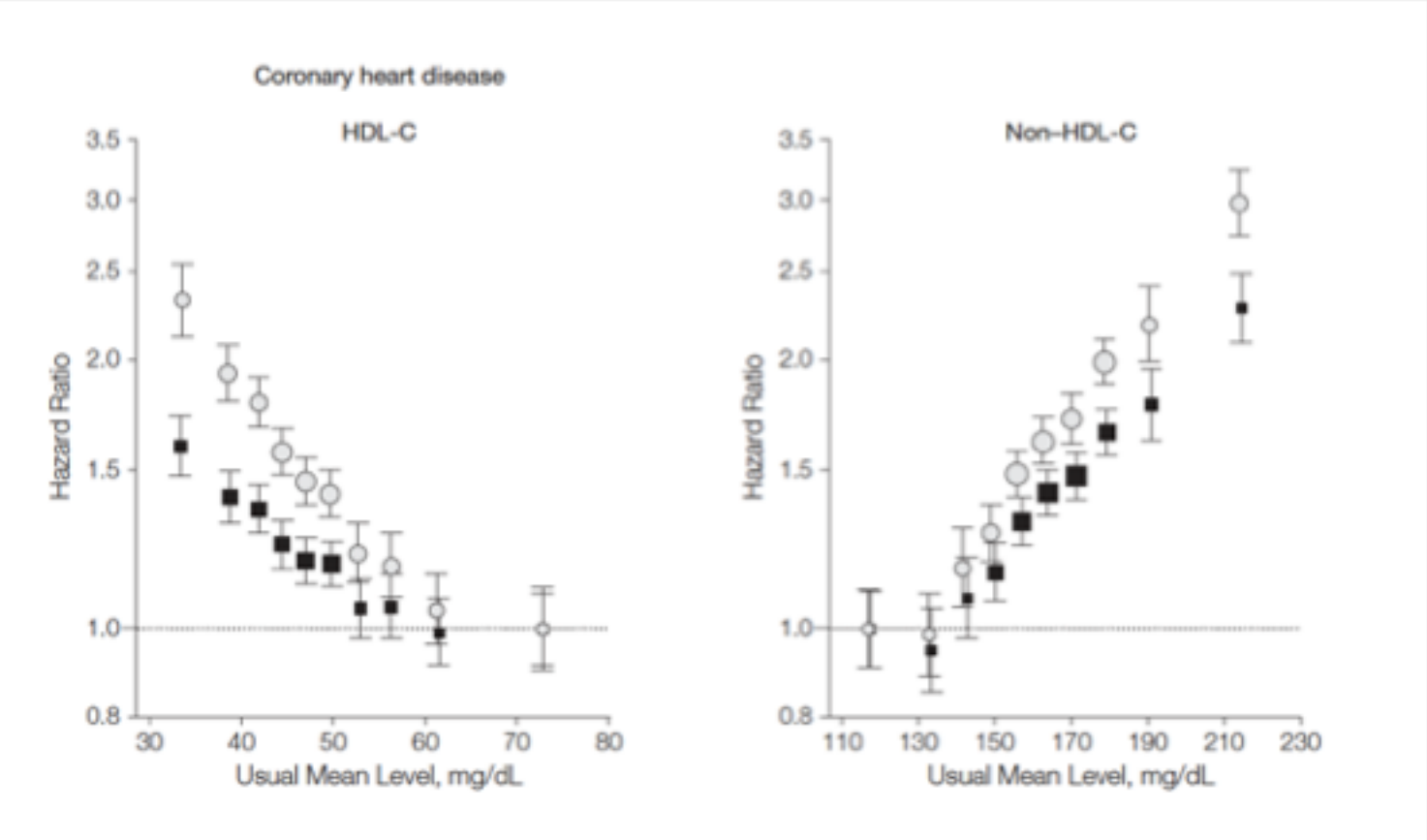
GWAS 11

Matti Pirinen
University of Helsinki
27.2.2019

CAUSAL INFERENCE



Emerging Risk Factor Collaboration JAMA 2009

- High LDL and low HDL cholesterol levels are **associated** with increased risk of heart disease
- Are these cholesterol levels **causal** for MI risk?
  - Can we decrease MI risk by decreasing LDL cholesterol, and/or by increasing HDL cholesterol?
  - Essential question for medicine !

