Introduction to Mendelian Randomization:

Using genes to inform causality

David Evans

The University Of Queensland AUSTRALIA | DIAMANTINA INSTITUTE | TRI partnering for better health | University of BRISTOL | MRC Integrative Epidemiology Unit | Children of the 90s | wellcome trust

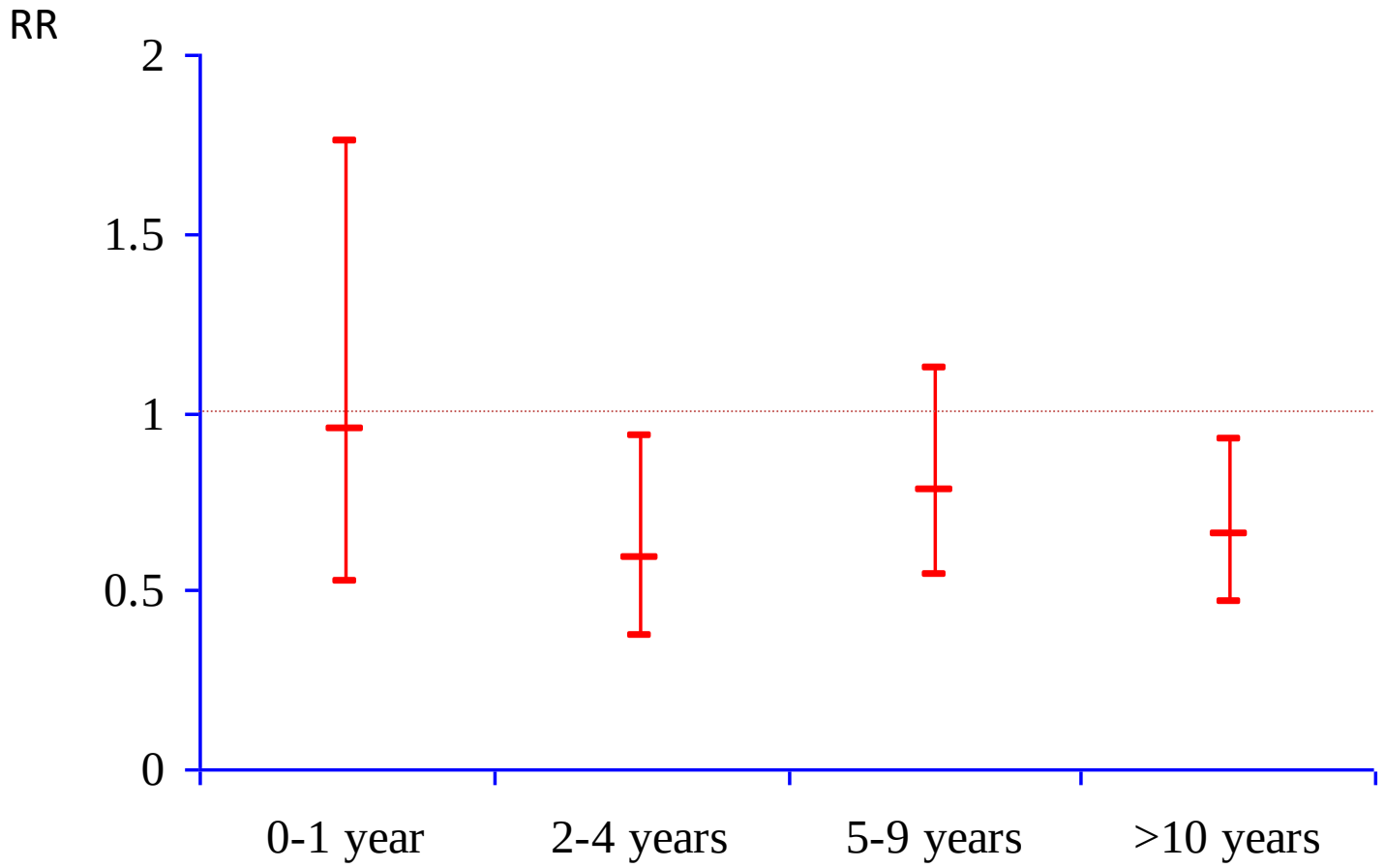
This Session

- Problems with observational data
- Randomized controlled trials
- Mendelian Randomization (MR):
  - How it works
  - Core assumptions
  - Calculating causal effect estimates
- MR example
- Limitations of MR

Problems with inferring causality in observational studies

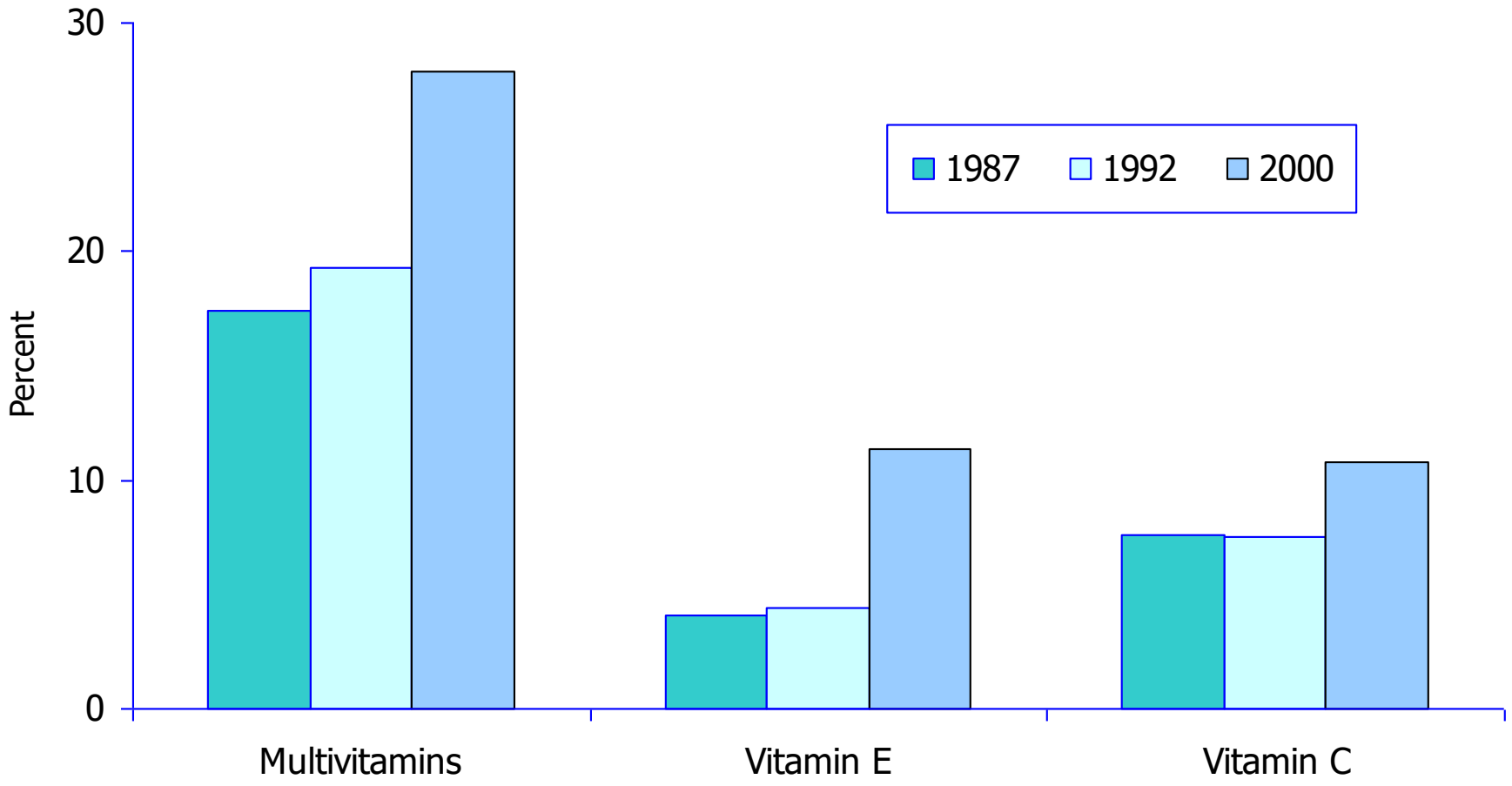
The Problem with Inferring Causality in Observational Studies

