** from the same underlying population.

Generate causal estimate

1. **The association of the SNP and the outcome**

Test for existence of an effect

Generate causal estimate

1. **The association of the SNP and the outcome**
2. **Two-stage least squares**
3. **The Wald estimator**

Test for existence of an effect

Estimate the size of the effect

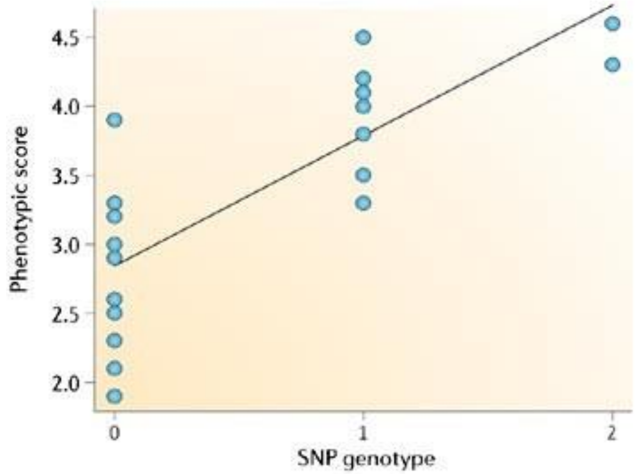
Calculating causal effect estimates Two-Stage Least Squares

A single sample of individuals with data on the SNP, the exposure and the outcome. Also known as “One sample MR”.

Manual calculation:

1. Regress exposure on SNP to get predicted values.
2. Regress outcome on **predicted** exposure (from 1st stage regression).

The regression coefficient from the second stage is the estimate of the causal effect of the exposure on the outcome.



Calculating causal effect estimates Two-Stage Least Squares

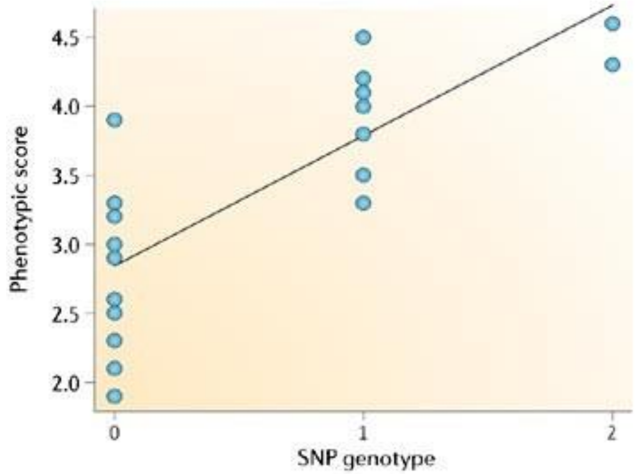
THE UNIVERSITY OF QUEENSLAND AUSTRALIA

A single sample of individuals with data on the SNP, the exposure and the outcome. Also known as “One sample MR”.

Manual calculation:

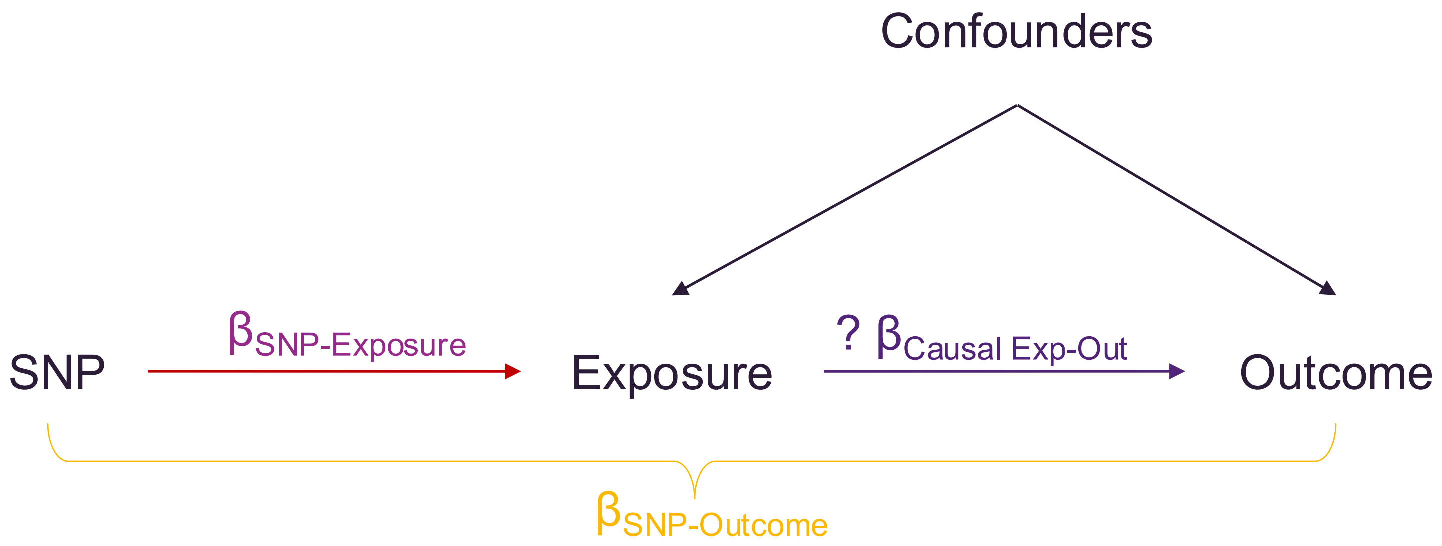
1. Regress exposure on SNP to get predicted values.
2. Regress outcome on **predicted** exposure (from 1st stage regression).

The regression coefficient from the second stage is the estimate of the causal effect of the exposure on the outcome.



