THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Introduction to Mendelian Randomization

Gunn-Helen Moen
DECRA Fellow

Emily Daubney
PhD student
IMB, University of Queensland

This session

- Problems with observational data
- Randomized controlled trials
- Mendelian Randomization (MR):
  - How it works
  - Core assumptions

- Calculating causal effect estimates
  - MR example

- Limitations of MR

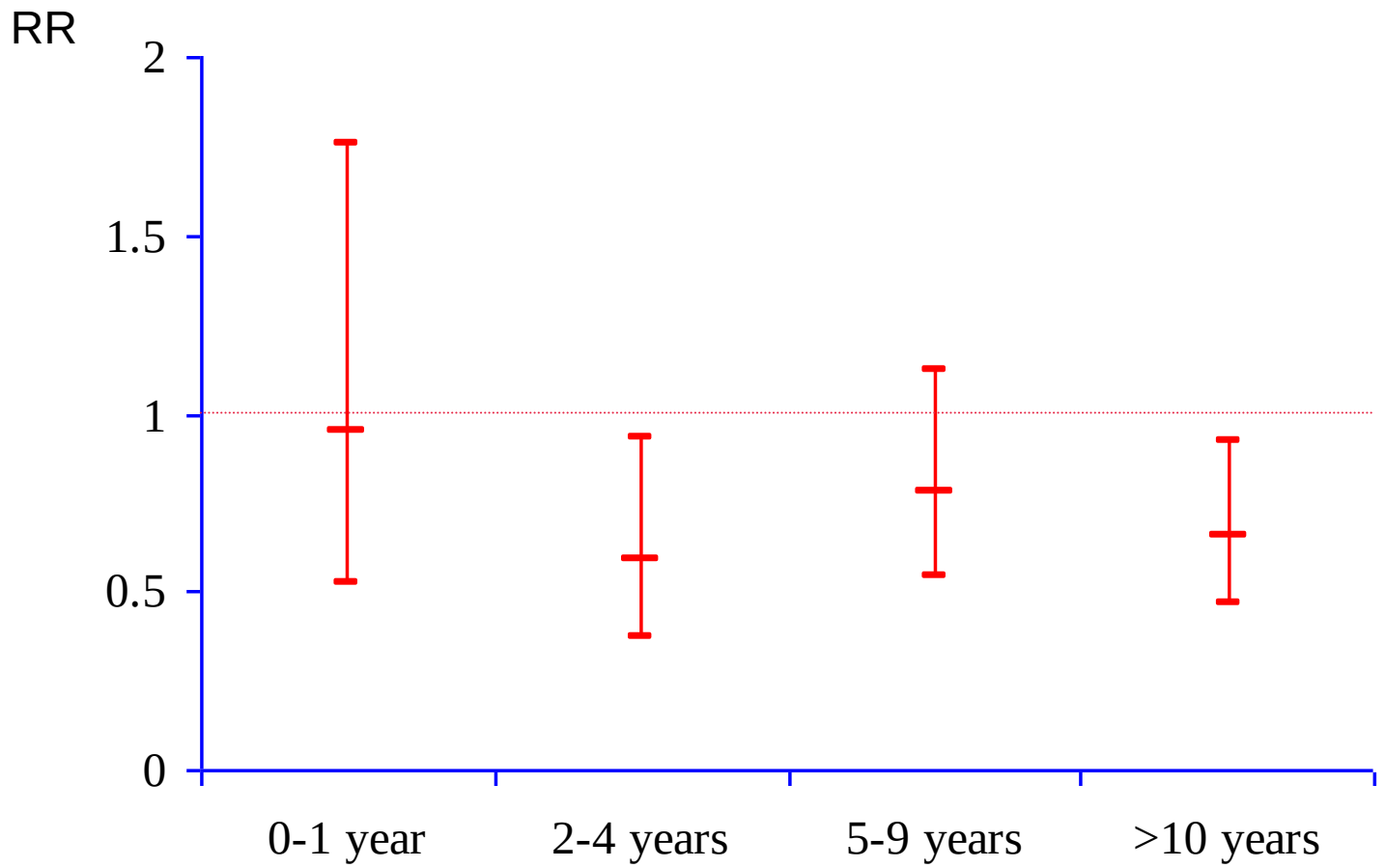
- MR sensitivity analysis
  - Inverse variance weighted MR
  - Heterogeneity tests
  - Multivariable MR
  - MR Egger

→ Protein MR in the UK Biobank with examples

Research question:

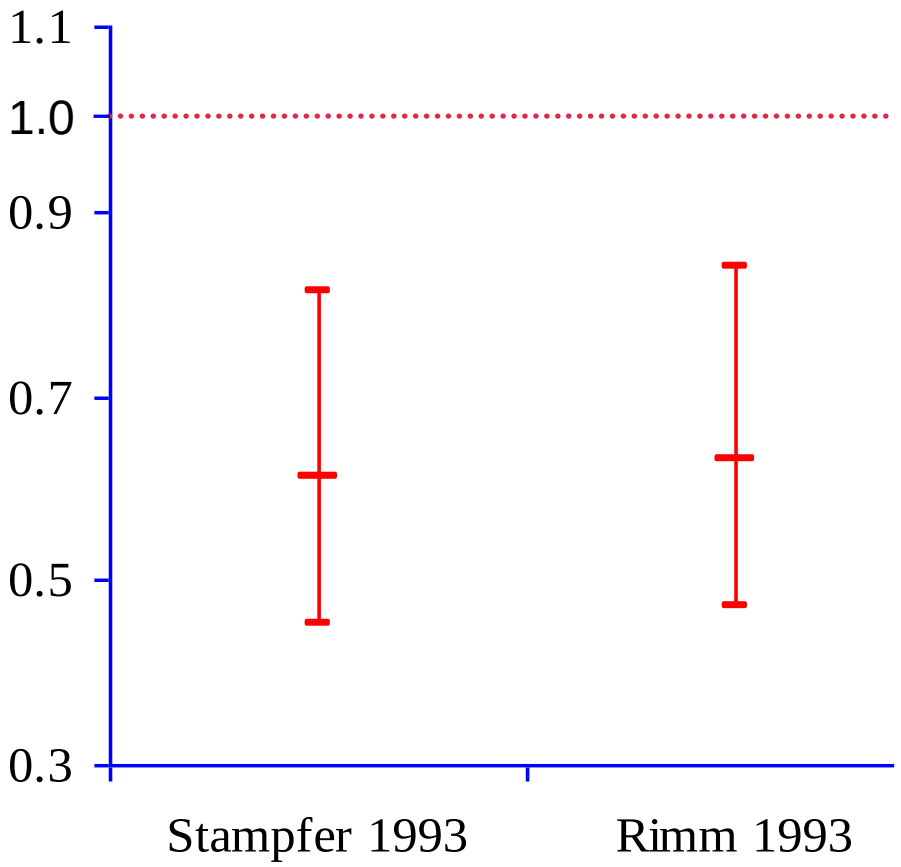
Does vitamin E reduce the risk of coronary heart disease?

CHD risk according to duration of current Vitamin E supplement use compared to no use



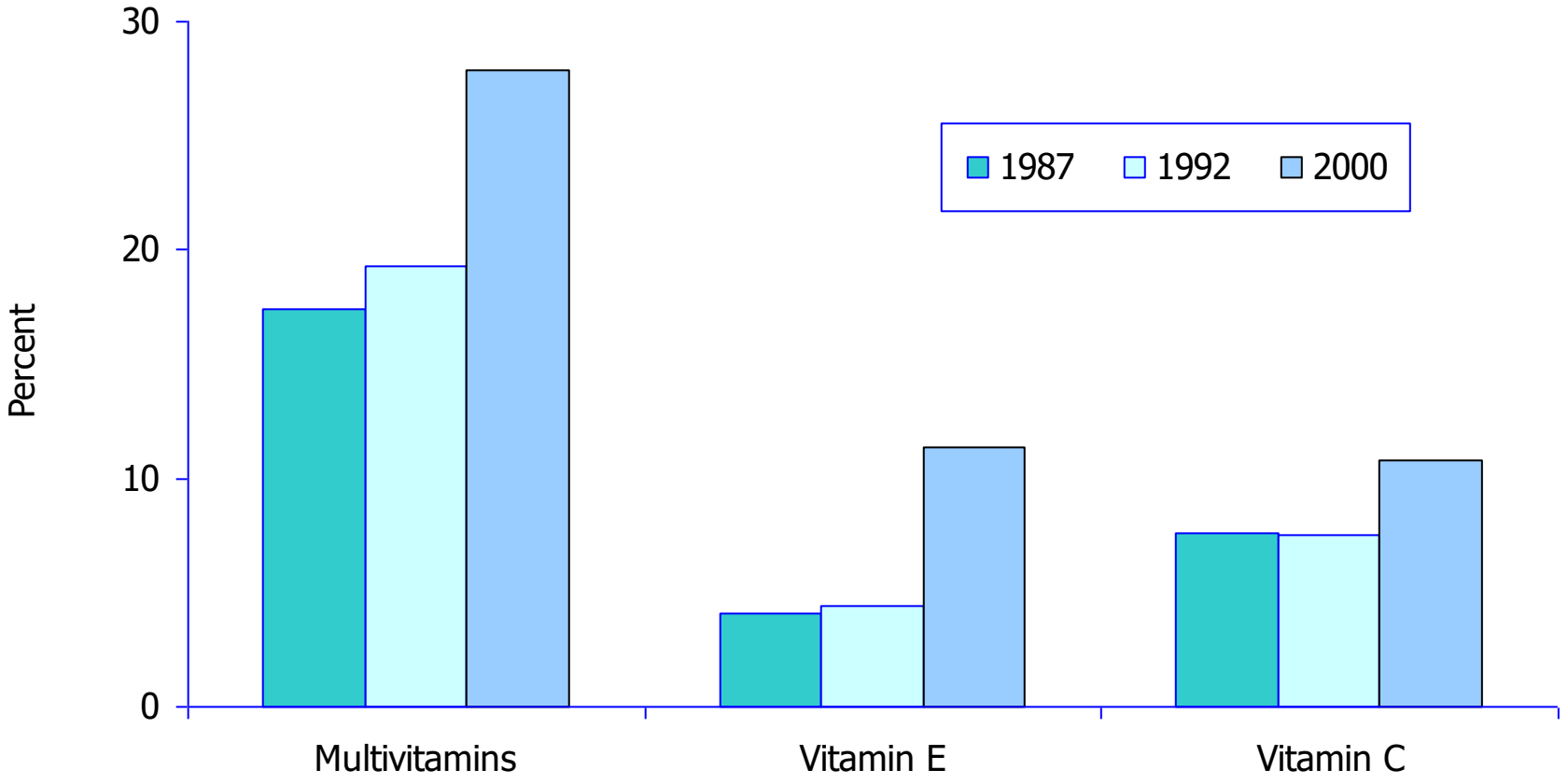
Rimm et al NEJM 1993; 328: 1450-6

Vitamin E supplement use and risk of Coronary Heart Disease



Stampfer et al NEJM 1993; 328: 144-9; Rimm et al NEJM 1993; 328: 1450-6;

