Evaluating Causality

David Evans
Institute for Molecular Bioscience
University of Queensland

Today

- Introduction to causality from an SEM perspective
- Mendelian randomization
- Extensions to causal modeling using Structural Equation Models

What is Causality?

What is Causality?

- Difficult to define!

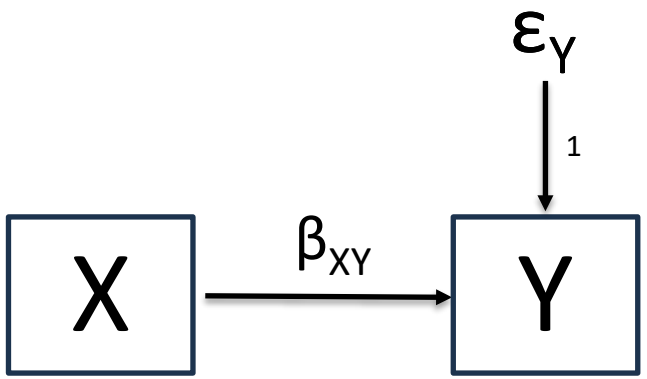
- Intervening on “X” produces a change in “Y” when all potential confounding variables are held constant

Modelling Causality in SEMs

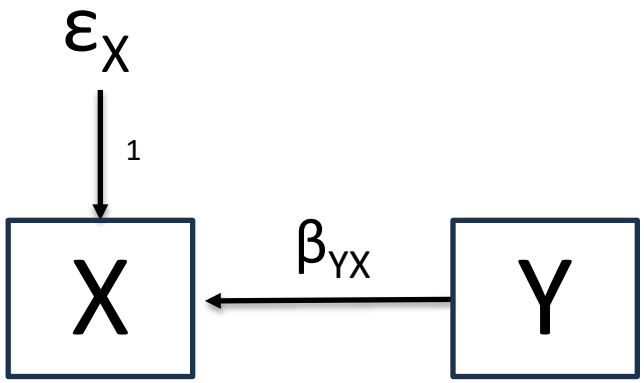
- Direction/Temporality
- Association
- (Pseudo)Isolation

Direction / Temporality

- Causation



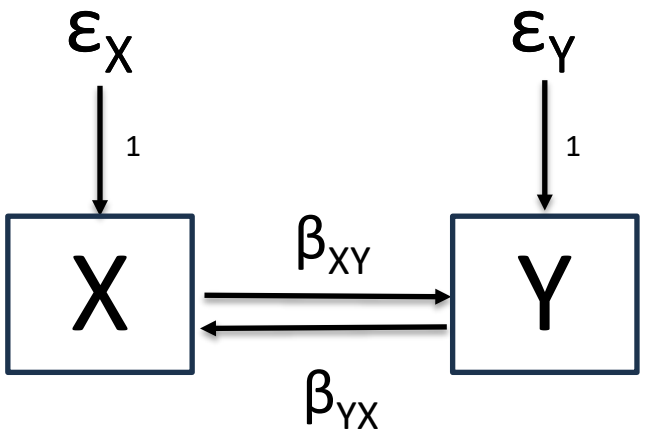
- Reverse Causation



- Genes come first!

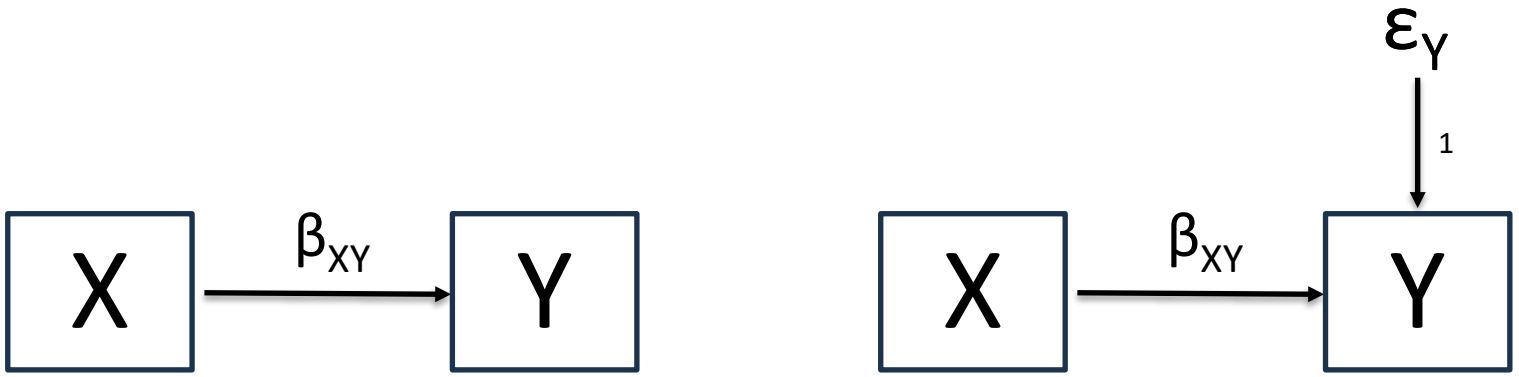
Direction / Temporality

- Reciprocal causation



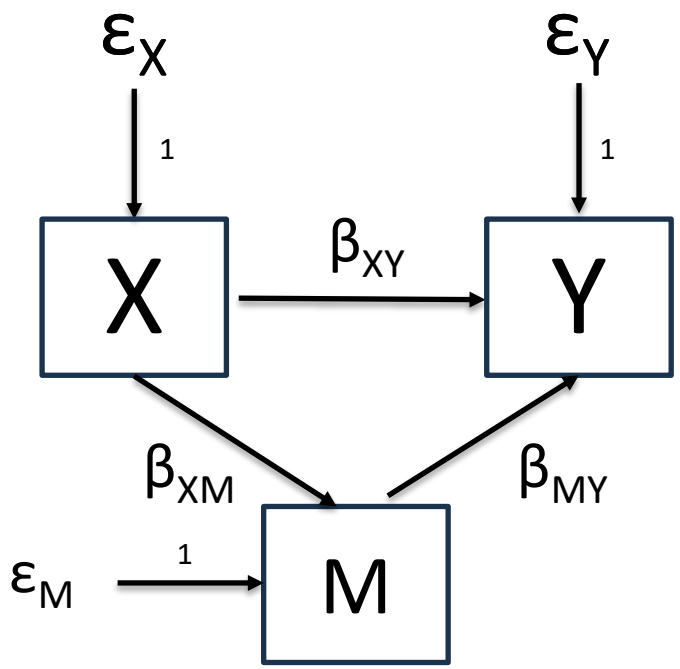
Association

- Probabilistic vs Deterministic relationships



Association

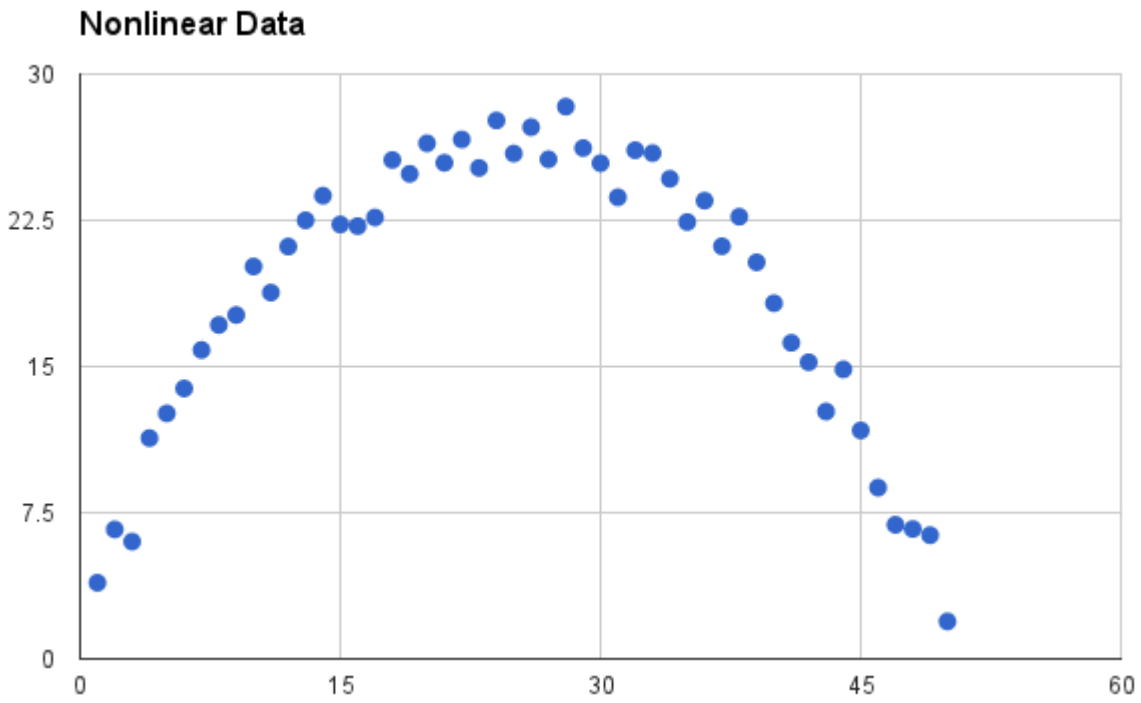
- Direct vs Indirect Effects vs Total Effects