This gives you: difference in outcome per unit change in (genetically-predicted) exposure

Calculating Causal Effect Estimates
Wald Estimator (Wald Ratio)



Where there is a linear relationship between SNP, exposure and outcome:

$\beta_{\text{SNP-Outcome}} = \beta_{\text{Causal Exp-Out}} \times \beta_{\text{SNP-Exposure}}$

$$\beta_{\text{Causal effect}} \text{ (Wald estimator)} = \frac{\beta_{\text{SNP-Outcome}}}{\beta_{\text{SNP-Exposure}}}$$

Calculating Causal Effect Estimates
Wald Estimator (Wald Ratio)

Wald estimator can be used in one sample ("One sample MR") as well as different samples ("Two sample MR")

Confounders

SNP $\xrightarrow[(0.5\text{ kg})]{\beta_{\text{SNP-Weight}}}$ Weight $\xrightarrow{?\ \beta_{\text{Causal Weight-BP}}}$ Blood pressure

$\underbrace{\text{SNP} \quad\quad \text{Weight} \quad\quad \text{Blood pressure}}_{\beta_{\text{SNP-Blood Pressure}}\ (0.9\text{ mmHg})}$

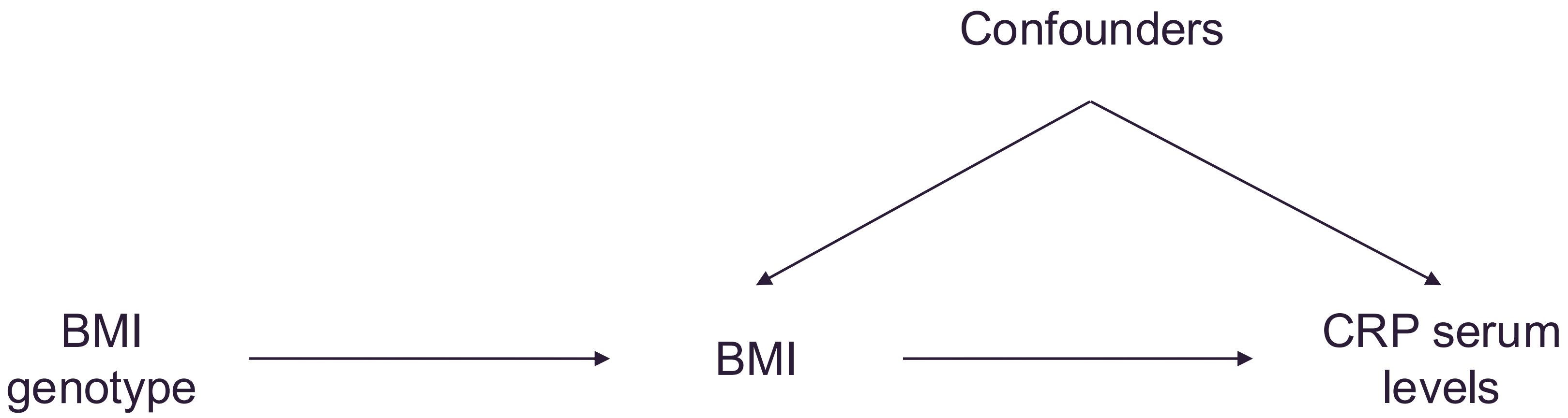
Where there is a linear relationship between SNP, exposure and outcome:

$$\beta_{\text{SNP-Outcome}} = \beta_{\text{Causal Exp-Out}} \times \beta_{\text{SNP-Exposure}}$$

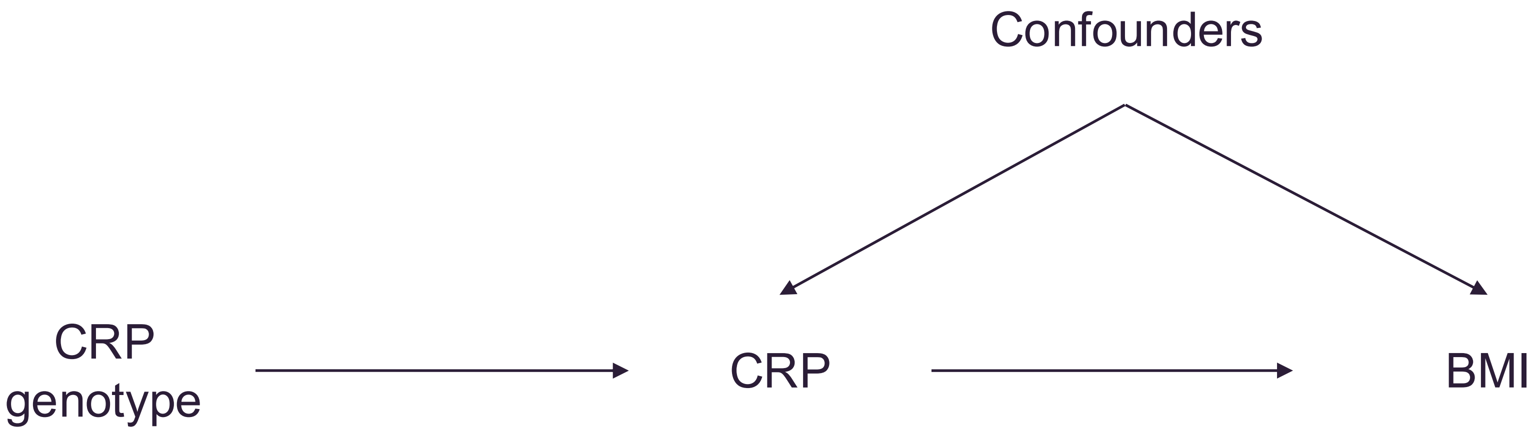
$$\beta_{\text{Causal effect}} \text{ (Wald estimator)} = \frac{\beta_{\text{SNP-Outcome}}}{\beta_{\text{SNP-Exposure}}}$$

$$\beta_{\text{Causal effect Weight-BP}} = \frac{0.9\text{ mmHg/allele}}{0.5\text{ kg/allele}} = 1.8\text{ mmHg/kg}$$