Use of vitamin supplements by US adults, 1987-2000



Source: Millen AE, Journal of American Dietetic Assoc 2004;104:942-950

Observational studies – potential problems

Reverse causation

Confounding

Observational studies – potential problems

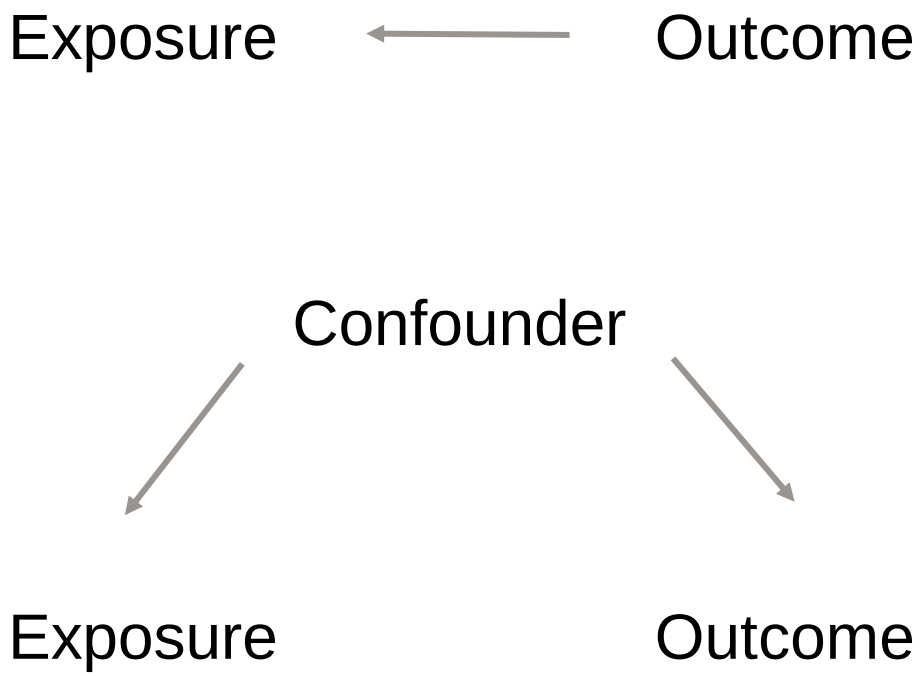
Reverse causation

Confounding

Observational studies – potential problems

Reverse causation

Confounding

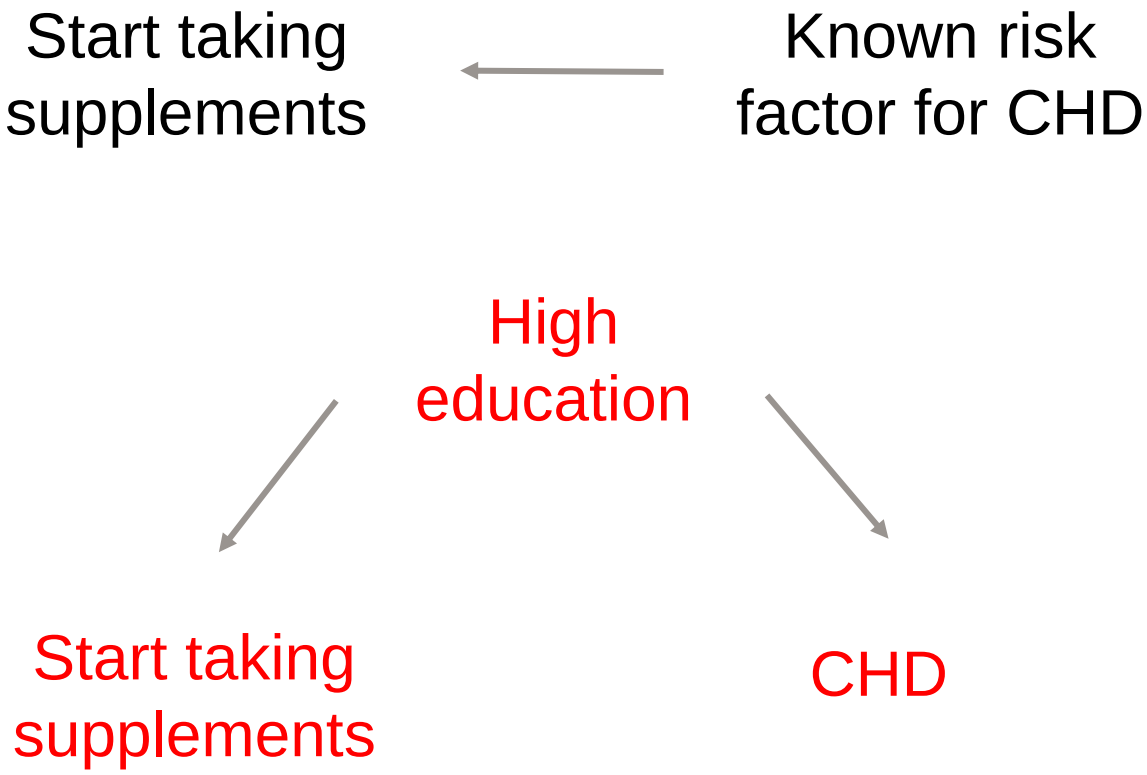


Observational studies – potential problems

Reverse causation

Confounding

- Measured and unmeasured!

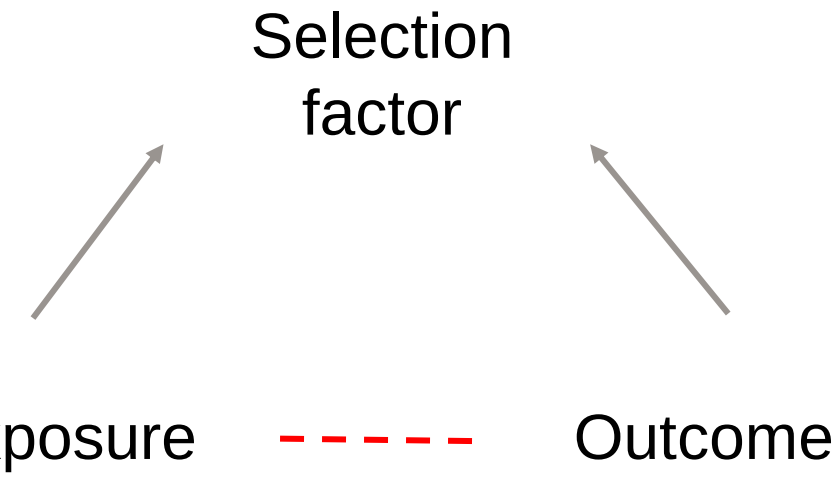


Observational studies – potential problems

Reverse causation

Confounding

Bias (e.g. Selection Bias)



JOURNAL ARTICLE

Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population

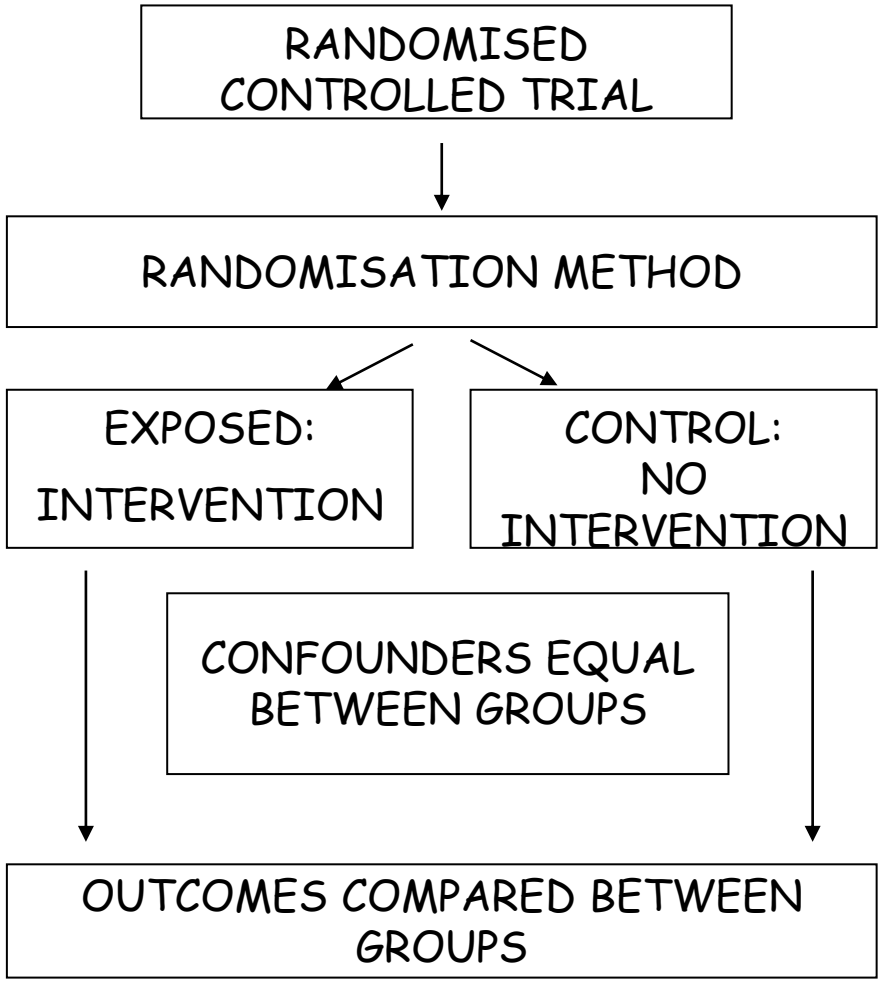
Anna Fry, Thomas J Littlejohns ✉, Cathie Sudlow, Nicola Doherty, Ligia Adamska, Tim Sprosen, Rory Collins, Naomi E Allen

American Journal of Epidemiology, Volume 186, Issue 9, 1 November 2017, Pages 1026–1034, https://doi.org/10.1093/aje/kwx246

Published: 21 June 2017 **Article history** ▾

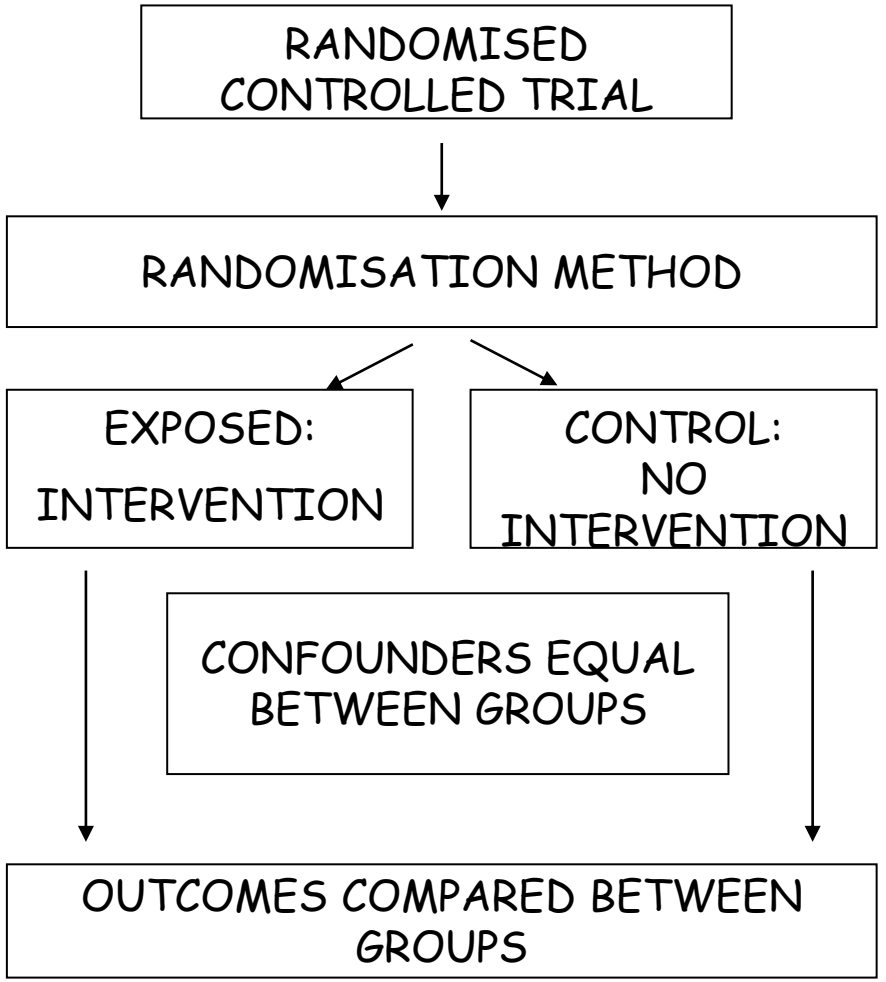
"5.5% participated in the baseline assessment … UK Biobank is not representative of the sampling population; there is evidence of a "healthy volunteer" selection bias"

