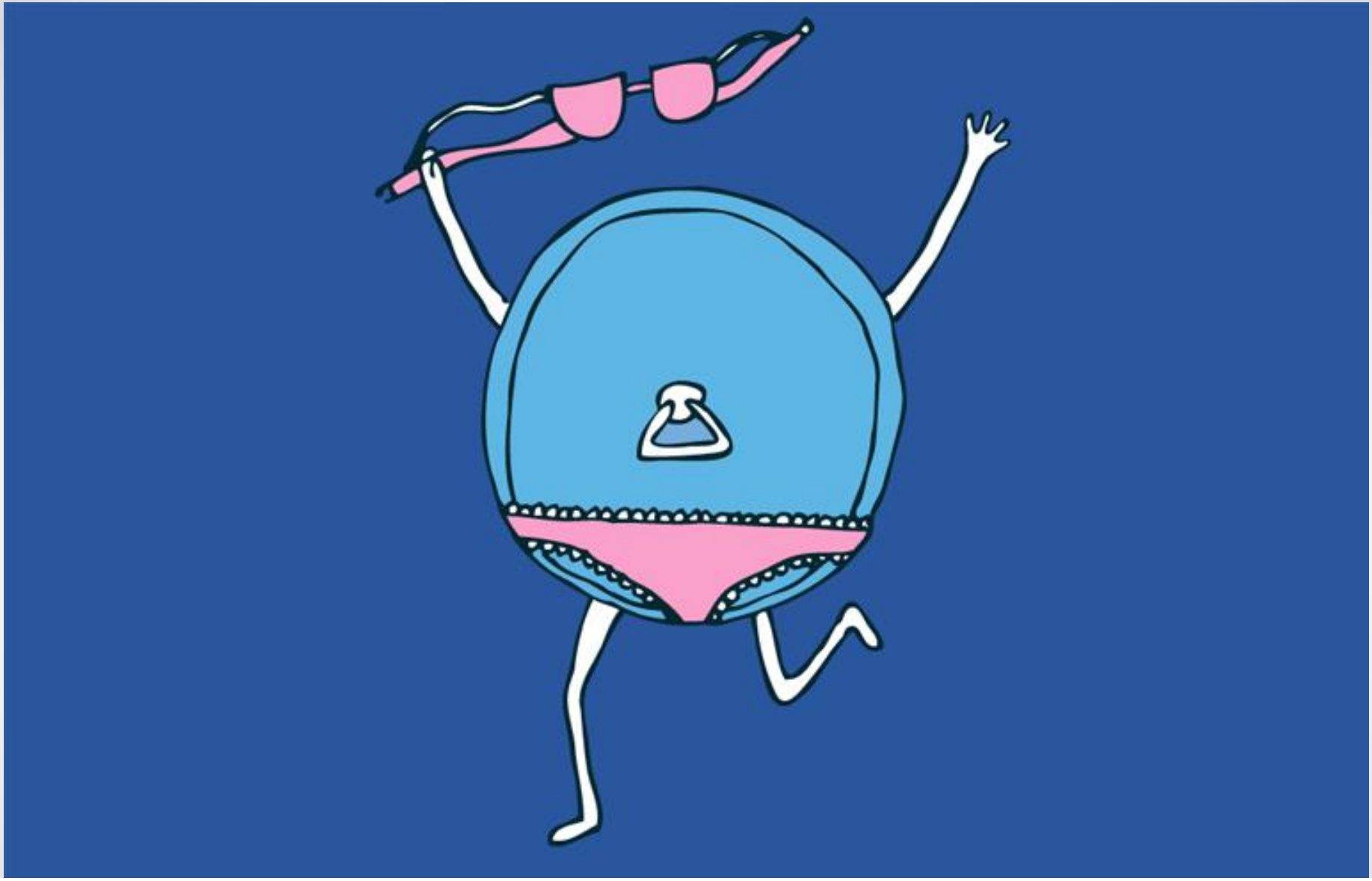
UNIVERSITY OF CAMBRIDGE

wellcome trust Fellow

Mendelian randomization: using genetics for guiding drug discovery

Stephen Burgess
MRC Biostatistics Unit and
Cardiovascular Epidemiology Unit,
University of Cambridge
PSI One Day Meeting, 29th January 2019

Twitter: @stevesphd
sb452@medschl.cam.ac.uk



www.crcpress.com/product/isbn/9781466573178

mlp pc game

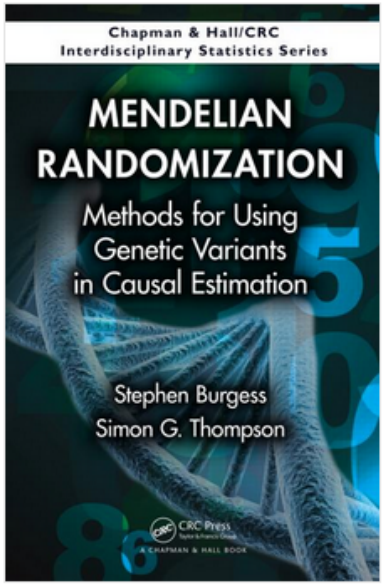
CRC Press
Taylor & Francis Group

United Kingdom | Wish List | My Account | Sign In / Register | Contact Us

SHOPPING CART

About Us | Resources | Textbooks | CRCnetBASE | Featured Authors | Search

Home / Statistics / Statistical Theory & Methods / Mendelian Randomization: M...



Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation

Stephen Burgess, Simon G. Thompson

Hardback
£38.99

February 26, 2015 **Forthcoming** by Chapman and Hall/CRC
Reference - 224 Pages - 22 B/W Illustrations
ISBN 9781466573178 - CAT# K16638
Series: Chapman & Hall/CRC Interdisciplinary Statistics

For Librarians Available on CRCnetBASE >>

FREE Standard Shipping!

Select Format:
Hardback

Quantity:
1

£38.99

Pre-Order

Add to Wish List

Description

Features

- Offers first-hand, in-depth guidance on Mendelian

Share this Title

Two-day course

- Wednesday 27th and Thursday 28th March 2019
- MRC Biostatistics Unit, Cambridge
- Details at www.mendelianrandomization.com

Observational data

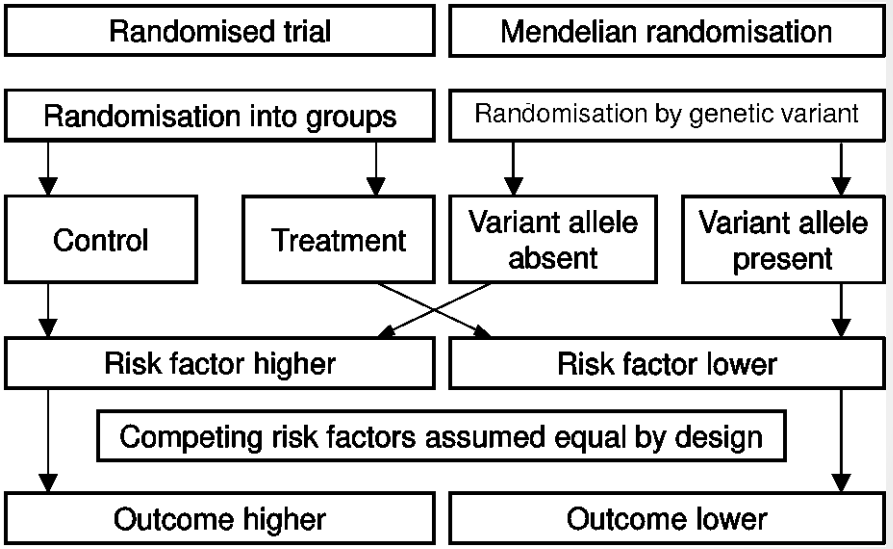
- Correlation is not causation
- Observed associations between a risk factor and an outcome may result from:
  - Confounding
  - Reverse causation

Randomized trials

- In a randomized trial, participants are randomly assigned to a treatment group
- Random allocation ensures all potential confounders are equally distributed (on average) between the groups
  - Unmeasured confounders
- Allocation of treatment gives a time-ordering
  - Reverse causation

