Mendelian randomization - an introduction

Bjørn Olav Åsvold

K.G. Jebsen Center for Genetic Epidemiology, NTNU

June 2022

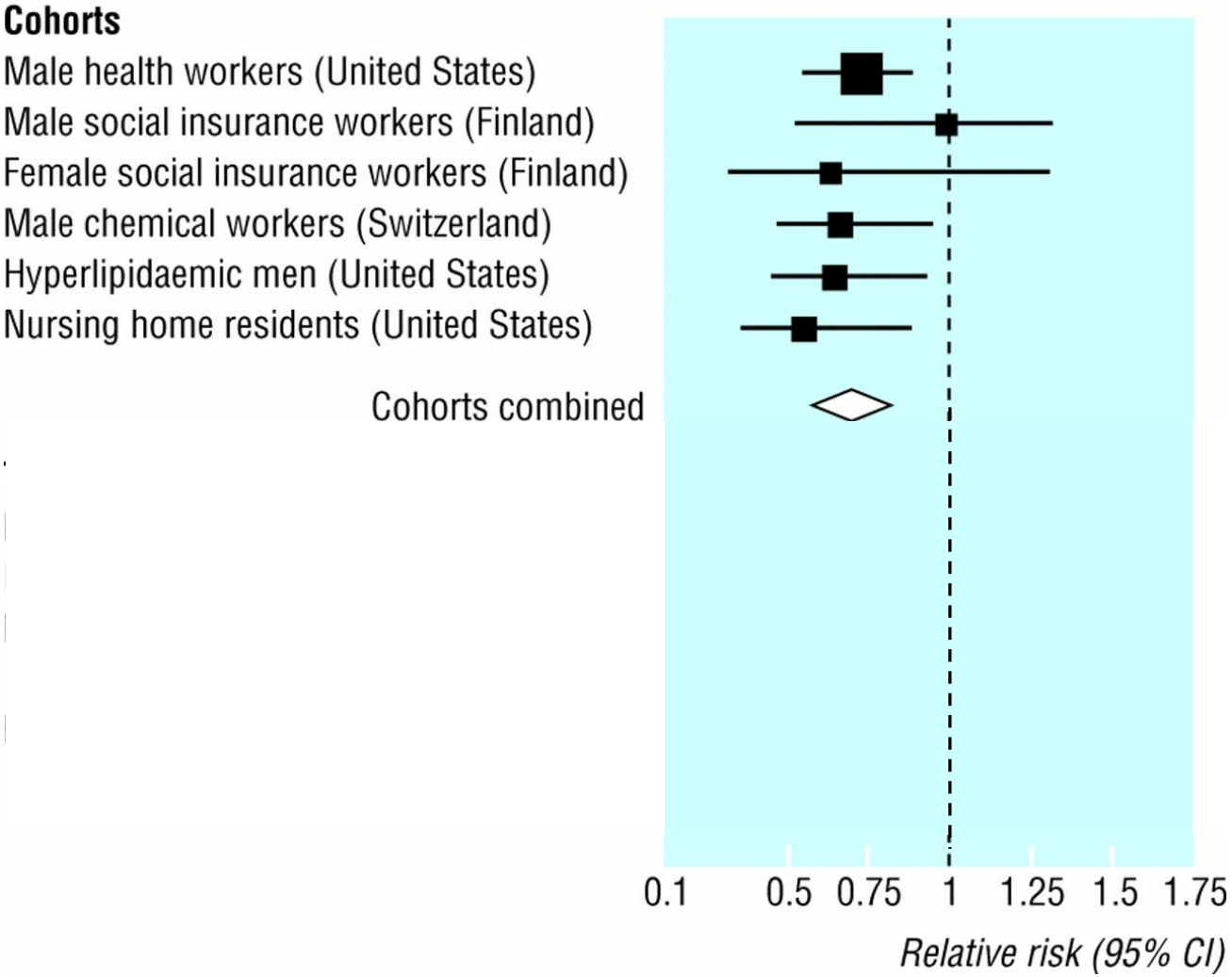
This lecture will cover

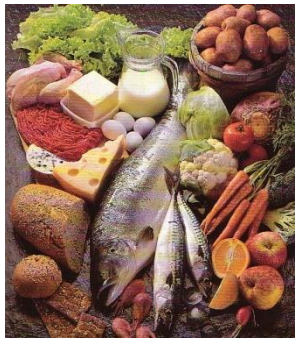
- The purposes of Mendelian randomization studies
- Principles underlying Mendelian randomization
- Assumptions of Mendelian randomization
- Strengths and limitations

What's the purpose of Mendelian randomization?

- In epidemiology, we often aim to examine causes of disease
- Causal inference is often challenged by the possibility of confounding or reverse causality
- Mendelian randomization: a study approach where genetic variants are used to examine the causal effects of modifiable exposures

The shortcomings of observational studies – one out of many examples: ß-carotene & CVD mortality