Randomized controlled trials



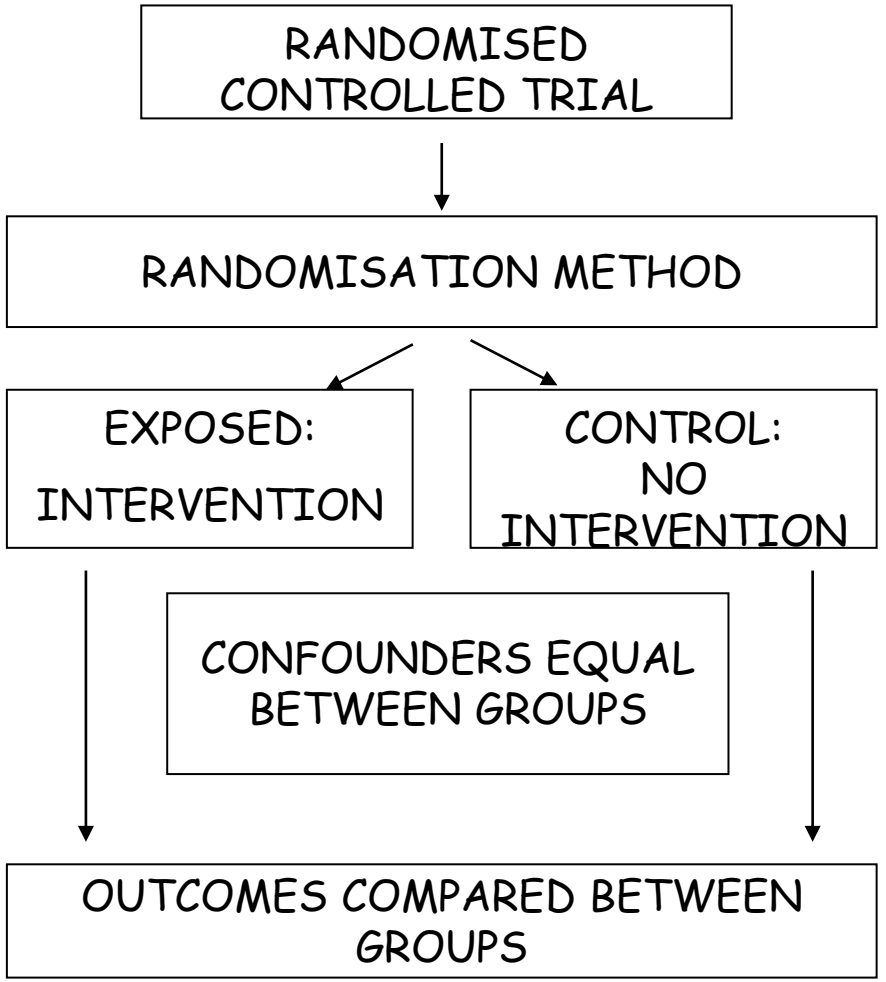
Randomized controlled trials

- Full control over the intervention and exposure of interest
- No confounding
- “Simple” statistical methods



Randomized controlled trials

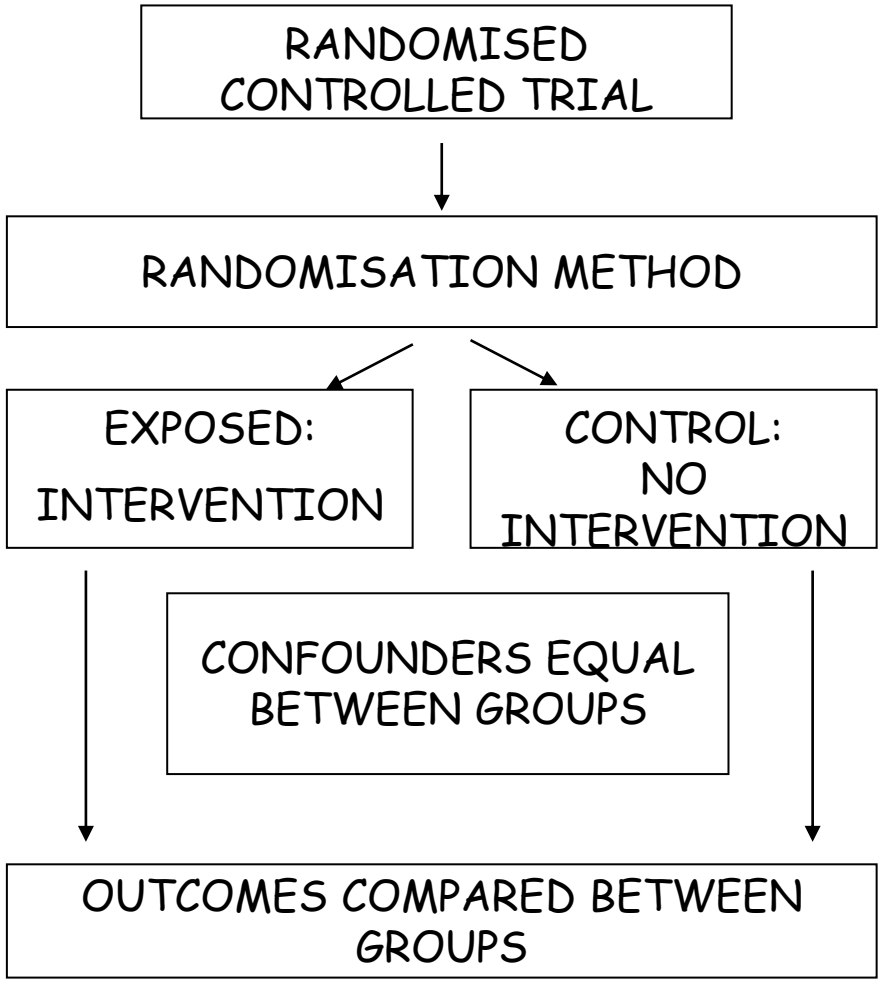
- Time consuming
- Expensive
- Difficult to study long term exposures
- Difficult to study diseases with long latency
- Generalizability



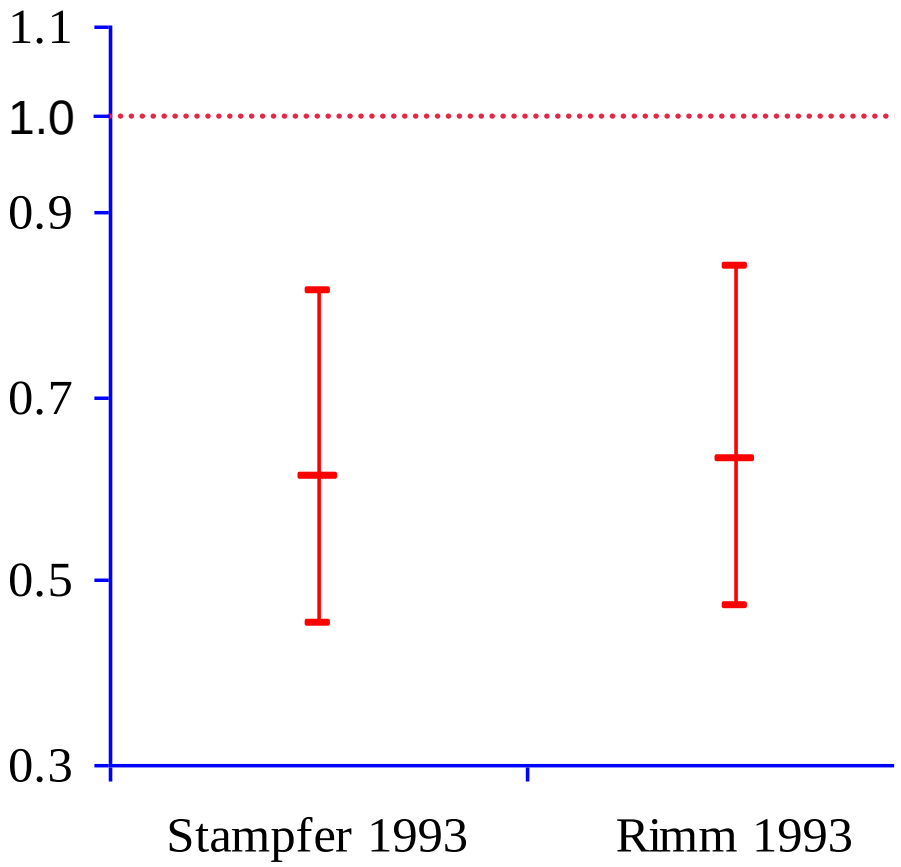
Randomized controlled trials

Not always ethical or practical:

- Toxic exposures, smoking, alcohol
- Pregnancy
- Children
- Individuals who are unable to give informed consent



Vitamin E supplement use and risk of Coronary Heart Disease



Stampfer et al NEJM 1993; 328: 144-9; Rimm et al NEJM 1993; 328: 1450-6; Eidelman et al Arch Intern Med 2004; 164:1552-6

Mendelian Randomization

Mendelian Randomization

What does it do?

- Assess causal relationship between two variables
- Estimate magnitude of causal effect

How does it do it?

- Using Mendel's laws of inheritance:

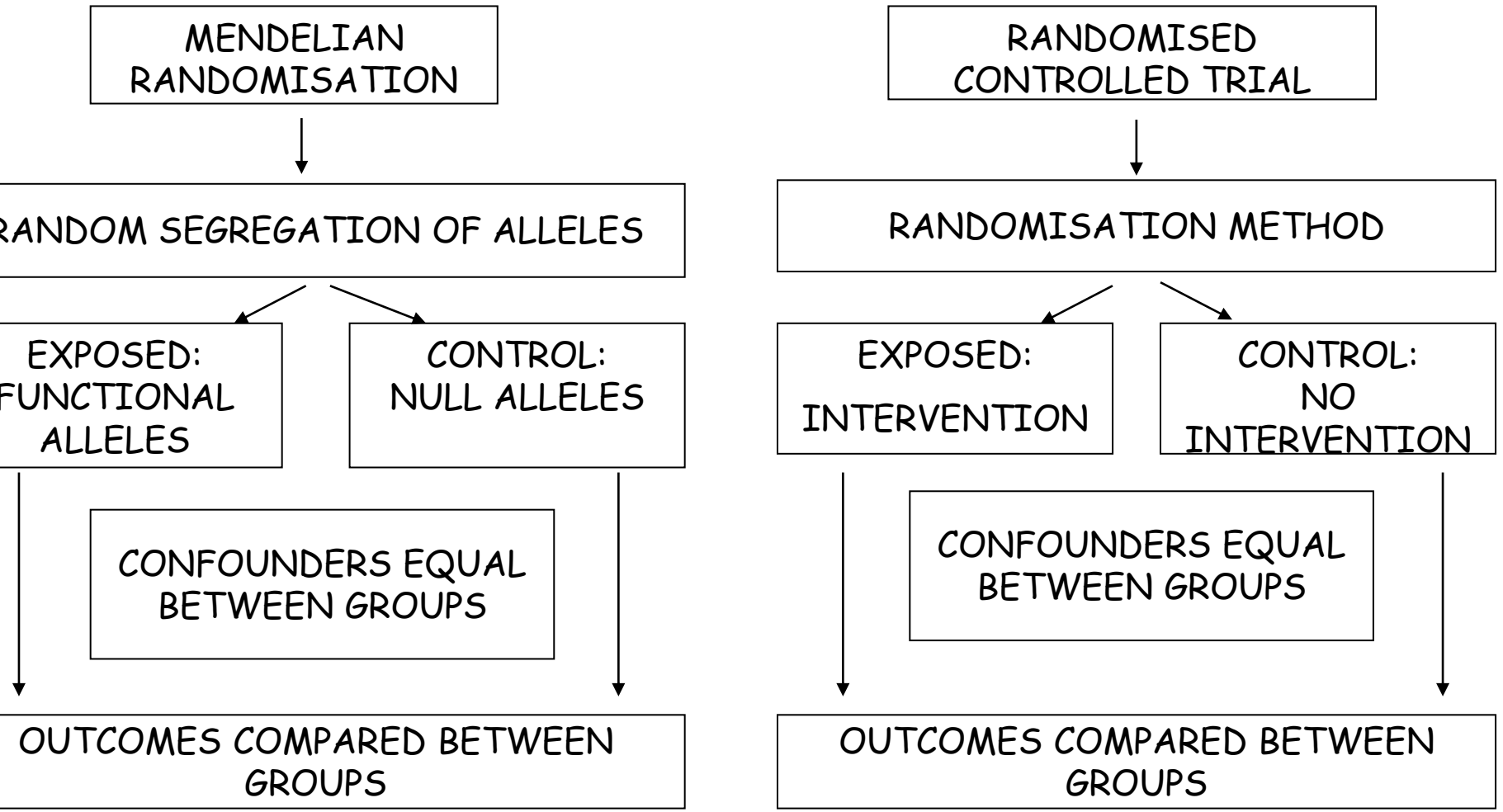
  1. Segregation: *alleles separate at meiosis and a randomly selected allele is transmitted to offspring*