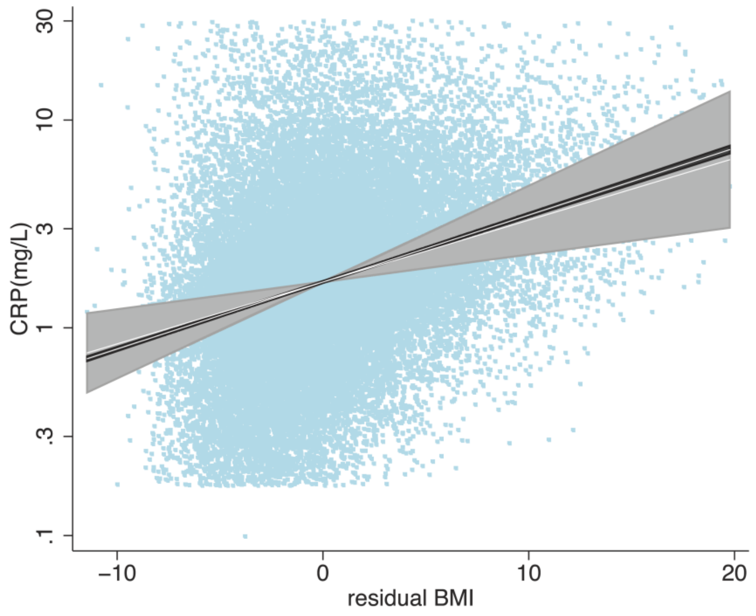
MR example: THE GOOD



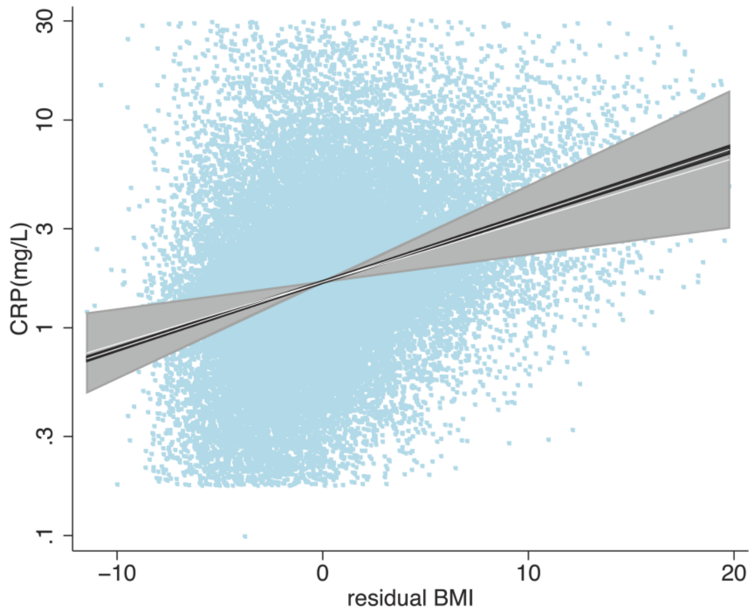
MR example: THE GOOD

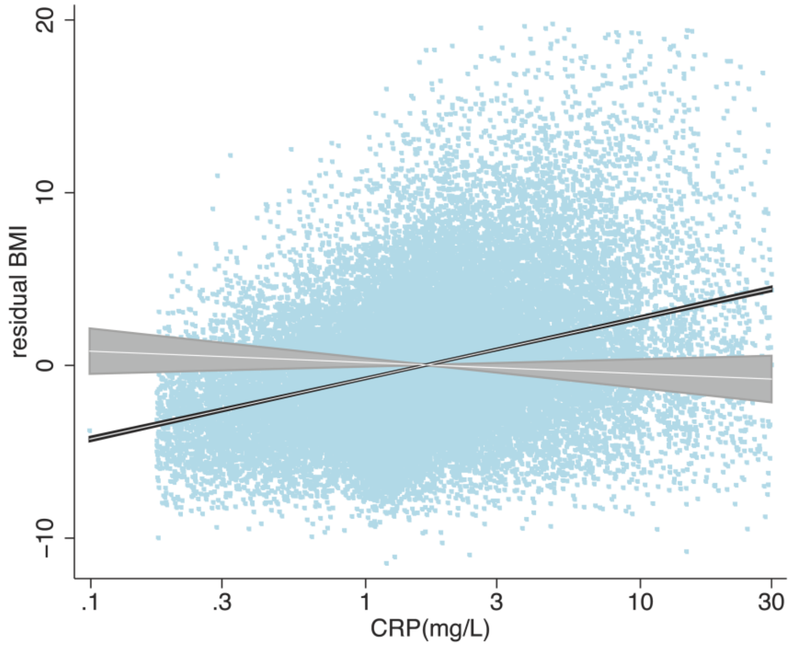


	Effect estimates				
Exposure ➜ Outcome	Observational association	Instrumental variable (MR)	P_{IV}	P_{diff}	F_{first}
BMI ➜ CRP	1.075 (1.073, 1.077)	1.06 (1.02, 1.11)	0.002	0.6	50.2



	Effect estimates				
Exposure → Outcome	Observational association	Instrumental variable (MR)	P_{IV}	P_{diff}	F_{first}
BMI → CRP	1.075 (1.073, 1.077)	1.06 (1.02, 1.11)	0.002	0.6	50.2
CRP → BMI	1.58 (1.53, 1.63)	-0.30 (-0.78, 0.18)	0.2	<0.00001	78.3





MR Example: THE BAD

MR Example: THE BAD

Nutrients. 2023 May; 15(9): 2091. PMCID: PMC10181479
Published online 2023 Apr 26. doi: 10.3390/nu15092091 PMID: 37432232

A Positive Causal Relationship between Noodle Intake and Metabolic Syndrome: A Two-Sample Mendelian Randomization Study

MR Example: THE BAD

Nutrients. 2023 May; 15(9): 2091.
Published online 2023 Apr 26. doi: 10.3390/nu15092091

PMCID: PMC10181479
PMID: 37432232

A Positive Causal Relationship between Noodle Intake and Metabolic Syndrome: A Two-Sample Men

> J Med Virol. 2023 Apr;95(4):e28722. doi: 10.1002/jmv.28722.