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



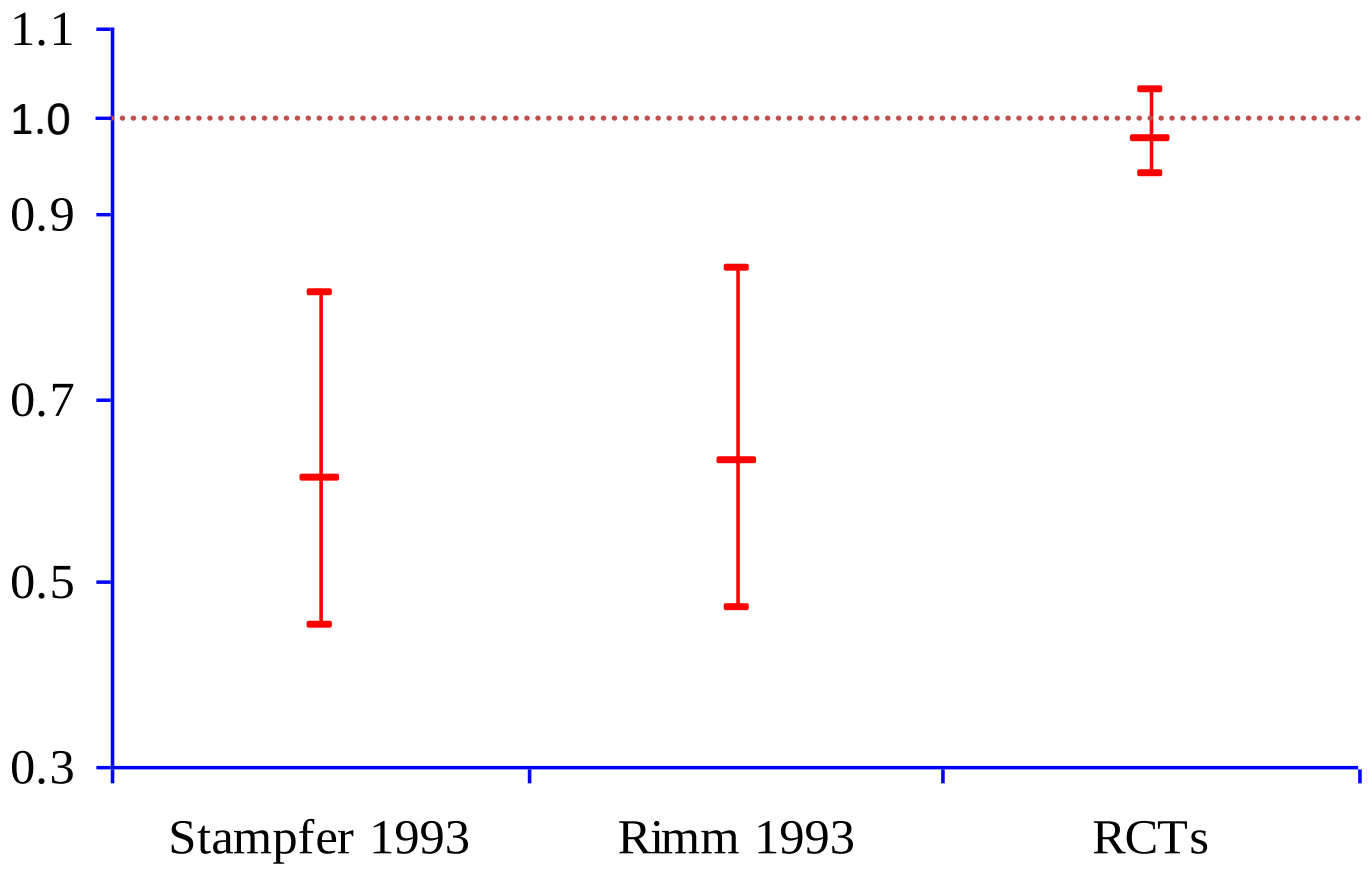
Rimm et al NEJM 1993; 328: 1450-6

Use of vitamin supplements by US adults, 1987-2000



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

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

MANY OTHER EXAMPLES

VITAMIN C, VITAMIN A, HRT,
MANY DRUG TARGETS…….

WHAT'S THE EXPLANATION?

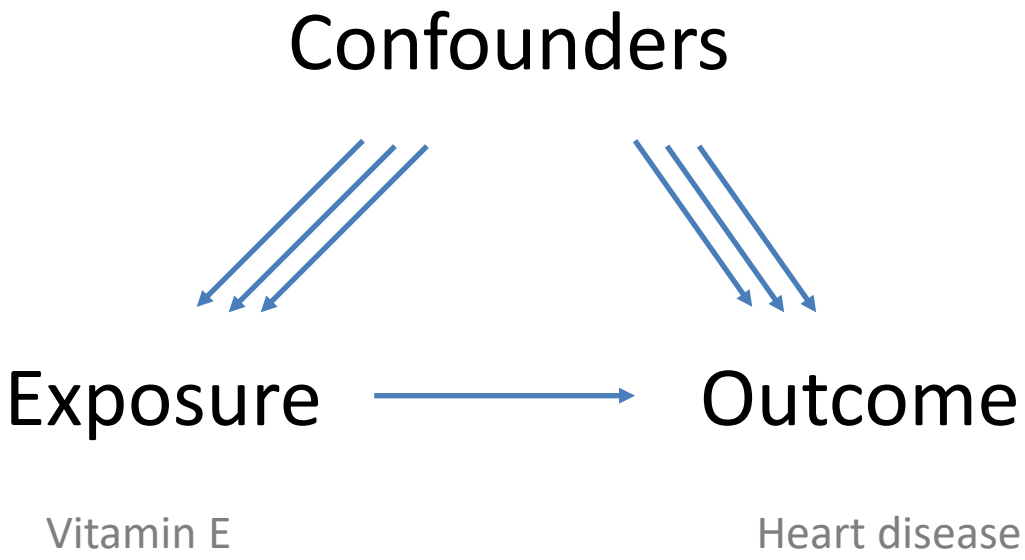
Vitamin E levels and confounding risk factors:

Childhood SES	↓
Manual social class	↓
No car access	↓
State pension only	↓
Smoker	↓
Obese	↓
Daily alcohol	↑
Exercise	↑
Low fat diet	↑
Height	↑
Leg length	↑

Women's Heart and Health Study
Lawlor et al, Lancet 2004

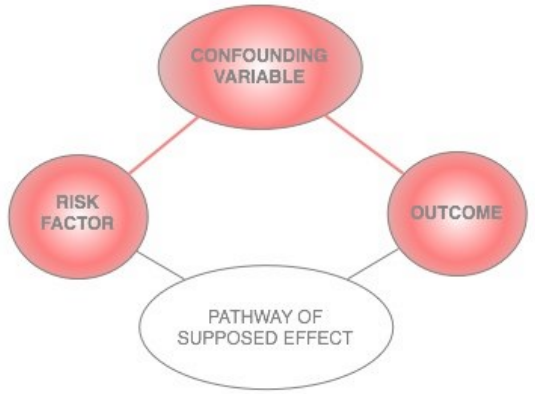
Confounding

Smoking, diet, alcohol, socioeconomic position….

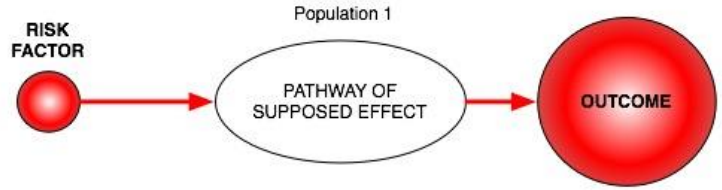


Classic limitations to "observational" science

- **Confounding**



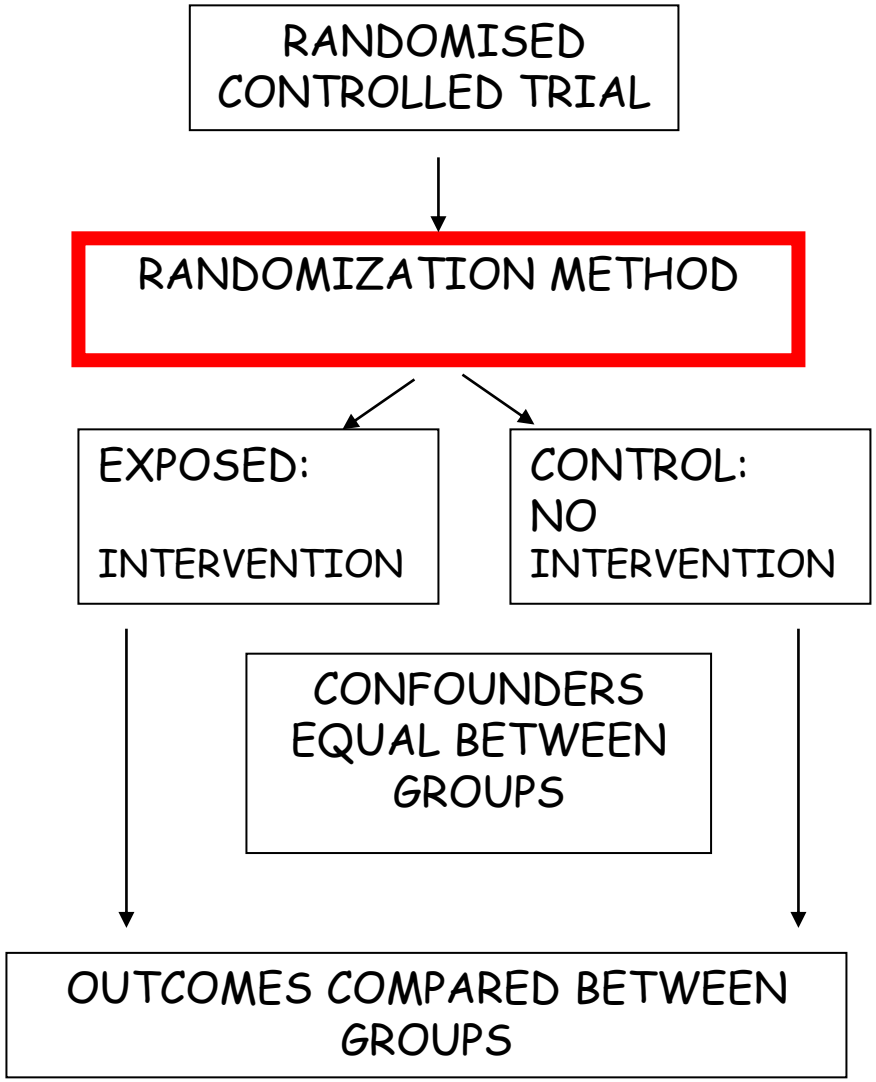
- **Reverse Causation**



- 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***

The Wide Applicability of MR

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

How does Mendelian randomization 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