- What if $\beta_{XY} = -(\beta_{XM} \times \beta_{MY})$

Association

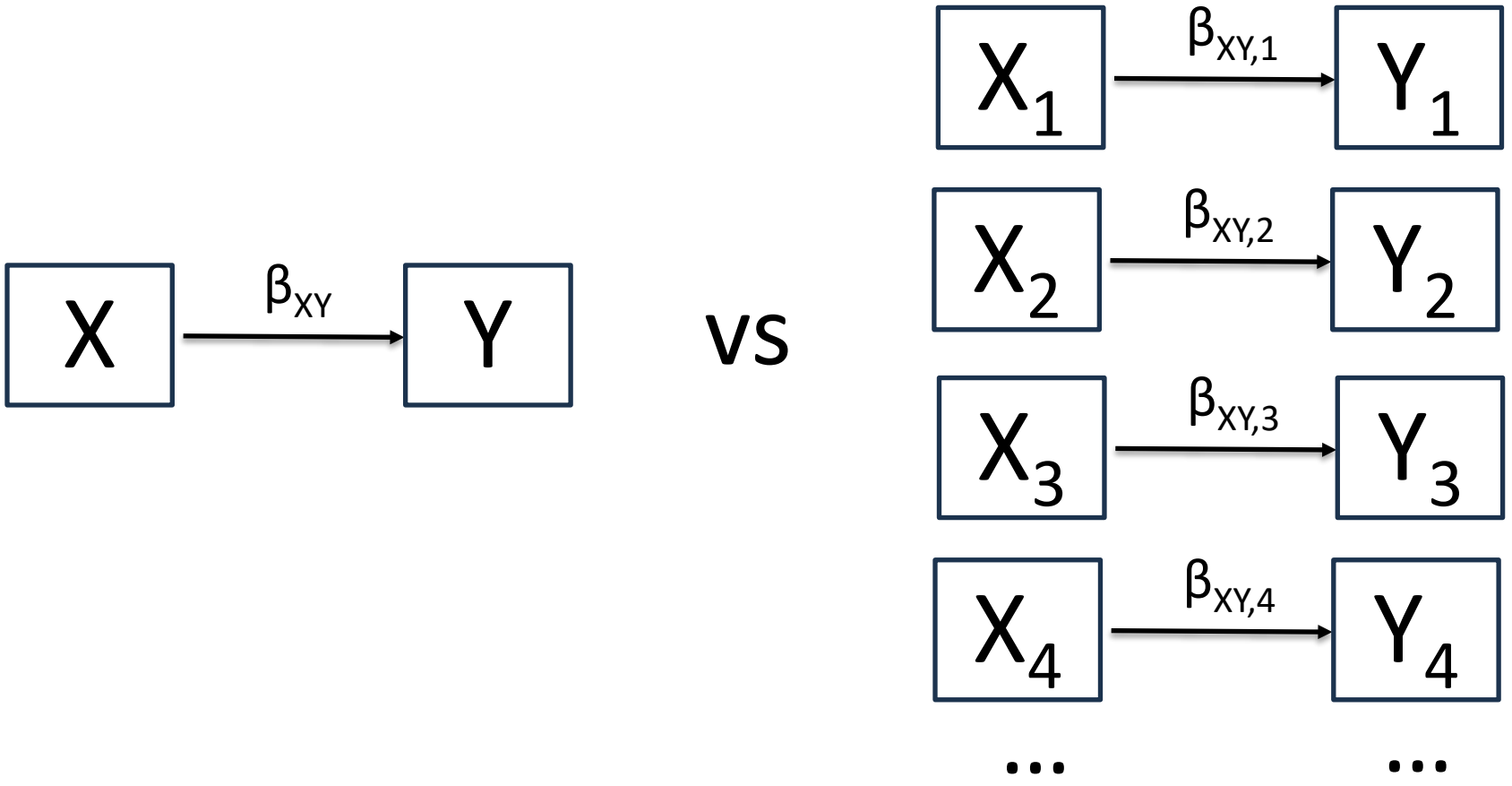
- Non-linear relationships



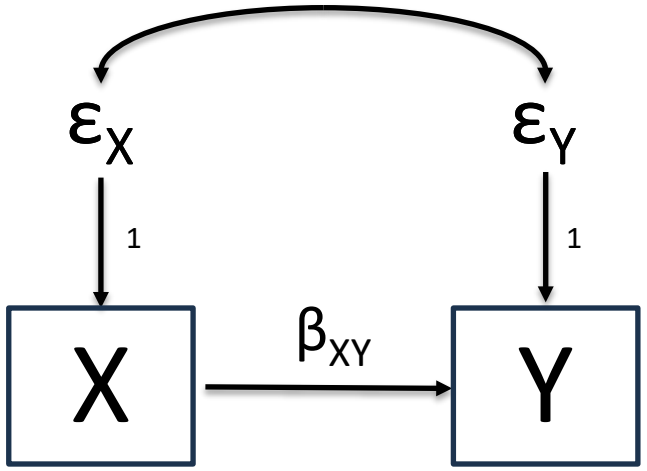
- SEM not so good. Transform?

Association

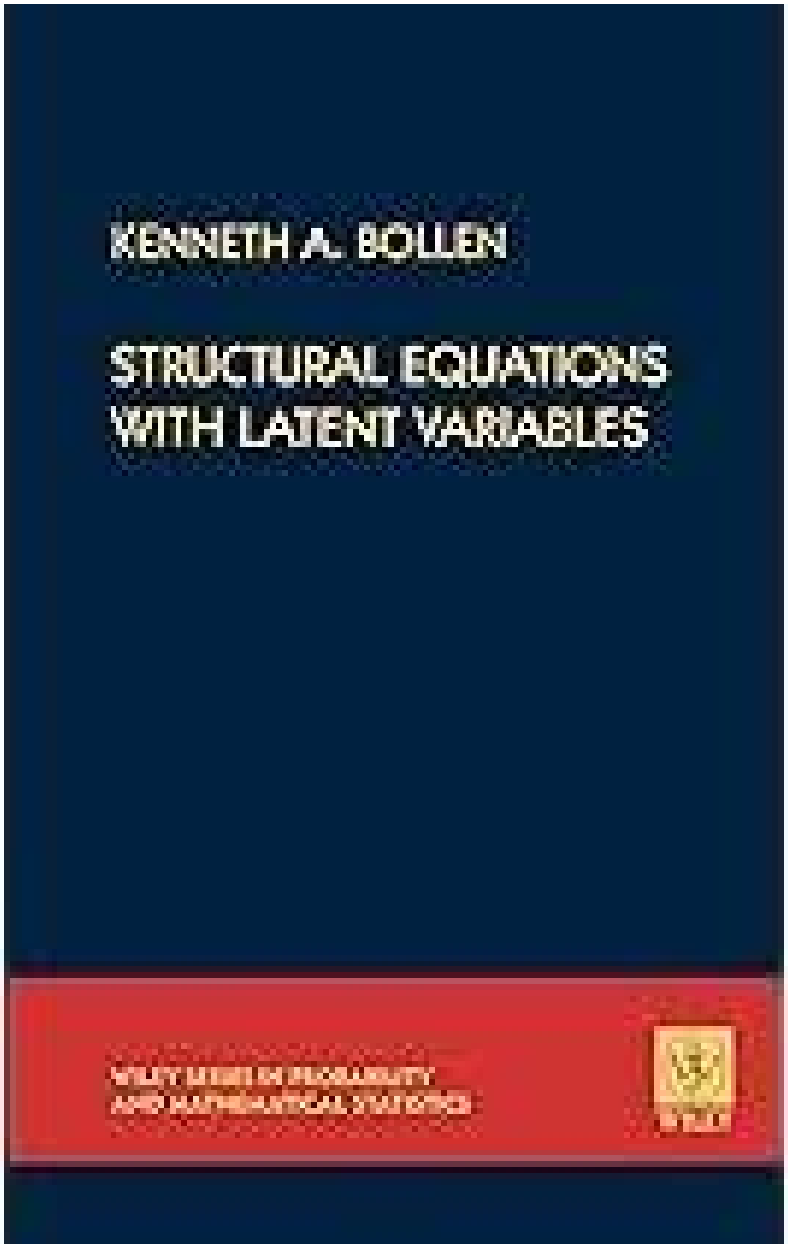
- Individual Effects vs Average Effects



(Pseudo)Isolation



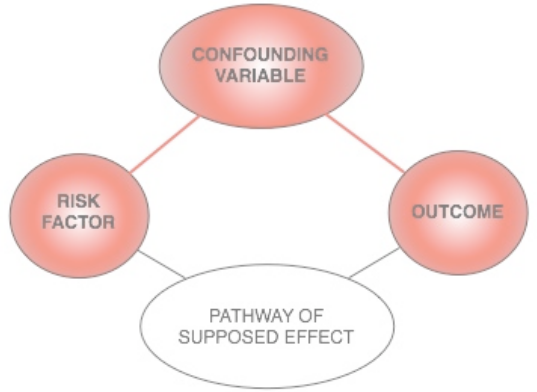
- Confounding nearly always an issue
  - Statistical control
  - Randomization
  - Instrumentation