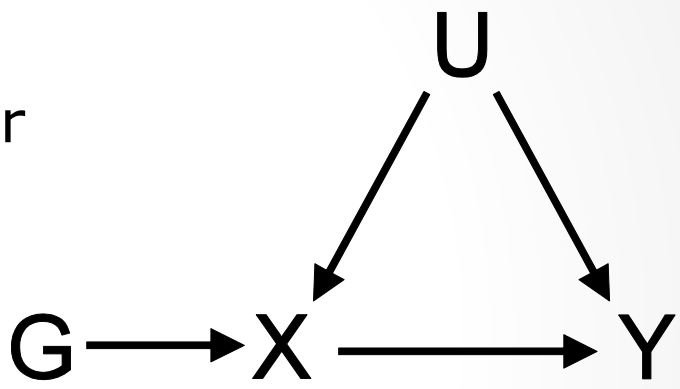
Assessment of causality

- Mendelian randomization is analogous to a randomized trial
- An association between the genetic variant and the outcome is indicative of a causal effect of the risk factor on the outcome



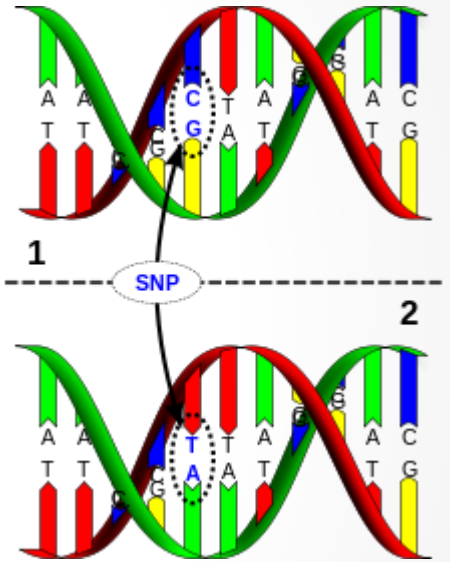
Instrumental variables

- An instrumental variable is:
  1. Associated with the risk factor of interest
  2. Not associated with any potentially confounding variable
  3. Only associated with the outcome via the risk factor
- Provides a natural experiment in observational data, similar to a randomized trial



Mendelian randomization

- Genetic variants are particularly suitable candidate instrumental variables
  - Scientific knowledge of genetic function
  - Specific association with traits
  - Genetic sequence is determined at conception
- Mendelian randomization is the use of genetic instrumental variables in observational data to obtain causal inferences

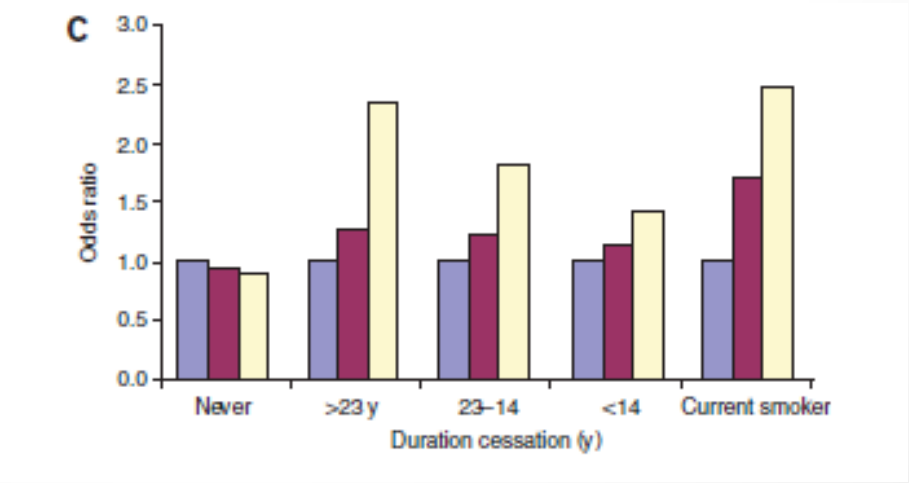


Role of Mendelian randomization in drug discovery

- Target discovery:
  - identify new targets
- Target validation: Offer insight into existing/putative targets:
  - increase or reduce enthusiasm
  - insight into safety profile
  - identify opportunities for repositioning
  - identify interactions between treatments
  - identify subgroups with greater efficacy
  - estimate impact in clinical trial

Genetics can recapitulate known causal associations

- GWAS of lung cancer identified a genomic region associated with lung cancer
  - Biological hypothesis: causal gene may regulate downstream signalling pathways and impact cell proliferation
  - Hint that SNP may not be associated in never smokers



- Genetics has "discovered" that smoking causes lung cancer!

nature Vol 452|3 April 2008|doi:10.1038/nature06846

LETTERS

A variant associated with nicotine dependence, lung cancer and peripheral arterial disease

Thorgeir E. Thorgeirsson[1]*, Frank Geller[1]*, Patrick Sulem[1]*, Thorunn Rafnar[1]*, Anna Wiste[1,2],

Table 1 | Genotype status and SQ level of 13,945 Icelandic smokers

Parameter	Genotype of rs1051730			Total *n* (frequency)	Frequency of T allele
	GG	GT	TT		
Cigarettes per day (SQ level)					
1 to 10 (0)	1,743	1,558	326	3,627 (0.260)	0.305
11 to 20 (1)	2,727	2,865	810	6,402 (0.459)	0.350
21 to 30 (2)	1,145	1,416	427	2,988 (0.214)	0.380
31 and more (3)	341	448	139	928 (0.067)	0.391
All levels (frequency)	5,956 (0.427)	6,287 (0.451)	1,702 (0.122)	13,945 (1.000)	0.347
Mean SQ level (mean ± s.d.)	1.01 ± 0.85	1.12 ± 0.86	1.22 ± 0.85	1.09 ± 0.86	

Table 4 | Association of rs1051730 allele T with LC and PAD

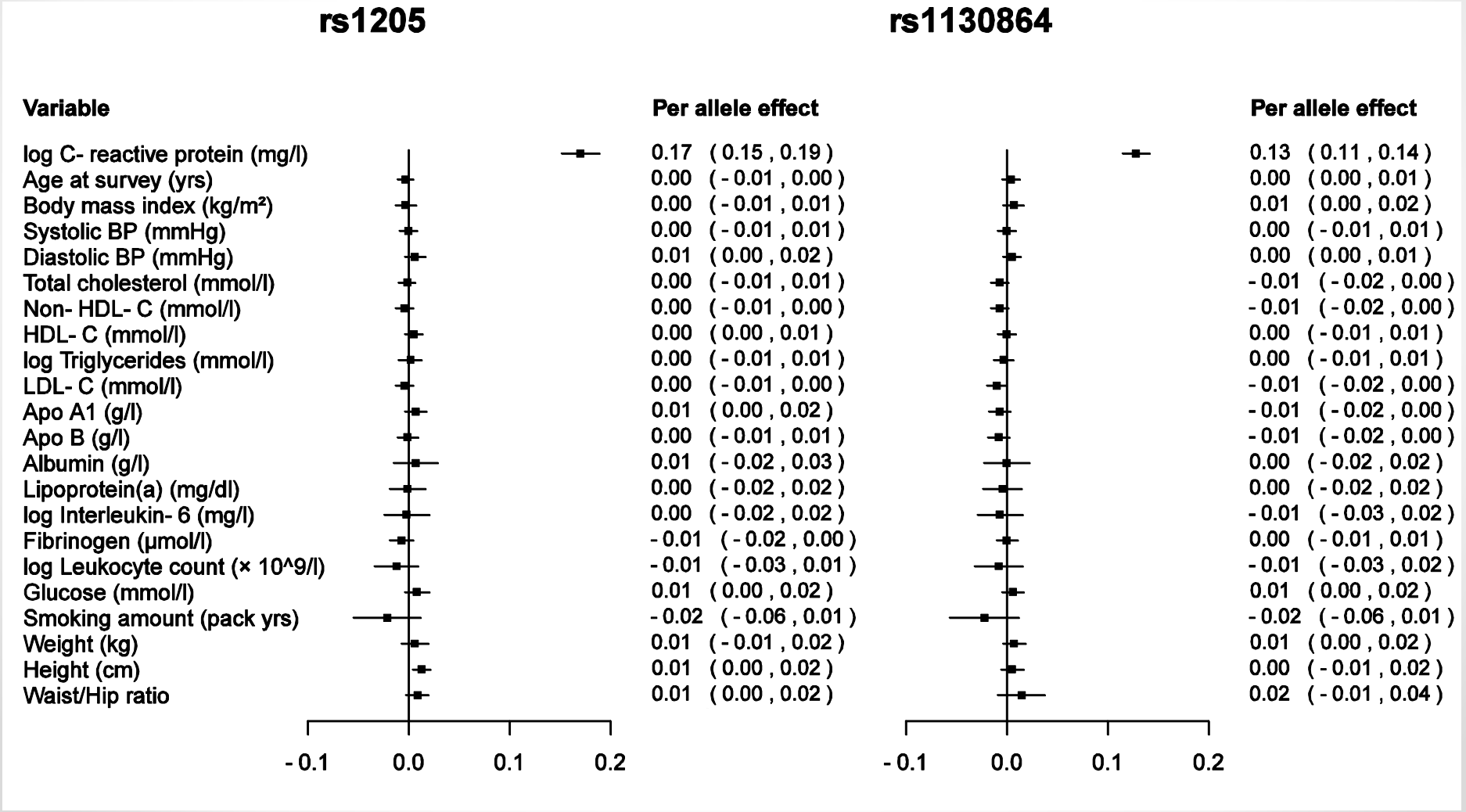
Study group	Controls		Cases		OR (95% CI)	*P*
	n	Frequency	*n*	Frequency		
LC						
Iceland	28,752	0.342	665	0.398	1.27 (1.13–1.43)	4.1×10^{-5}
Spain	1,474	0.390	269	0.483	1.46 (1.22–1.76)	5.4×10^{-5}
The Netherlands	2,018	0.314	90	0.350	1.18 (0.86–1.61)	0.31
Foreign combined	3,492	-	359	-	1.38 (1.18–1.62)	6.6×10^{-5}
All combined	32,244	-	1,024	-	1.31 (1.19–1.44)	1.5×10^{-8}
PAD						
Iceland	28,752	0.342	1,503	0.379	1.18 (1.09–1.27)	5.3×10^{-5}
New Zealand	435	0.274	441	0.337	1.35 (1.10–1.65)	0.0041
Austria	403	0.352	457	0.395	1.20 (0.99–1.46)	0.068
Sweden	140	0.304	172	0.331	1.14 (0.81–1.60)	0.46
Italy	234	0.378	165	0.412	1.15 (0.86–1.54)	0.33
Foreign combined	1,212	-	1,235	-	1.23 (1.09–1.39)	5.9×10^{-4}
All combined	29,964	-	2,738	-	1.19 (1.12–1.27)	1.4×10^{-7}