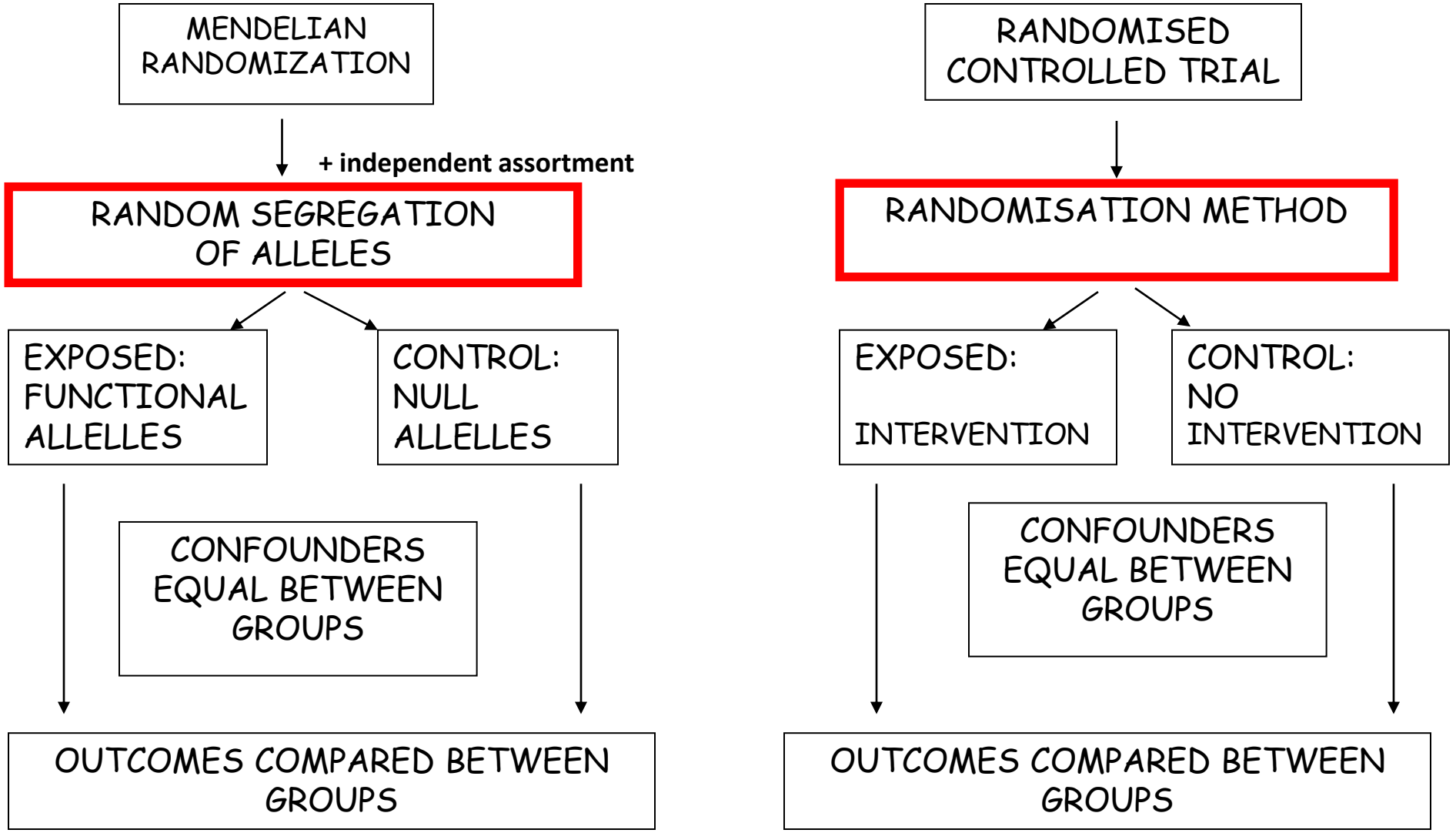
Mendel's Laws of Inheritance

Mendel in 1862

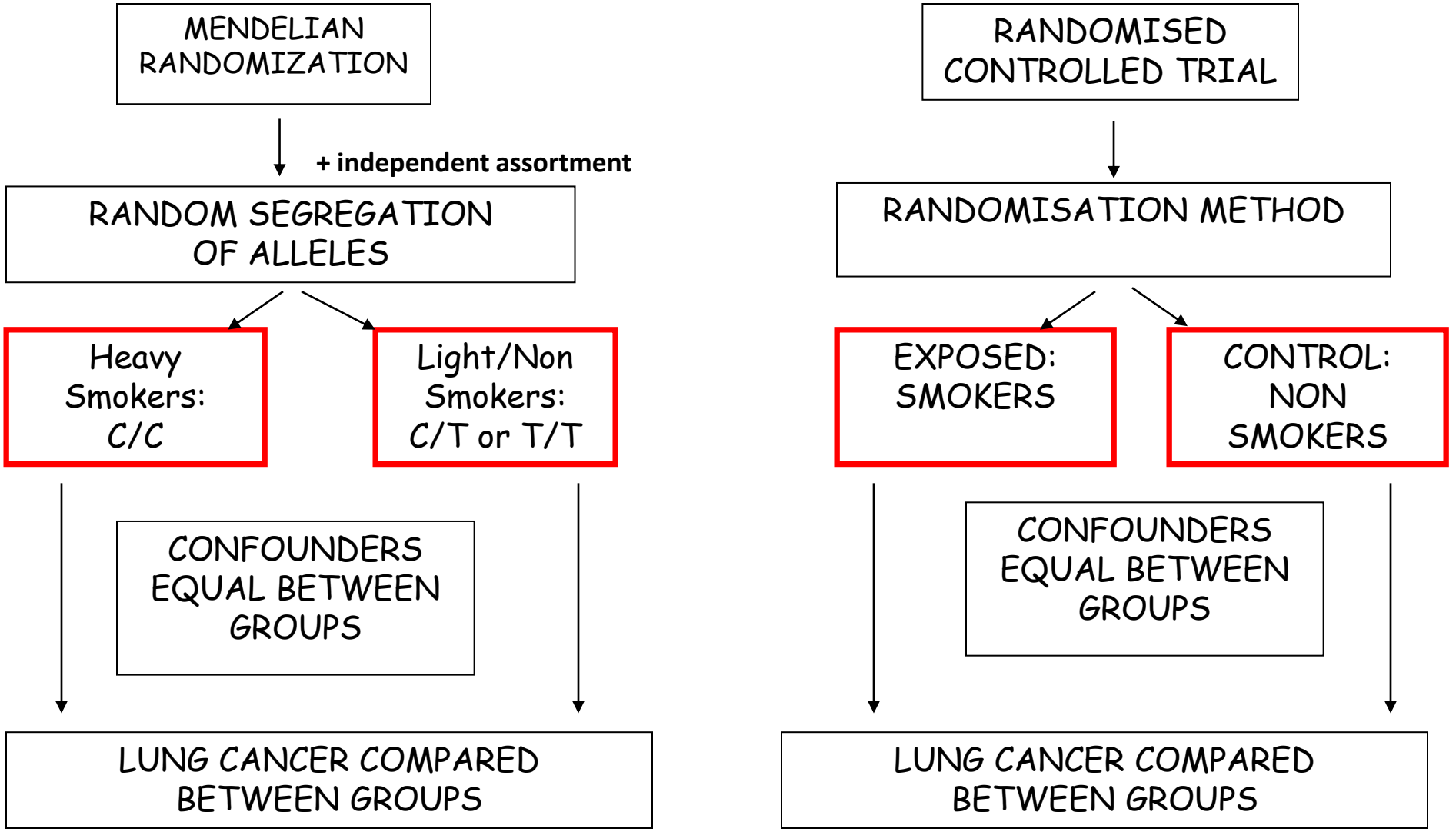
1. **Segregation:** alleles separate at meiosis and a randomly selected allele is transmitted to offspring