  2. Independent assortment: *alleles for separate traits are transmitted independently of one another*

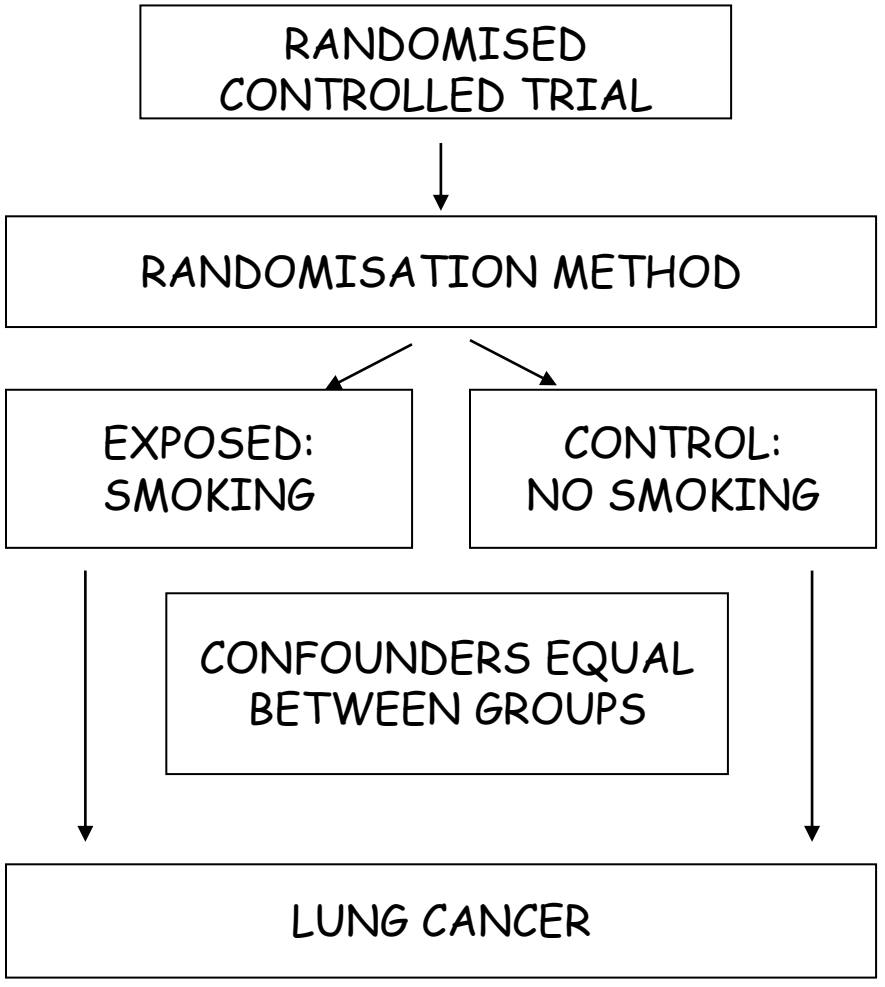
What do we need?

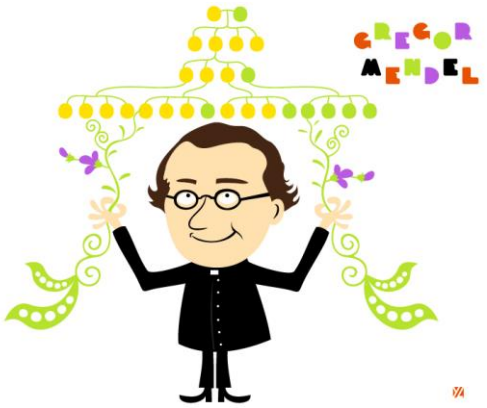
- Observational studies with genetic information