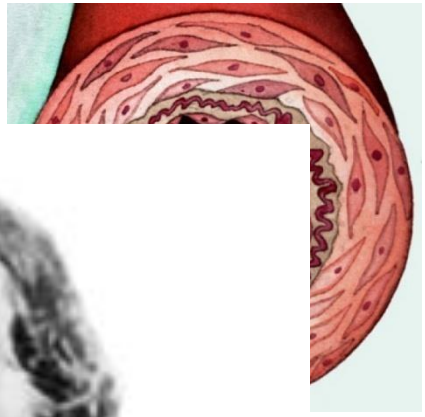


Confounders

Exposure → Disease

Cardiovascular mortality

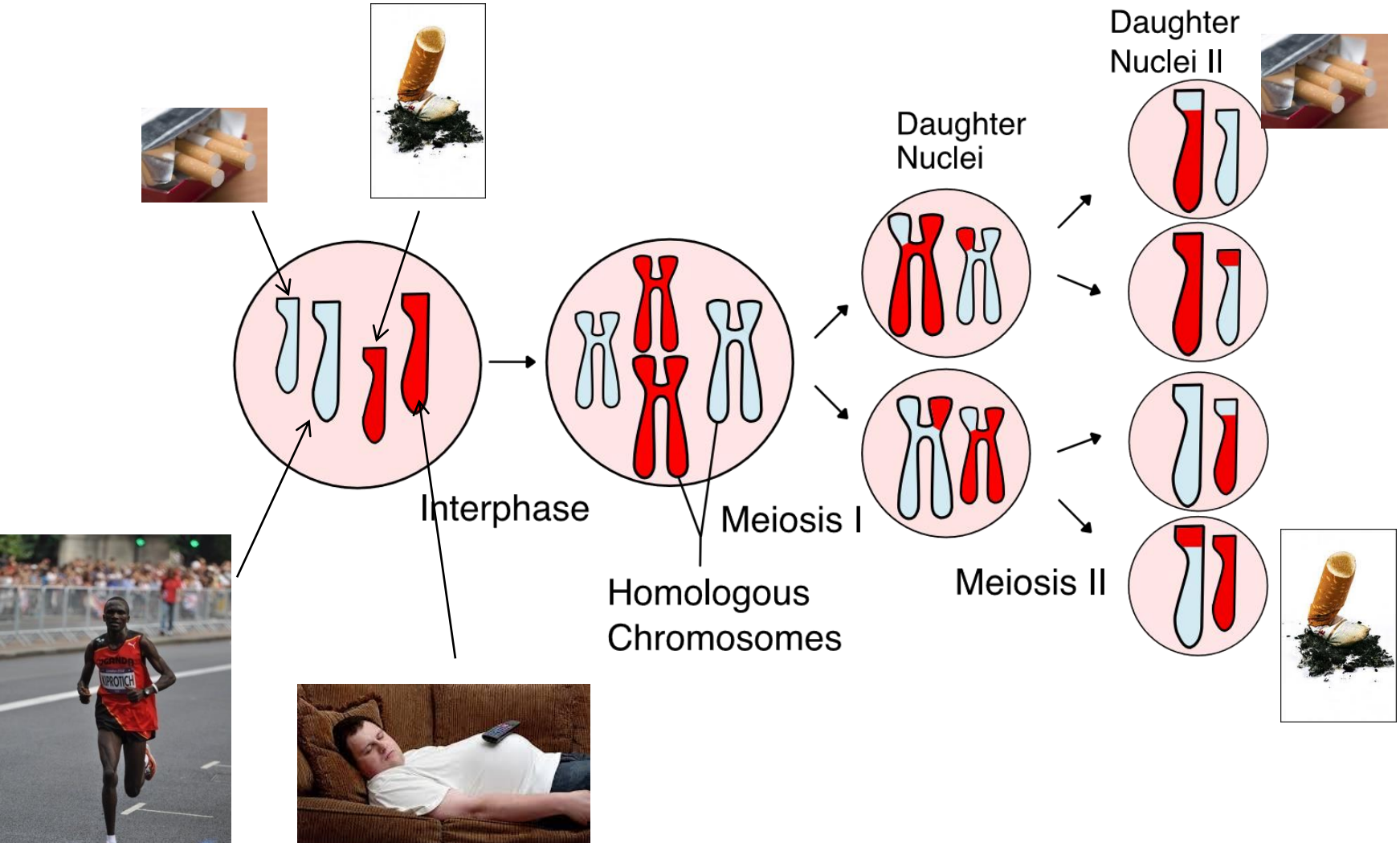
Example: CRP levels and coronary heart disease



CR eart disease

- Mendelian randomization: Genetic variants are used as proxies – instrumental variables – for modifiable risk factors in observational studies.

- Relies on Mendel’s laws of inheritance:
- Generally, genes are randomly assorted from parents to offspring, and inheritance of one trait is independent of inheritance of other traits.


Daughter
Nuclei II
Daughter
Nuclei
Interphase
Meiosis I
Homologous
Chromosomes
Meiosis II

- Mendelian randomization: Genetic variants are used as proxies – instrumental variables – for modifiable risk factors in observational studies

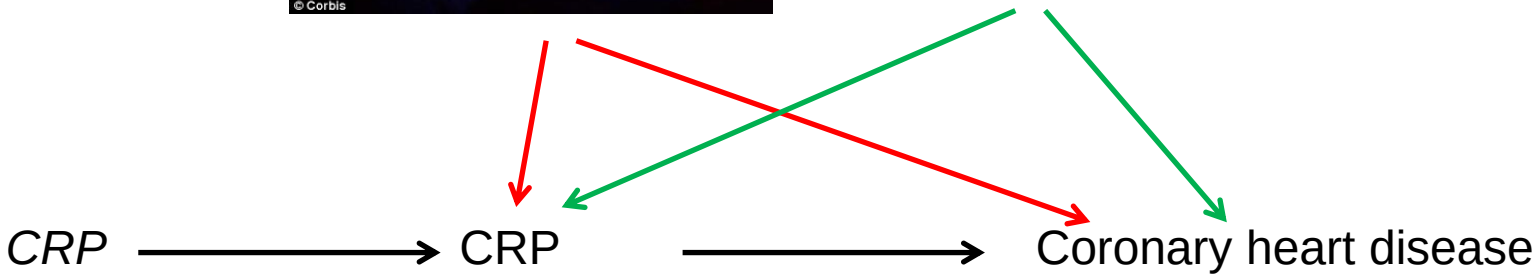
- Relies on Mendel's laws of inheritance:
- Generally, genes are randomly assorted from parents to offspring, and inheritance of one trait is independent of inheritance of other traits.

CRP level

- Genetic variants associated with a specific trait will not be liable to confounding by behavioual, socioeconomic and physiological factors.

obesity, smoking, preclinical CHD

Example: CRP levels and coronary heart disease



The NEW ENGLAND JOURNAL of MEDICINE

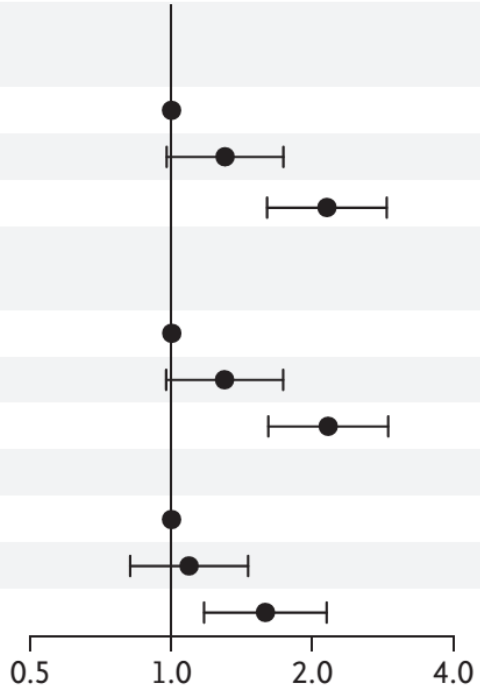
ORIGINAL ARTICLE

Genetically Elevated C-Reactive Protein and Ischemic Vascular Disease

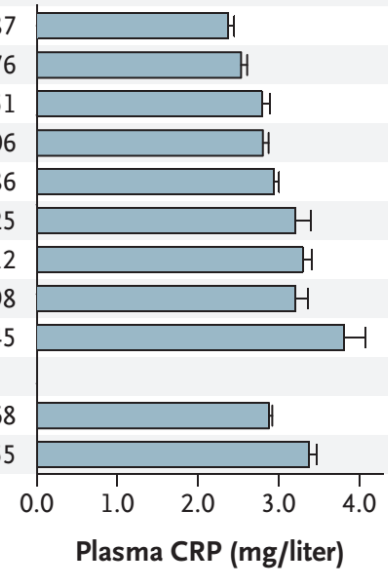
Jeppe Zacho, M.D., Anne Tybjærg-Hansen, M.D., D.M.Sc.,
Jan Skov Jensen, M.D., D.M.Sc., Peer Grande, M.D., D.M.Sc.,
Henrik Sillesen, M.D., D.M.Sc., and Børge G. Nordestgaard, M.D., D.M.Sc.