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

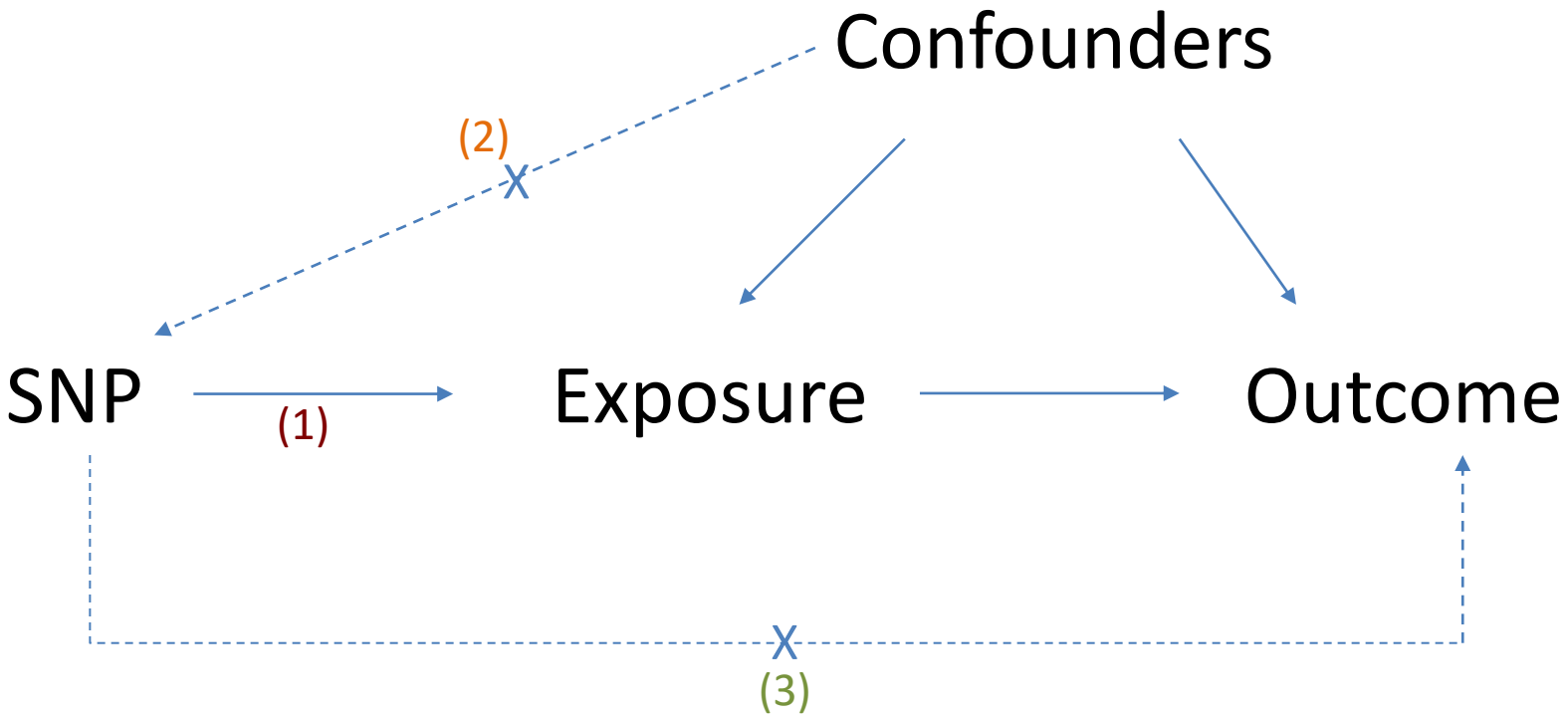
Mendelian randomization and RCTs



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

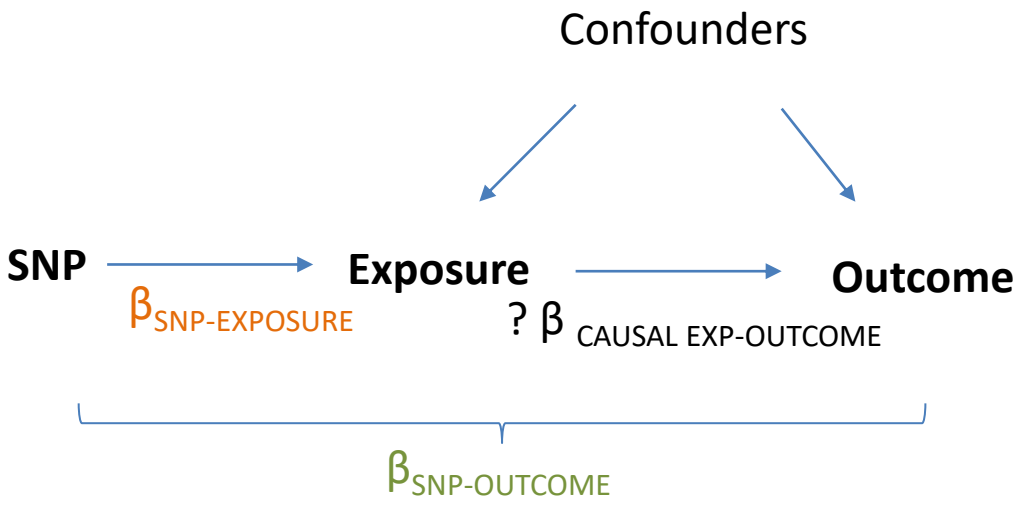
Why are genetic associations special?

- 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

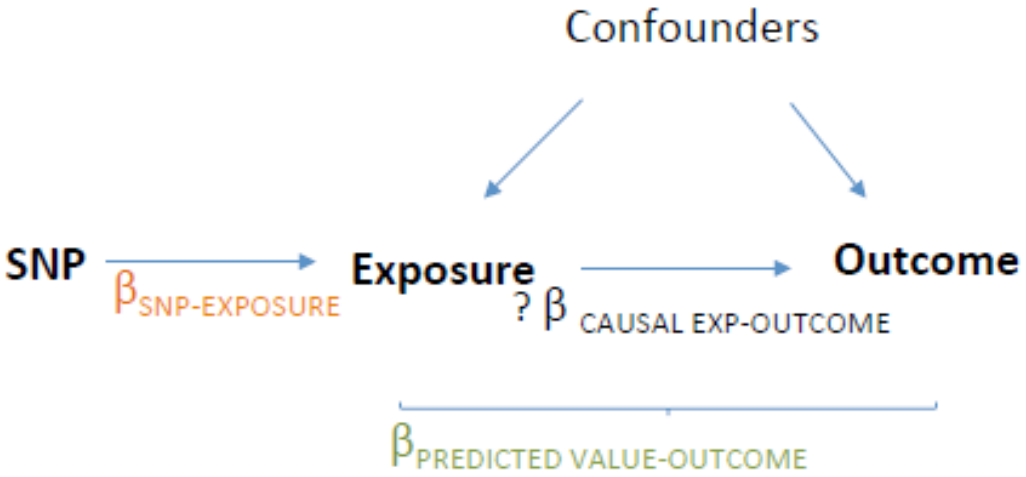
Calculating Causal Effect Estimates

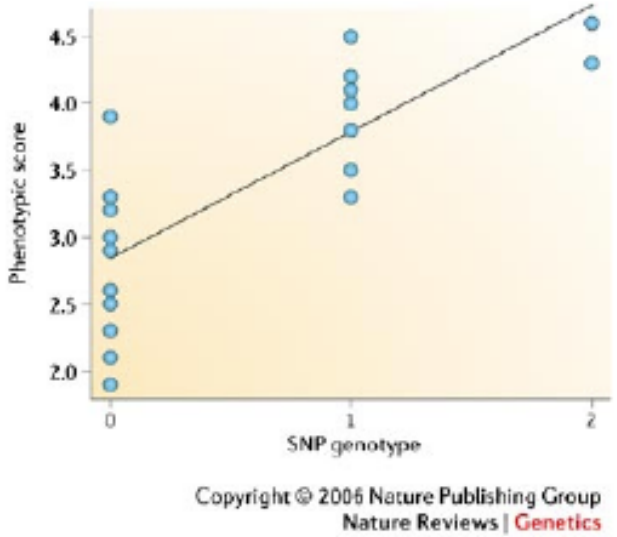


After SNP identified robustly associated with exposure of interest:

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

Calculating Causal Effect Estimates





Two-stage Least Squares (2SLS):

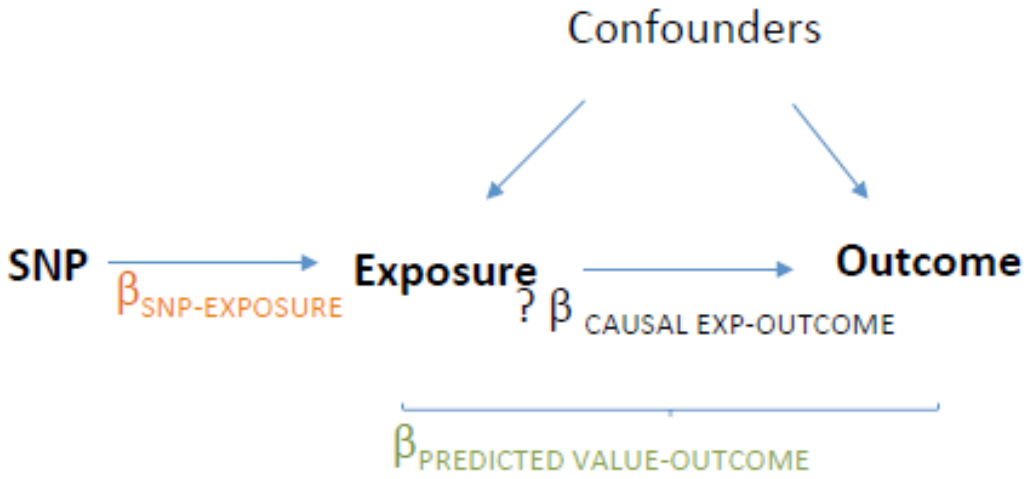
(1) Regress exposure on SNP & obtain predicted values

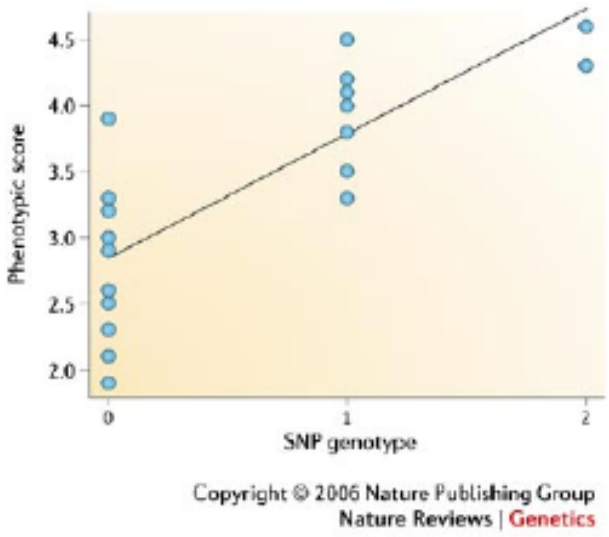
(2) Regress outcome on **predicted** exposure (from 1st stage regression)

(3) Adjust standard errors

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

Calculating Causal Effect Estimates





Two-stage Least Squares (2SLS):