Mendelian Randomization

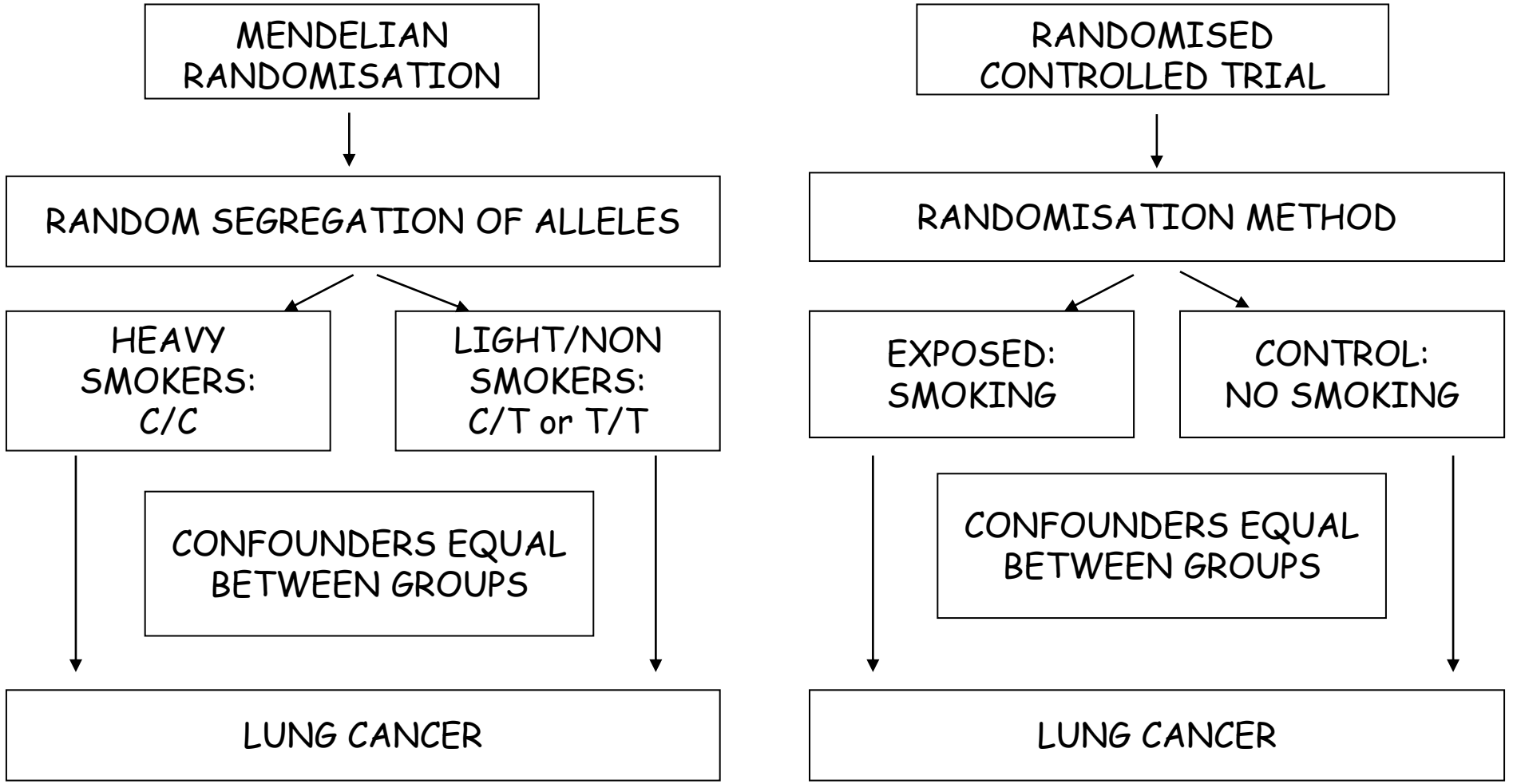


Mendelian Randomization

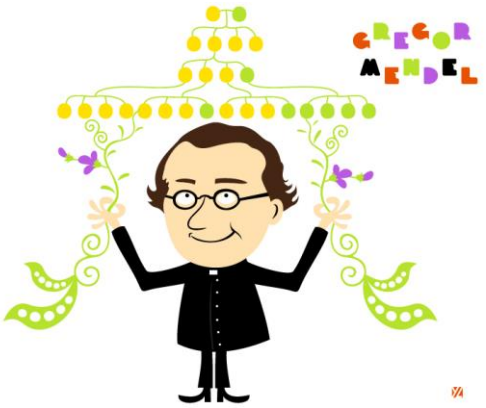




Mendelian Randomization



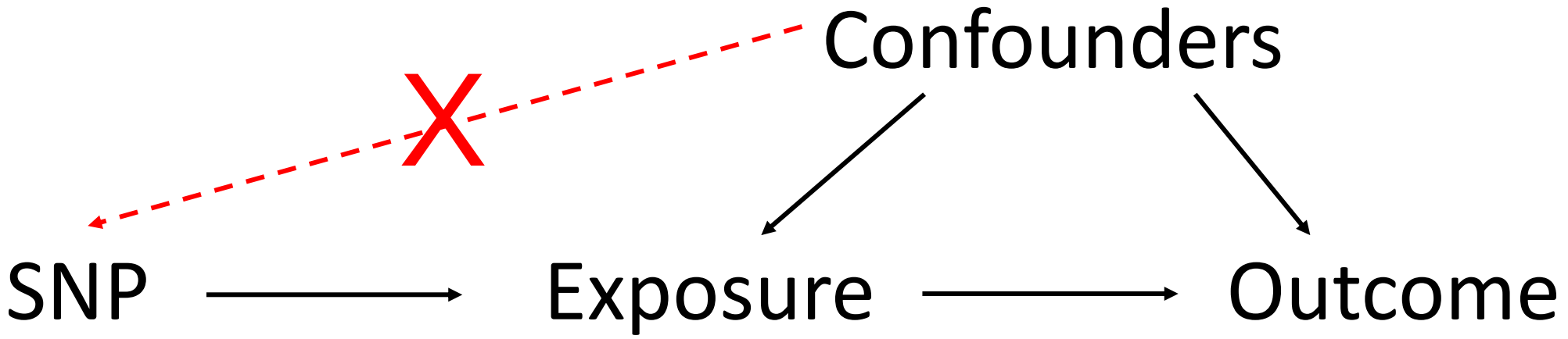
Assumptions

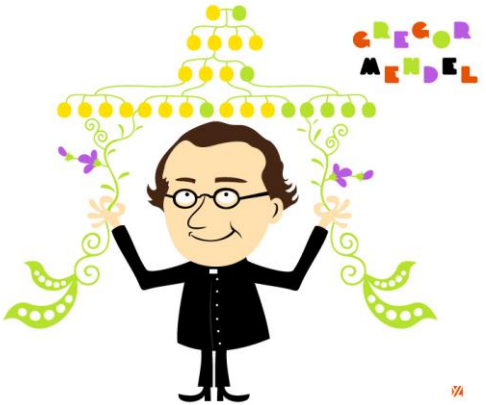


SNP ⟶ Exposure

(1) SNP is associated with the exposure

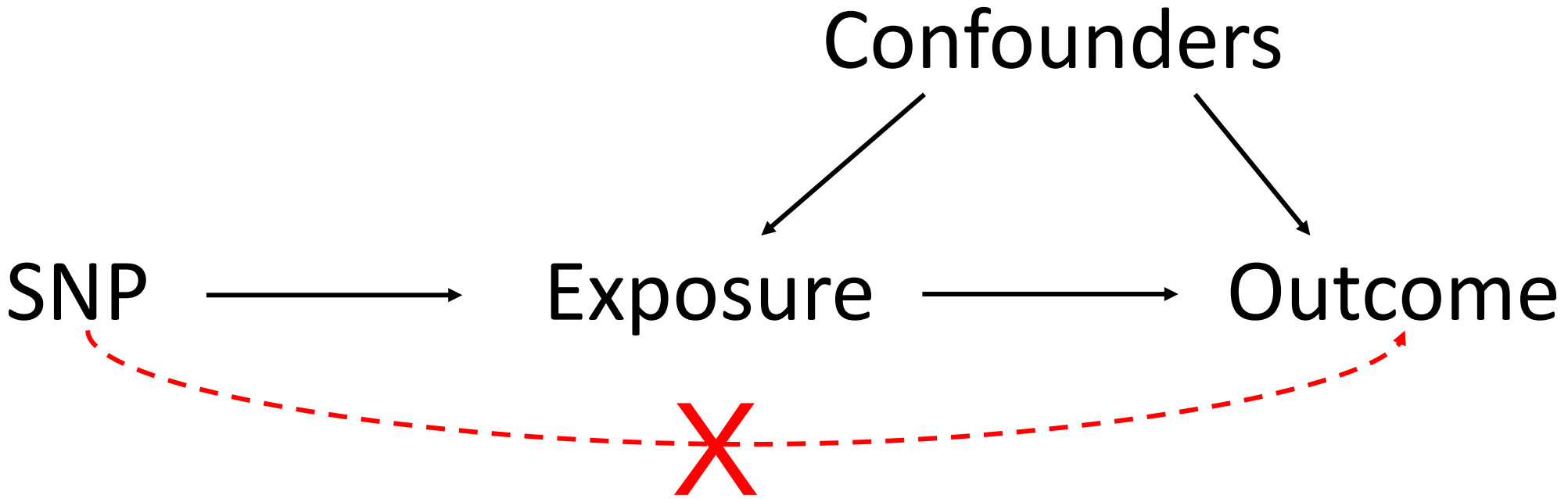
Assumptions

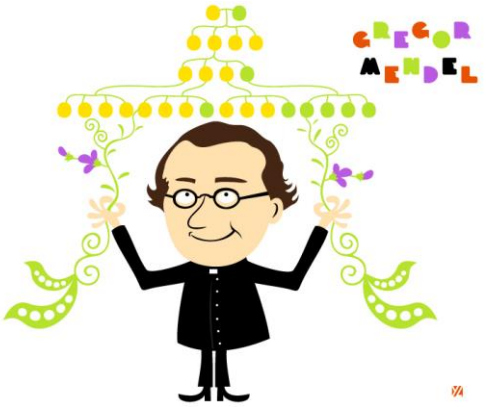




(2) SNP is not associated with confounding variables

Assumptions

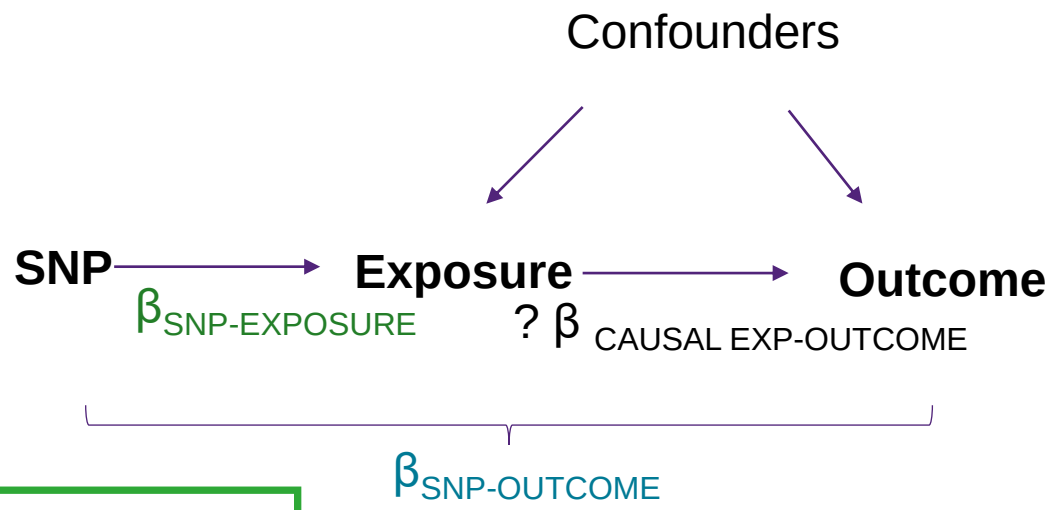




(3) SNP only potentially associated with outcome through the exposure

Calculating causal effects

Calculating Causal Effect Estimates

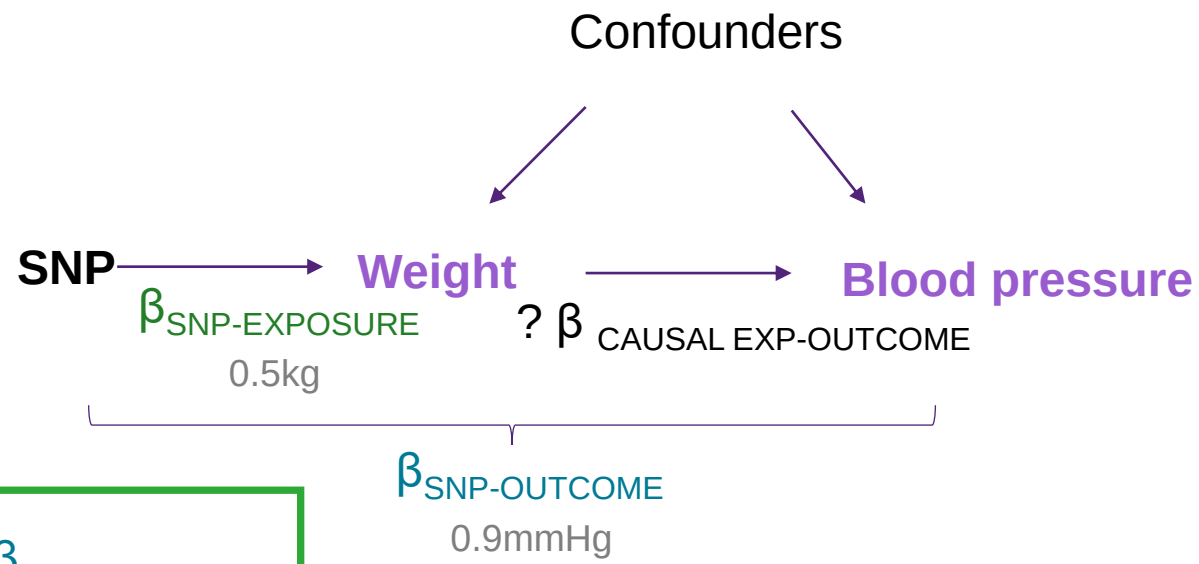


Wald Estimator: $\dfrac{\beta_{SNP\text{-}OUTCOME}}{\beta_{SNP\text{-}EXPOSURE}}$

$\beta_{SNP\text{-}OUTCOME} = \beta_{CAUSAL\ EXP\text{-}OUTCOME} \times \beta_{SNP\text{-}EXPOSURE}$

Can be performed in different samples (2 sample MR analysis)

Calculating Causal Effect Estimates



Wald Estimator: $\dfrac{\beta_{SNP\text{-}OUTCOME}}{\beta_{SNP\text{-}EXPOSURE}}$

= change in outcome per unit change in exposure

BP and weight:

$\dfrac{0.9\ mmHg/allele}{0.5\ kg/allele}$

=1.8 mmHg/kg

Can be performed in different samples (2 sample MR analysis)

Limitations to Mendelian Randomization

- The existence of instruments
- Population stratification
- **Power (also "weak instrument bias")**
- Pleiotropy

Power and Weak Instruments

Power:

- Genetic variants explain very small amounts of phenotypic variance in a 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

Using Multiple Genetic Variants as Instruments

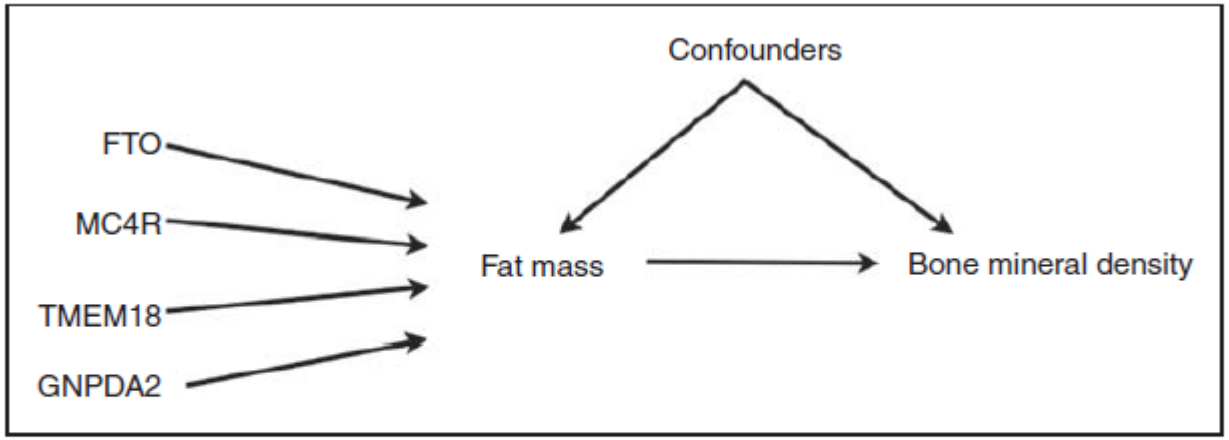


Figure 1. DAG for a Mendelian randomisation analysis using four genetic variants as instrumental variables for the effect of fat mass on bone mineral density.