CAUSAL INFERENCE

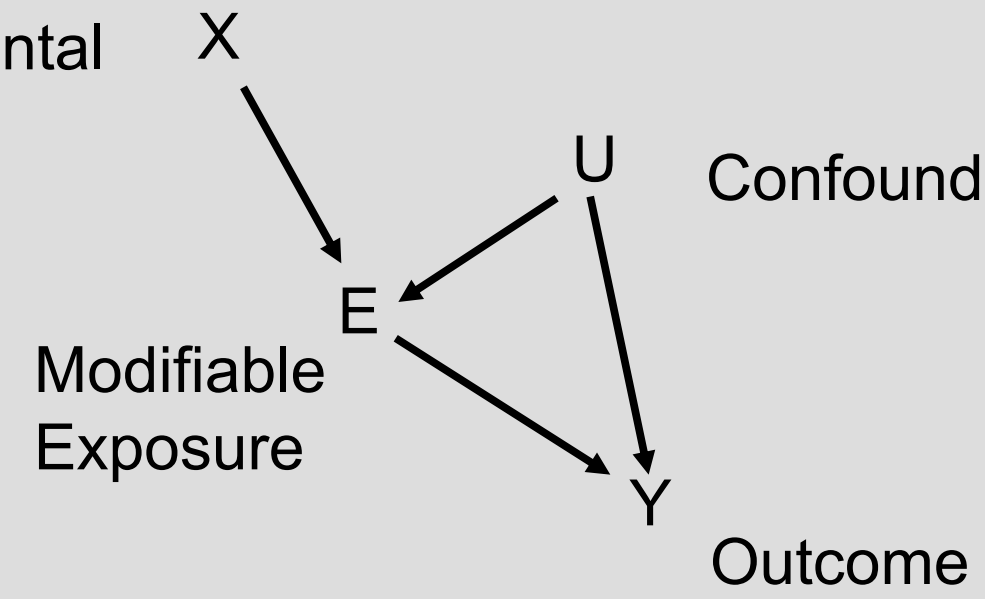
- Causality difficult or impossible to get from observational studies
  - Observed correlation does not mean causation
    - Confounders (unmeasured variables that affect the outcome AND the putative effect)
  - Reverse causation (outcome affects the putative effect)
- Randomized clinical trials are the gold standard
  - But expensive, take long time and not always possible (smoking & lung cancer)

MENDELIAN RANDOMIZATION (MR)

- I want to know whether by decreasing LDL-C level risk of heart disease is decreased
- I know from a lipid GWAS that individuals that carry a minor allele of SNP ‘S' have on average lower levels of LDL than others
- Are carriers of minor allele at SNP ‘S' protected from heart disease?
  - Why is there no reverse causation?
  - Under which assumption is there no confounding?

ASSUMPTIONS OF MR

SNP, ("Instrumental variable")