KENNETH A. BOLLEN
STRUCTURAL EQUATIONS WITH LATENT VARIABLES
WILEY SERIES IN PROBABILITY AND MATHEMATICAL STATISTICS

The Problems with Observational Studies and the need for Randomized Controlled Trials

Classic limitations to "observational" science

- **Confounding**

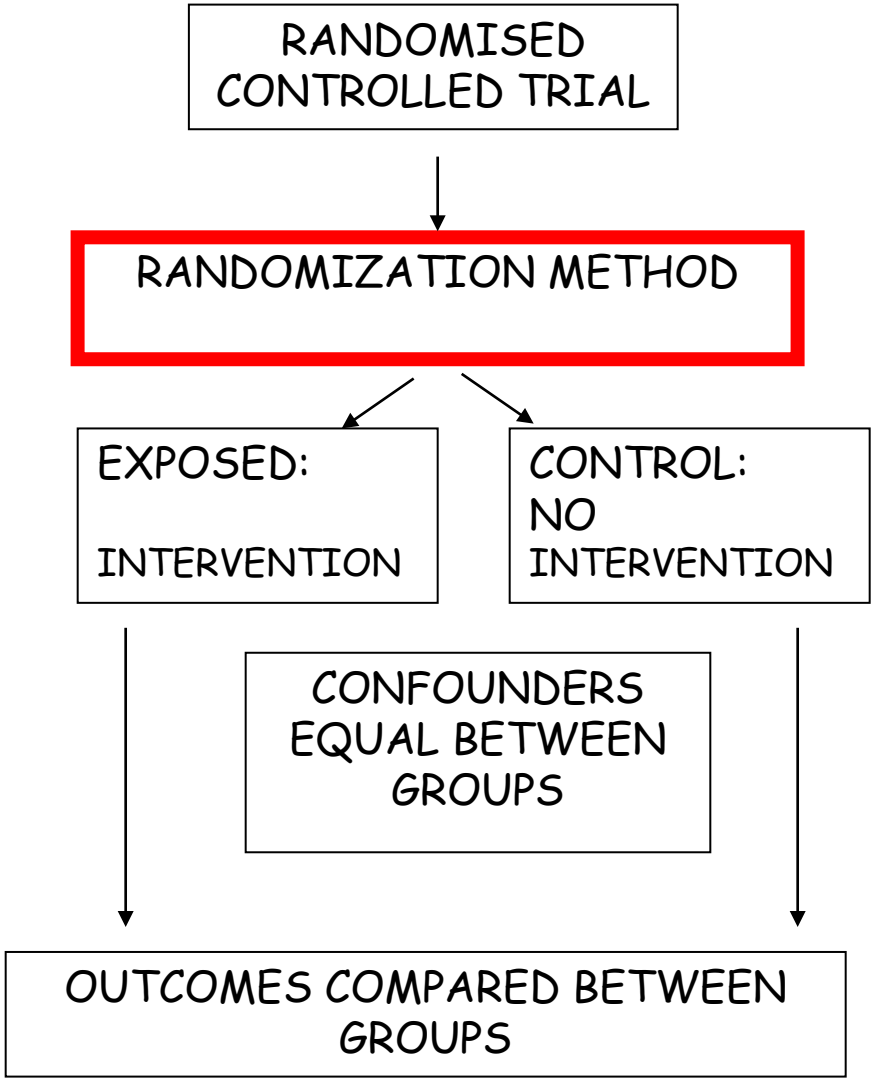


- **Reverse Causation**

- Bias

RCTs: the Gold Standard in Inferring Causality

Randomization makes causal inference possible



The Need for Observational Studies

- **Randomized Controlled Trials (RCTs):**
  - Not always ethical or practically feasible eg anything toxic
  - Expensive, requires experimentation in humans
  - Impractical for long follow up times
  - Should only be conducted on interventions that show very strong observational evidence in humans

- **Observational studies:**
  - Association between environmental exposures and disease measured in observational designs (non-experimental)
    eg case-control studies or cohort studies
  - Reliably assigning causality in these types of studies is ***very limited***

How does Mendelian randomization(MR) work?

What does MR do?

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

How does it do this?