Example: C-reactive protein and coronary heart disease

- C-reactive protein is observationally associated with coronary heart disease risk
- But that does not necessarily mean that it is a causal risk factor
- Inflammation hypothesis:
  - Does inflammation lead to heart disease?
  - Are inflammatory biomarkers raised in response to pre-clinical disease?
  - Are both processes driven by another mechanism?
- Can use common genetic variation in the *CRP* gene region to assess question of causation

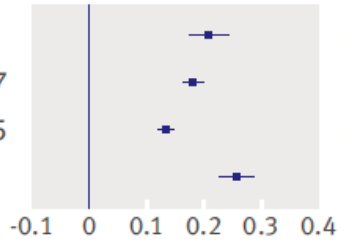
Example: CRP and CHD risk

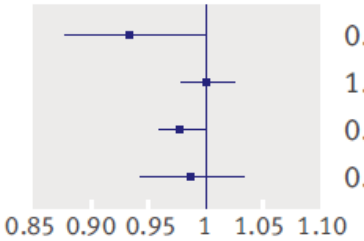
- Assessing associations with measured covariates



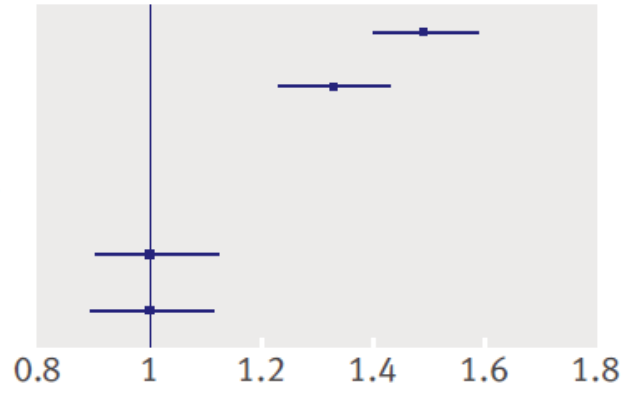
Example: CRP and CHD risk

Single nucleotide polymorphism	Allele frequency*	No of studies/cases /participants†	Per allele higher mean ln CRP (95% CI), mg/L	Per allele higher mean ln CRP (95% CI), mg/L	Per allele risk ratio for CHD (95% CI)	Per allele risk ratio for CHD (95% CI)
rs3093077	0.06	19/15 133/96 807		0.21 (0.17 to 0.24)		0.93 (0.87 to 1.00)
rs1205	0.67	43/40 527/172 567		0.18 (0.16 to 0.20)		1.00 (0.98 to 1.02)
rs1130864	0.30	41/37 145/157 905		0.13 (0.12 to 0.15)		0.98 (0.96 to 1.00)
rs1800947	0.94	31/31 636/93 507		0.26 (0.23 to 0.29)		0.99 (0.94 to 1.03)





	Risk ratio* (95% CI) for CHD per 1 SD higher ln CRP (mg/L)	Risk ratio* (95% CI) for CHD per 1 SD higher ln CRP (mg/L)
Circulating usual concentrations of CRP		
Adjusted for age, sex, and ethnicity		1.49 (1.40 to 1.59)
Further adjusted†		1.33 (1.23 to 1.43)
Genetically raised concentrations of CRP‡		
SNP analyses		1.00 (0.90 to 1.13)
Haplotype analyses		1.00 (0.89 to 1.12)



Taxonomy of MR studies

1. Risk factor is a protein biomarker

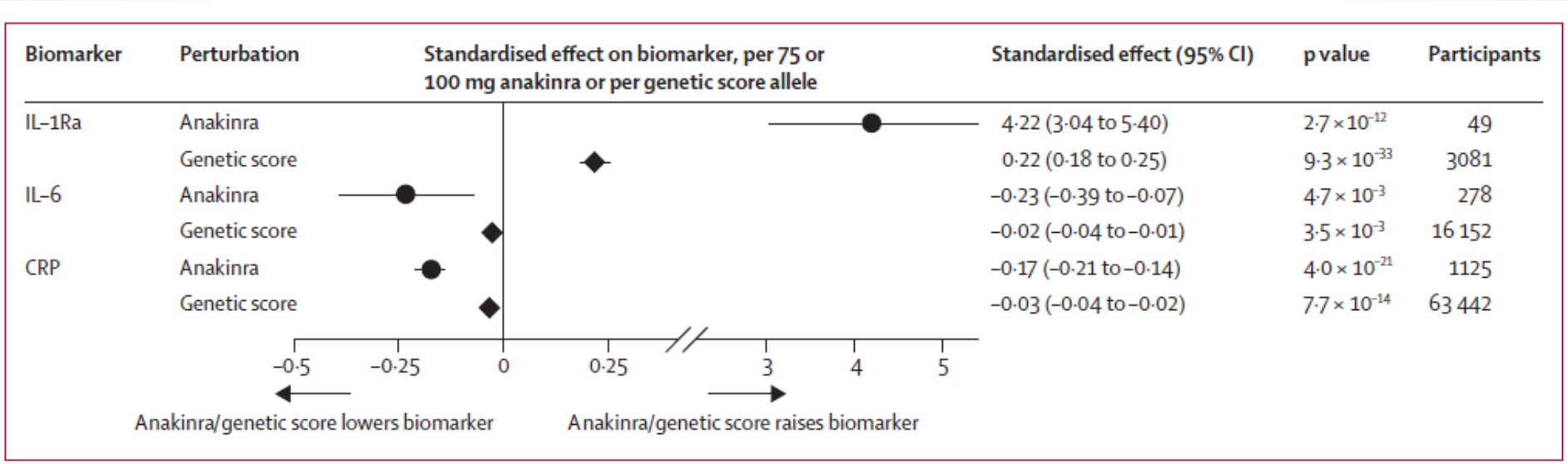
- One or a small number of genetic variants
- Usually one gene region
- Gene region has strong biological link with risk factor
- Often, we can develop drugs to act on the same biological pathway
- The most plausible assessment of causation

Example: interleukin-1 and cardiovascular outcomes

- Interleukin-1 (IL1) is another inflammatory biomarker
- We consider two genetic variants in the *IL1RN* gene region that encodes the IL-1 receptor antagonist
- Anakinra, the recombinant form of IL-1 receptor antagonist, is used in the treatment of rheumatoid arthritis