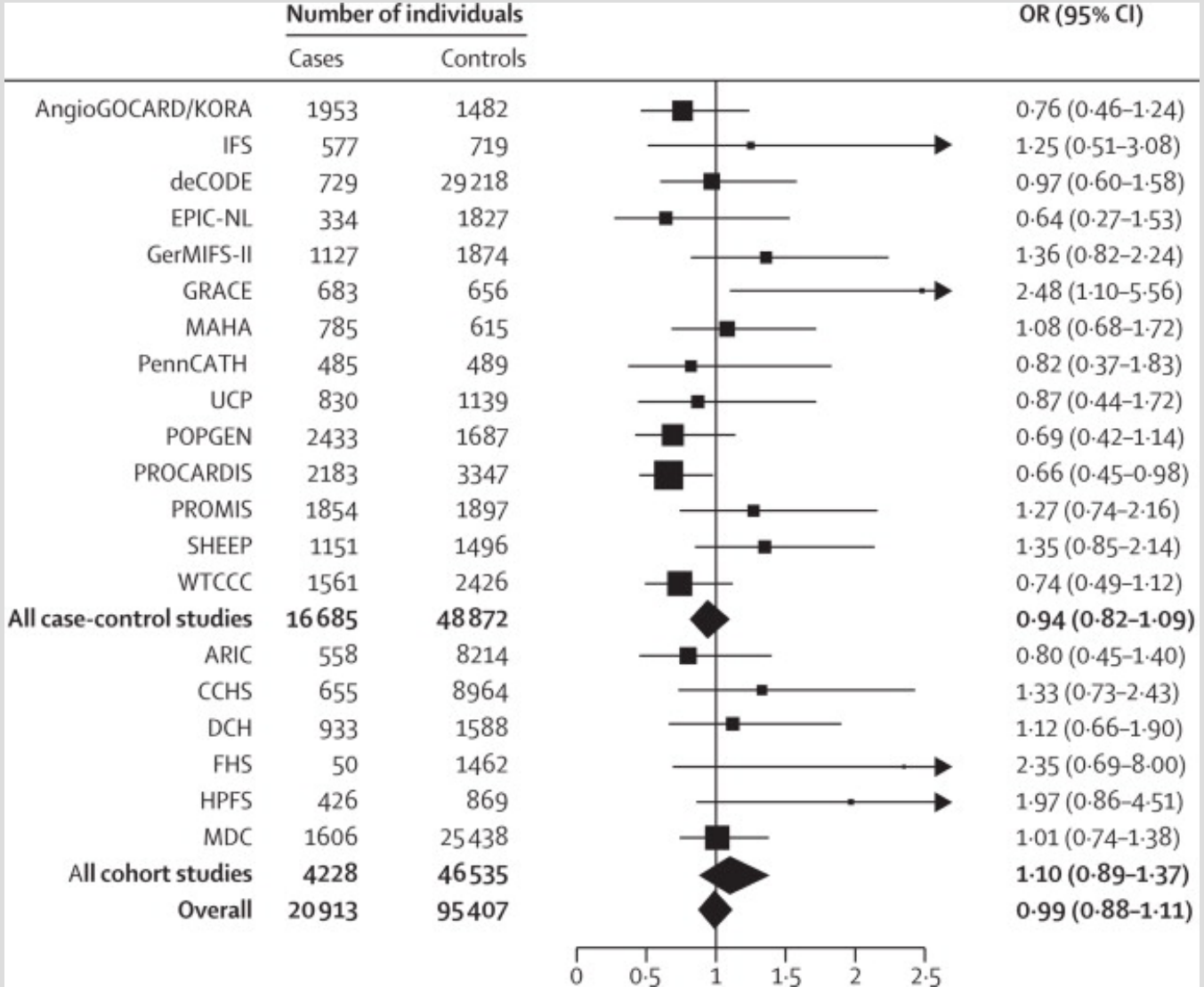
e.g.
X is a lipid SNP
E is HDL-C level
Y is heart event
U is diet

1. X is associated with E
2. X is independent of U
3. X is independent of Y given E and U

VOIGHT ET AL. 2012 LANCET

- Background: High plasma HDL cholesterol is associated with reduced risk of myocardial infarction, but whether this association is causal is unclear. Exploiting the fact that genotypes are randomly assigned at meiosis, are independent of non-genetic confounding, and are unmodified by disease processes, mendelian randomisation can be used to test the hypothesis that the association of a plasma biomarker with disease is causal.
- Methods: We performed two mendelian randomisation analyses. First, we used as an instrument a single nucleotide polymorphism (SNP) in the endothelial lipase gene (LIPG Asn396Ser) and tested this SNP in 20 studies (20 913 myocardial infarction cases, 95 407 controls). Second, we used as an instrument a genetic score consisting of 14 common SNPs that exclusively associate with HDL cholesterol and tested this score in up to 12 482 cases of myocardial infarction and 41 331 controls. As a positive control, we also tested a genetic score of 13 common SNPs exclusively associated with LDL cholesterol.