By harnessing Mendel's laws of inheritance

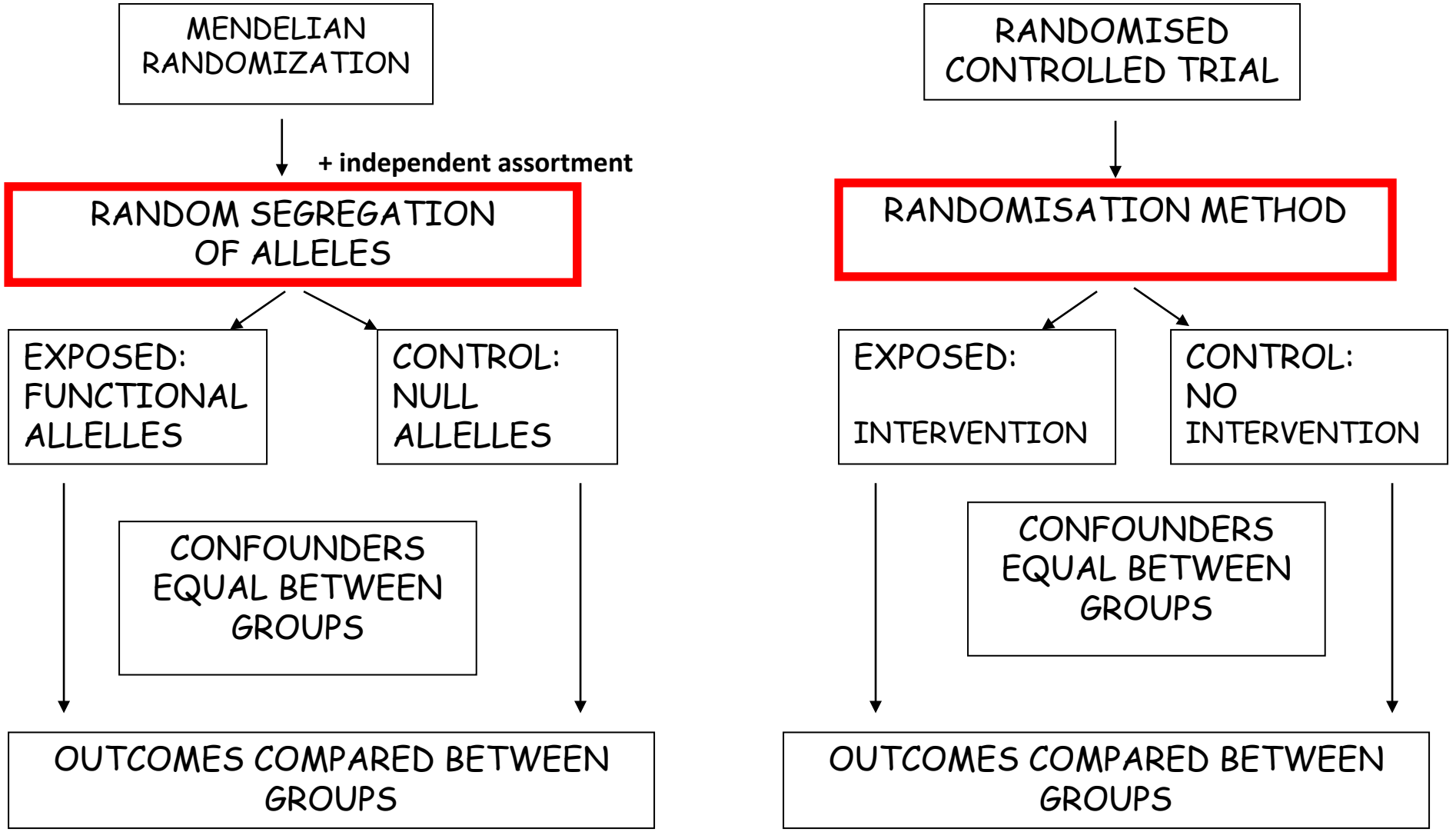
Mendel's Laws of Inheritance

Mendel in 1862

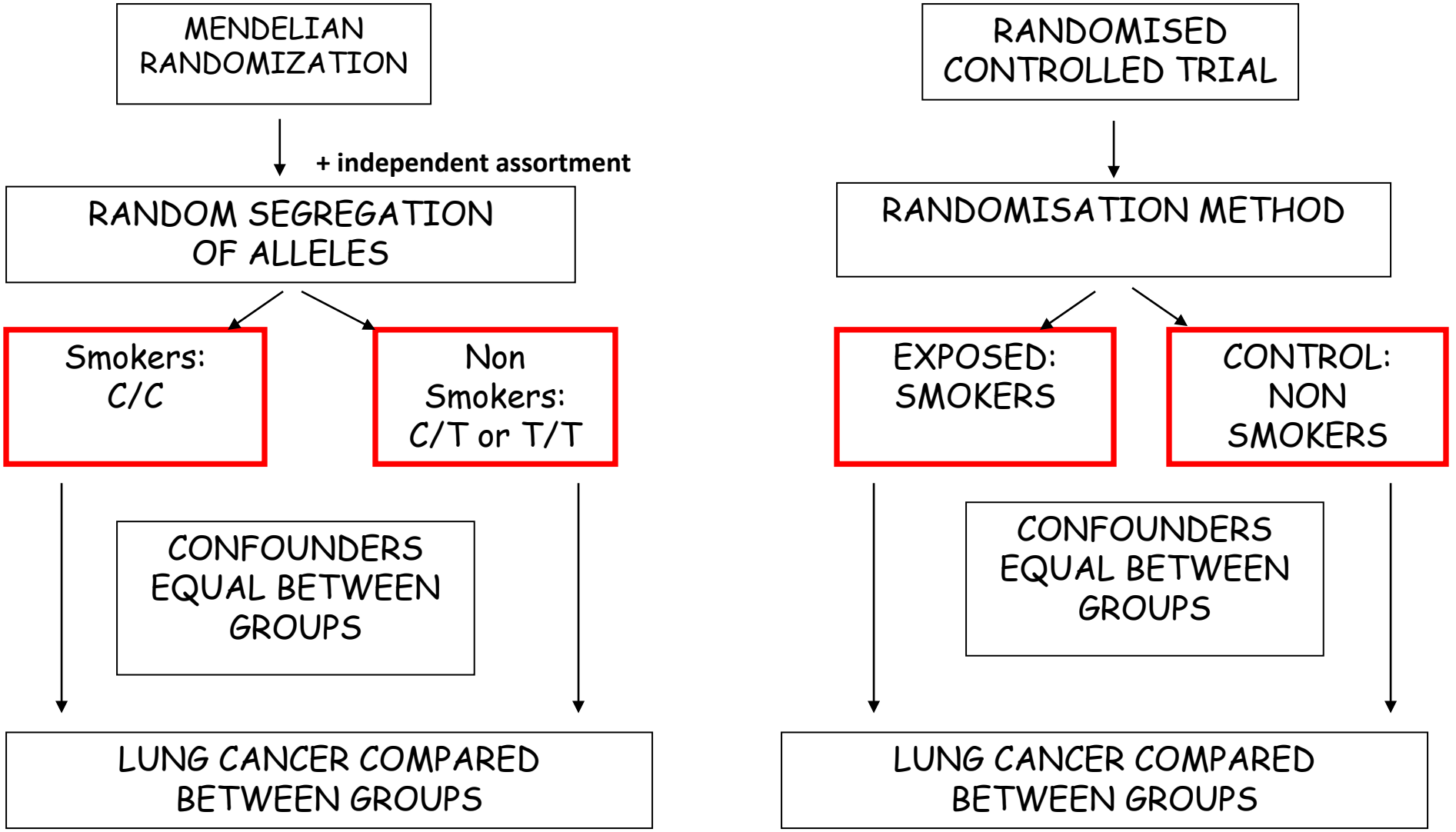
1. **Segregation:** alleles separate at meiosis and a randomly selected allele is transmitted to offspring

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

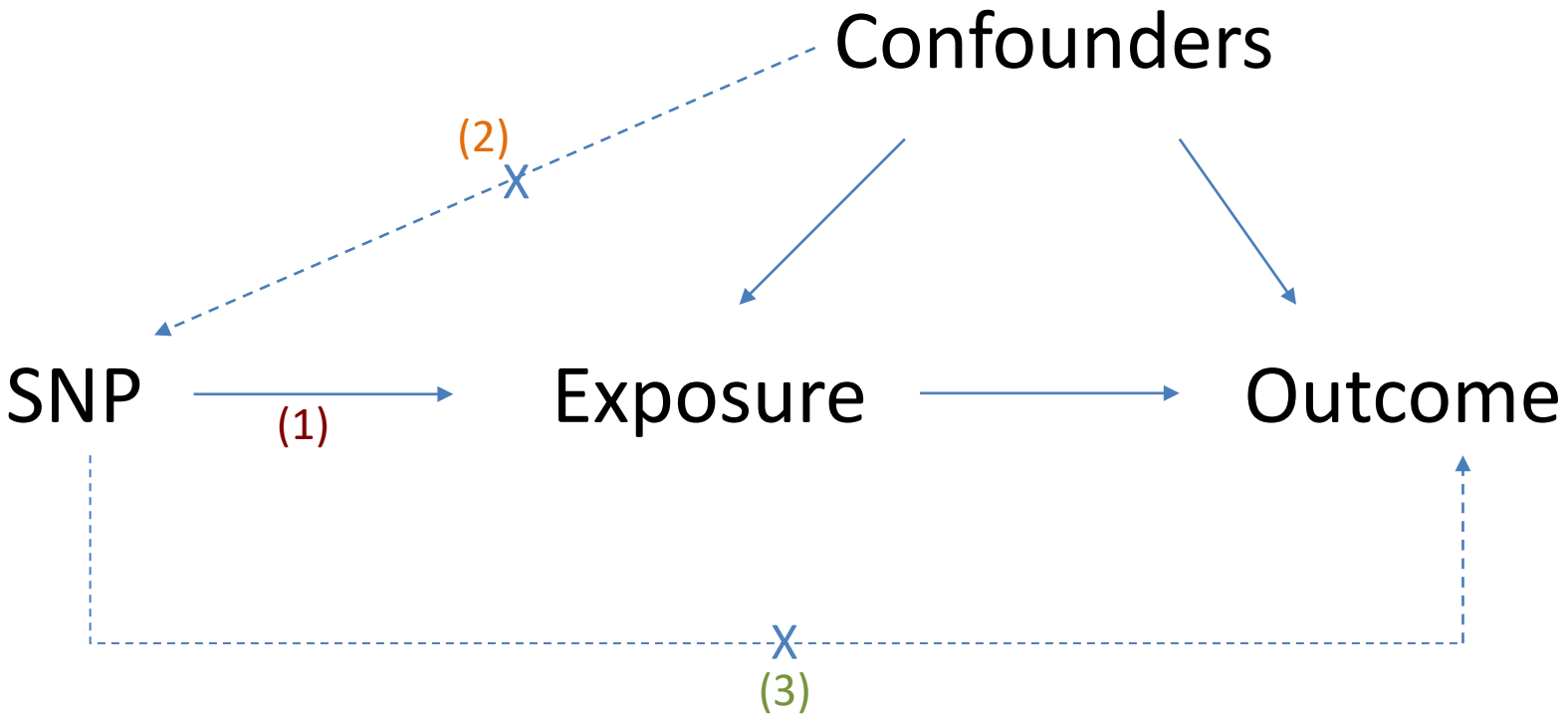
Mendelian randomization and RCTs



Mendelian randomization: Smoking and Lung Cancer



Mendelian Randomization: 3 Core Assumptions



(1) SNP is associated with the exposure

(2) SNP is NOT associated with confounding variables

(3) SNP ONLY associated with outcome through the exposure

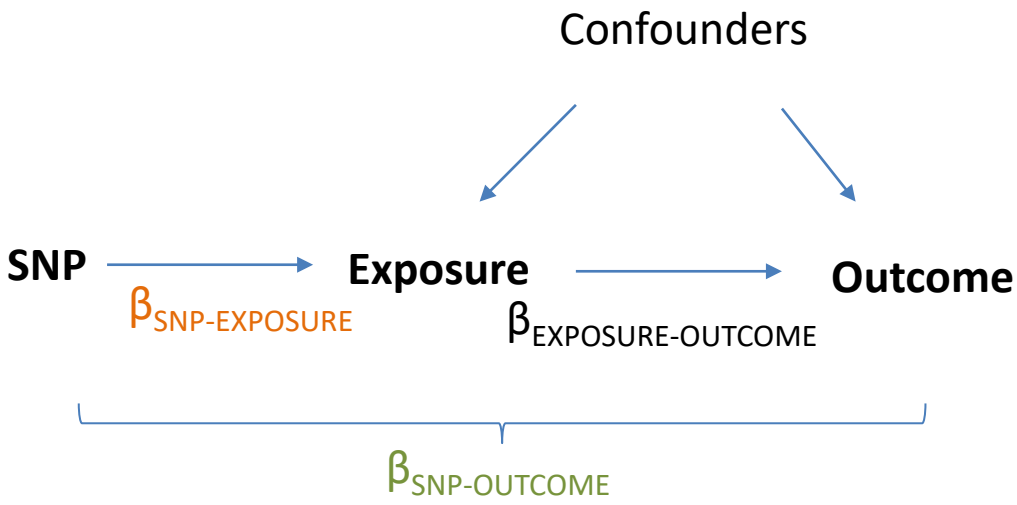
Why are genetic associations special?