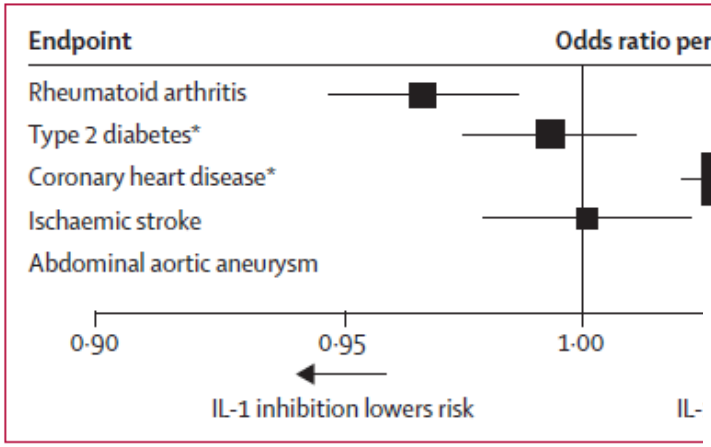
Example: IL1 and CVD risk

- We can compare known effects of anakinra on intermediate traits to the associations of the genetic variants with these traits

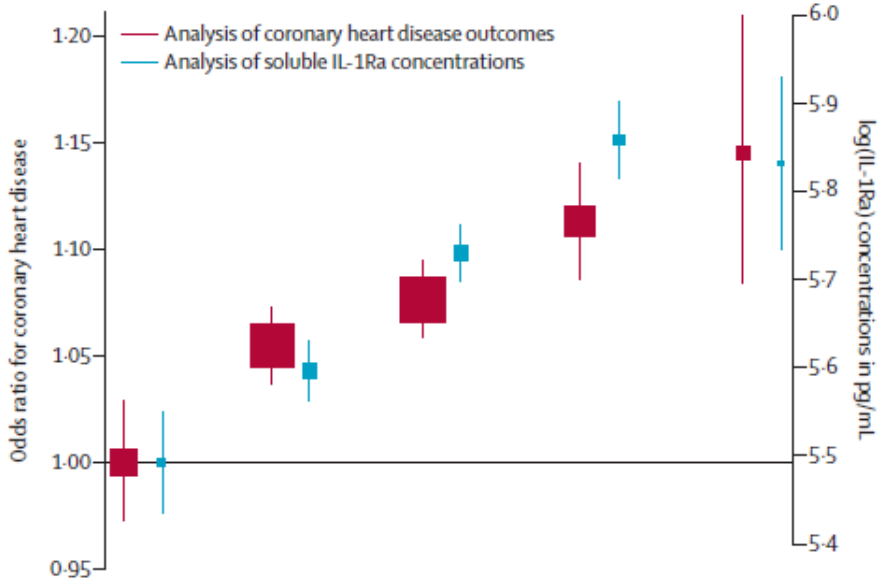


Freitag et al. "Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a Mendelian randomisation analysis", Lancet D+E 2015

Example: IL1 and CVD risk

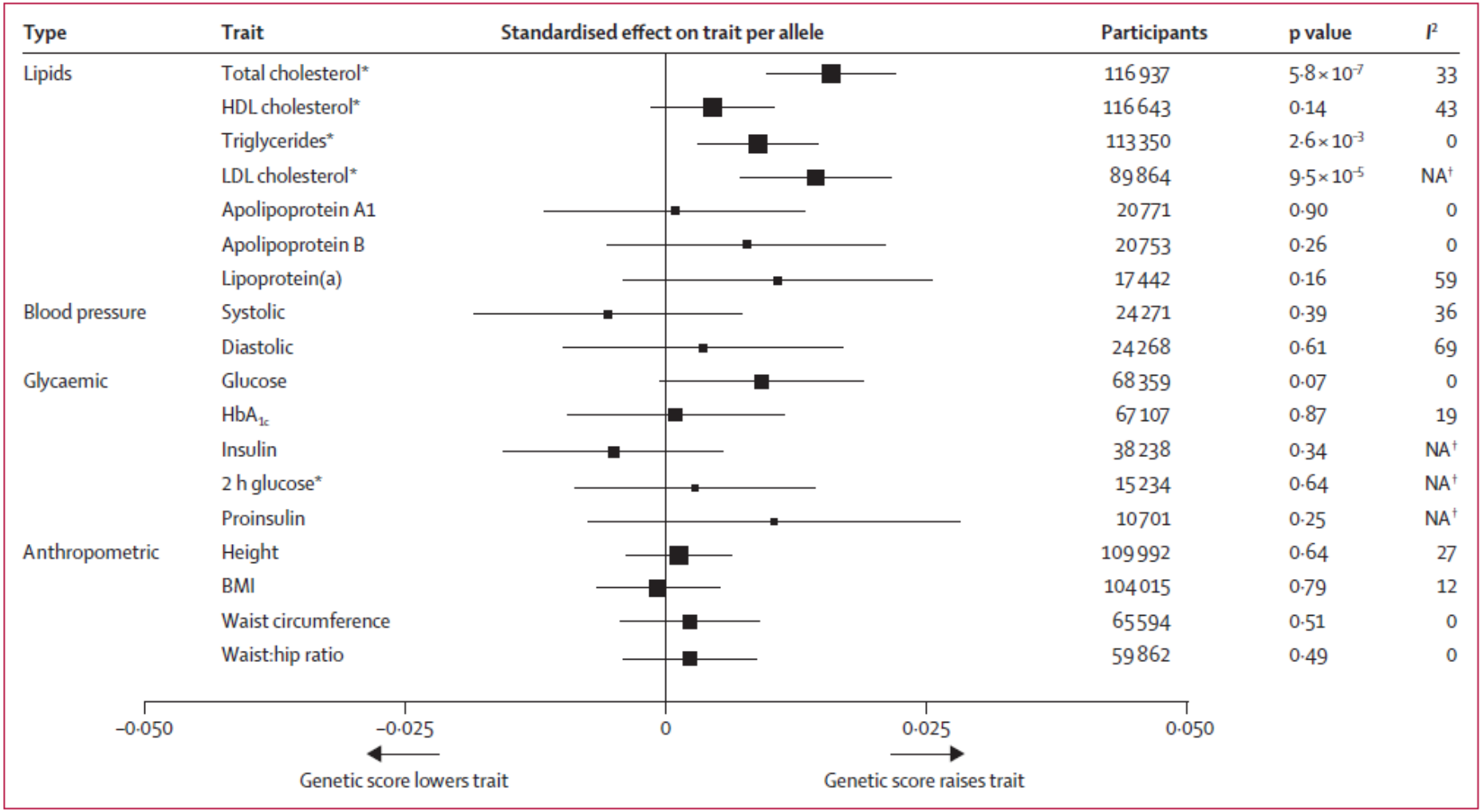


Endpoint	Odds ratio (95% CI)	p value	Cases	Controls
Rheumatoid arthritis	0·97 (0·95–0·99)	$9·9 \times 10^{-4}$	14 361	43 923
Type 2 diabetes*	0·99 (0·97–1·01)	0·47	18 715	61 692
Coronary heart disease*	1·03 (1·02–1·04)	$3·9 \times 10^{-10}$	70 532	126 374
Ischaemic stroke	1·00 (0·98–1·02)	0·92	12 389	62 004
Abdominal aortic aneurysm	1·08 (1·04–1·12)	$1·8 \times 10^{-5}$	4682	38 739



Analysis of coronary heart disease outcomes						
Cases	8121	23 029	25 679	11 446	2257	70 532
Controls	16 960	41 198	43 912	20 203	4101	126 374
Analysis of soluble IL-1Ra concentrations						
Controls	332	990	1084	564	111	3081