Causal effects of COVID-19 on cancer risk: A Mendelian randomization study

Jia Li [1], Haocheng Bai [1], Hao Qiao [2], Chong Du [1], Peizhuo Yao [1], Yu Zhang [1], Yifan Cai [1], Yiwei Jia [1], Xinyu Wei [1], Chaofan Li [1], Xuanyu Liu [1], Weiwei Wang [1], Shiyu Sun [1], Cong Feng [1], Yijian Hu [1], Zhangjian Zhou [1], Shuqun Zhang [1], Yinbin Zhang [1]

MR Example: THE BAD

Nutrients. 2023 May; 15(9): 2091. PMCID: PMC10181479
Published online 2023 Apr 26. doi: 10.3390/nu15092091 PMID: 37432232

A Positive Causal Relationship between Noodle Intake and Metabolic Syndrome: A Two-Sample Men[...]

J Med Virol. 2023 Apr;95(4):e28722. doi: 10.1002/jmv.28722.

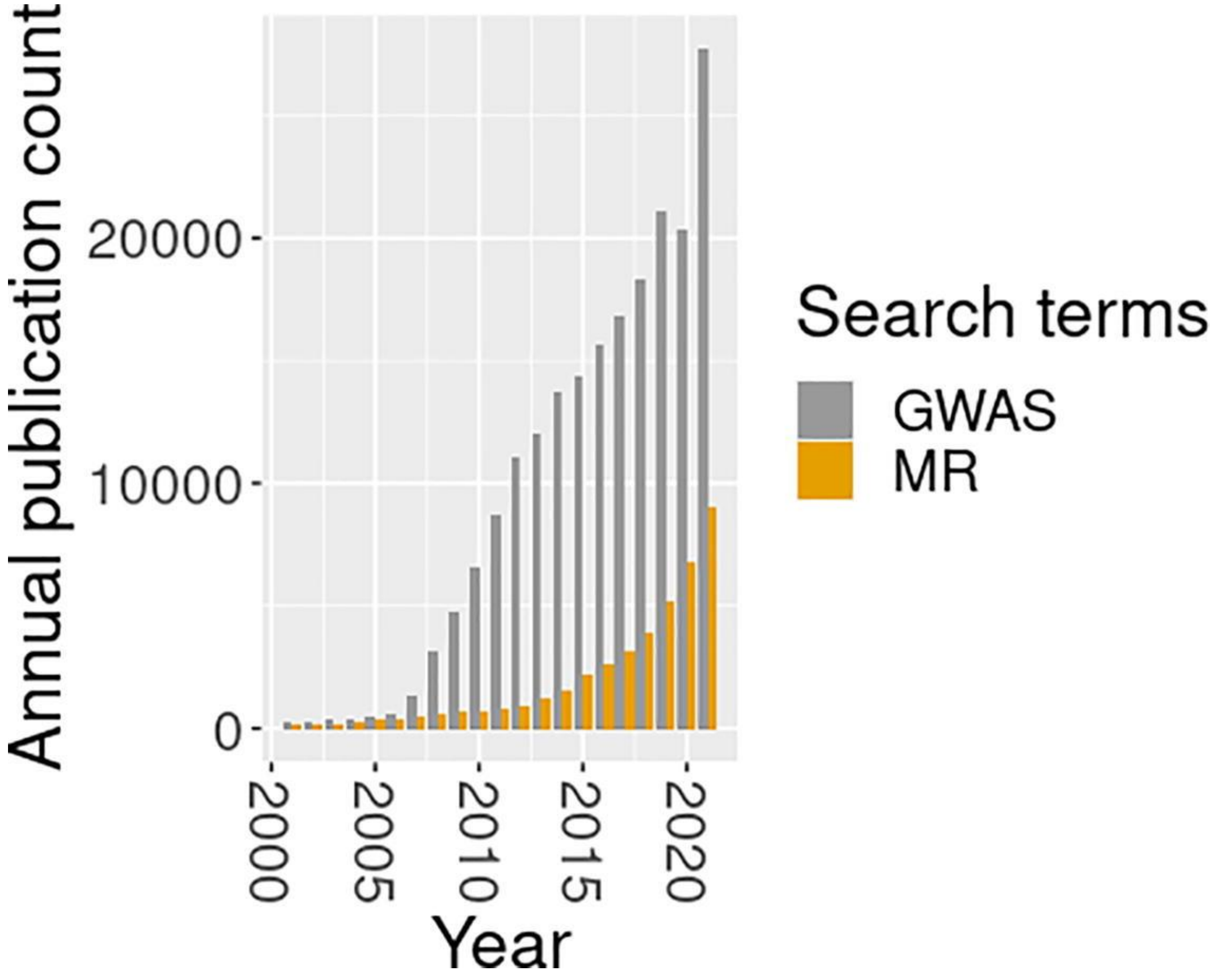
Causal effects of COVID-19 on cancer risk: A Mendelian randomization study

Jia Li [1], Haocheng Bai [1], Hao Qiao [2], Chong Du [1], Peizhuo Yao [1], Yu Zhang [1], Yifan Cai [1], Yiwei Jia [1], Xinyu Wei [1], Chaofan Li [1], Xuanyu Liu [1], Weiwei Wang [1], Shiyu Sun [1], Cong Feng [1], Yijian Hu [1], Zhangjian Zhou [1], Shuqun Zhang [1], Yinbin Zhang [1]

Ann Transl Med. 2021 Feb; 9(3): 263. PMCID: PMC7940946
doi: 10.21037/atm-20-3063 PMID: 33708890

Habitual consumption of alcohol with meals and lung cancer: a Mendelian randomization study

MR Example: THE BAD



Limitations of MR

Reasons for failing to observe a SNP-outcome association despite a real causal association existing

Power and weak instrument bias

Power:

- Genetic variants explain very small amounts of phenotypic variance in a given trait.
- VERY large sample sizes are generally required.