VOIGHT ET AL. 2012 LANCET



	Number of individuals		OR (95% CI)
	Cases	Controls	
AngioGOCARD/KORA	1953	1482	0·76 (0·46–1·24)
IFS	577	719	1·25 (0·51–3·08)
deCODE	729	29218	0·97 (0·60–1·58)
EPIC-NL	334	1827	0·64 (0·27–1·53)
GerMIFS-II	1127	1874	1·36 (0·82–2·24)
GRACE	683	656	2·48 (1·10–5·56)
MAHA	785	615	1·08 (0·68–1·72)
PennCATH	485	489	0·82 (0·37–1·83)
UCP	830	1139	0·87 (0·44–1·72)
POPGEN	2433	1687	0·69 (0·42–1·14)
PROCARDIS	2183	3347	0·66 (0·45–0·98)
PROMIS	1854	1897	1·27 (0·74–2·16)
SHEEP	1151	1496	1·35 (0·85–2·14)
WTCCC	1561	2426	0·74 (0·49–1·12)
All case-control studies	**16685**	**48872**	**0·94 (0·82–1·09)**
ARIC	558	8214	0·80 (0·45–1·40)
CCHS	655	8964	1·33 (0·73–2·43)
DCH	933	1588	1·12 (0·66–1·90)
FHS	50	1462	2·35 (0·69–8·00)
HPFS	426	869	1·97 (0·86–4·51)
MDC	1606	25438	1·01 (0·74–1·38)
All cohort studies	**4228**	**46535**	**1·10 (0·89–1·37)**
Overall	**20913**	**95407**	**0·99 (0·88–1·11)**

Association of LIPG Asn396Ser with myocardial infarction in 116,320 participants from 20 studies. In each study, the HDL-cholesterol-raising serine allele was modelled.

VOIGHT ET AL. 2012 LANCET

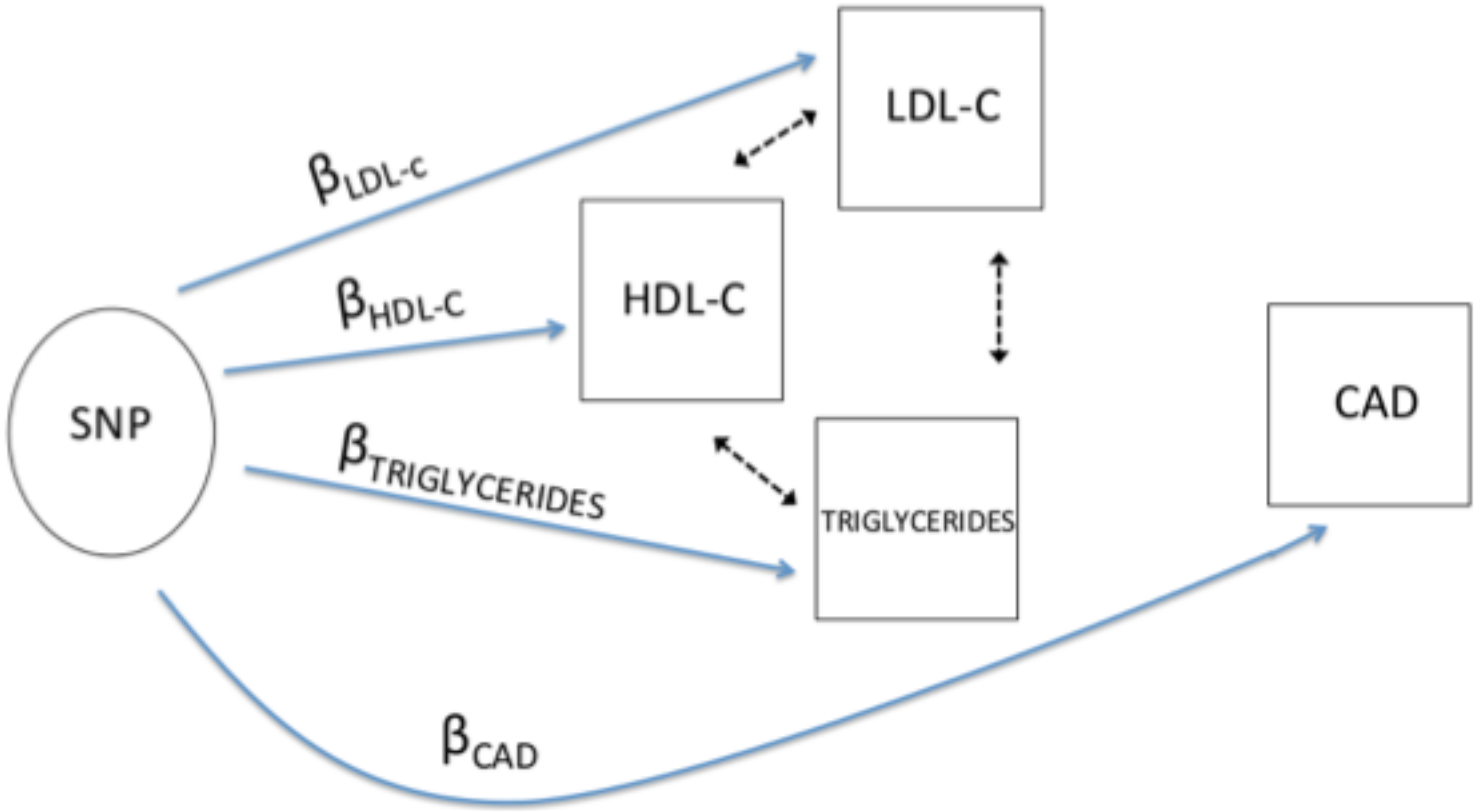
	Odds ratio (95% CI) per SD increase in plasma lipid based on observational epidemiology	Odds ratio (95% CI) per SD increase in plasma lipid conferred by genetic score
LDL cholesterol	1·54 (1·45–1·63)	2·13 (1·69–2·69), p=2e−10
HDL cholesterol	0·62 (0·58–0·66)	0·93 (0·68–1·26), p=0·63