Example: IL1 and CVD risk

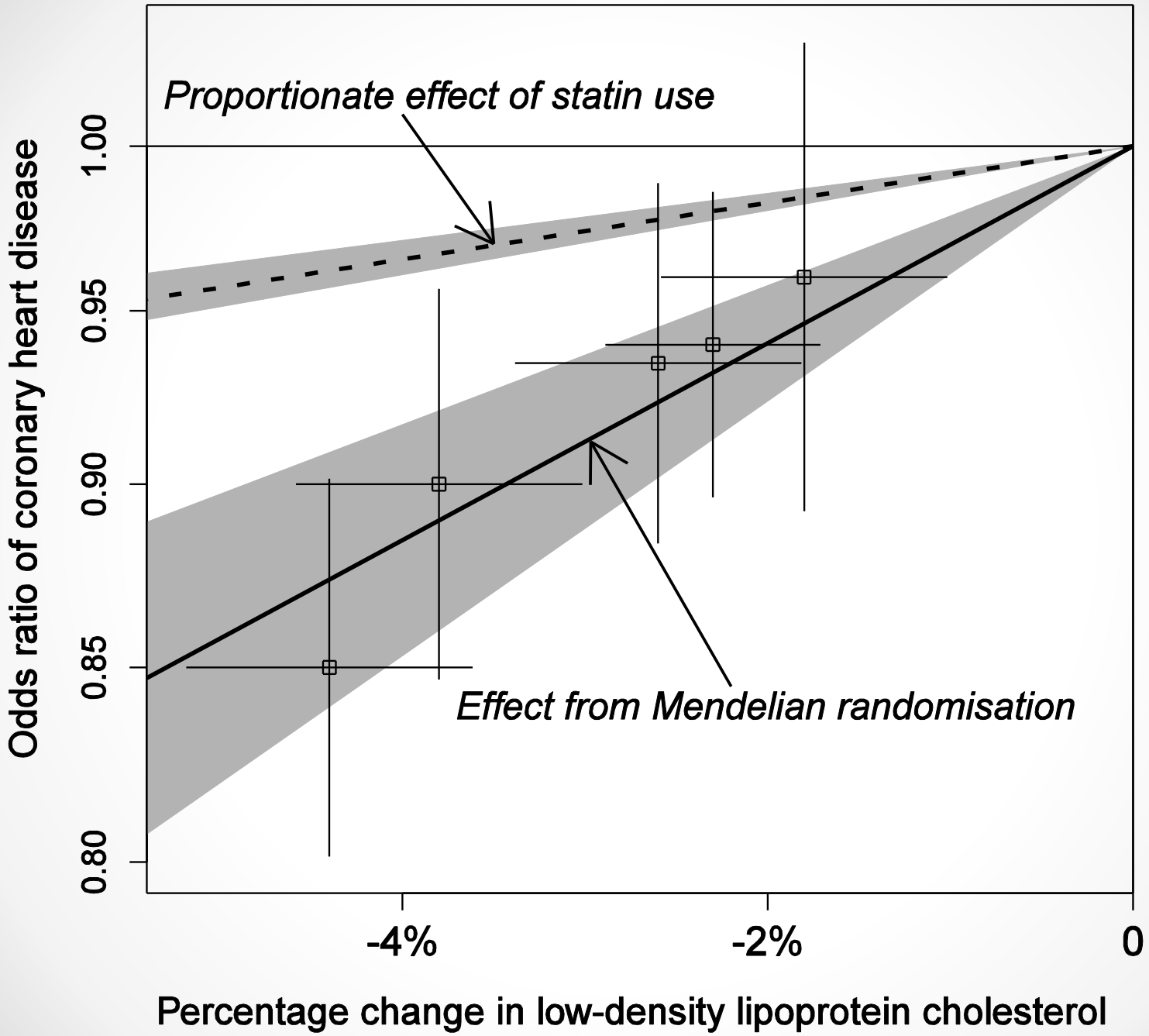


Interpretation of a causal effect

- A Mendelian randomization estimate represents the change in the outcome from a change in the risk factor corresponding to a life-long difference in the usual levels of the risk factor from conception
- This typically differs from the effect of lowering/raising the risk factor in practice, such as by a clinical/pharmacological intervention
- Why might estimates differ?
- When does this matter?

Long-term versus short-term

- Genetic effects are typically life-long, clinical interventions are often short-term
  - Atherosclerosis is a chronic condition
  - We would expect the impact of a one-day lowering in LDL-cholesterol levels to be less than that of a one-year intervention, or a five-year intervention
  - So we would expect the impact of genetically lowered LDL-c to be greater even than that of a five-year intervention
  - In contrast, we may expect there to be less difference between the MR estimate and trial estimate for the risk factor of blood pressure


Proportionate effect of statin use
Effect from Mendelian randomisation
Odds ratio of coronary heart disease
1.00
0.95
0.90
0.85
0.80
-4%
-2%
0
Percentage change in low-density lipoprotein cholesterol

Long-term versus short-term

- For lipids, the causal estimate from Mendelian randomization of the effect of LDL-c on CHD risk is 3-3.5 times the proportionate effect of statins on CHD risk
  - (assuming 5 year treatment in primary prevention: 25% lower lipids, 30% lower CHD risk)
- For blood pressure, the MR estimate is 2.5 times the effect of SBP/DBP on CHD risk

Usual versus pathological levels

- Is CRP causal for coronary heart disease risk?
  - C-reactive protein (CRP) is an acute phase reactant
  - CRP concentrations increase sharply during disease episodes
  - Genetic associations with CRP are usually estimated in healthy individuals
  - There are suggestions that genetic predictors of "usual" CRP levels are also associated with pathological concentrations
  - However, the primary question addressed in a Mendelian randomization analysis is: "do long-term elevated levels of CRP cause CHD?" and not "do acute levels of CRP cause CHD?"
  - Long-term question more relevant for disease prevention? (eg polypill)

Lipoprotein(a) and CHD risk

- Lipoprotein(a) is highly heritable
- Concentration differs by 1000 fold between individuals
- *LPA* gene region explains up to 90% of the variance in Lp(a)
  - I am somewhat sceptical of this claim, as several Lp(a) assays are isoform-sensitive, but certainly highly (60%+) heritable
- Compounds to lower Lp(a) have passed Phase 2 trials
- These compounds lower Lp(a) by up to 90%

Lipoprotein(a) and CHD risk

Genetically Elevated and Increased Risk c

Pia R. Kamstrup, MD, PhD
Anne Tybjærg-Hansen, MD, DMSc
Rolf Steffensen, MD
Børge G. Nordestgaard, MD, DMSc

Context Hi
dial infarctior

Objective
causal.

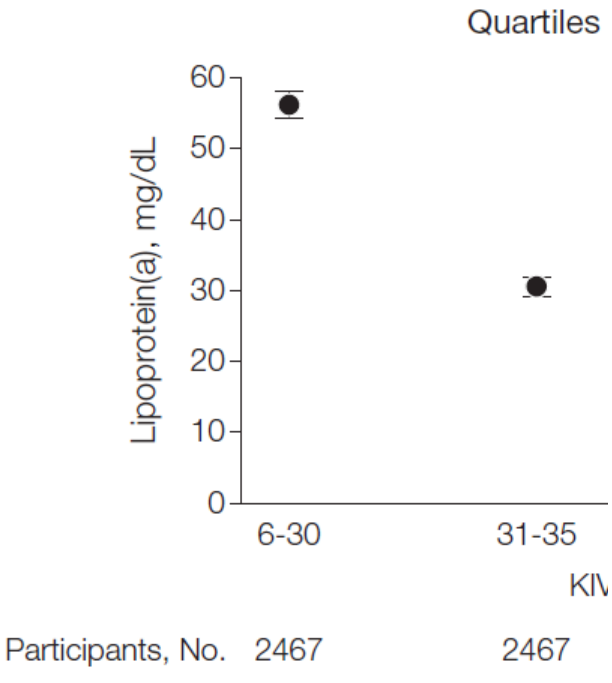


Figure 3. Risk of Myocardial Infarction by Quartiles of Apolipoprotein(a) KIV-2 Repeats in the CCHS, CGPS, and CIHDS

CCHS

Quartile	KIV-2 Repeats, No. of Repeats	Participants, No.	Events, No.
1	6-30	2160	181
2	31-35	2159	158
3	36-40	2159	137
4 [Reference]	41-99	2159	123

Adjusted for Age and Sex: *P* <.001; 0.8 1 2 4; HR (95% CI)
Multivariably Adjusted: *P* <.001; 0.8 1 2 4; HR (95% CI)

CGPS

Quartile	KIV-2 Repeats, No. of Repeats	Controls, No.	Cases, No.
1	6-30	7016	287
2	31-35	7064	246
3	36-40	7095	219
4 [Reference]	41-99	7076	234

P = .005; 0.8 1 2 4; OR (95% CI)
P = .005; 0.8 1 2 4; OR (95% CI)

CIHDS

Quartile	KIV-2 Repeats, No. of Repeats	Controls, No.	Cases, No.
1	13-30	291	325
2	31-35	304	311
3	36-40	302	313
4 [Reference]	41-96	333	282

P = .03; 0.8 1 2 4; OR (95% CI)
P = .01; 0.8 1 2 4; OR (95% CI)

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Genetic Variants Associated with Lp(a) Lipoprotein Level and Coronary Disease