Weak instruments:

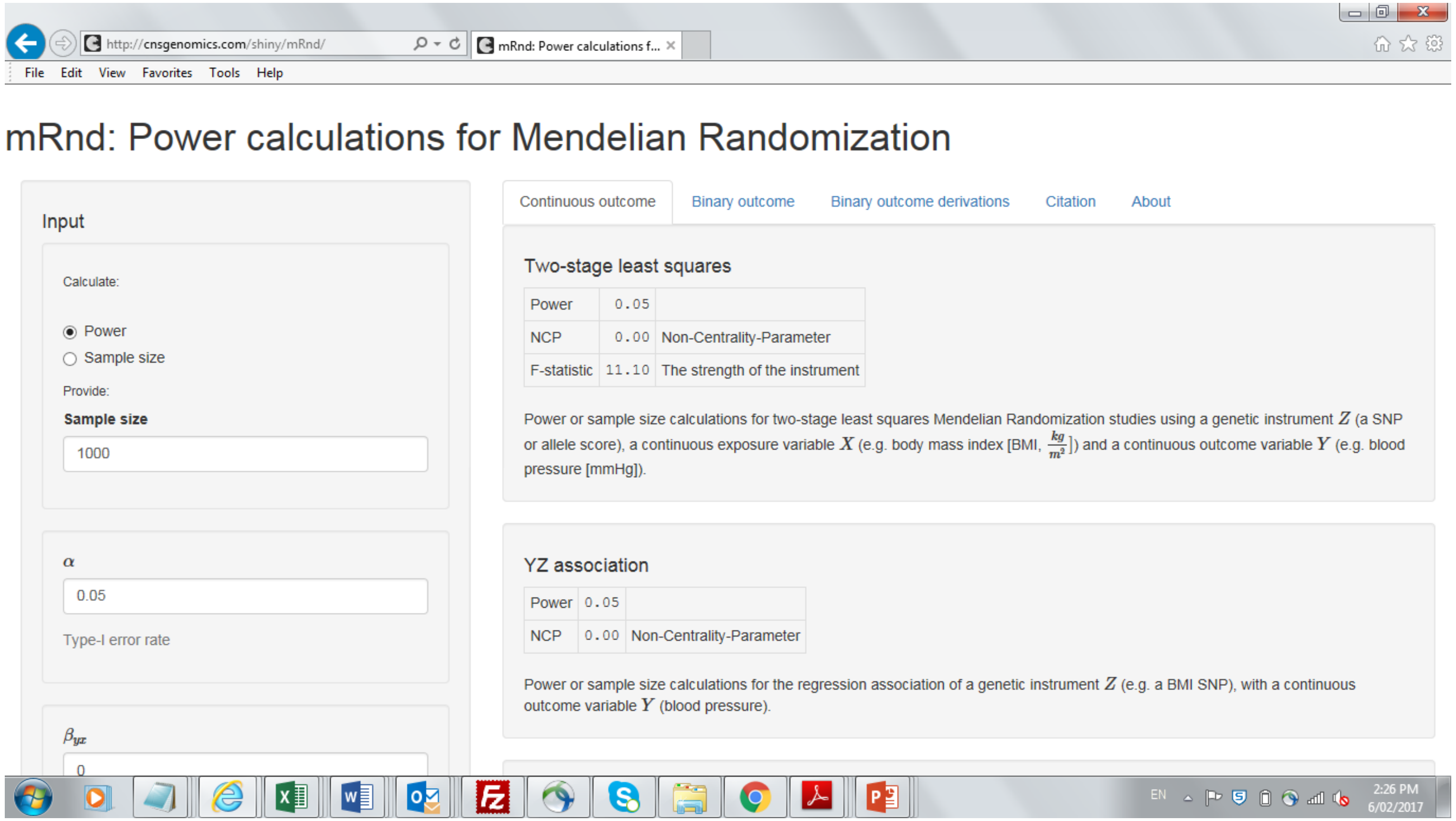
- Genetic variants that are weak proxies for the exposure.
- Results in biased causal estimates from MR.

Different impact of the bias from weak instruments:

- **One-Sample MR:** to the confounded estimate**.**
- **Two-Sample MR:** to the null.

Reasons for failing to observe a SNP-outcome association despite a real causal association existing

Power



Using Multiple Genetic Variants as Instruments

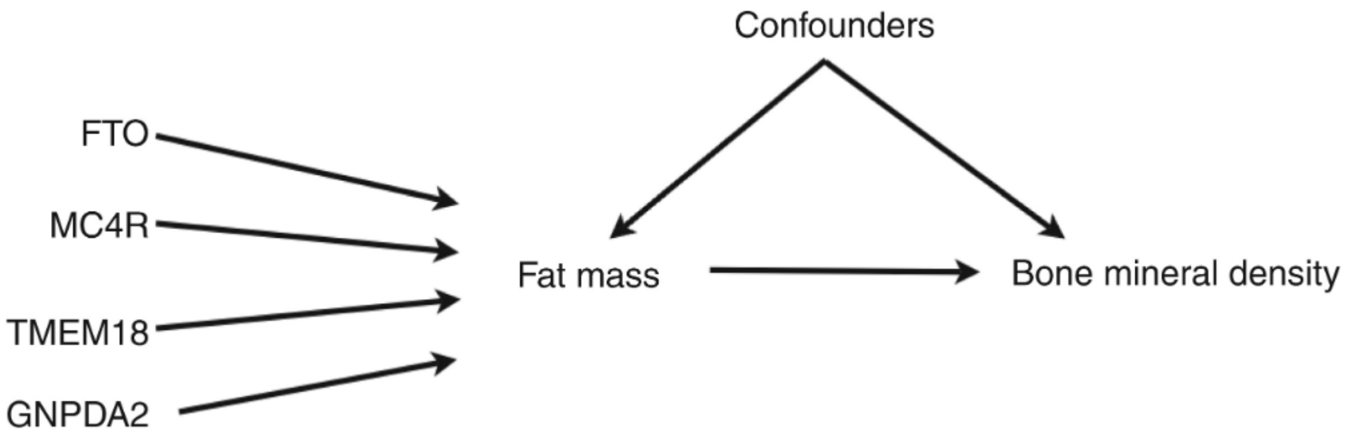


Figure 1. DAG for a Mendelian randomisation analysis using four genetic variants as instrumental variables for the effect of fat mass on bone mineral density.