A Ischemic Heart Disease

CRP Level	No. of Participants	No. of Events	P for Trend
Adjusted for age, sex, and statin use			<0.001
<1 mg/liter	659	50	
1–3 mg/liter	5370	738	
>3 mg/liter	1905	433	
Adjusted for age, sex, statin use, and genotype			<0.001
<1 mg/liter	659	50	
1–3 mg/liter	5370	738	
>3 mg/liter	1905	433	
Multifactorially adjusted			<0.001
<1 mg/liter	652	50	
1–3 mg/liter	5294	721	
>3 mg/liter	1842	415	



Genotype				Change in CRP Level %	No. of Partici-pants	Mean ±SE
CRP genotype combination						
1. AA	CC	CC	TT	0	3,887	
2. GA	CC	CC	TT	7	6,876	
3. GG	CC	CC	TT	18	3,051	
4. GA	CT	CT	TT	18	7,006	
5. GG	CT	CT	TT	25	6,486	
6. GA	CC	CA	GT	35	1,125	
7. GG	TT	TT	TT	38	3,412	
8. GG	CC	CA	GT	36	998	
9. GG	CT	AT	GT	64	1,045	
3–6				22	17,668	
7–9				43	5,455	



- Higher CRP levels were associated with increased risk of coronary heart disease,
- but genetically determined higher CRP levels were not associated with increased risk of CHD.
- This strongly suggests that elevated CRP levels do not cause coronary heart disease,
- and that the observed association between CRP levels and risk of coronary heart disease is explained by confounding or reverse causality.

Randomized study?

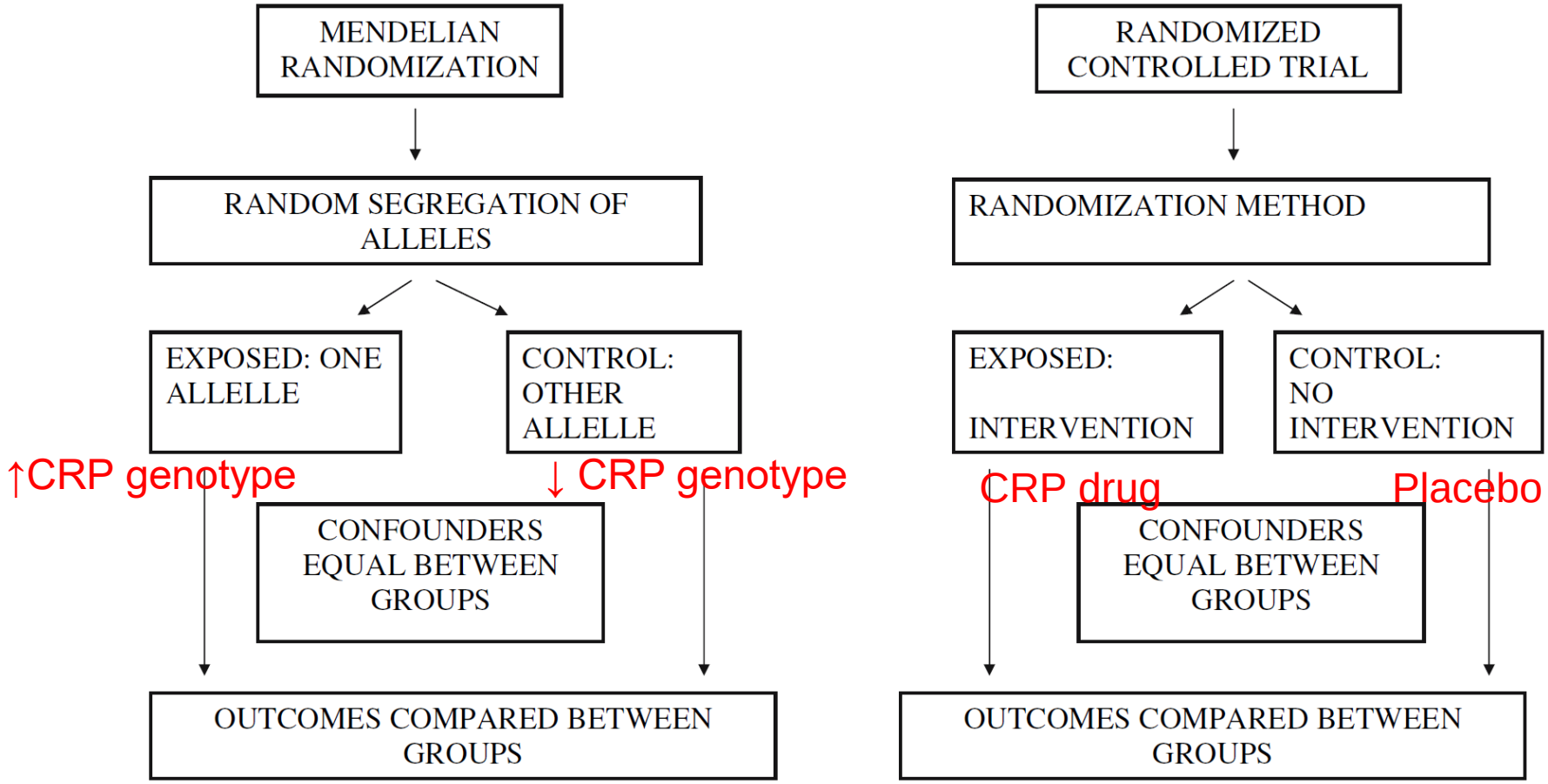
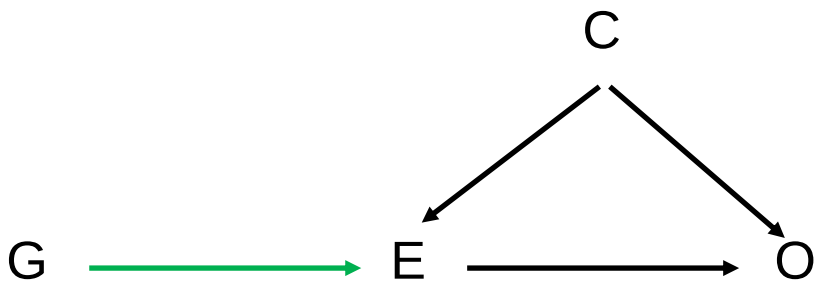