Robert Clarke, F.R.C.P., John F. Peden, Ph.D., Jemma C. Hopewell, Ph.D., Theodosios Kyriakou, Ph.D., Anuj Goel, M.Sc., Simon C. Heath, Ph.D., Sarah Parish, D.Phil., Simona Barlera, M.S., Maria Grazia Franzosi, Ph.D., Stephan Rust, Ph.D., Derrick Bennett, Ph.D., Angela Silveira, Ph.D., Anders Malarstig, Ph.D., Fiona R. Green, Ph.D., Mark Lathrop, Ph.D., Bruna Gigante, M.D., Karin Leander, Ph.D., Ulf de Faire, M.D., Udo Seedorf, Ph.D., Anders Hamsten, F.R.C.P., Rory Collins, F.R.C.P., Hugh Watkins, F.R.C.P., and Martin Farrall, F.R.C.Path., for the PROCARDIS Consortium*

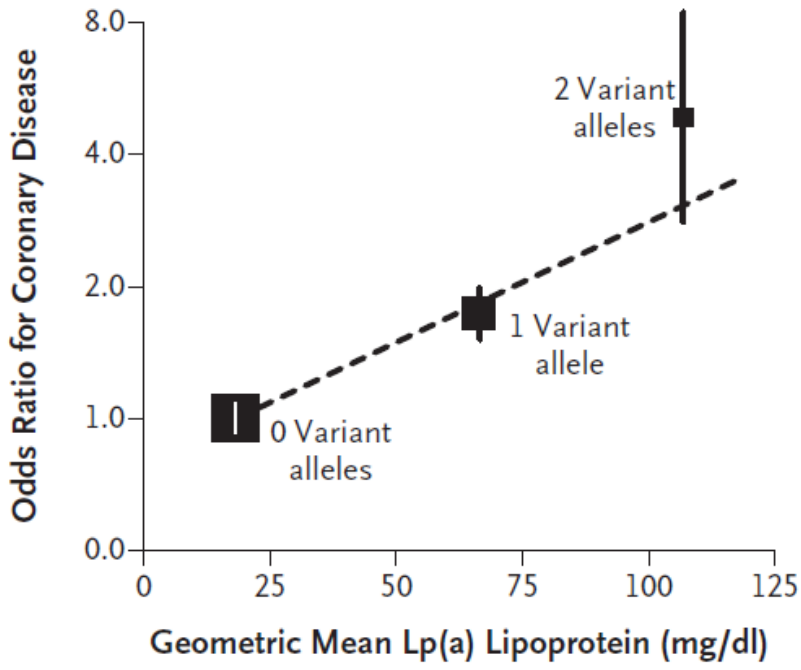
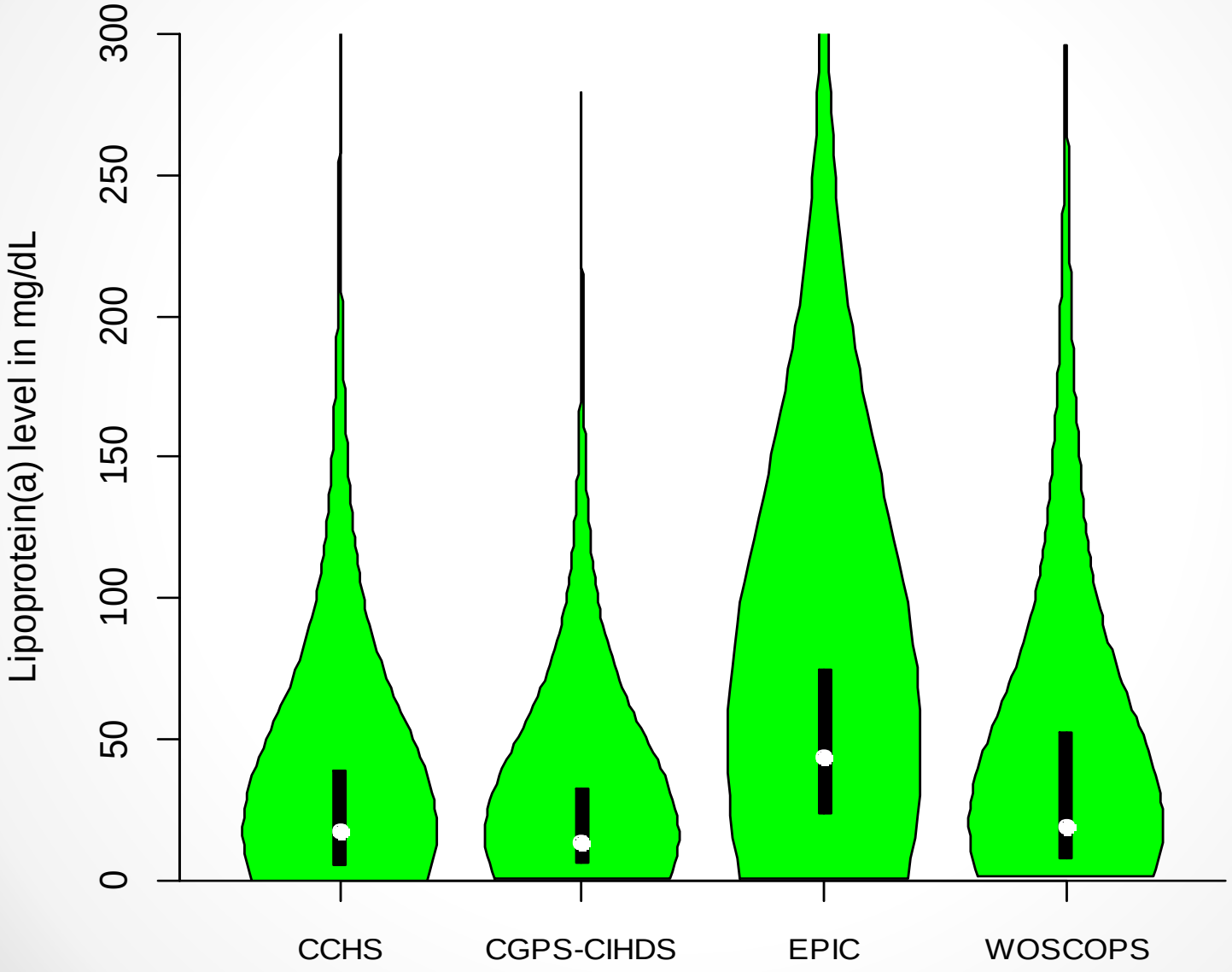


Figure 3. Association of the *LPA* Genotype Score with the Lp(a) Lipoprotein Level and the Risk of Coronary Disease in the PROCARDIS Cohort.