- Robustness to confounding due to Mendel's laws:
  - Law of segregation: inheritance of an allele is random and independent of environment etc
  - Law of independent assortment: genes for different traits segregate independently (assuming not in LD)

- The direction of causality is known – always from SNP to trait
- Genetic variants are **potentially** very good instrumental variables
- Using genetic variants as IVs is a special case of IV analysis, known as Mendelian randomization

Calculating causal effect estimates

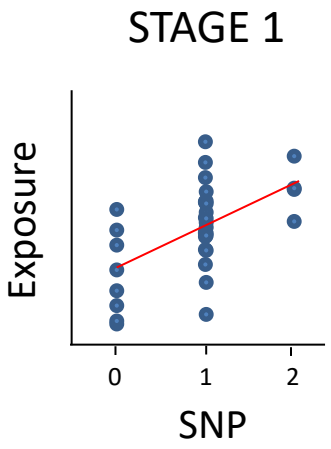
Calculating Causal Effect Estimates

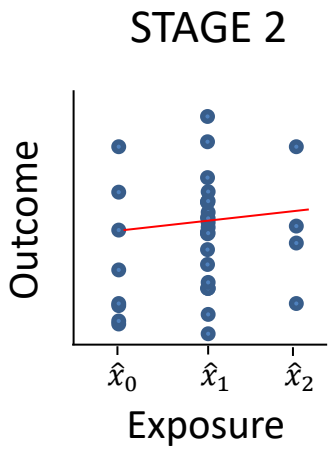


After SNP identified robustly associated with exposure of interest:

- Two-stage least-squares (TSLS) regression
- Wald Estimator

Calculating Causal Effect Estimates





(1) Regress exposure on SNP and obtain predicted values of the exposure (STAGE 1)

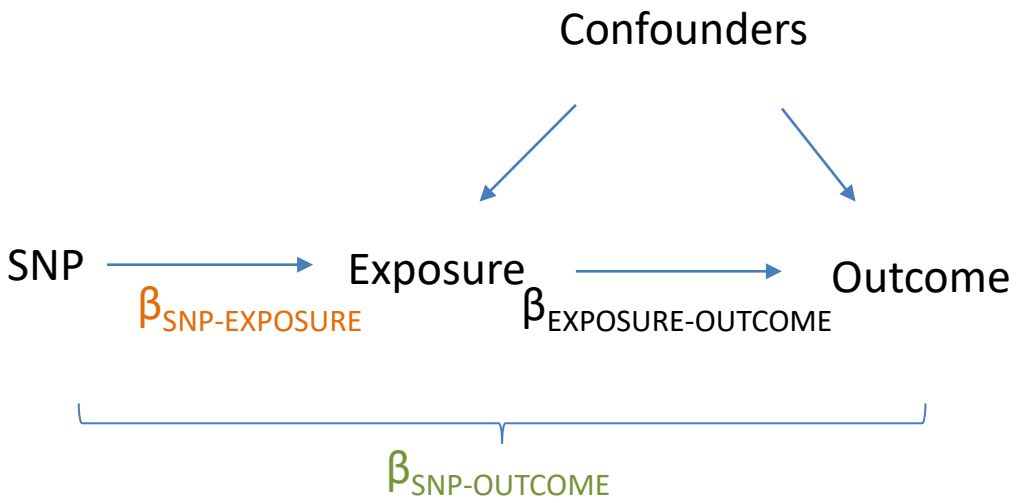
(2) Regress outcome on predicted values of the exposure (STAGE 2)

(3) Adjust standard errors

(4) Slope of 2nd stage regression is the estimate of the causal effect

*Needs to be done in the one sample ("One sample MR")

Calculating Causal Effect Estimates



Causal effect by Wald Estimator* :

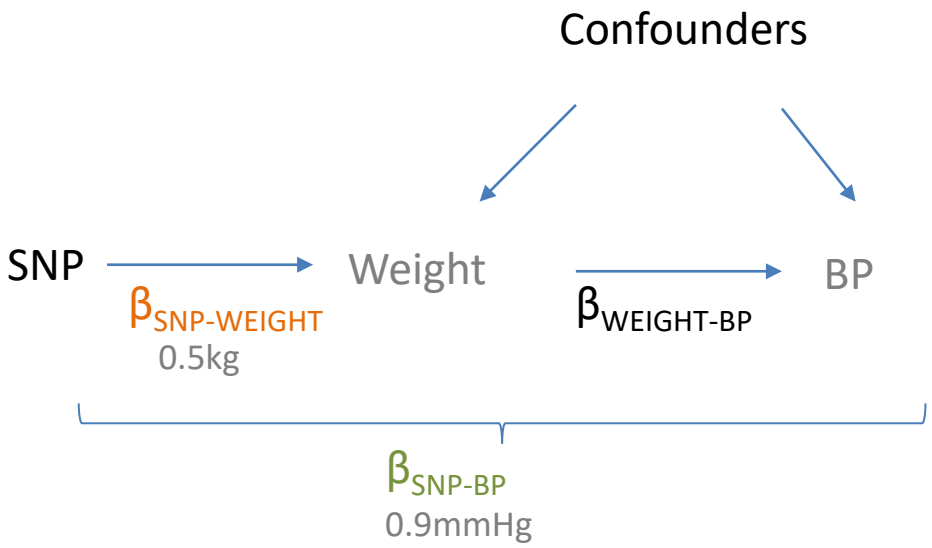
$$\frac{\hat{\beta}_{SNP\text{-}OUTCOME}}{\hat{\beta}_{SNP\text{-}EXPOSURE}}$$

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

*Can be used in different samples ("Two sample MR")

*Approximate SEs can be obtained by: $\mathrm{SE}(\beta_{\mathrm{EXPOSURE-OUTCOME}}) \cong \frac{\sigma_{\mathrm{SNP-OUTCOME}}}{\beta_{\mathrm{SNP-EXPOSURE}}}$

Calculating Causal Effect Estimates



Causal effect by Wald Estimator* :

$$\frac{\hat{\beta}_{SNP\text{-}OUTCOME}}{\hat{\beta}_{SNP\text{-}EXPOSURE}}$$

= change in outcome per unit change in exposure

BP and weight:

$$\frac{0.9 \text{ mmHg/allele}}{0.5 \text{ kg/allele}}$$

=1.8 mmHg/kg

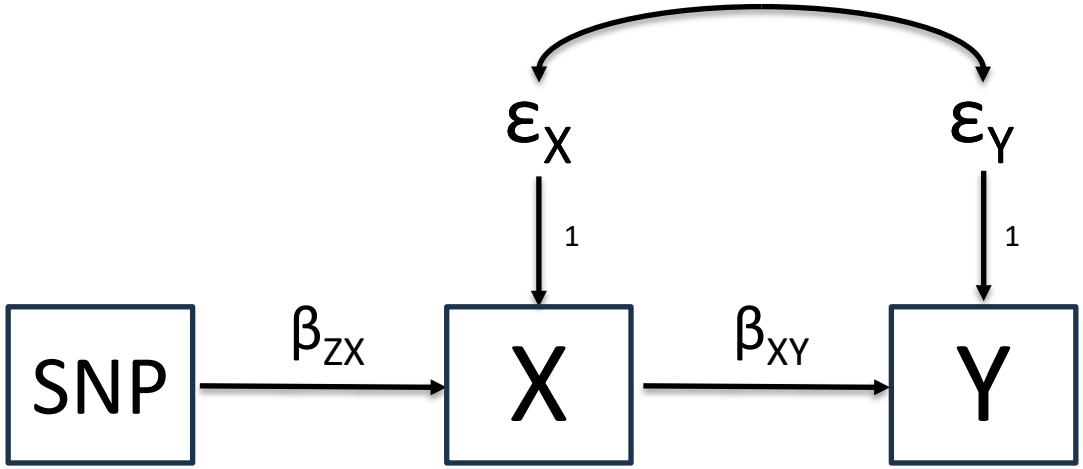
*Can be used in different samples ("Two sample MR")

*Approximate SEs can be obtained by: $\mathrm{SE}(\beta_{\mathrm{EXPOSURE-OUTCOME}}) \cong \frac{\sigma_{\mathrm{SNP-OUTCOME}}}{\beta_{\mathrm{SNP-EXPOSURE}}}$

MR can also be performed using just the results from GWAS

- Also known as two-sample MR, SMR, or MR with summary data etc
- Advantages:
  - The data is readily available, non-disclosive, free, open source
  - The exposure and outcome might not be measured in the same sample
  - The sample size of the outcome variable, key to statistical power, is not limited by requiring overlapping measures of the exposure
- Disadvantages:
  - Some extensions of MR not possible, e.g. non-linear MR, use of GxE for negative controls, various sensitivity analyses