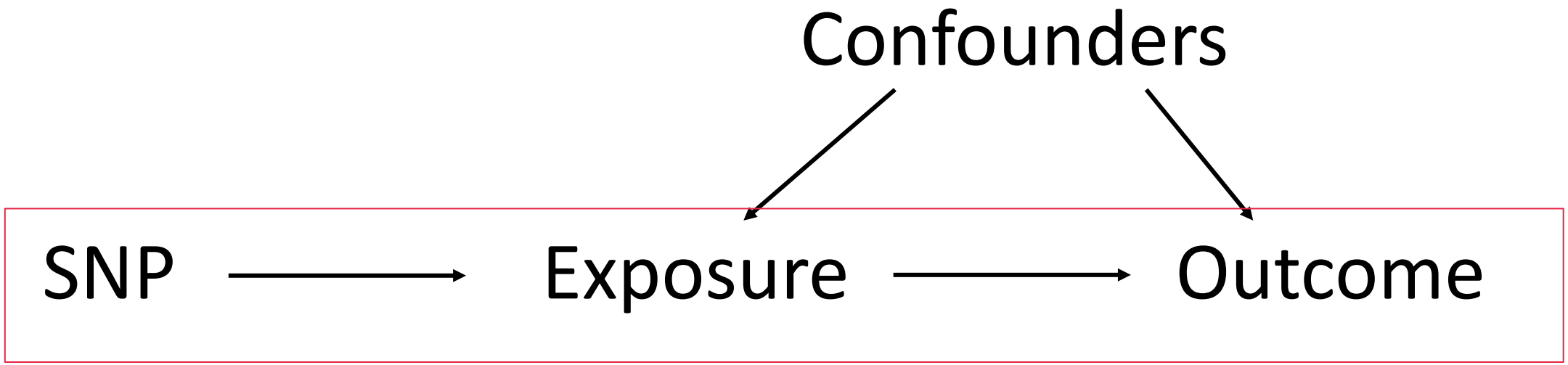
Palmer et al (2011) Stat Method Res

- Genetic Score
- Test multiple variants individually and meta-analyse

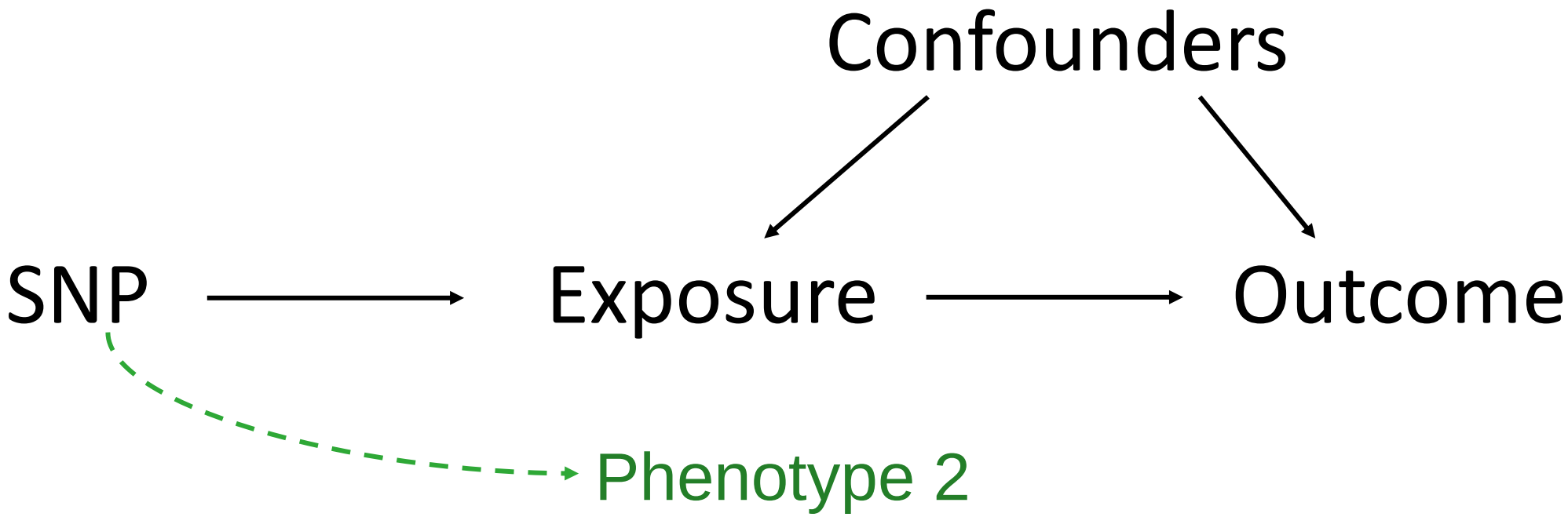
Limitations to Mendelian Randomization

- The existence of instruments
- Population stratification
- Power (also “weak instrument bias”)
- **Pleiotropy**
  - The phenomenon in which a single locus affects two or more traits
  - **Vertical pleiotropy** is observed when a trait influenced by genetic factors has, in turn, influenced another trait by acting as a mediator
  - **Horizontal pleiotropy** occurs when the genetic variant used to proxy the exposure influences the outcome outside of its effect on the exposure

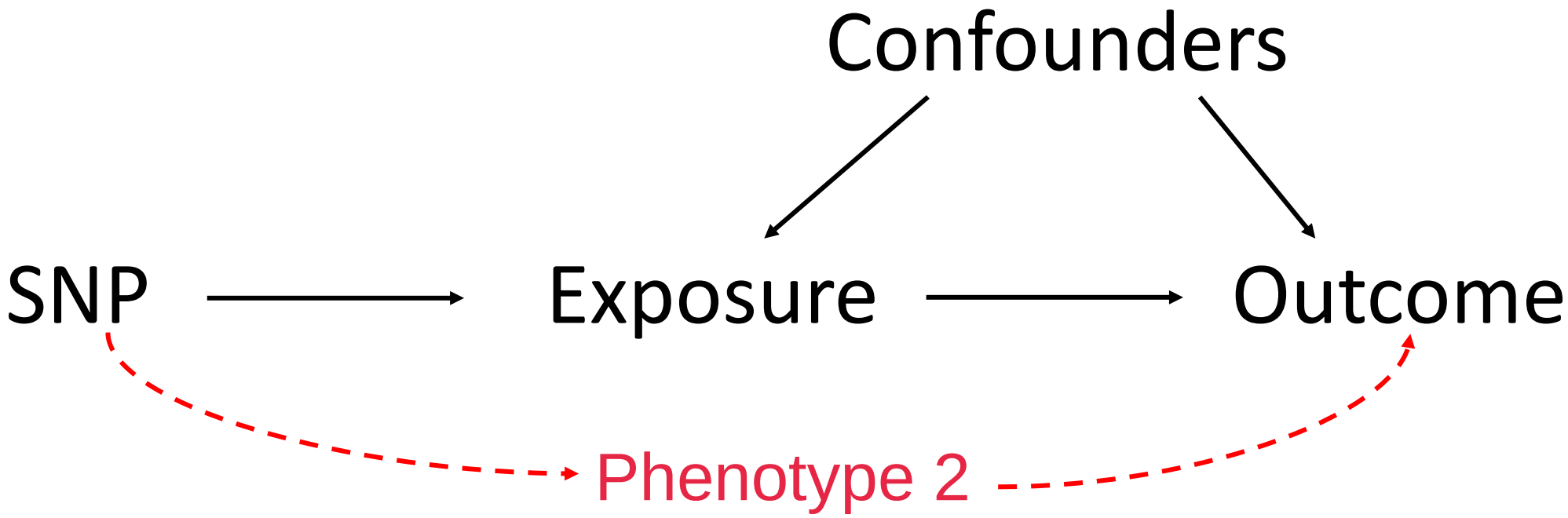
Pleiotropy



Pleiotropy

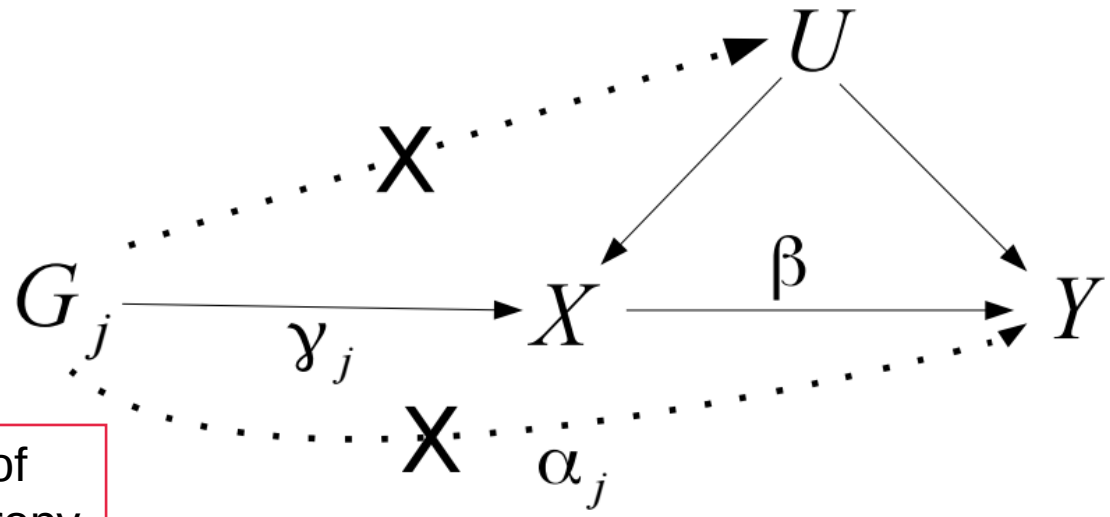


Pleiotropy



Two Sample MR:
Single Variants – What happens when we have pleiotropy?

Wald Estimator: $\frac{\beta_{SNP\text{-}OUTCOME}}{\beta_{SNP\text{-}EXPOSURE}}$



Single variant Wald estimate:

$$\beta_j = \beta + \frac{\alpha_j}{\gamma_j}$$

Strength of the pleiotropy

Strength of the association between the instrument and exposure

Summary

Mendelian Randomization (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

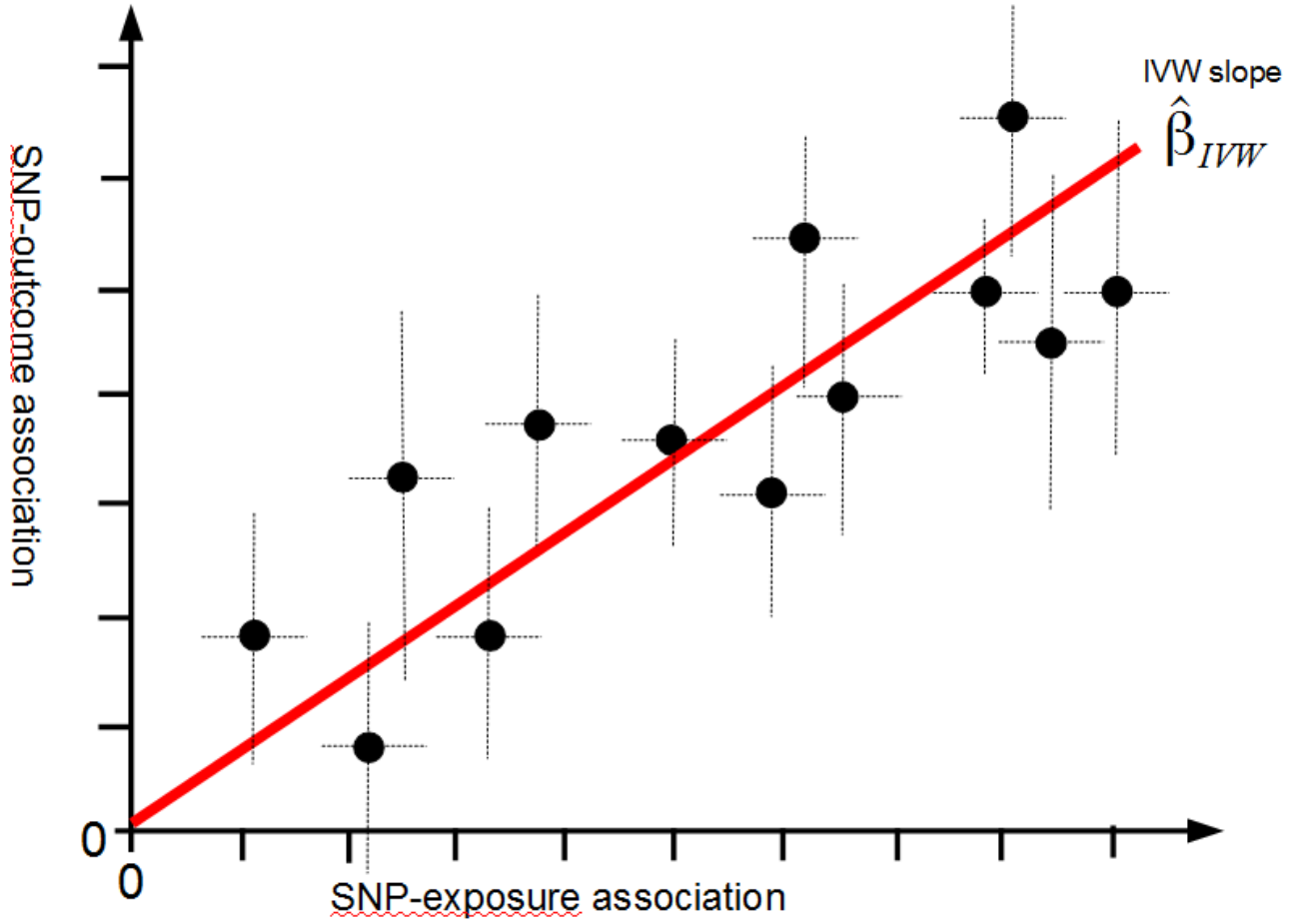
Sensitivity analysis

Fixed Effects Inverse variant weighted (IVW) MR

IVW is equivalent to a weighted regression of SNP-outcome effects on SNP-exposure effects passing through the origin.