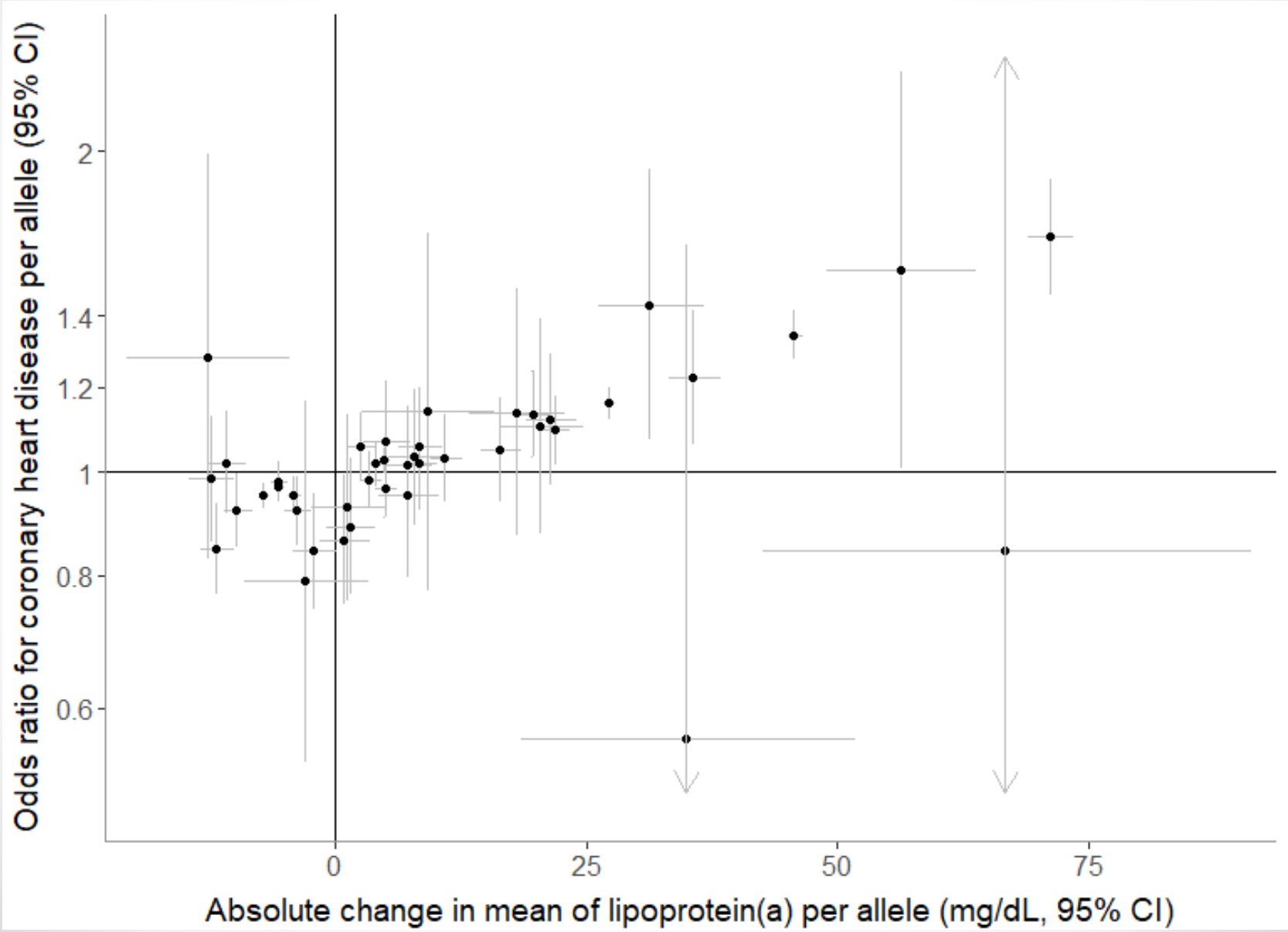
Lipoprotein(a) and CHD risk

- However, several previous treatments that lower Lp(a) by 30-40% have not been successful in Phase 3 trials
- To what extent must we lower Lp(a) to improve disease risk?
- We considered ultra-fine mapping in the LPA gene region

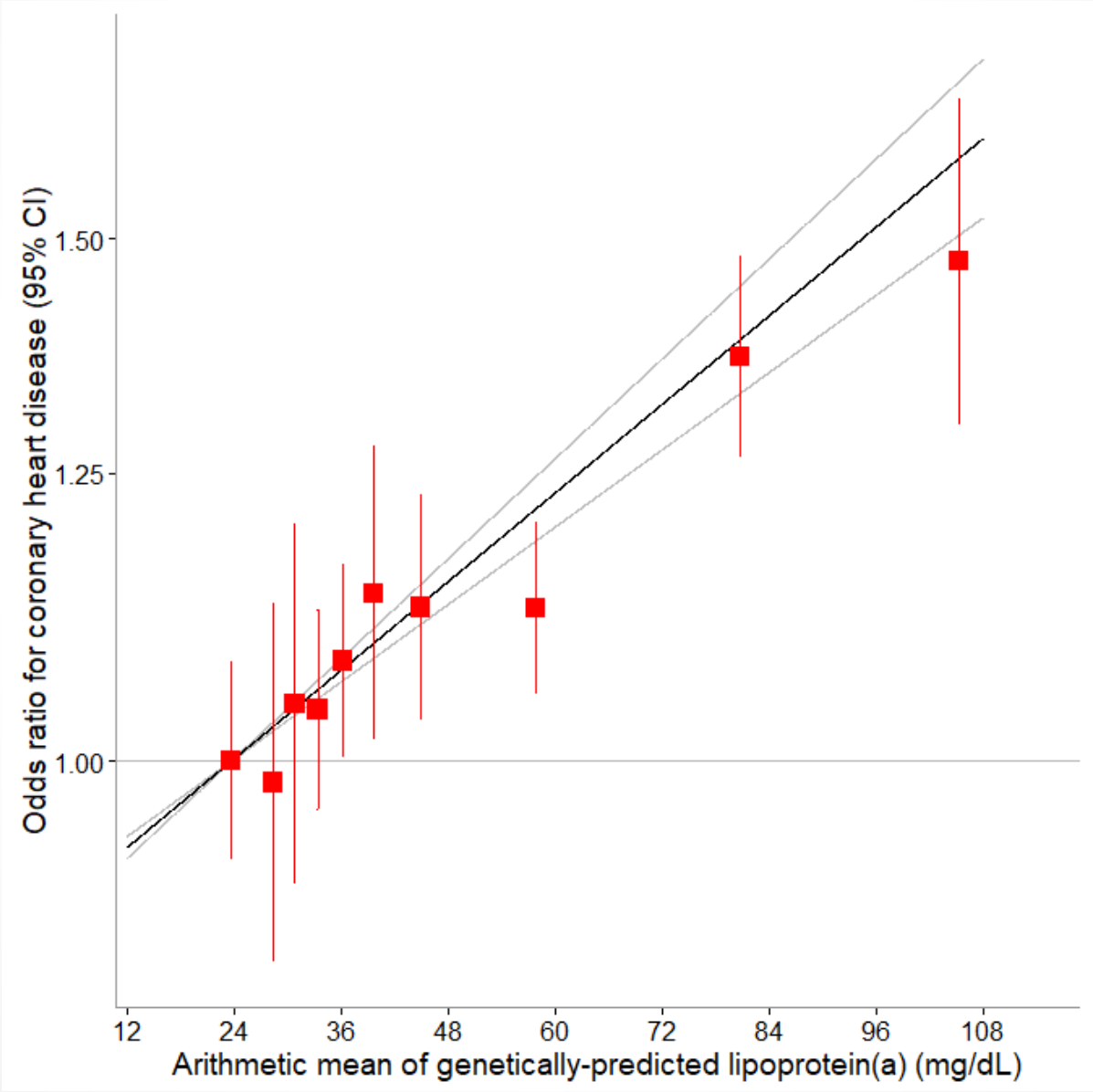
Lipoprotein(a) and CHD risk



Lipoprotein(a) and CHD risk



Lipoprotein(a) and CHD risk



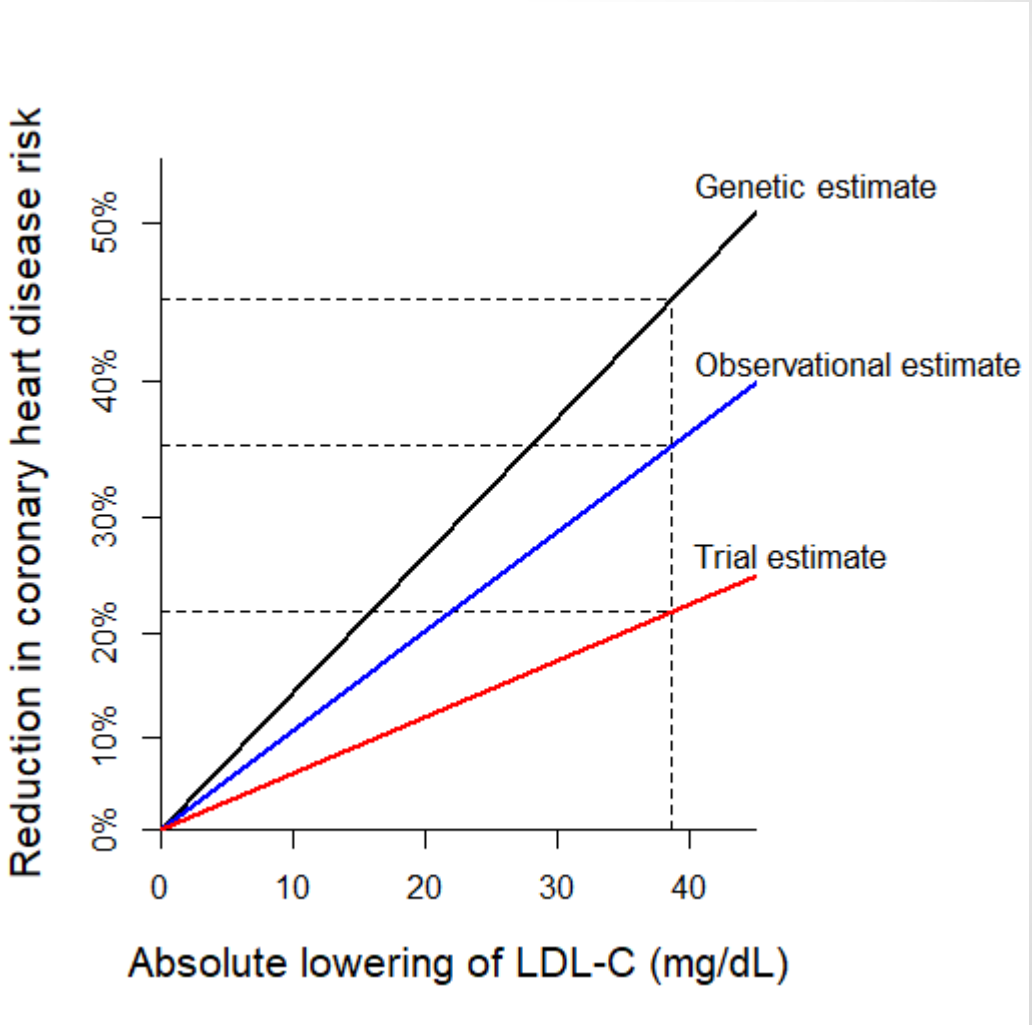
Lipoprotein(a) and CHD risk

- Linear relationship between Lp(a) and log-CHD risk
- Makes more sense to model Lp(a) on absolute scale rather than log-transformed scale
- Changes in CHD risk are proportional to absolute change in LDL-cholesterol concentration
- Need to trim/winsorize to account for outliers

- But MR estimate represents effect of life-long intervention in Lp(a)…
- … how to find effect of short-term lowering?