(1) Regress exposure on SNP & obtain predicted values

(2) Regress outcome on **predicted** exposure (from 1st stage regression)

(3) Adjust standard errors

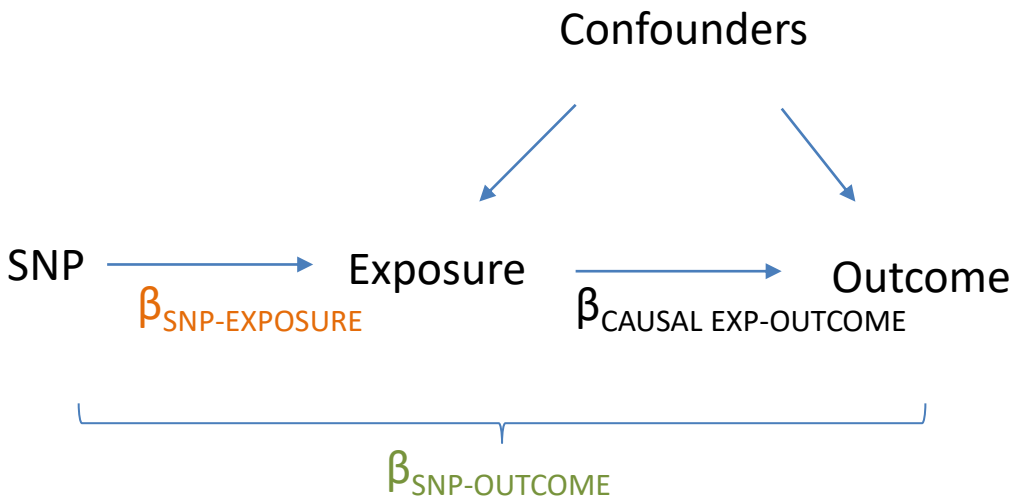
This gives you: difference in outcome per unit change in (genetically-predicted) exposure

Genetically determined exposure → "randomized" → can ascribe causality

(if assumptions are met)

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

Calculating Causal Effect Estimates



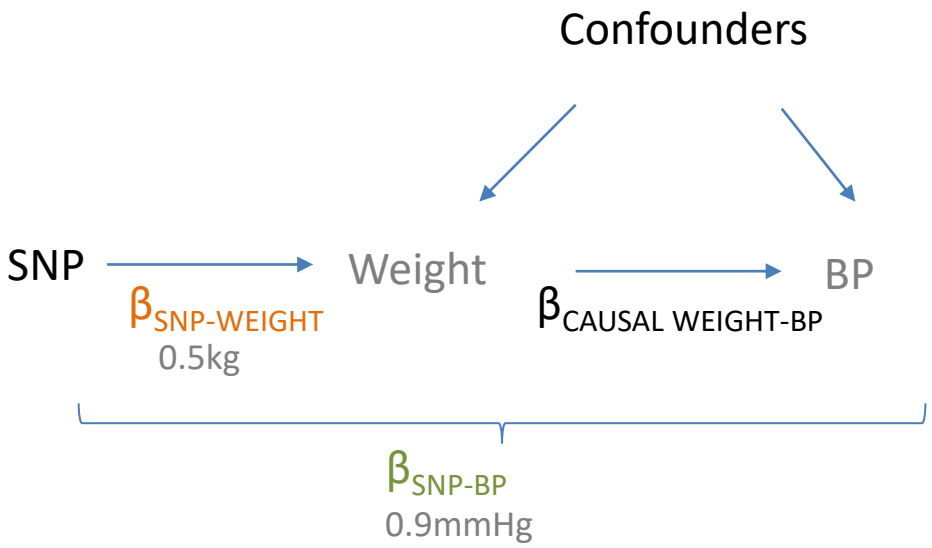
Causal effect by Wald Estimator* :

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

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

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

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:

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

=1.8 mmHg/kg

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

MR can also be performed using just the results from GWAS

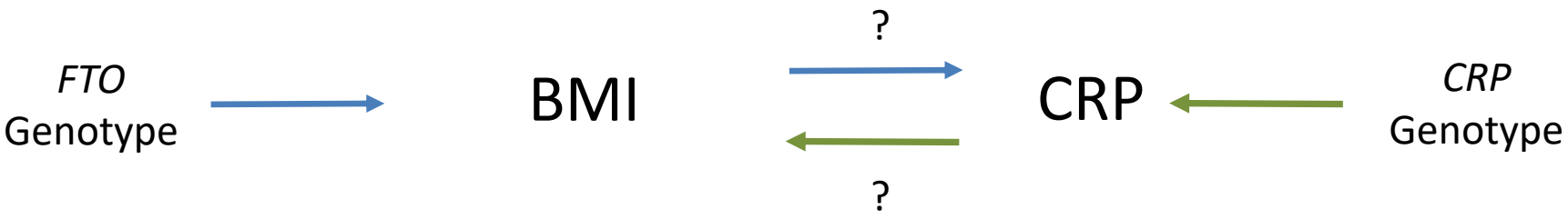
- Also known as two-sample MR, SMR, or MR with summary data etc
- Advantages:
  - The data is readily available, non-disclosive, free, open source
  - The exposure and outcome might not be measured in the same sample
  - The sample size of the outcome variable, key to statistical power, is not limited 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

An Example using Mendelian randomization

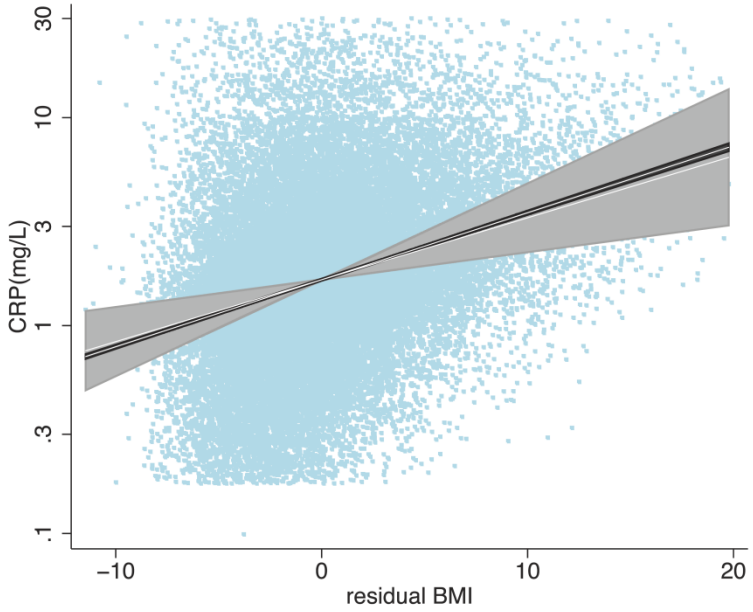
MR Example using CRP

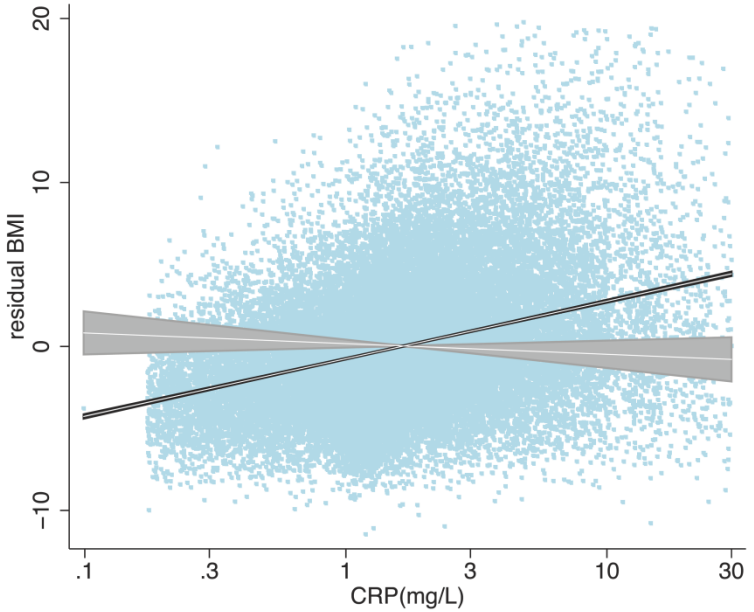
- C-Reactive Protein (CRP) is a biomarker of inflammation
- It is associated with BMI, metabolic syndrome, CHD and a number of other diseases
- It is unclear whether these observational relationships are causal or due to confounding or reverse causality
- This question is important from the perspective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

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



	Effect estimates				
Outcome / explanatory variable	**Observational**	**Instrumental variable**	P_{IV}	P_{diff}	F_{first}
CRP/BMI	1.075 (1.073, 1.077)	1.06 (1.02, 1.11)	0.002	0.6	50.2





Limitations to Mendelian randomization

Limitations to Mendelian Randomization

1- Population stratification

2- Canalisation (“Developmental compensation”)

3- The existence of instruments

4- Power and “weak instrument bias”