Fig. 7 Mendelian randomization and randomized controlled trial designs compared, after Hingorani and Humphries (2005)

Ebrahim & Davey Smith, Hum Genet 2008;123:15-33

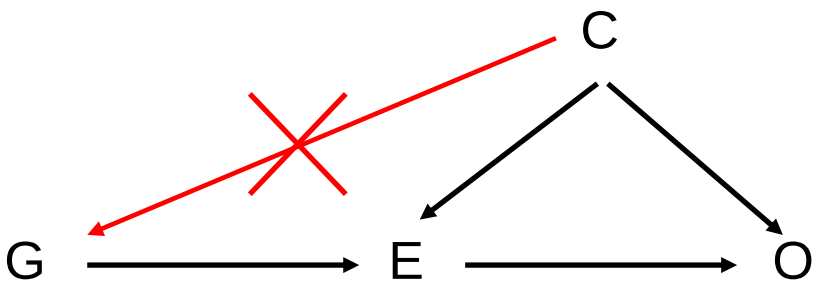
3 key assumption of MR studies

1. The genetic instrument is associated with the exposure (the relevance assumption).



3 key assumption of MR studies

2. There is no association between the genetic instrument and confounders of the exposure-outcome relation – in other words: the genetic instrument shares no common causes with the outcome (the independence assumption).

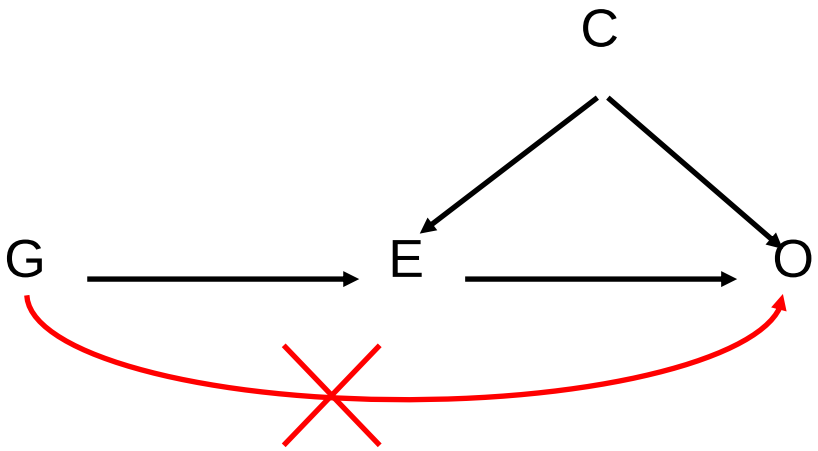


Such associations can occur due to

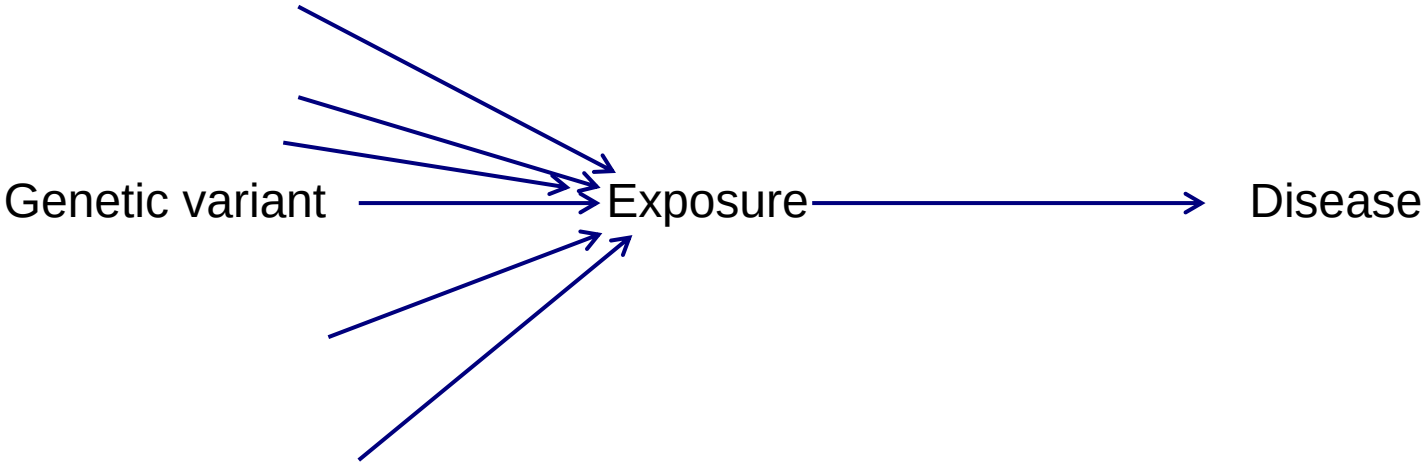
- Weak instrument bias
- Horizontal pleiotropy
- Linkage disequilibrium among gene loci
- Population stratification

3 key assumption of MR studies

3. There is no association between the genetic instrument and the outcome except through the exposure of interest (the exclusion restriction assumption).



This assumption is violated if there is horizontal pleiotropy: a genetic variant affects multiple traits through separate pathways


Genetic variant
Exposure
Disease

MR study designs

- We need to estimate 2 associations:
- Gene-exposure association: G-X
- Gene-outcome association: G-Y
- We usually estimate the causal effect: G-Y / G-X

- One-sample MR: G-X and G-Y estimated in the same study sample
- Two-sample MR: G-X and G-Y estimated in different samples

- Bidirectional MR: Both causal directions examined
- Multivariable MR: Several correlated exposures examined
- Two-step MR: Mediation analysis
- Gene x environment interaction
- …..

How to select the genetic instruments?