- The SNPs are weighted by 1/SE_SNP_outcome
- The slope is the estimate of the causal effect

Assumes no pleiotropy.



Heterogeneity

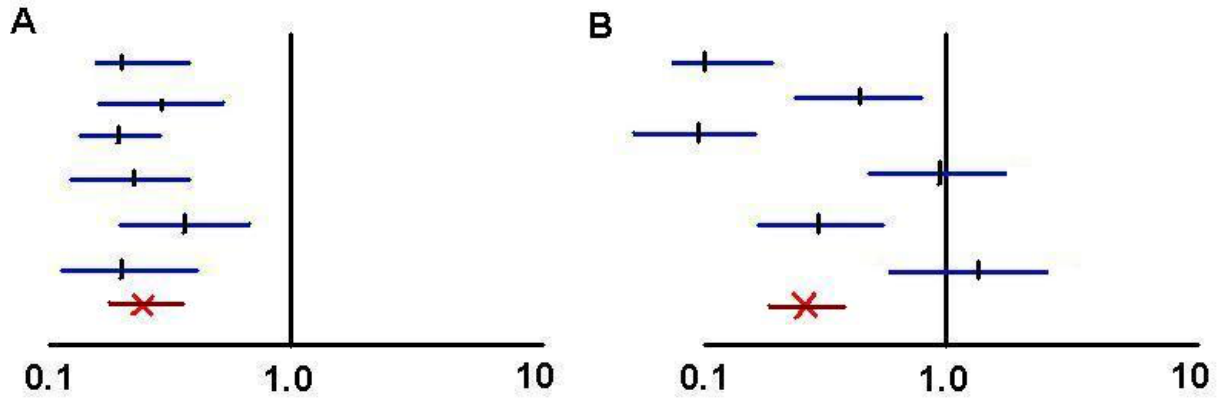
- We expect that each SNP represents an independent study, and each should give an unbiased (if imprecise) estimate of the causal effect of x on y

- 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

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

n=6 instruments
Expect Q = 5 if there is no heterogeneity
Q is chi-square distributed with n-1 degrees of freedom

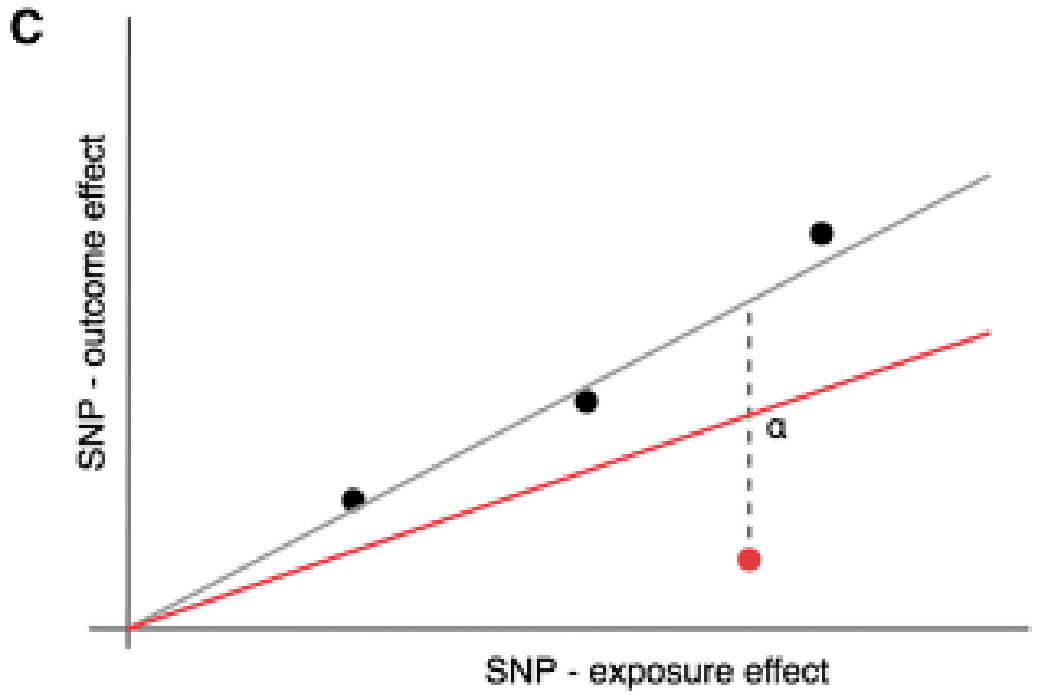


Option 1: Remove outliers

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

However:

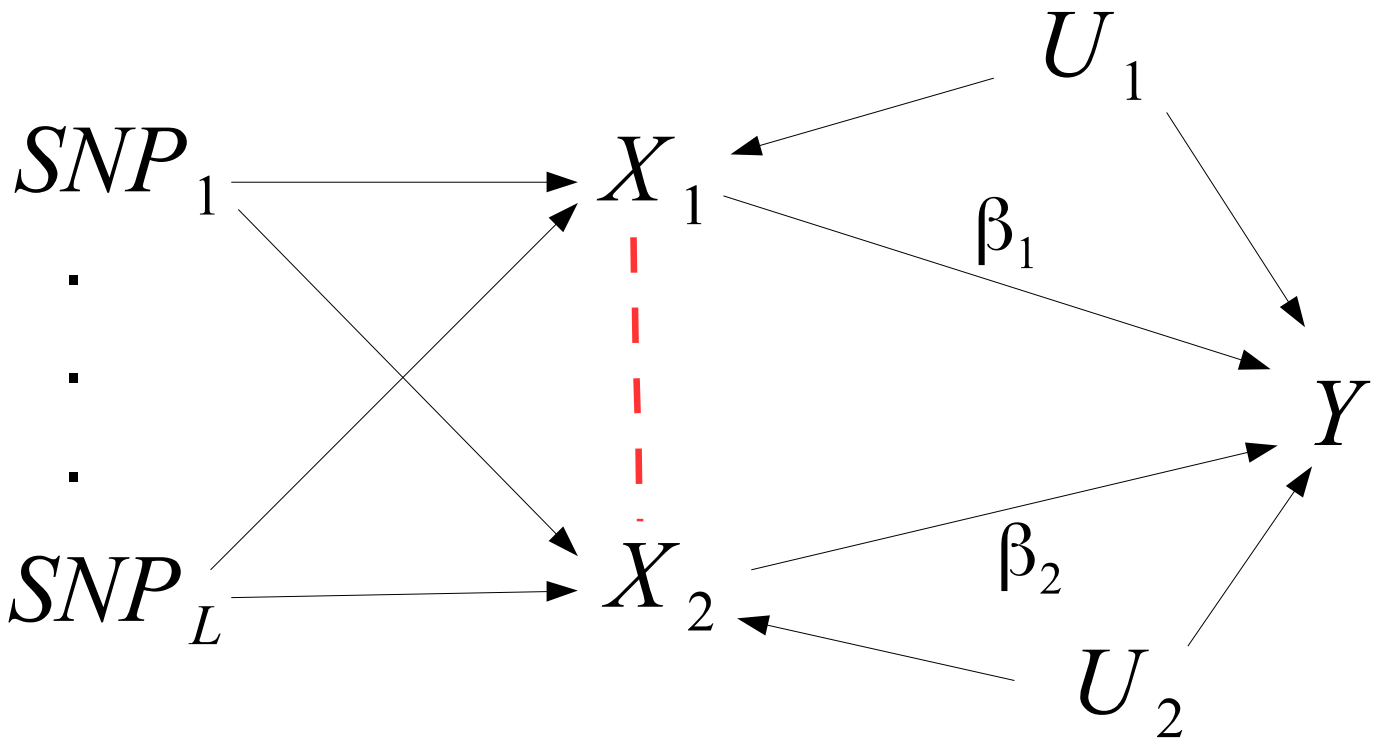
- Cherry picking – remove outliers will artificially provide a more precise estimate
- What if the outlier is the only valid instrument, and all the others are invalid?



Option 2: Multivariable MR

- We are testing for whether X1 has an influence on Y
- We know that some instruments for X1 also have influences on X2
  - This opens up the possibility of horizontal pleiotropy biasing our estimate

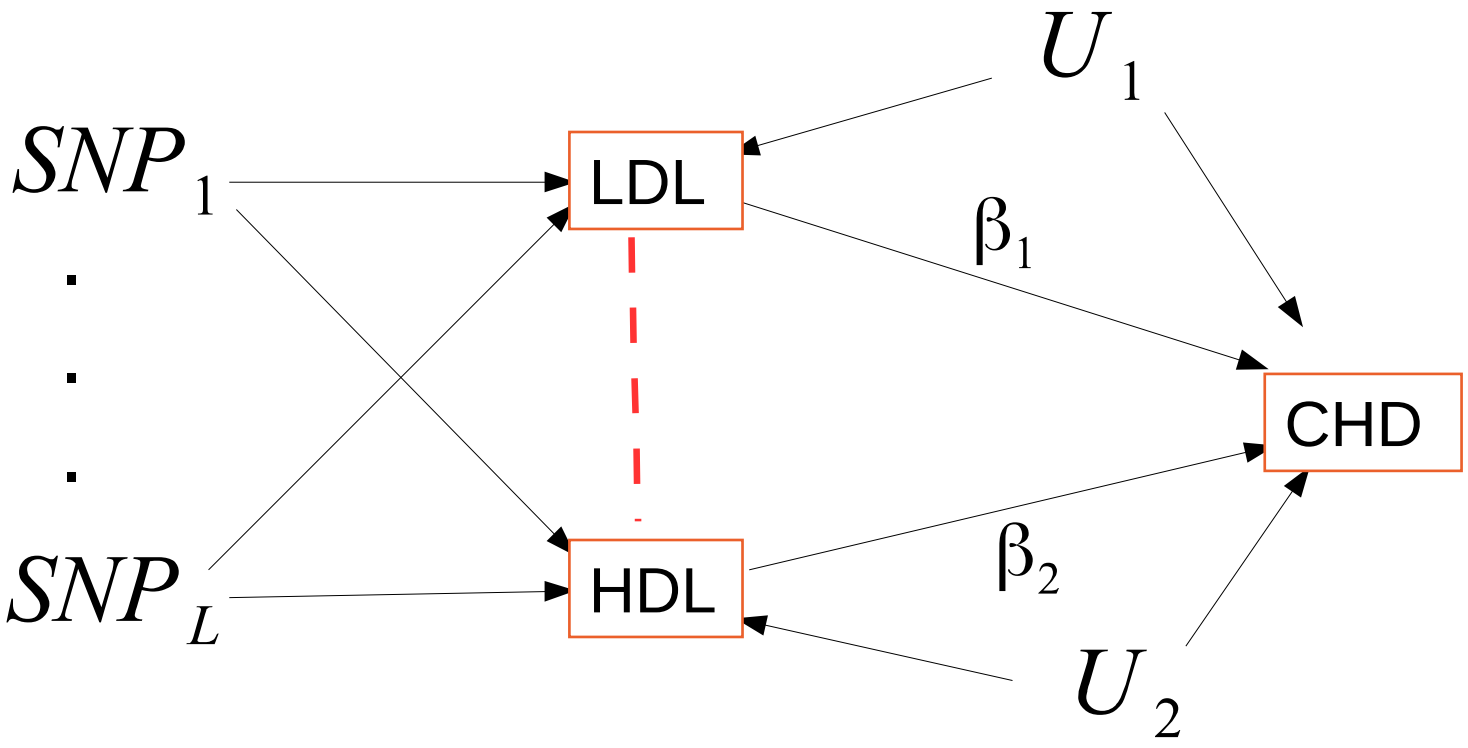
What is the X1-Y association adjusting for X2?