5- 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

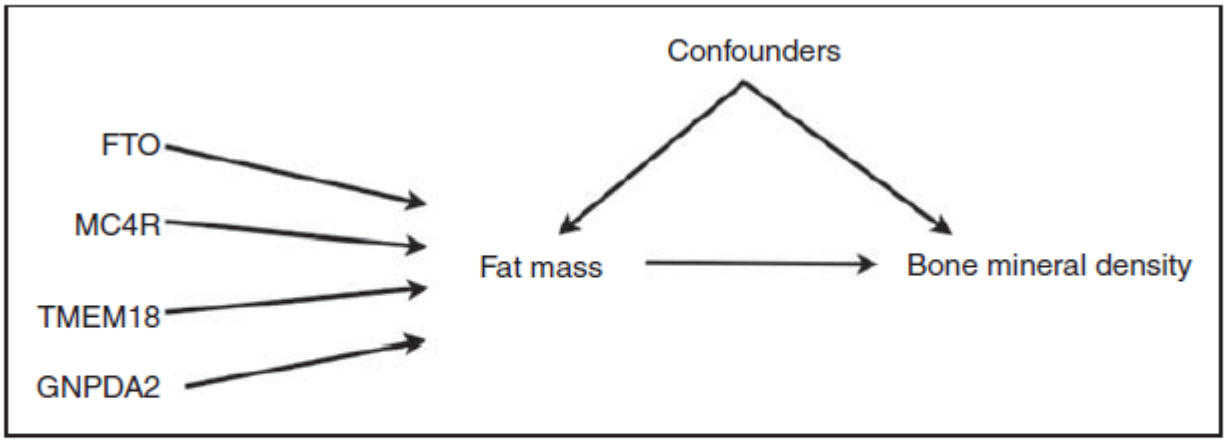
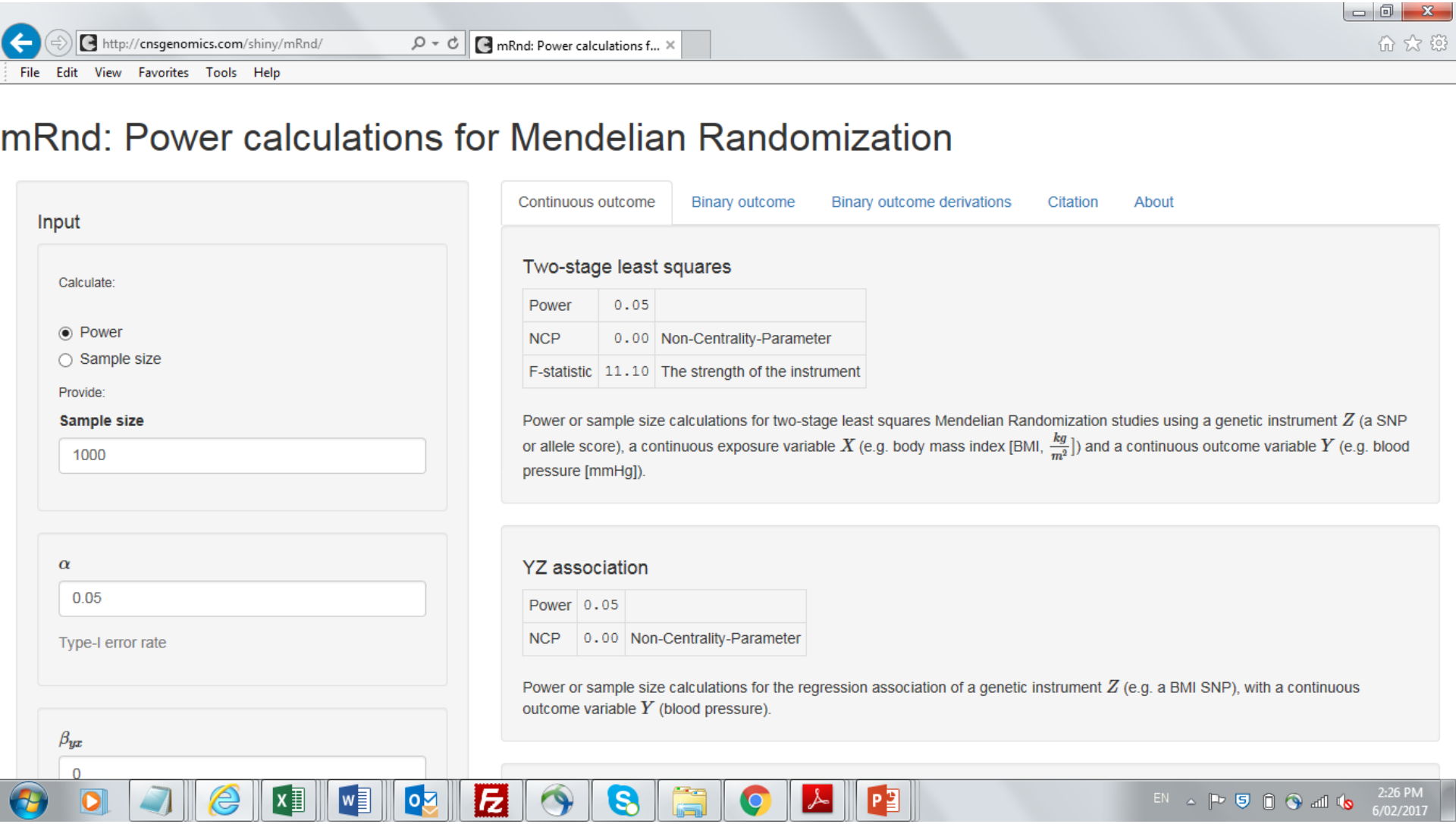


Figure 1. DAG for a Mendelian randomisation analysis using four genetic variants as instrumental variables for the effect of fat mass on bone mineral density.

Palmer et al (2011) Stat Method Res

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

Calculating Power in Mendelian Randomization Studies



Limitations to Mendelian Randomization

1- Population stratification

2- Canalisation (“Developmental compensation”)

3- The existence of instruments

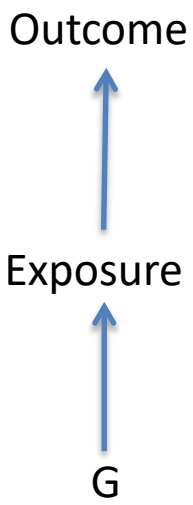
4- Power (also “weak instrument bias”)

5- Pleiotropy

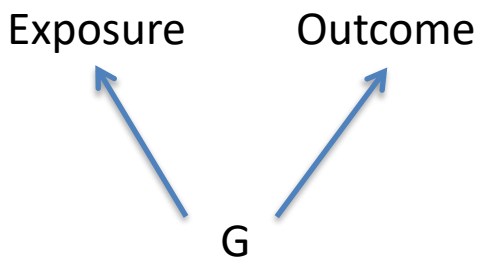
Pleiotropy

- Genetic variant influences more than one trait

- Horizontal vs Vertical pleiotropy



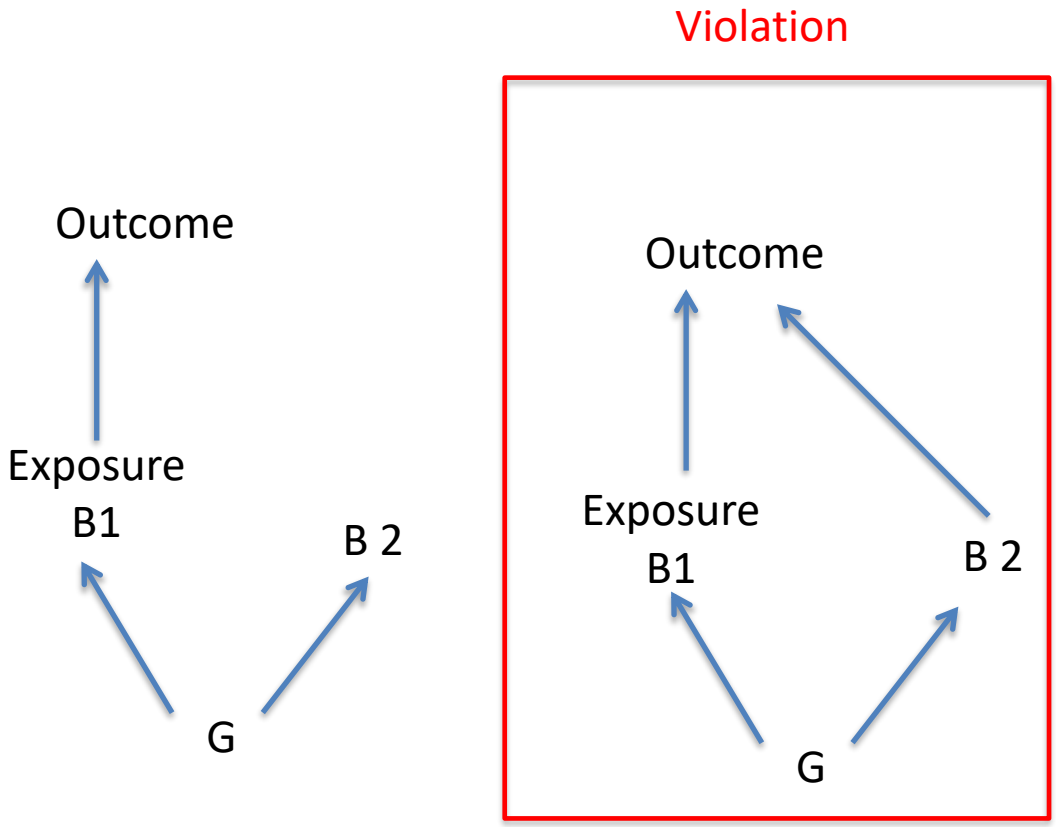
Vertical Pleiotropy



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 <u>affects your outcome</u>