- Two general approaches:

1. Select SNPs with proven or plausible biological effect on the target exposure

2. Select SNPs from GWAS

Some limitations of MR – and possible solutions

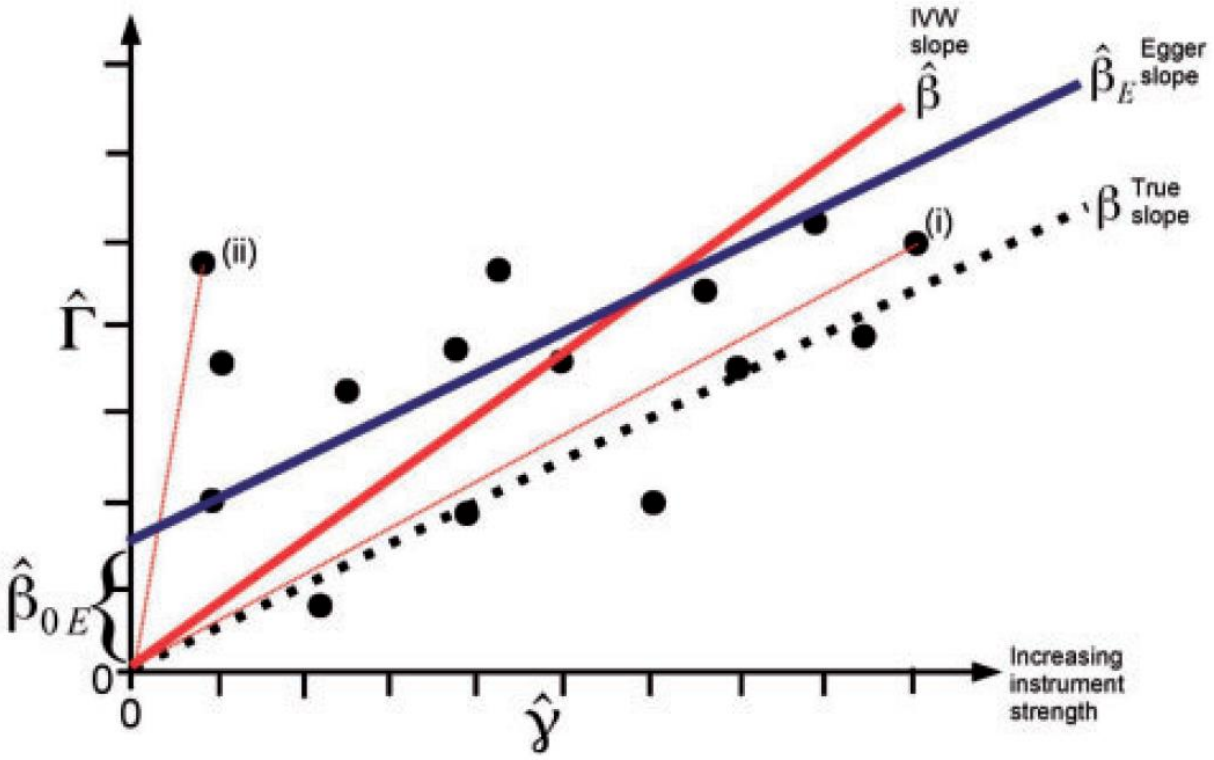
- Lack of suitable instruments
  - More GWAS
- Low power
  - Combine multiple SNPs (genetic risk score or meta-analysis), larger *N*
- Winner's curse (overestimation of SNP-exposure association)
- Poor biological understanding – possibilities for pleiotropy
  - Select variants with known biology
  - If multiple variants: numerous sensitivity analyses available
- Trait heterogeneity (SNPs associated with multiple dimensions of a single trait)
- Time-varying exposures
- Effects on disease risk vs. disease progression
- Dynastic effects, population stratification, assortative mating
  - Within-family MR

Some Mendelian randomization analyses

(sensitivity analyses to examine / correct for pleiotropy)

- Inverse-variance weighted MR: main analysis, meta-analysis (weighted average) of ratio estimates, assumes balanced pleiotropy
- MR-Egger regression: allows for directional pleiotropy, but has other assumptions and lower power
- Weighted median: valid if the majority of the information comes from valid instruments
- Weighted mode: assumes that the mode of instruments is valid
- MR-PRESSO, MR-Lasso, MR-Robust: assume that there is a small number of invalid instruments (outliers)
- MR-RAPS
- Contamination mixture
- MRMix

See for example: Slob & Burgess, Genetic Epidemiology 2020;44(4):313-29.

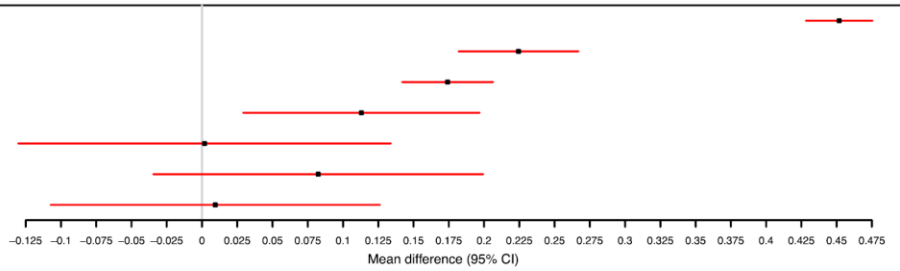


Bowden et al., IJE 2015

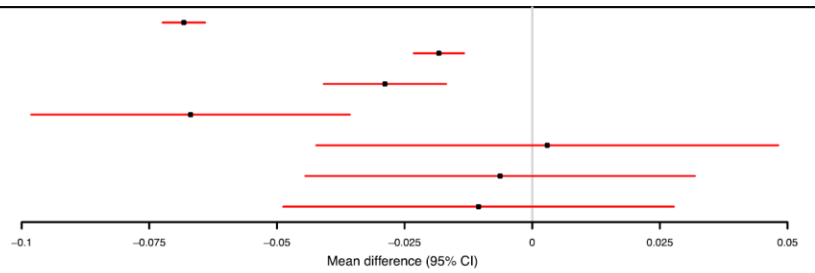
Some limitations of MR – and possible solutions

- Lack of suitable instruments
  - More GWAS
- Low power
  - Combine multiple SNPs (genetic risk score or meta-analysis), larger *N*
- Winner's curse (overestimation of SNP-exposure association)
- Poor biological understanding – possibilities for pleiotropy
  - Select variants with known biology
  - If multiple variants: numerous sensitivity analyses available
- Trait heterogeneity (SNPs associated with multiple dimensions of a single trait)
- Time-varying exposures
- Effects on disease risk vs. disease progression
- Dynastic effects, population stratification, assortative mating
  - Within-family MR

Effect of height on education	*N*	Mean difference (95% CI)
IPD Ordinary least squares	61,008	0.45 (0.43 – 0.48)
IPD OLS family FE	61,008	0.22 (0.18 – 0.27)
IPD MR–PRS unrelateds	354,836	0.17 (0.14 – 0.21)
IPD MR–PRS siblings	61,008	0.11 (0.03 – 0.20)
IPD MR–PRS siblings family FE	61,008	0.00 (–0.13 – 0.13)
2SMR IVW siblings	61,008	0.08 (–0.03 – 0.20)
2SMR IVW siblings – split sample	61,008	0.01 (–0.11 – 0.13)