Option 2: Multivariable MR

- We are testing for whether **LDL** has an influence on **CHD**
- We know that some instruments for **LDL** also have influences on **HDL**
- This opens up the possibility of horizontal pleiotropy biasing our estimate

What is the LDL-CHD association adjusting for HDL?

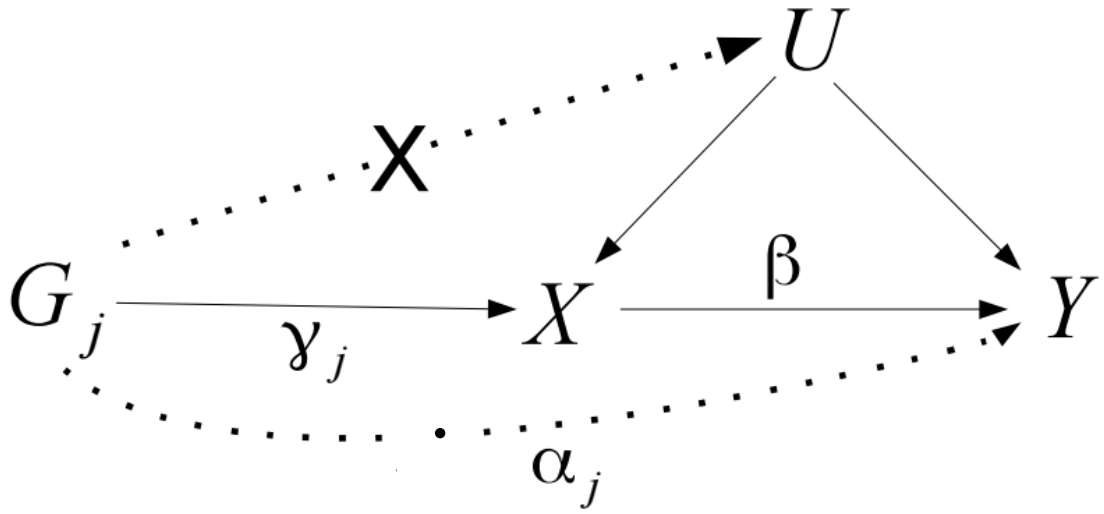


MR Egger Regression

In Mendelian Randomization when multiple genetic variants are being used as IVs, Egger regression can:

- Identify the presence of ‘directional’ pleiotropy
  (biasing the IV estimate)

- provide a less biased causal estimate
  (in the presence of pleiotropy)

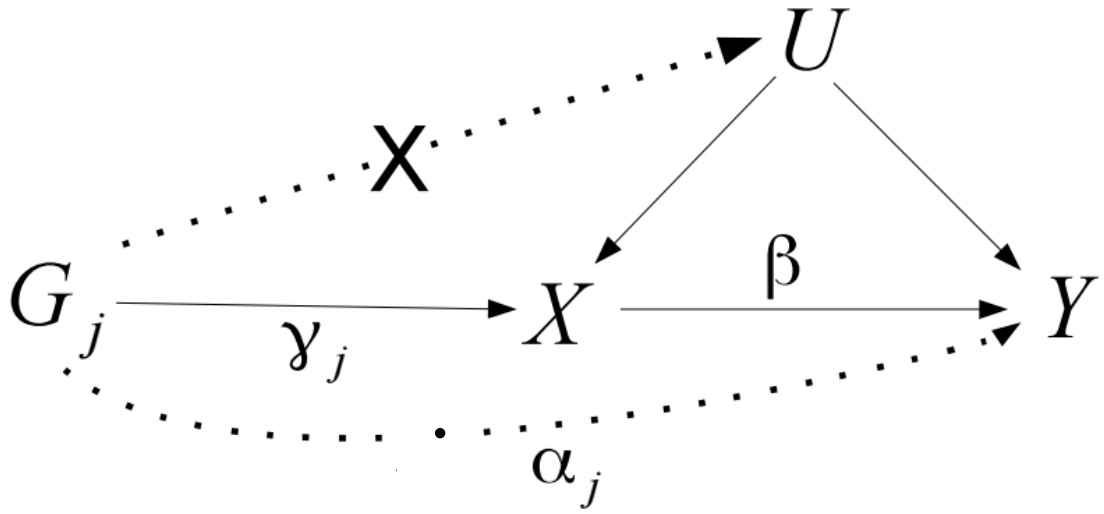


InSIDE Assumption

InSIDE:

INstrument Strength Independent of Direct Effect

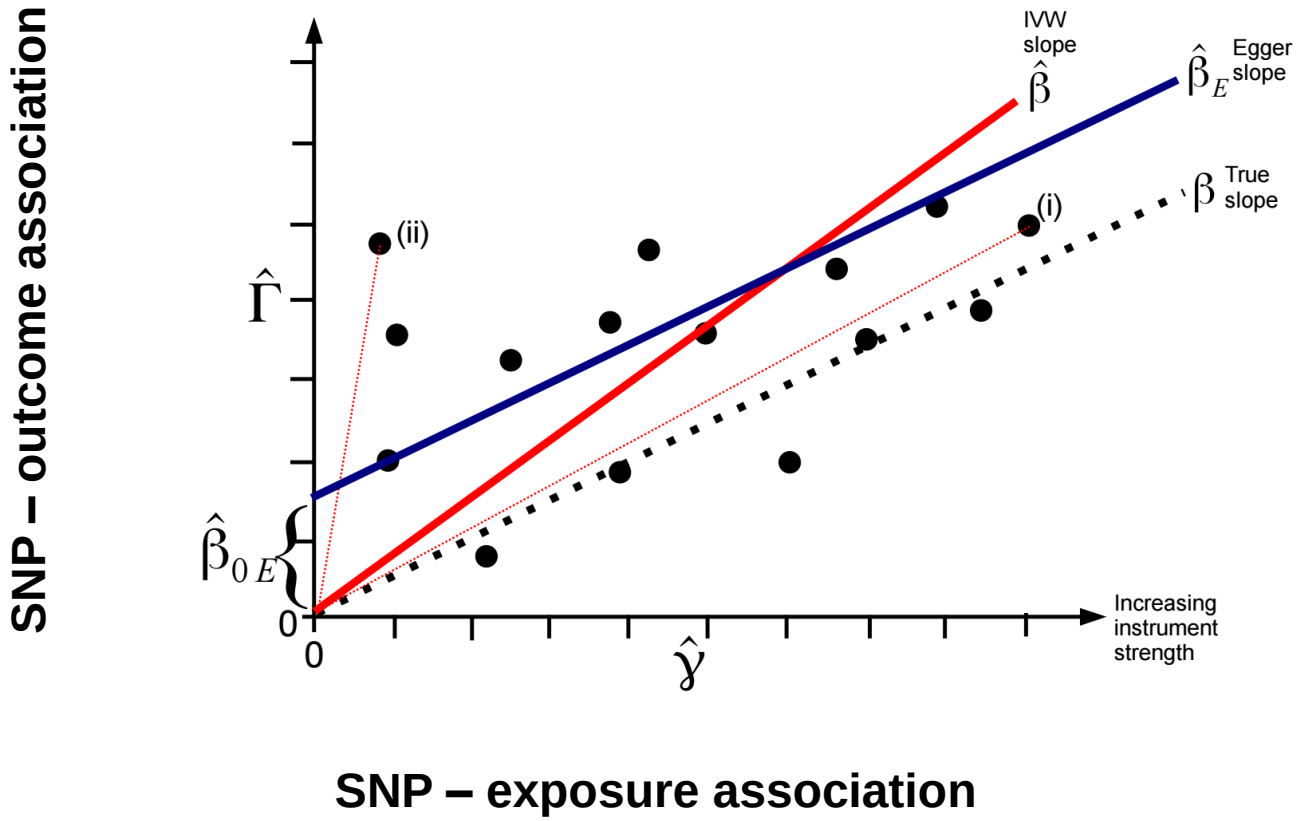
Correlation between the SNP exposure and the direct effect of the SNP on the outcome is zero.



$\hat{\alpha}_j$ is independent of its denominator, $\hat{\gamma}_j$.

ALL INVALID INSTRUMENTS INSIDE ASSUMPTION SATISFIED

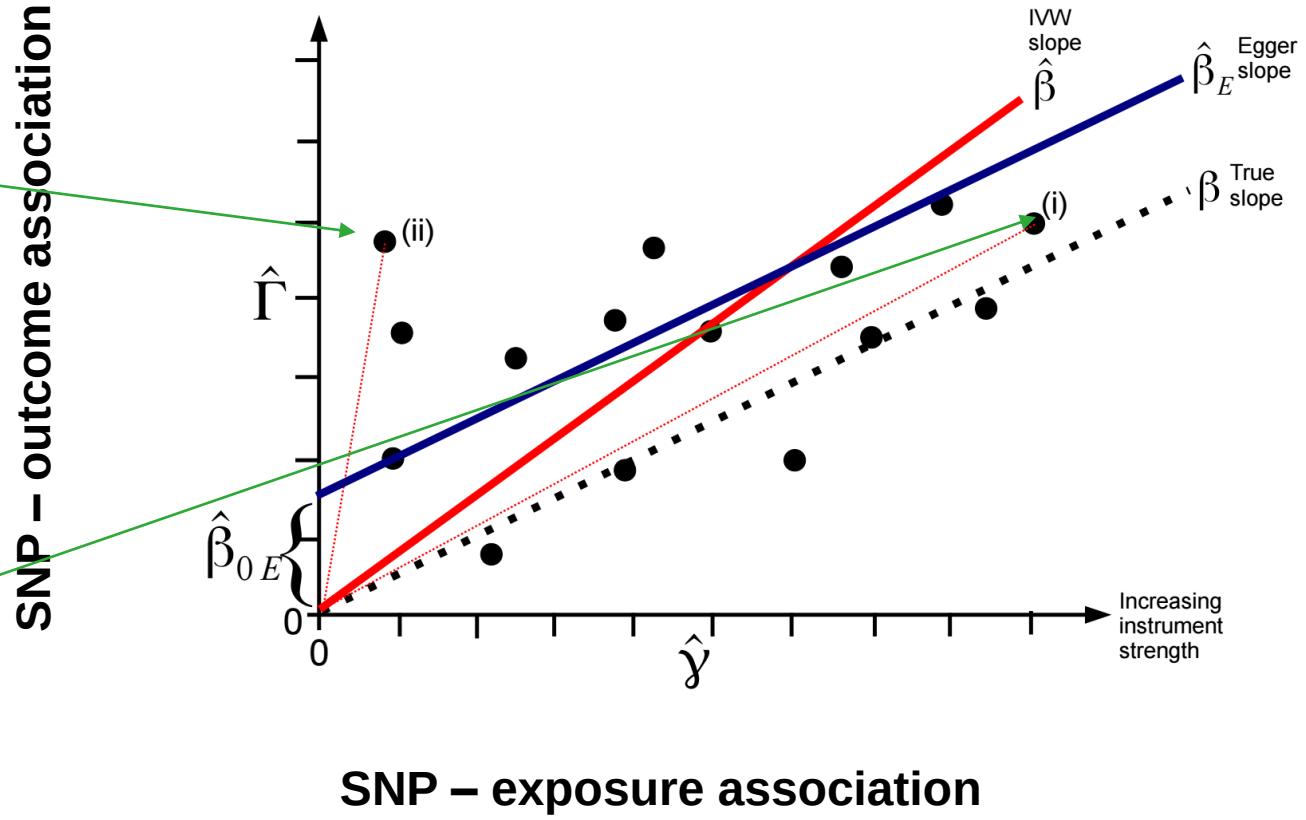
- Egger's test assesses whether the intercept term is significantly different from zero.
- The estimated values of the intercept can be interpreted as the average pleiotropic effect across all genetic variants.
- An intercept term different from zero indicates directional pleiotropy



Another example of instrument strength

Weak instrument with big effect on outcome

Strong instrument with smaller effect on outcome



Other sensitivity analysis

- Weighted Median
- Simple and Weighted Mode
- +++

Next up, Proteomics in UK Biobank!

Thanks to Dave Evans for help preparing the slides