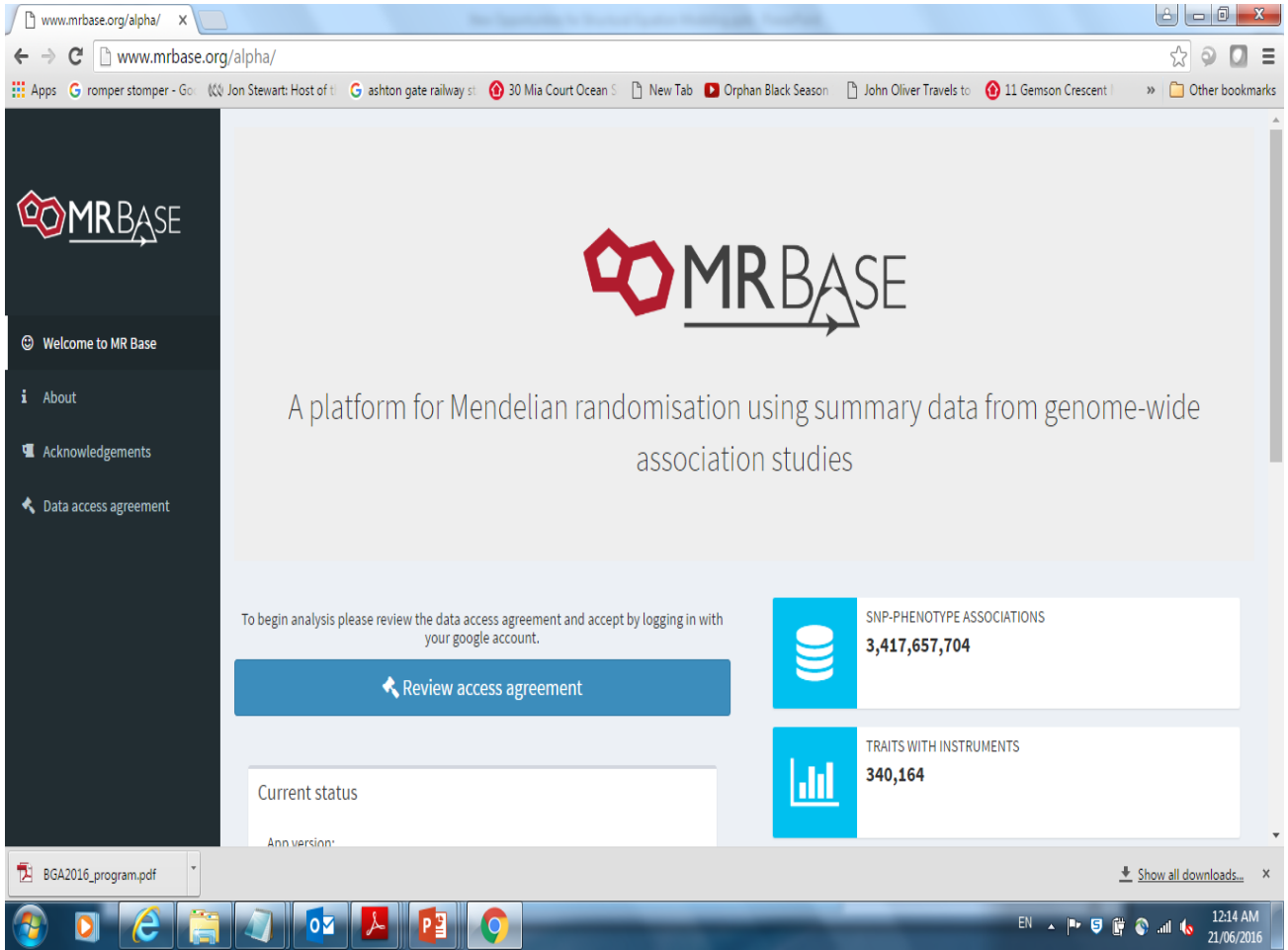
MR Base

Jie “Chris” Zheng

http://www.mrbase.org/

Gib Hemani

Phil Haycock



MR BASE

- Welcome to MR Base
- About
- Acknowledgements
- Data access agreement

Logged in as
David Evans
epxde@bristol.ac.uk

- **Perform MR analysis**
  - **Choose exposures**
  - Choose outcomes
  - Run MR
- Quick SNP lookup

Choosing instruments for the exposure

To use two sample MR to estimate the causal effect of an exposure on an outcome, the first step is to identify SNPs that are robustly associated with the exposure. These summary statistics for these SNPs can be taken from a sample from which there is no data on the outcome.

Please provide instruments by choosing from one of the data sources below, or by uploading your own data. You can choose multiple exposures to be analysed, and multiple instruments per exposure.

Choose instruments

Select exposure source

- (•) Manual file upload
- ( ) NHGRI-EBI GWAS catalog
- ( ) MR Base GWAS catalog
- ( ) Gene expression QTLs
- ( ) Protein level QTLs
- ( ) Metabolite level QTLs
- ( ) Methylation level QTLs

Manual file upload

The file must be a plain text file.

To do simple SNP look ups it must have at least one column with the header **SNP** .

To do an MR analysis it must have the following column headers:

- **SNP** - rs IDs of the instruments for the exposure
- **beta** - effect sizes for each SNP
- **se** - standard errors
- **effect_allele** - Effect allele

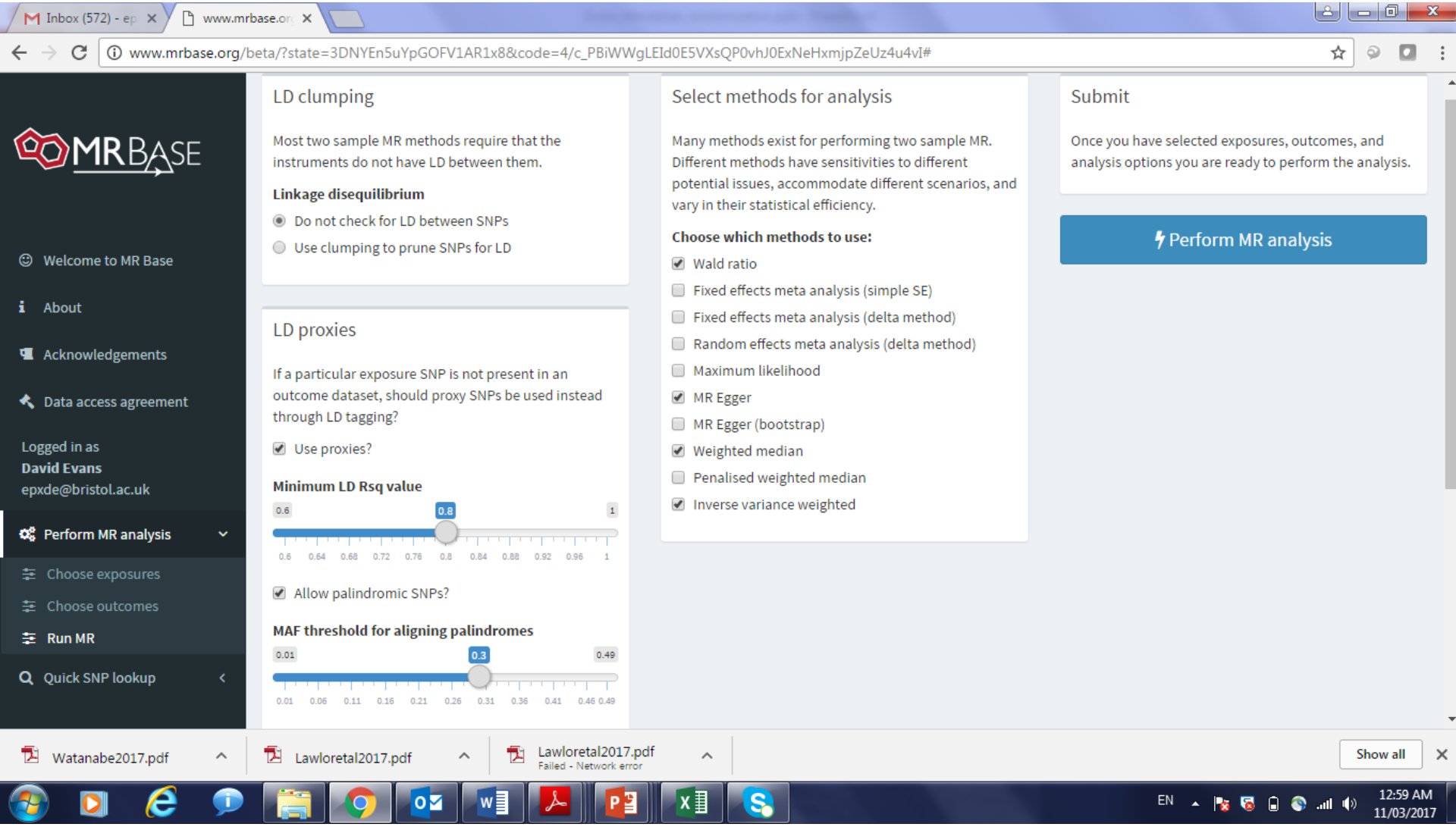
It's useful to have these columns too:

- **other_allele** - Other allele
- **eaf** - Effect allele frequency

You can see an example file here: telomere_length.txt

Upload plain text file *Preview of uploaded table*

Browse... No file s...