MR as an SEM



- SNP is associated with exposure (X)
- SNP is not associated with latent confounders
- SNP only potentially influences outcome (Y) through exposure (X)

The Wide Applicability of MR

- Traditional Observational Epidemiological Studies
- Behavior Genetics and the Social Sciences
- Molecular Studies
- Pharmacogenomics

An Example using Mendelian randomization

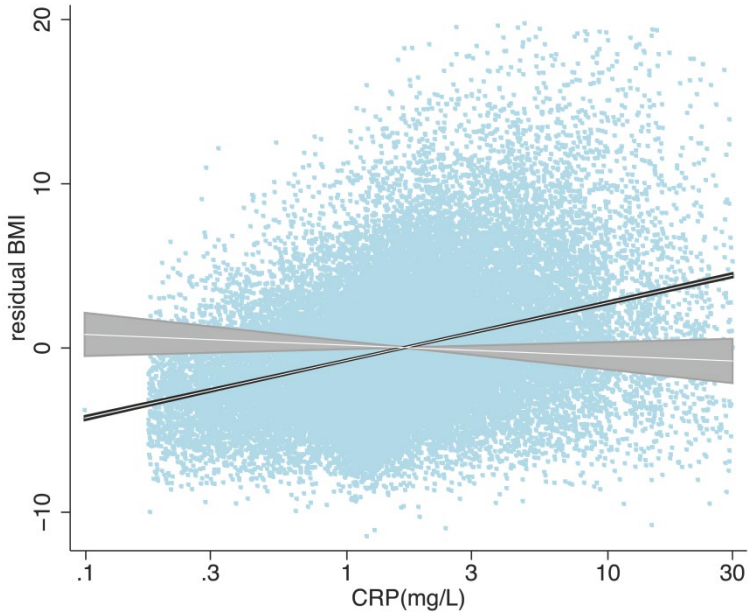
MR Example using CRP

- C-Reactive Protein (CRP) is a biomarker of inflammation
- It is associated with BMI, metabolic syndrome, CHD and a number of other diseases
- It is unclear whether these observational relationships are causal or due to confounding or reverse causality
- This question is important from the perspective of intervention and drug development

"Bi-directional Mendelian Randomization": Testing causality and reverse causation

- **NB.** Note that the CRP SNP is an excellent instrument because (1) it is strongly related to CRP levels, and (2) because it is in the CRP gene itself, its effect is more likely to be mediated through changes in the level of CRP (i.e. less potential for horizontal pleiotropy)

	Effect estimates				
	Observational	**MR Estimate**	**P_{IV}**	**P_{diff}**	**F_{first}**
CRP -> BMI	1.58 (1.53 – 1.62)	-0.30 (-0.78 – 0.18)	0.2	<0.00001	78.3



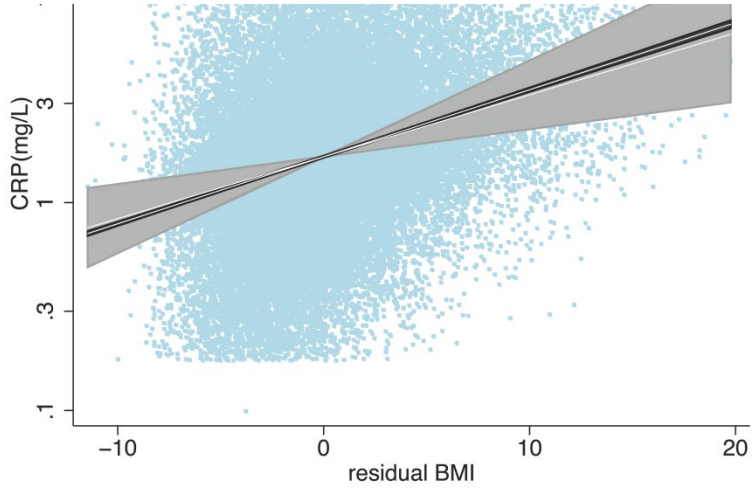
P_{IV}- Test of whether MR causal effect estimate different from zero

P_{diff}- Test of whether Observational and MR causal effect estimates are different from each other

F_{first}- Test of how strong the instrument for CRP is

“Bi-directional Mendelian Randomization”: Testing causality and reverse causation

	Effect estimates				
	Observational	**MR Estimate**	**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



P_{IV}- Test of whether MR causal effect estimate different from zero

P_{diff}- Test of whether Observational and MR causal effect estimates are different from each other

F_{first}- Test of how strong the instrument for BMI is

Limitations to Mendelian randomization

Some Limitations to Mendelian Randomization

1- Population stratification

2- The existence of instruments

3- Power and "weak instrument bias"

4- Pleiotropy

Power and Weak Instruments

- 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:
  - **Single Sample MR:** to the confounded estimate
  - **Two-Sample MR:** to the null

Using Multiple Genetic Variants as Instruments

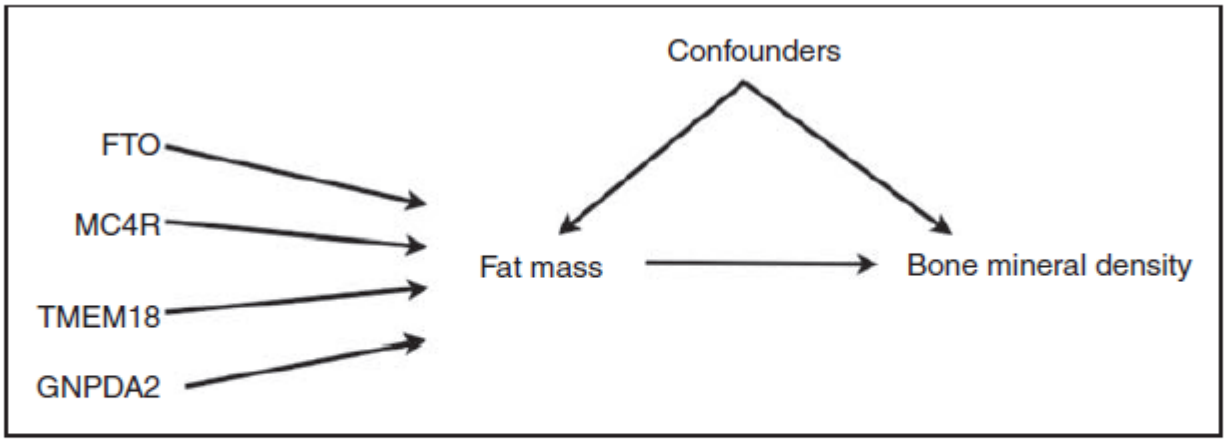
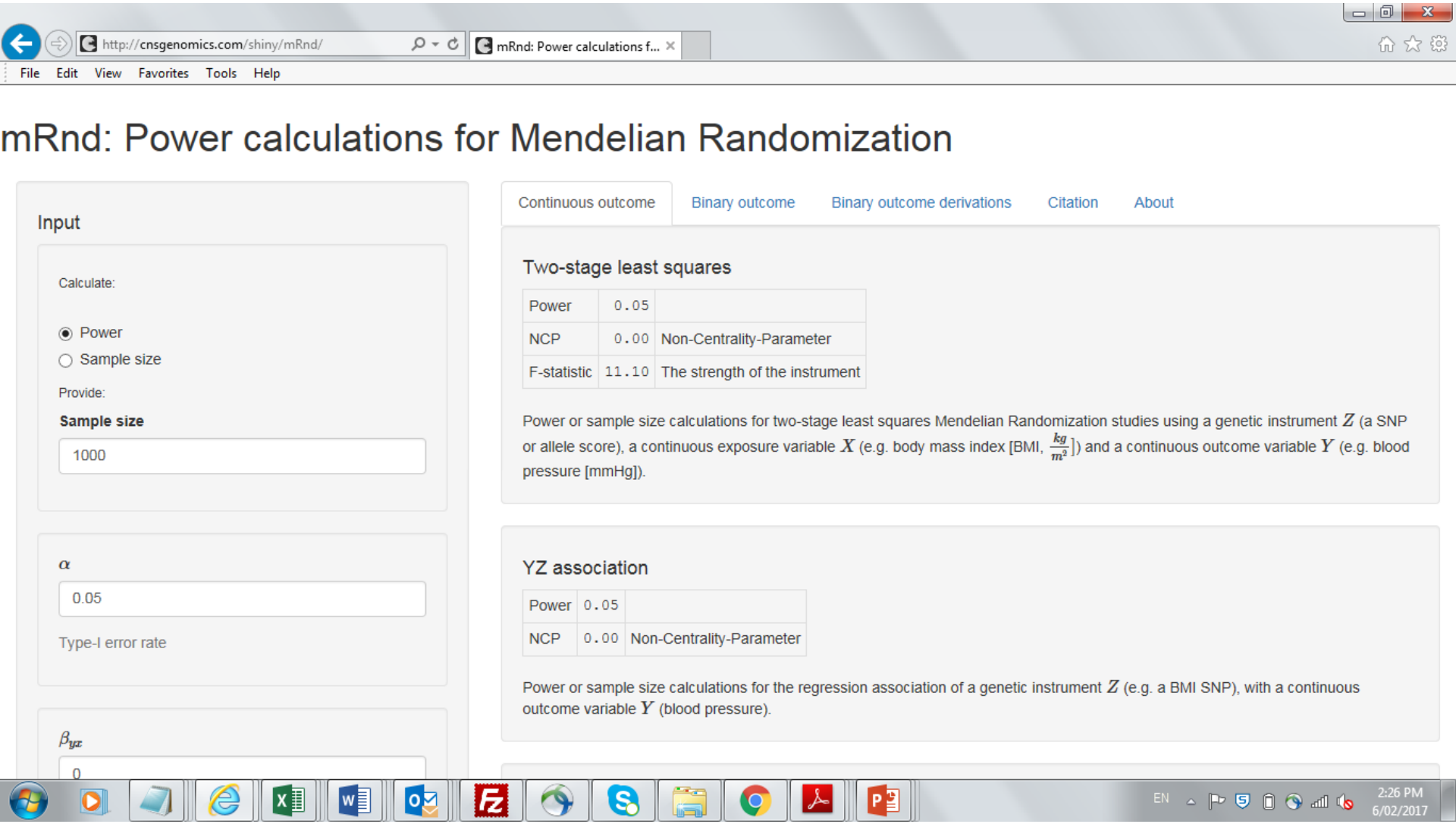