Effect of BMI on education	*N*	Mean difference (95% CI)
IPD Ordinary least squares	61,008	–0.07 (–0.07 – –0.06)
IPD OLS family FE	61,008	–0.02 (–0.02 – –0.01)
IPD MR–PRS unrelateds	354,836	–0.03 (–0.04 – –0.02)
IPD MR–PRS siblings	61,008	–0.07 (–0.10 – –0.04)
IPD MR–PRS siblings family FE	61,008	0.00 (–0.04 – 0.05)
2SMR IVW siblings	61,008	–0.01 (–0.04 – 0.03)
2SMR IVW siblings – split sample	61,008	–0.01 (–0.05 – 0.03)



Brumpton et al., Nature Communications 2020

Another use of MR:
Identify intended and unintended drug effects

NTNU
K.G. Jebsen Center for
Genetic Epidemiology

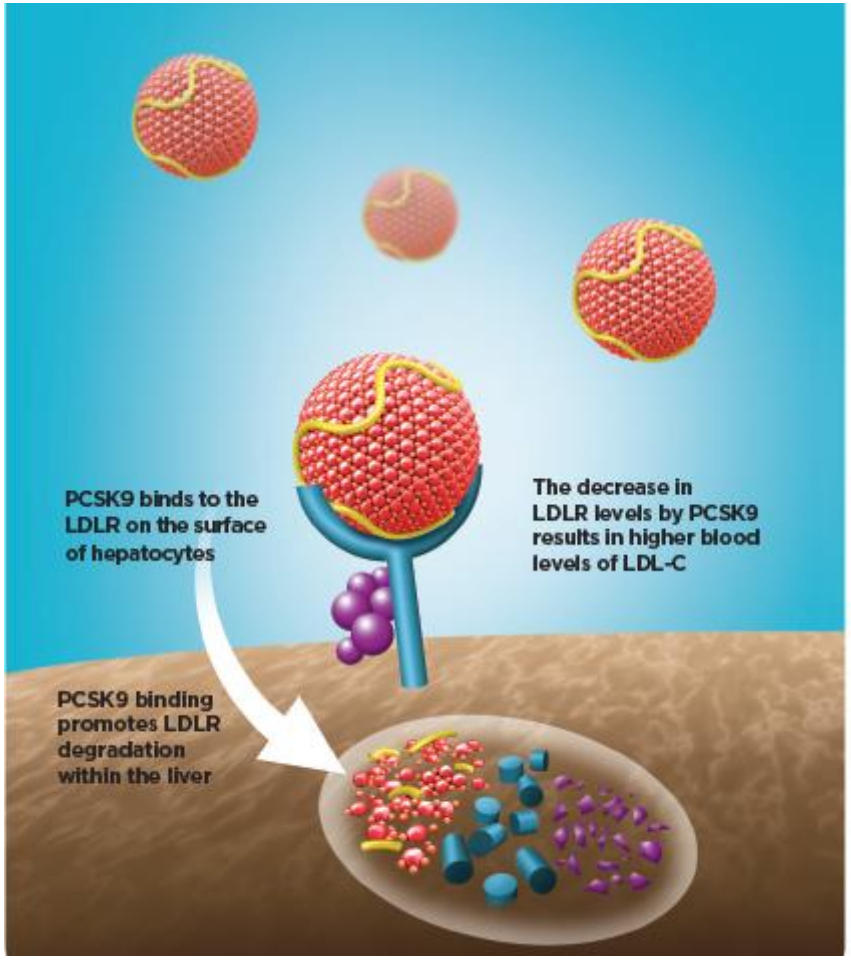
What side effects can we anticipate from long-term PCSK9 inhibition?

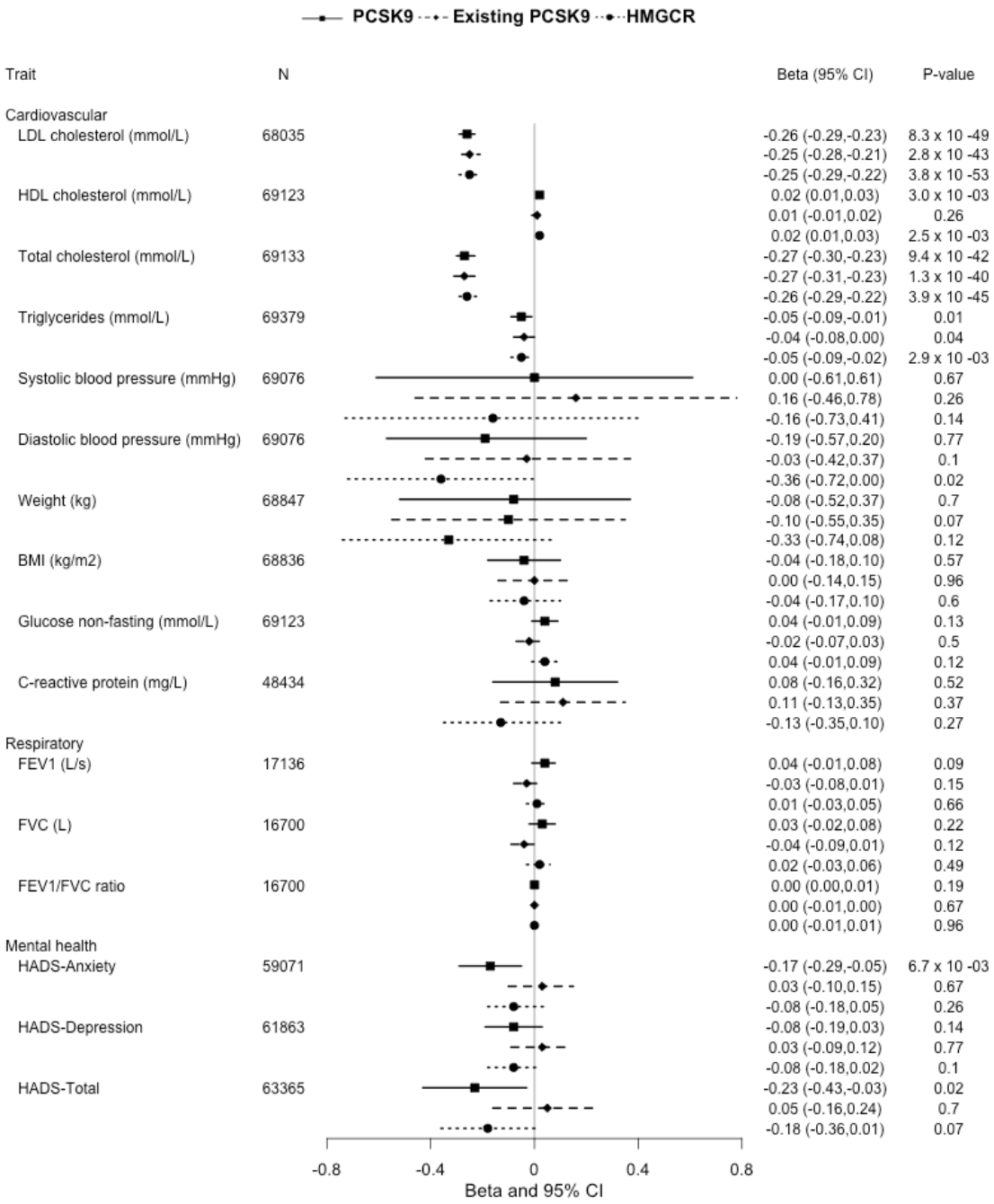
PCSK9 inhibitors: New, effective lipid-lowering drugs that reduce the risk of cardiovascular events