Inbox (572) - ep
www.mrbase.org
www.mrbase.org/beta/?state=3DNYEn5uYpGOFV1AR1x8&code=4/c_PBiWWgLEId0E5VXsQP0vhJ0ExNeHxmjpZeUz4u4vI#
MR BASE
Welcome to MR Base
About
Acknowledgements
Data access agreement
Logged in as
David Evans
epxde@bristol.ac.uk
Perform MR analysis
Choose exposures
Choose outcomes
Run MR
Quick SNP lookup
LD clumping
Most two sample MR methods require that the instruments do not have LD between them.
Linkage disequilibrium
Do not check for LD between SNPs
Use clumping to prune SNPs for LD
LD proxies
If a particular exposure SNP is not present in an outcome dataset, should proxy SNPs be used instead through LD tagging?
Use proxies?
Minimum LD Rsq value
0.6 0.8 1
0.6 0.64 0.68 0.72 0.76 0.8 0.84 0.88 0.92 0.96 1
Allow palindromic SNPs?
MAF threshold for aligning palindromes
0.01 0.3 0.49
0.01 0.06 0.11 0.16 0.21 0.26 0.31 0.36 0.41 0.46 0.49
Select methods for analysis
Many methods exist for performing two sample MR. Different methods have sensitivities to different potential issues, accommodate different scenarios, and vary in their statistical efficiency.
Choose which methods to use:
Wald ratio
Fixed effects meta analysis (simple SE)
Fixed effects meta analysis (delta method)
Random effects meta analysis (delta method)
Maximum likelihood
MR Egger
MR Egger (bootstrap)
Weighted median
Penalised weighted median
Inverse variance weighted
Submit
Once you have selected exposures, outcomes, and analysis options you are ready to perform the analysis.
Perform MR analysis
Watanabe2017.pdf
Lawloretal2017.pdf
Lawloretal2017.pdf
Failed - Network error
Show all
EN
12:59 AM
11/03/2017


Inbox (572) - epxde@bris
www.mrbase.org/beta/?s
www.mrbase.org/beta/?state=3DNYEn5uYpGOFV1AR1x8&code=4/c_PBiWWgLEId0E5VXsQP0vhJ0ExNeHxmjpZeUz4u4vI#
Run MR
MR Results
Quick SNP lookup
5e-08
Perform clumping
Display columns
ID
Trait
Note
First author
Consortium
Number of cases
Number of controls
Sample size
Number of variants
Year
PubmedID
Access
Category
Population
Priority
Sd
Sex
Subcategory
Unit
Search: ldl
ID Trait Note First author Consortium Number of cases Number of controls Sample size Number of variants Year PubmedID Access Category Po
300 300 LDL cholesterol Willer CJ GLGC 173082 2437752 2013 24097068 public Risk factor
781 781 LDL cholesterol Metabo-chip Willer CJ GLGC 83198 120251 2013 24097068 public Risk factor
880 880 Total cholesterol in large LDL L.LDL.C Kettunen 21552 11871461 2016 27005778 public Metabolites
881 881 Cholesterol esters in large VLDL L.LDL.CE Kettunen 19273 11820655 2016 27005778 public Metabolites
Watanabe2017.pdf
Lawloretal2017.pdf
Lawloretal2017.pdf
Failed - Network error
Show all
EN
1:10 AM
11/03/2017

Display columns

- [ ] ID
- [x] Trait
- [x] Note
- [x] First author
- [x] Consortium
- [x] Number of cases
- [x] Number of controls
- [x] Sample size
- [x] Number of variants
- [x] Year
- [ ] PubmedID
- [ ] Access
- [ ] Category
- [ ] Population
- [ ] Priority
- [ ] Sd
- [ ] Sex
- [x] Subcategory
- [ ] Unit

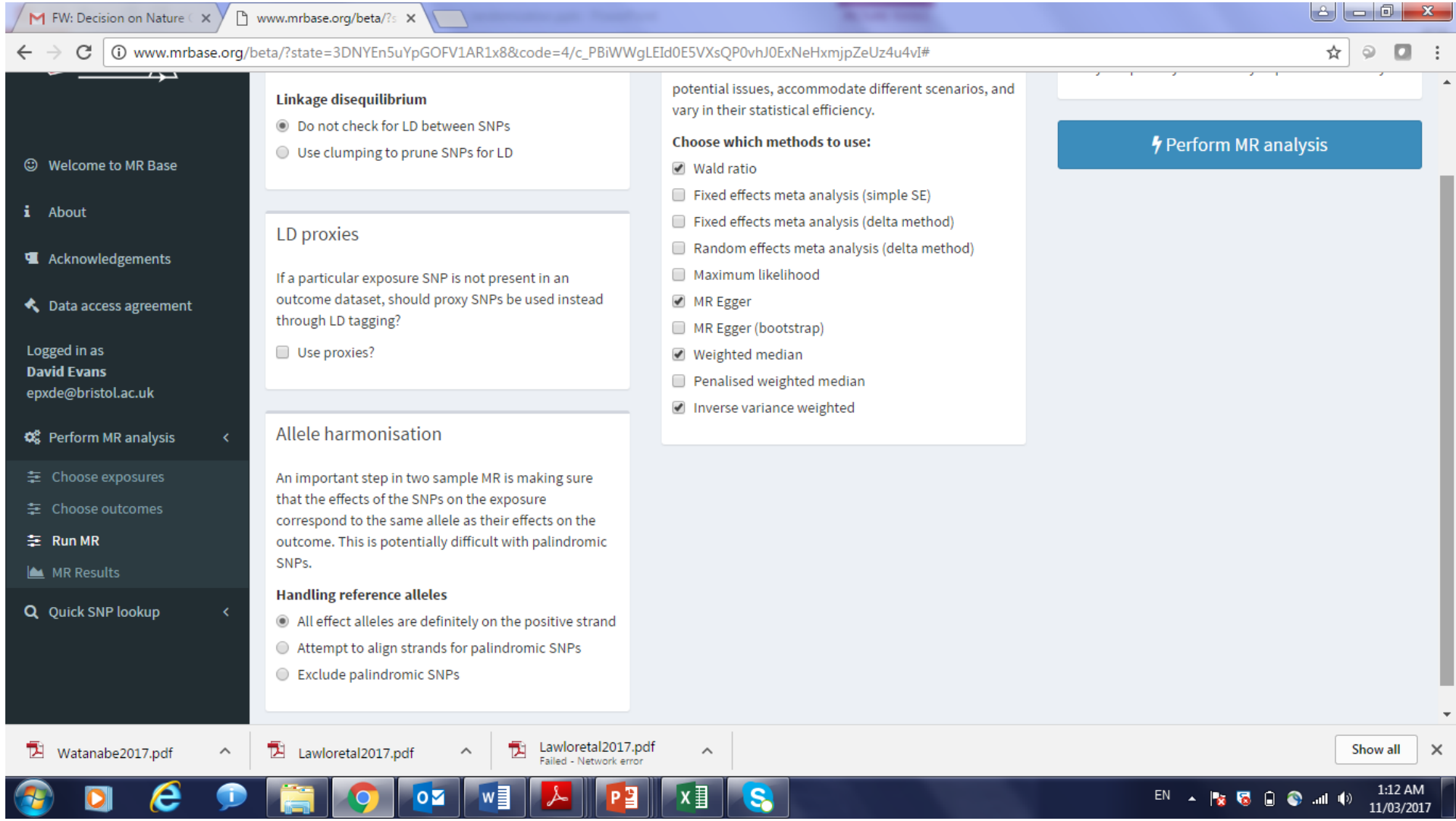
Show 10 entries

Search: coro

	Trait	Note	First author	Consortium	Number of cases	Number of controls	Sample size	Number of variants	Year	Subcategory
6	Coronary heart disease		Peden	C4D	15420	15062	30482	540233	2011	Cardiovascular
7	**Coronary heart disease**		**Nikpay**	**CARDIoGRAMplusC4D**	**60801**	**123504**	**184305**	**9455779**	**2015**	**Cardiovascular**
8	Coronary heart disease		Schunkert H	CARDIoGRAM	22233	64762	86995	2420361	2011	Cardiovascular
9	Coronary heart disease		Deloukas	CARDIoGRAMplusC4D	63746	130681	194427	79129	2013	Cardiovascular
	All	All	All	All	All	All	All	All	All	All

Showing 1 to 4 of 4 entries (filtered from 1,033 total entries)

Previous 1 Next


FW: Decision on Nature
www.mrbase.org/beta/?s
www.mrbase.org/beta/?state=3DNYEn5uYpGOFV1AR1x8&code=4/c_PBiWWgLEId0E5VXsQP0vhJ0ExNeHxmjpZeUz4u4vI#
Welcome to MR Base
About
Acknowledgements
Data access agreement
Logged in as
David Evans
epxde@bristol.ac.uk
Perform MR analysis
Choose exposures
Choose outcomes
Run MR
MR Results
Quick SNP lookup
Linkage disequilibrium
Do not check for LD between SNPs
Use clumping to prune SNPs for LD
LD proxies
If a particular exposure SNP is not present in an outcome dataset, should proxy SNPs be used instead through LD tagging?
Use proxies?
Allele harmonisation
An important step in two sample MR is making sure that the effects of the SNPs on the exposure correspond to the same allele as their effects on the outcome. This is potentially difficult with palindromic SNPs.
Handling reference alleles
All effect alleles are definitely on the positive strand
Attempt to align strands for palindromic SNPs
Exclude palindromic SNPs
potential issues, accommodate different scenarios, and vary in their statistical efficiency.
Choose which methods to use:
Wald ratio
Fixed effects meta analysis (simple SE)
Fixed effects meta analysis (delta method)
Random effects meta analysis (delta method)
Maximum likelihood
MR Egger
MR Egger (bootstrap)
Weighted median
Penalised weighted median
Inverse variance weighted
Perform MR analysis
Watanabe2017.pdf
Lawloretal2017.pdf
Lawloretal2017.pdf
Failed - Network error
Show all
EN
1:12 AM
11/03/2017

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.
- Davies et al (2018). Reading Mendelian randomization studies: a guide, glossary, and checklist for clinicians. BMJ, Jul 12, 362:k601.
- 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.
- Zheng et al. (2017). Recent developments in Mendelian randomization studies. Curr Epidemiol Rep, 4(4), 330-345.