Figure 1. DAG for a Mendelian randomisation analysis using four genetic variants as instrumental variables for the effect of fat mass on bone mineral density.

Palmer et al (2011) Stat Method Res

- Allelic scores
- Testing multiple variants individually
- Meta-analyse individual SNPs

Calculating Power in Mendelian Randomization Studies



Some Limitations to Mendelian Randomization

1- Population stratification

2- The existence of instruments

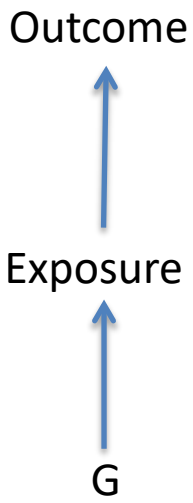
3- Power and “weak instrument bias”

4- Pleiotropy

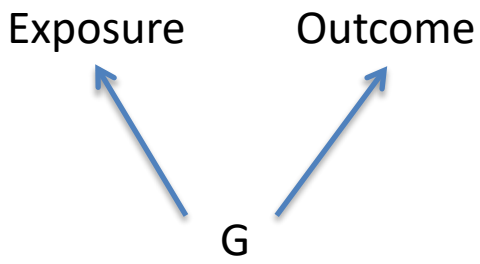
Pleiotropy

- Genetic variant influences more than one trait

- Horizontal vs Vertical pleiotropy



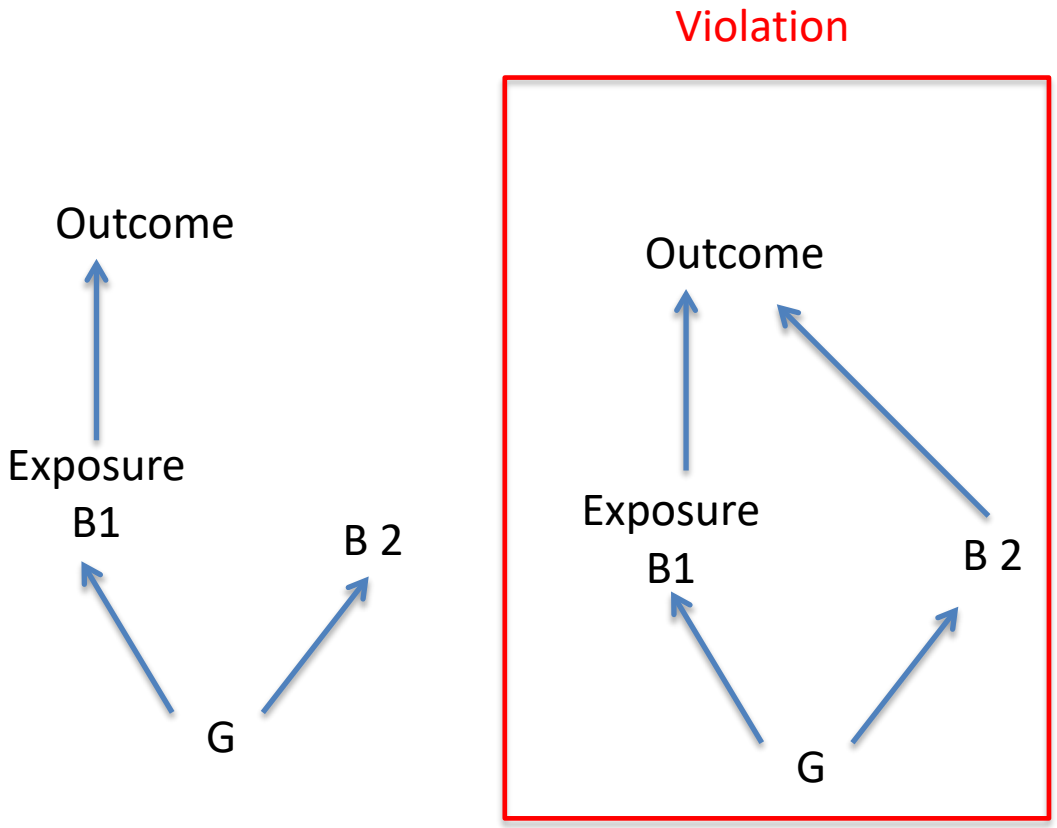
Vertical Pleiotropy



Horizontal Pleiotropy

Pleiotropy

- Genetic variant influences more than one trait
- Pleiotropy only violates MR's assumptions if it involves a pathway outside that of the exposure and is a pathway that affects your outcome