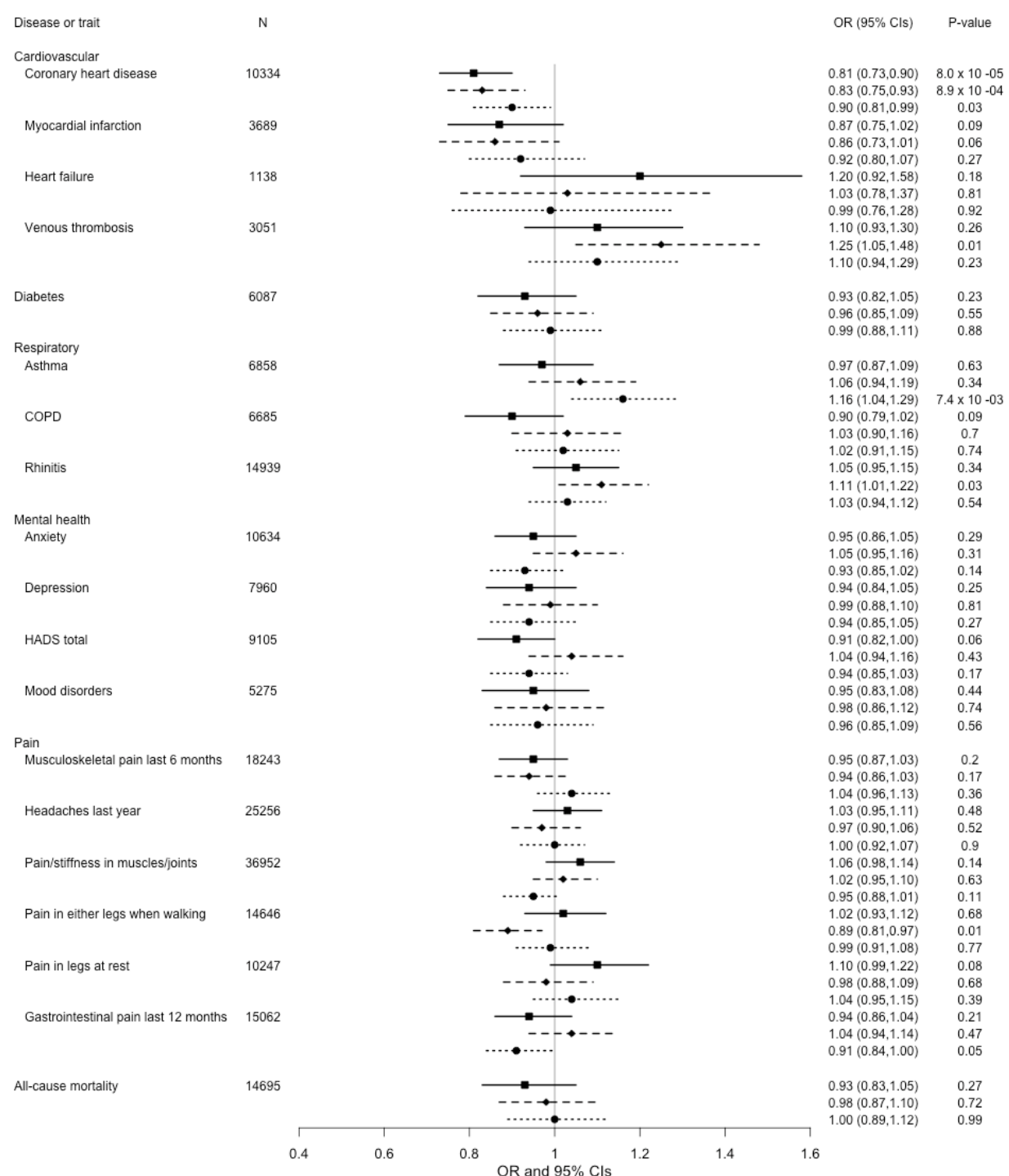
Is long-term PCSK9 inhibition hazardous?

We lack long-term RCT evidence

Ben Brumpton et al.,
Circ Genom Precis Med 2019




PCSK9
Existing PCSK9
HMGCR
Trait
N
Beta (95% CI)
P-value
Cardiovascular
LDL cholesterol (mmol/L) 68035 -0.26 (-0.29,-0.23) 8.3 x 10 -49
-0.25 (-0.28,-0.21) 2.8 x 10 -43
-0.25 (-0.29,-0.22) 3.8 x 10 -53
HDL cholesterol (mmol/L) 69123 0.02 (0.01,0.03) 3.0 x 10 -03
0.01 (-0.01,0.02) 0.26
0.02 (0.01,0.03) 2.5 x 10 -03
Total cholesterol (mmol/L) 69133 -0.27 (-0.30,-0.23) 9.4 x 10 -42
-0.27 (-0.31,-0.23) 1.3 x 10 -40
-0.26 (-0.29,-0.22) 3.9 x 10 -45
Triglycerides (mmol/L) 69379 -0.05 (-0.09,-0.01) 0.01
-0.04 (-0.08,0.00) 0.04
-0.05 (-0.09,-0.02) 2.9 x 10 -03
Systolic blood pressure (mmHg) 69076 0.00 (-0.61,0.61) 0.67
0.16 (-0.46,0.78) 0.26
-0.16 (-0.73,0.41) 0.14
Diastolic blood pressure (mmHg) 69076 -0.19 (-0.57,0.20) 0.77
-0.03 (-0.42,0.37) 0.1
-0.36 (-0.72,0.00) 0.02
Weight (kg) 68847 -0.08 (-0.52,0.37) 0.7
-0.10 (-0.55,0.35) 0.07
-0.33 (-0.74,0.08) 0.12
BMI (kg/m2) 68836 -0.04 (-0.18,0.10) 0.57
0.00 (-0.14,0.15) 0.96
-0.04 (-0.17,0.10) 0.6
Glucose non-fasting (mmol/L) 69123 0.04 (-0.01,0.09) 0.13
-0.02 (-0.07,0.03) 0.5
0.04 (-0.01,0.09) 0.12
C-reactive protein (mg/L) 48434 0.08 (-0.16,0.32) 0.52
0.11 (-0.13,0.35) 0.37
-0.13 (-0.35,0.10) 0.27
Respiratory
FEV1 (L/s) 17136 0.04 (-0.01,0.08) 0.09
-0.03 (-0.08,0.01) 0.15
0.01 (-0.03,0.05) 0.66
FVC (L) 16700 0.03 (-0.02,0.08) 0.22
-0.04 (-0.09,0.01) 0.12
0.02 (-0.03,0.06) 0.49
FEV1/FVC ratio 16700 0.00 (0.00,0.01) 0.19
0.00 (-0.01,0.00) 0.67
0.00 (-0.01,0.01) 0.96
Mental health
HADS-Anxiety 59071 -0.17 (-0.29,-0.05) 6.7 x 10 -03
0.03 (-0.10,0.15) 0.67
-0.08 (-0.18,0.05) 0.26
HADS-Depression 61863 -0.08 (-0.19,0.03) 0.14
0.03 (-0.09,0.12) 0.77
-0.08 (-0.18,0.02) 0.1
HADS-Total 63365 -0.23 (-0.43,-0.03) 0.02
0.05 (-0.16,0.24) 0.7
-0.18 (-0.36,0.01) 0.07
-0.8
-0.4
0
0.4
0.8
Beta and 95% CI