Lipoprotein(a) and CHD risk

- We assume that Lp(a) and LDL-cholesterol are similar
- In that ratio of life-long (MR) to short-term effect is assumed to be same between Lp(a) and LDL-cholesterol



Lipoprotein(a) and CHD risk

Approach 1:

Step 1:

$$\frac{\text{MR estimate for lipoprotein(a) per 10mg/dL lowering}}{\text{MR estimate for LDL−cholesterol per 10mg/dL lowering}} = \text{Ratio of atherogenenicity}$$

Step 2:

Proposed magnitude of lipoproprotein(a) reduction in mg/dL × Ratio of atherogenenicity = Equivalent reduction in LDL−cholesterol in mg/dL

Step 3:

$$\exp\left(\frac{\text{Equivalent reduction in LDL−cholesterol}}{38.67} \times \log(0.78)\right) =$$

Predicted odds ratio for CHD risk reduction in short−term trial

Approach 2:

Step 1:

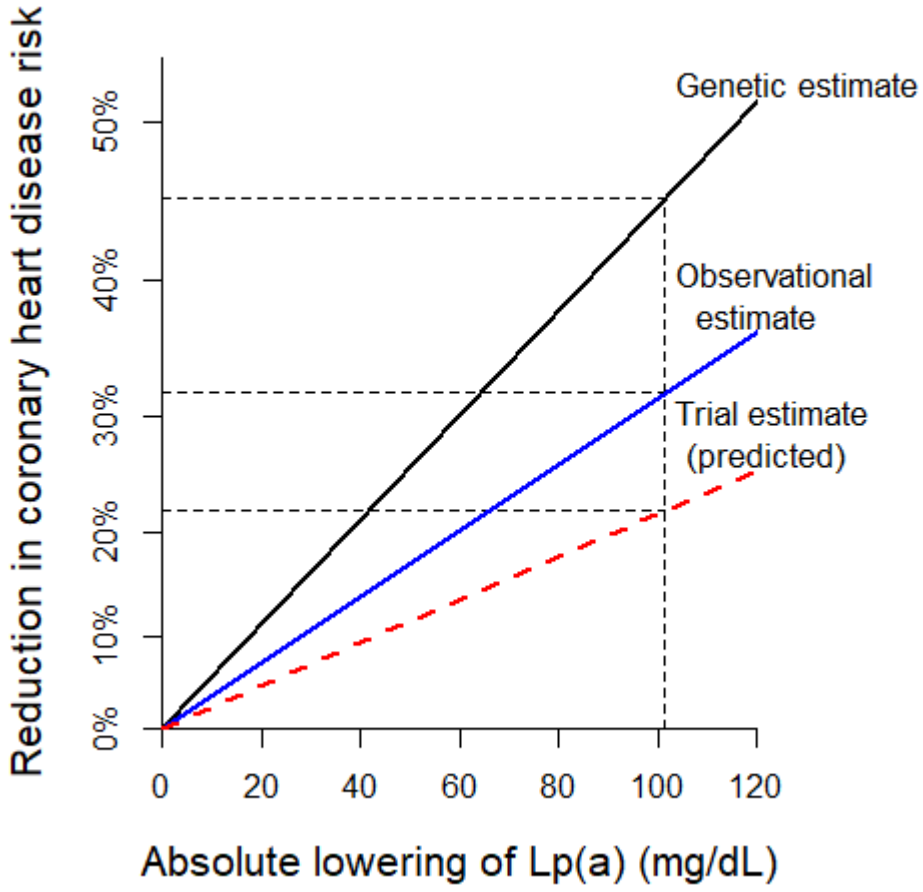
$$\frac{\text{Trial estimate for LDL−cholesterol per 1mmol/L lowering}}{\text{MR estimate for LDL−cholesterol per 1mmol/L lowering}} = \text{Ratio of short−term to life−long effects}$$

Step 2:

exp(MR estimate for lipoprotein(a) lowering × Ratio of short−term to life−long effects) = Predicted odds ratio for CHD risk reduction in short−term trial

Lipoprotein(a) and CHD risk

- Based on the MR estimates, a ~100 mg/dL change in Lp(a) is equivalent to a 38.67mg/dL (1mmol/L) change in LDL-cholesterol
- Hence short-term 100 mg/dL change in Lp(a) will have similar effect to statins in a short-term trial



Lipoprotein(a) and CHD risk

Table 2: Expected Clinical Benefit of Lowering Lp(a)

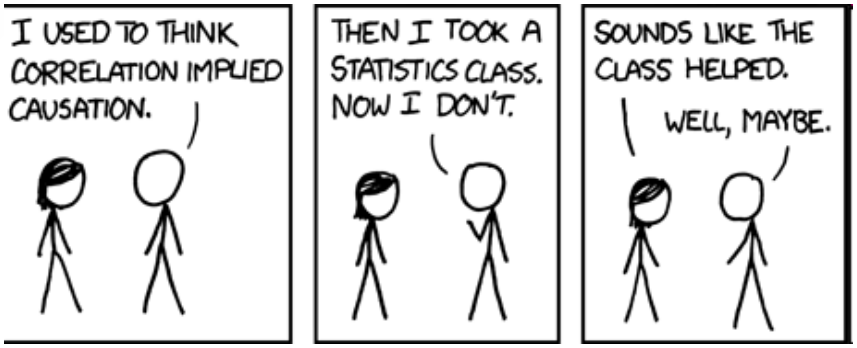
Reduction in Lp(a) (mg/dL)	Reduction in LDL-C for equivalent CHD risk reduction (mg/dL)[1]	Estimated life-long proportional risk reduction due to genetically decreased exposure (%)[2]	Estimated short-term proportional risk reduction in randomized trial (%)[3]
120	45.7 (34.1, 65.4)	51.1 (45.5, 56.2)	27.7 (20.9, 37.5)
100	38.1 (28.4, 54.5)	44.9 (39.7, 49.8)	23.7 (17.8, 32.4)
80	30.5 (22.7, 43.6)	38.0 (33.2, 42.3)	19.4 (14.5, 26.9)
50	19.0 (14.2, 27.3)	25.8 (22.3, 29.1)	12.6 (9.3, 17.8)
30	11.4 (8.5, 16.4)	16.4 (14.1, 18.7)	7.8 (5.7, 11.1)
20	7.6 (5.7, 10.9)	11.3 (9.6, 12.9)	5.3 (3.8, 7.5)
10	3.8 (2.8, 5.5)	5.8 (4.9, 6.7)	2.7 (1.9, 3.9)
5	1.9 (1.4, 2.7)	2.9 (2.5, 3.4)	1.3 (1.0, 1.9)

Lipoprotein(a) and CHD risk

- Only ~1% of the population have Lp(a) >100mg/dL
- For someone with Lp(a) 20mg/dL, a 30-40% reduction would not provide sufficient benefit
- Hence need to target those with high Lp(a)
- Drugs may be more effective in African descent individuals, as they have higher average Lp(a) concentrations
- Will be ineffective in East Asians
- But, the lifelong effect of Lp(a) >160mg/dL is similar OR to FH – and similar prevalence
- Screening strategy (genetic or phenotypic)

Summary

- Genetic variants can be used to obtain causal inferences about modifiable risk factors from observational data
- Mendelian randomization investigations are becoming easier to perform and more powerful due to summarized data
- Question marks:
  - how to choose variants?
  - how to interpret?
  - how to prioritize?
- Quality of genetic variants as instrumental variables (genetic proxies) is crucial to the quality of evidence
- Estimand: what are we estimating?


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Thank you!