Creating allelic scores using multiple genetic variants.

Testing multiple variants individually and then meta-analysing individual SNPs.

Reasons for detecting a causal SNP-outcome when it does not exist

Population Stratification:

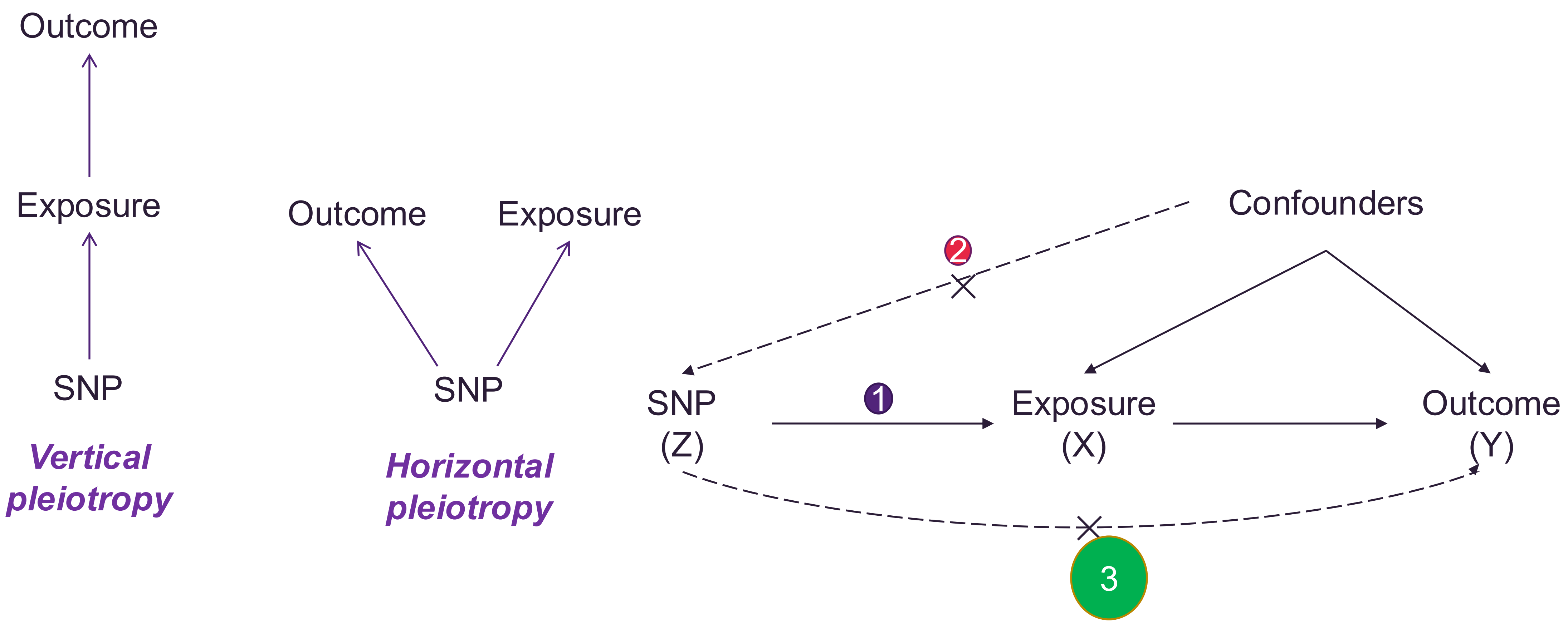
- Creates genetic confounding.
- Assumption 2 is violated.

Overlapping discovery GWAS and MR estimation samples.

Pleiotropy

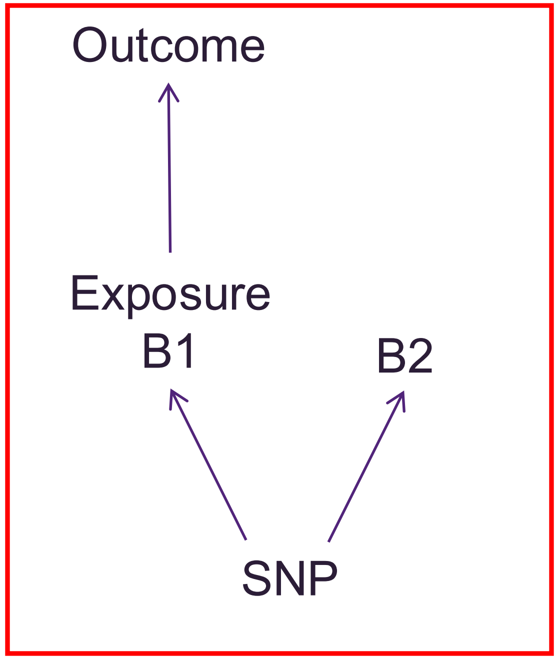
- Multiple phenotypic effects.
- Assumption 3 is violated.