Both have clear observational association with MI,
but only LDL-C has evidence of causality from genetics

VOIGHT ET AL. 2012 LANCET

- In summary, our results showed that polymorphisms related to plasma LDL cholesterol were consistently associated with risk of myocardial infarction, whereas this was not the case for variants related to plasma HDL cholesterol.
- A polymorphism in the endothelial lipase gene and a genetic score of 14 common SNPs that specifically raised HDL cholesterol were not associated with myocardial infarction, suggesting that some genetic mechanisms that raise HDL cholesterol do not lower risk of myocardial infarction.
- Hence, interventions (lifestyle or pharmacological) that raise plasma HDL cholesterol cannot be assumed ipso facto to lead to a corresponding benefit with respect to risk of myocardial infarction.

DO ET AL. 2013 NATURE GENETICS



- Triglycerides and HDL-C and LDL-C highly correlated and SNPs affect several traits
  - Mendelian randomisation assumptions are not met
- With many SNPs we can regress the effect on the outcome (here CAD) on the effects on TG, HDL and LDL and see which are relevant predictors in the JOINT model

DO ET AL. 2013 NATURE GENETICS

Outcome	Predictor	Covariate	β	s.e.m.	P
		–	0.41	0.039	4×10^{-20}
β_{CAD}	$\beta_{LDL\text{-}C}$	$\beta_{HDL\text{-}C}$	0.38	0.039	9×10^{-19}
		$\beta_{triglycerides}$	0.40	0.034	1×10^{-23}
		$\beta_{HDL\text{-}C}$, $\beta_{triglycerides}$	0.38	0.034	2×10^{-22}
		–	−0.18	0.052	0.0006
β_{CAD}	$\beta_{HDL\text{-}C}$	$\beta_{LDL\text{-}C}$	−0.12	0.041	0.005
		$\beta_{triglycerides}$	−0.09	0.048	0.057
		$\beta_{LDL\text{-}C}$, $\beta_{triglycerides}$	−0.04	0.037	0.35
		–	0.44	0.074	2×10^{-8}
β_{CAD}	$\beta_{triglycerides}$	$\beta_{LDL\text{-}C}$	0.42	0.057	5×10^{-12}
		$\beta_{HDL\text{-}C}$	0.36	0.074	3×10^{-6}
		$\beta_{LDL\text{-}C}$, $\beta_{HDL\text{-}C}$	0.36	0.057	1×10^{-9}

Residuals for β_{CAD} were calculated after adjustment of a SNP's effect on the denoted lipid trait. A total of 185 SNPs identified from GWAS for LDL-C, HDL-C and triglycerides were included in regression analysis. $\beta_{LDL\text{-}C}$, $\beta_{HDL\text{-}C}$ and $\beta_{triglycerides}$ represent the effect sizes for a SNP on LDL-C, HDL-C and triglycerides, respectively, in the GWAS meta-analysis for lipids. Regression was performed with the predictor variable of the effect size on lipid traits (β estimate from predictor column) and the outcome variable of residual CAD effect size after adjusting for covariates.

DO ET AL. 2013 NATURE GENETICS

- In summary, we use common polymorphisms and employ a statistical framework to dissect causal influences among a set of correlated biomarkers. By applying this framework to a correlated set of plasma lipid measures and CAD risk, we suggest a causal role of triglyceride-rich lipoproteins in the development of CAD.
- (See Notes 11 for this analysis.)