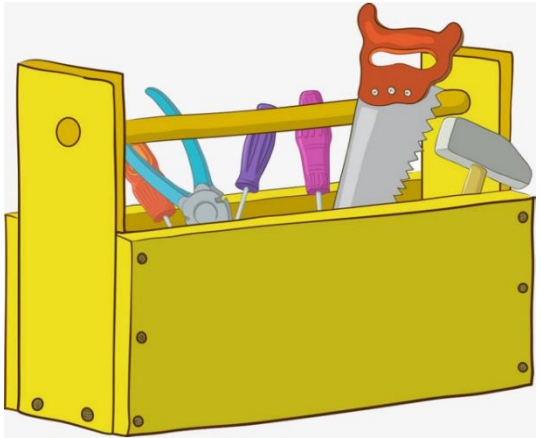

Disease or trait
N
OR (95% CIs)
P-value
Cardiovascular
Coronary heart disease
10334
0.81 (0.73,0.90)
8.0 x 10 -05
0.83 (0.75,0.93)
8.9 x 10 -04
0.90 (0.81,0.99)
0.03
Myocardial infarction
3689
0.87 (0.75,1.02)
0.09
0.86 (0.73,1.01)
0.06
0.92 (0.80,1.07)
0.27
Heart failure
1138
1.20 (0.92,1.58)
0.18
1.03 (0.78,1.37)
0.81
0.99 (0.76,1.28)
0.92
Venous thrombosis
3051
1.10 (0.93,1.30)
0.26
1.25 (1.05,1.48)
0.01
1.10 (0.94,1.29)
0.23
Diabetes
6087
0.93 (0.82,1.05)
0.23
0.96 (0.85,1.09)
0.55
0.99 (0.88,1.11)
0.88
Respiratory
Asthma
6858
0.97 (0.87,1.09)
0.63
1.06 (0.94,1.19)
0.34
1.16 (1.04,1.29)
7.4 x 10 -03
COPD
6685
0.90 (0.79,1.02)
0.09
1.03 (0.90,1.16)
0.7
1.02 (0.91,1.15)
0.74
Rhinitis
14939
1.05 (0.95,1.15)
0.34
1.11 (1.01,1.22)
0.03
1.03 (0.94,1.12)
0.54
Mental health
Anxiety
10634
0.95 (0.86,1.05)
0.29
1.05 (0.95,1.16)
0.31
0.93 (0.85,1.02)
0.14
Depression
7960
0.94 (0.84,1.05)
0.25
0.99 (0.88,1.10)
0.81
0.94 (0.85,1.05)
0.27
HADS total
9105
0.91 (0.82,1.00)
0.06
1.04 (0.94,1.16)
0.43
0.94 (0.85,1.03)
0.17
Mood disorders
5275
0.95 (0.83,1.08)
0.44
0.98 (0.86,1.12)
0.74
0.96 (0.85,1.09)
0.56
Pain
Musculoskeletal pain last 6 months
18243
0.95 (0.87,1.03)
0.2
0.94 (0.86,1.03)
0.17
1.04 (0.96,1.13)
0.36
Headaches last year
25256
1.03 (0.95,1.11)
0.48
0.97 (0.90,1.06)
0.52
1.00 (0.92,1.07)
0.9
Pain/stiffness in muscles/joints
36952
1.06 (0.98,1.14)
0.14
1.02 (0.95,1.10)
0.63
0.95 (0.88,1.01)
0.11
Pain in either legs when walking
14646
1.02 (0.93,1.12)
0.68
0.89 (0.81,0.97)
0.01
0.99 (0.91,1.08)
0.77
Pain in legs at rest
10247
1.10 (0.99,1.22)
0.08
0.98 (0.88,1.09)
0.68
1.04 (0.95,1.15)
0.39
Gastrointestinal pain last 12 months
15062
0.94 (0.86,1.04)
0.21
1.04 (0.94,1.14)
0.47
0.91 (0.84,1.00)
0.05
All-cause mortality
14695
0.93 (0.83,1.05)
0.27
0.98 (0.87,1.10)
0.72
1.00 (0.89,1.12)
0.99
0.4
0.6
0.8
1
1.2
1.4
1.6
OR and 95% CIs

PCSK9 inhibition

- Genetically determined, life-long PCSK9 inhitibion leads to lower LDL cholesterol and lower risk of coronary heart disease.
- We did not find convincing evidence of side effects.
- Such a study does not replace the need for RCTs,
- but gives information about anticipated effects and side effects long before evidence from long-term RCTs is available.

	MR Study	RCT
Cost	Low cost, mostly use of existing data	High cost, generating data
Time	Quick and efficient	Very intensive and lengthy
Population	Very large sample size, usually representative of target population	Small sample size, usually not representative of target population
Exposure	Exposed to genetic variant at birth by random	Exposed to the drug by random
Outcomes	Multiple outcomes at once, hypothesis generating, can study life time diseases	Specific outcome, hypothesis testing, difficult to study diseases with long latent period
Drug Effects	Both common and uncommon Short-term and life-time effects	Mostly common Mostly short-term effects
Ethics	No ethical issues, no exposure to the drug	Ethical issue due to exposure to a new drug

Summary

- Mendelian randomization is a way of utilizing genetic variation to study causal effects of modifiable exposures
- Mendelian randomization is not generally prone to confounding, is immune to reverse causation, and can inform about effects of life-long exposure,
- but has other limitations and relies on assumptions that are only partially testable
- MR dictionary:

https://mr-dictionary.mrcieu.ac.uk/