Pleiotropy: Genetic variant influences more than one trait

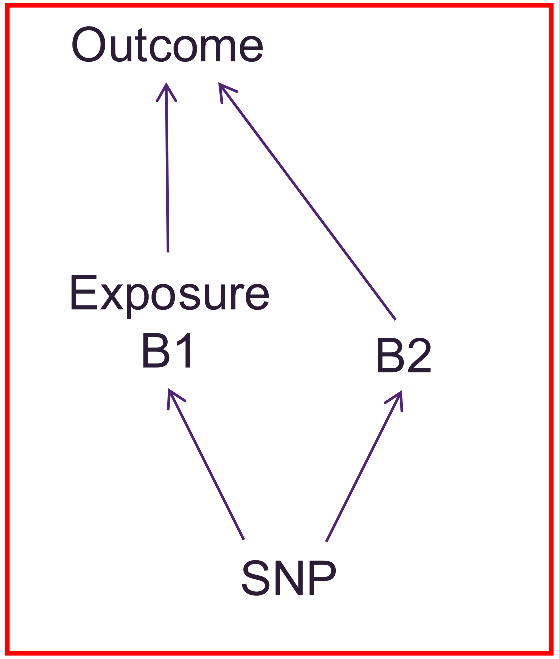


Horizontal Pleiotropy

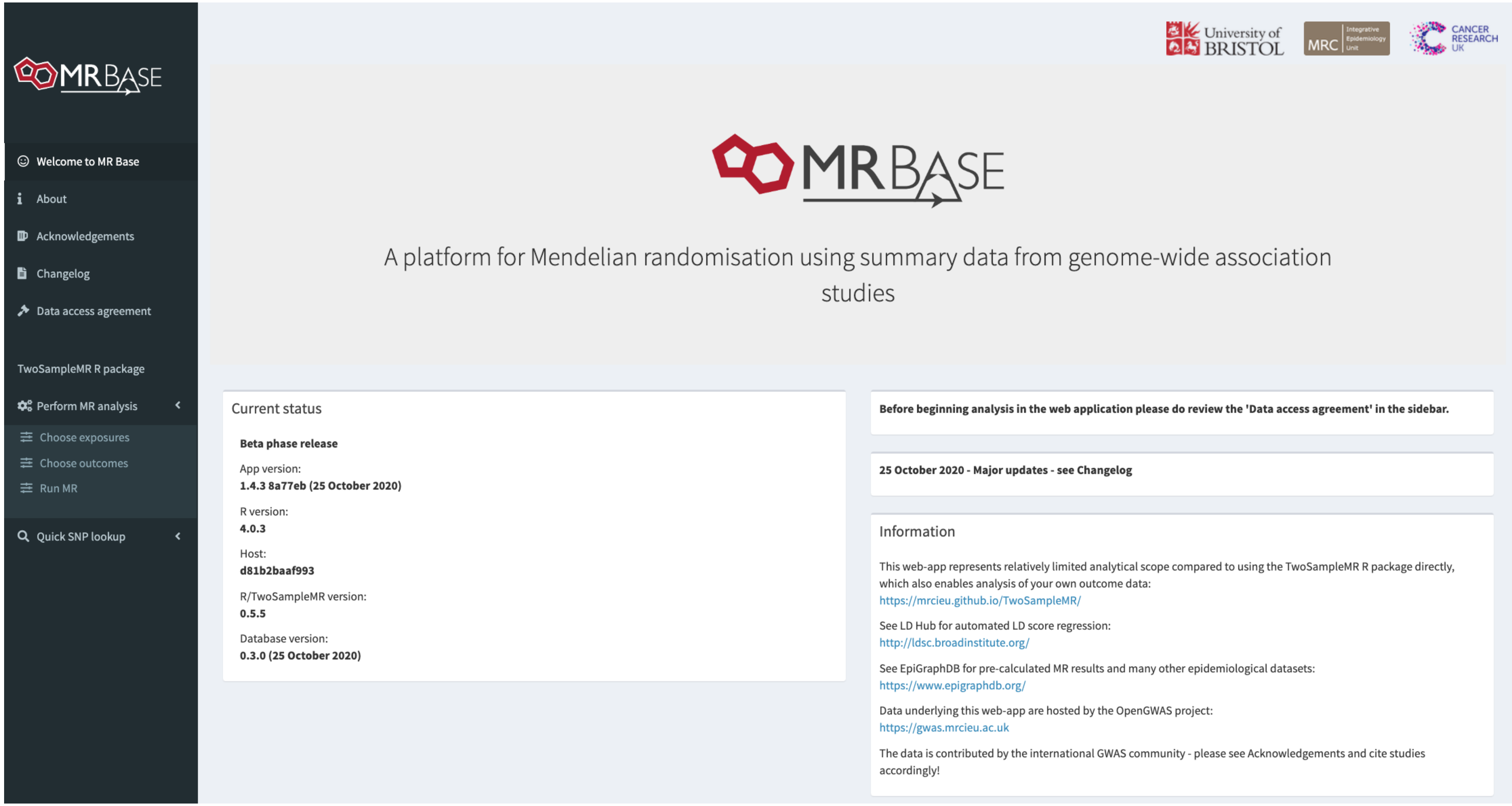
Pleiotropy only violates MR's assumptions if it involves a pathway outside that of the exposure and is a pathway that <u>affects your outcome.</u>