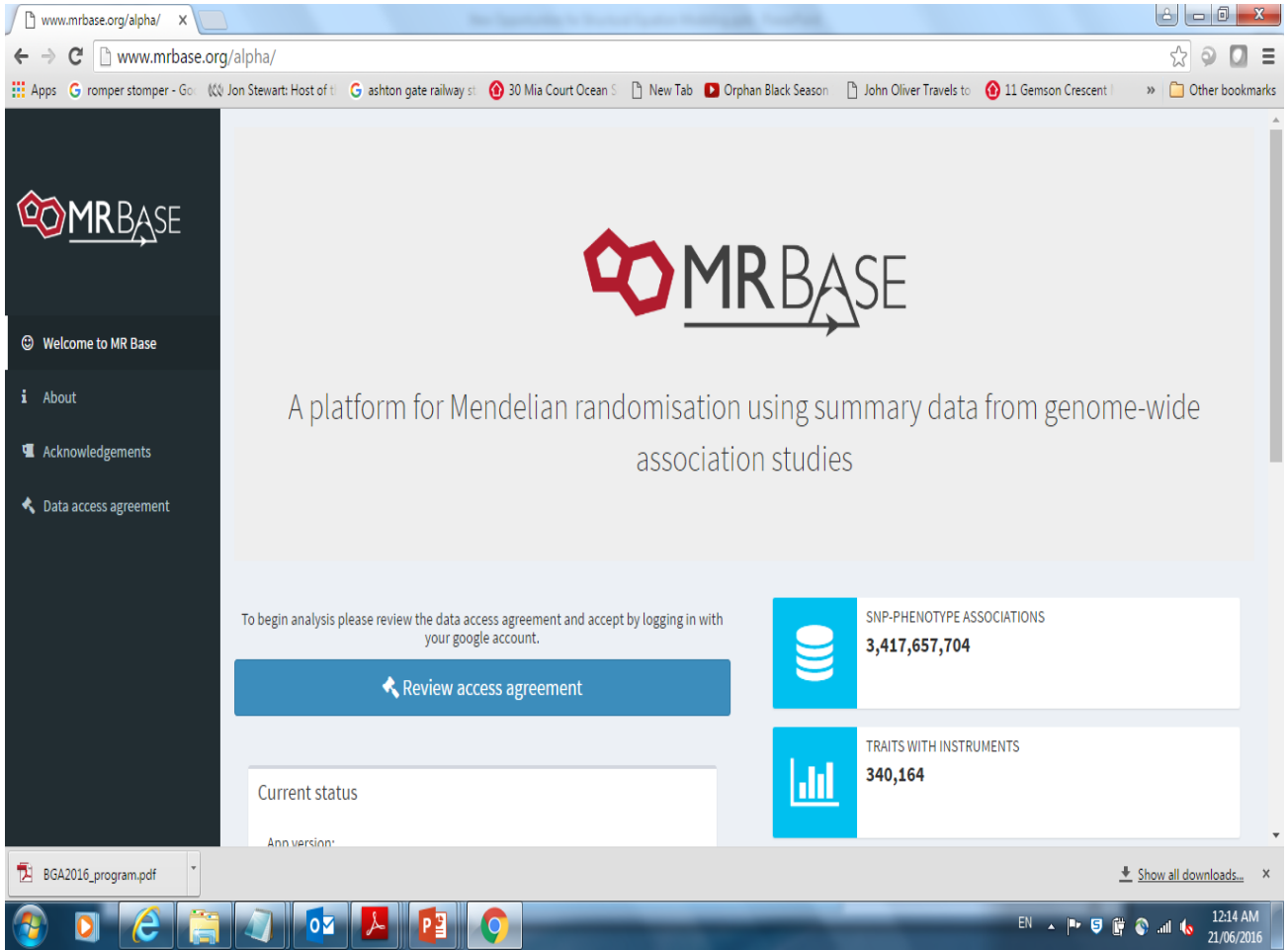
MR Base

http://www.mrbase.org/

Jie “Chris” Zheng

Gib Hemani

Phil Haycock



Triangulation in Science

Repeating experiments is not enough

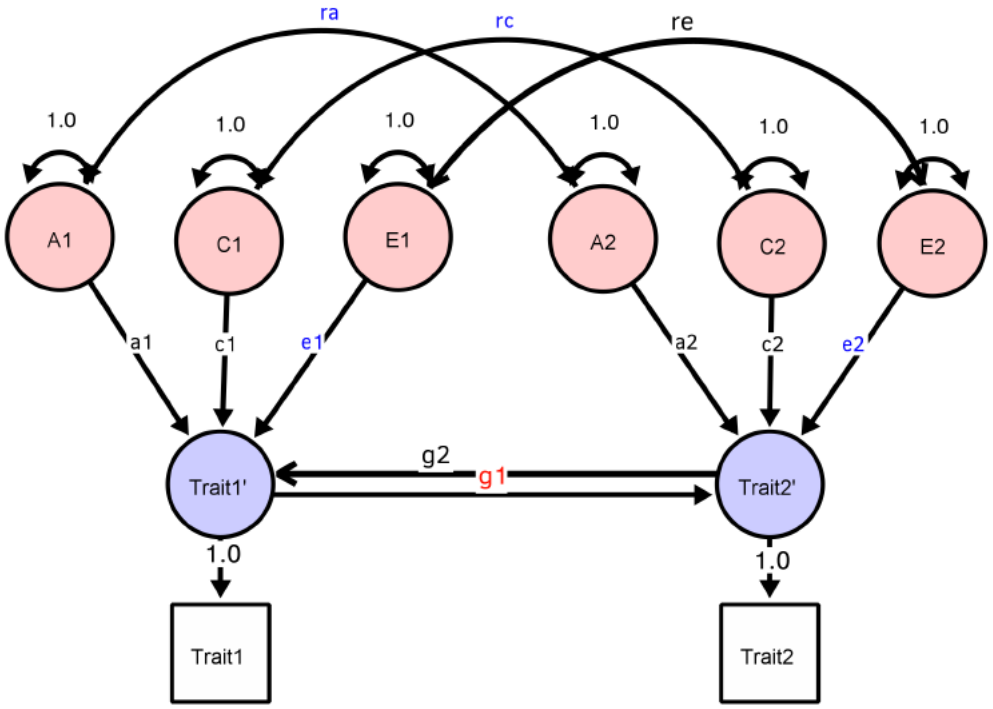
Verifying results requires disparate lines of evidence — a technique called triangulation. **Marcus R. Munafò** and **George Davey Smith** explain.

Nature (2018)

Useful References

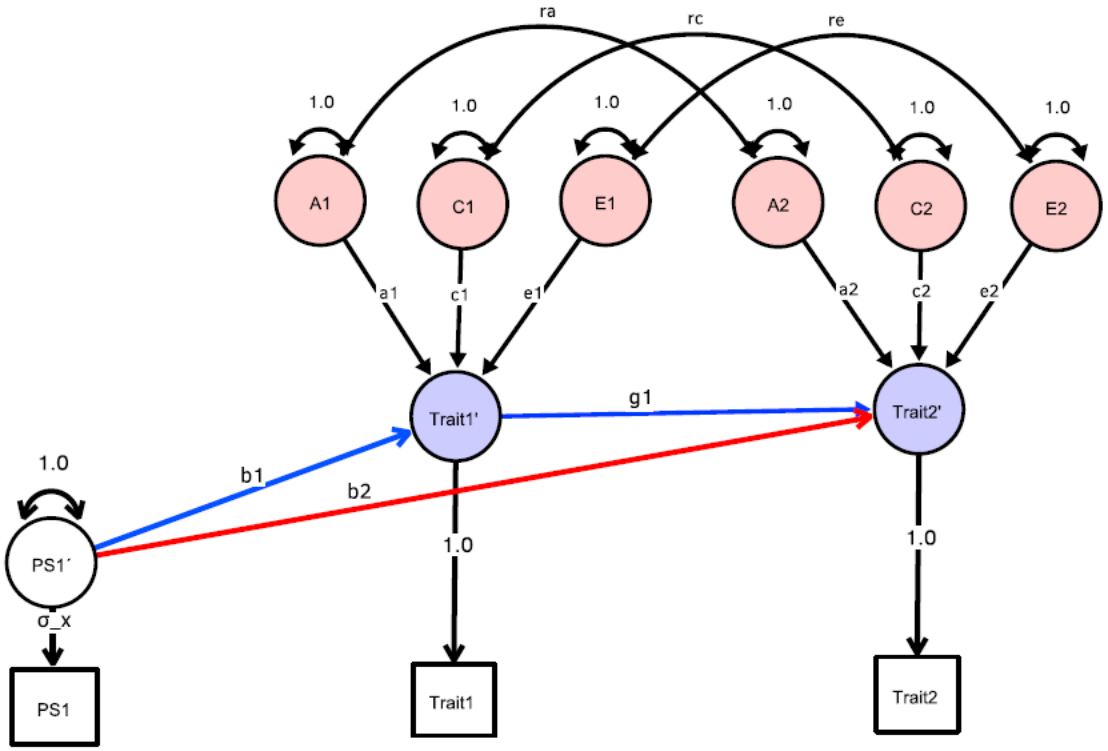
- Brion et al (2013). Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol*, 42(5), 1497-501.
- Davey-Smith & Hemani (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*, 23(1), R89-98.
- Davey-Smith & Ebrahim (2003). "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? IJE, 32, 1-22.
- Evans & Davey-Smith (2015). Mendelian randomization: New applications in the coming age of hypothesis free causality. Annu Rev Genomics Hum Genet, 16, 327-50.
- Hemani et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. Elife, May 30, 7, e34408.
- Munafo & Davey Smith (2018). Robust research needs many lines of evidence. Nature, 553(7689), 399-401.
- Zheng et al. (2017). Recent developments in Mendelian randomization studies. Curr Epidemiol Rep, 4(4), 330-345.

Classic Direction of Causation Model

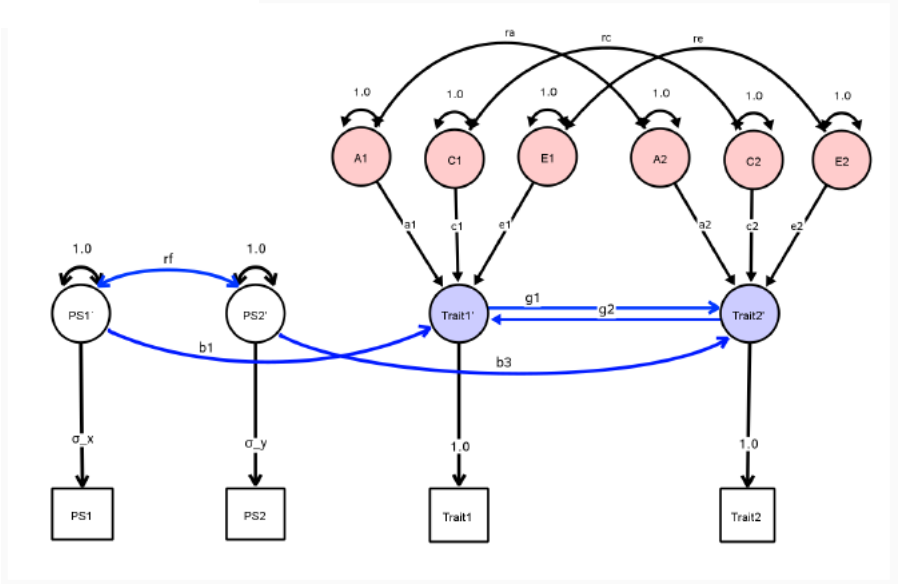


Madur Singh

MR Direction of Causation Model



Luis Silva Castro de Araujo






THANK YOU!