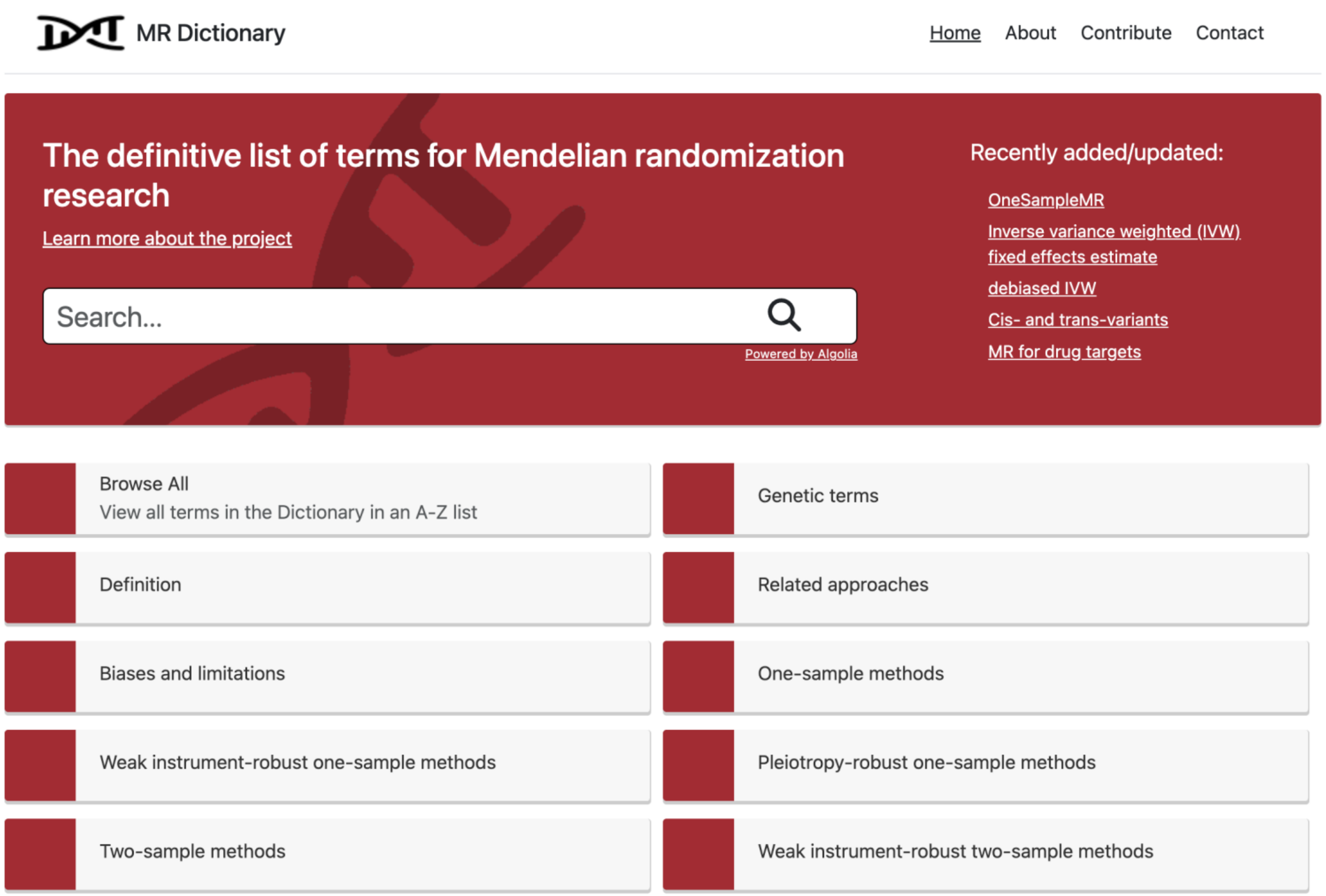
Violation



MR Base


MR BASE
Welcome to MR Base
About
Acknowledgements
Changelog
Data access agreement
TwoSampleMR R package
Perform MR analysis
Choose exposures
Choose outcomes
Run MR
Quick SNP lookup
University of BRISTOL
MRC Integrative Epidemiology Unit
CANCER RESEARCH UK
MR BASE
A platform for Mendelian randomisation using summary data from genome-wide association studies
Current status
Beta phase release
App version:
1.4.3 8a77eb (25 October 2020)
R version:
4.0.3
Host:
d81b2baaf993
R/TwoSampleMR version:
0.5.5
Database version:
0.3.0 (25 October 2020)
Before beginning analysis in the web application please do review the 'Data access agreement' in the sidebar.
25 October 2020 - Major updates - see Changelog
Information
This web-app represents relatively limited analytical scope compared to using the TwoSampleMR R package directly, which also enables analysis of your own outcome data:
https://mrcieu.github.io/TwoSampleMR/
See LD Hub for automated LD score regression:
http://ldsc.broadinstitute.org/
See EpiGraphDB for pre-calculated MR results and many other epidemiological datasets:
https://www.epigraphdb.org/
Data underlying this web-app are hosted by the OpenGWAS project:
https://gwas.mrcieu.ac.uk
The data is contributed by the international GWAS community - please see Acknowledgements and cite studies accordingly!

MR Dictionary



Conclusion

- MR uses genetic variants as proxies of modifiable exposures and can overcome some key limitations of observational studies.
- MR can reliably test for causal relationships, provided that three key assumptions are met.
- SNPs with known functional consequences increase the value of MR studies:
  - Less likely to violate the assumptions.
  - Increased biological understanding of the SNP -> exposure associations.
- Effect sizes are likely to be small, so sample sizes need to be very large.

Useful references

George Davey Smith, Gibran Hemani, Mendelian randomization: genetic anchors for causal inference in epidemiological studies, *Human Molecular Genetics*, Volume 23, Issue R1, 15 September 2014, Pages R89–R98, https://doi.org/10.1093/hmg/ddu328

Brion, Marie-Jo A et al. "Calculating statistical power in Mendelian randomization studies." *International journal of epidemiology* vol. 42,5 (2013): 1497-501. doi:10.1093/ije/dyt179

Davey-Smith & Ebrahim (2003). "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? IJE, 32. 1-22

Davies et al (2018). Reading Mendelian randomization studies: a guide. glossary, and checklist for clinicians. BMJ. Jul 12, 362:601

Evans & Davey-Smith (2015). Mendelian randomization: New applications in the coming age of hypothesis free causality. Annu Rev Genomics Hum Genet, 16, 327-50

Sanderson, E., Glymour, M.M., Holmes, M.V. *et al.* Mendelian randomization. *Nat Rev Methods Primers* **2**, 6 (2022). https://doi.org/10.1038/s43586-021-00092-5

Contact

Daisy Crick

d.crick@uq.edu.au
+61 412 345 678