D. Lawlor & R. Harbord
U. Bristol
July '07

Approaches to estimating **causal** effects

- Medical statistics and epidemiology
  - Measure all the confounders, then control for them in regression models
  - Deal with reverse causality in prospective studies (early years removed) (b/c the deaths = people who were sick already)
  - Often difficult to eliminate unmeasured/residual confounding or reverse causality as an explanation for findings
- Econometrics
  - Find an **instrumental variable**, whose association with the outcome is only via the risk factor of interest
  - Such variables are often difficult to identify

13

Instrumental variables

- A statistical approach that could help to provide better causal inference in observational epidemiology by:
  - Dealing with confounders
  - Dealing with reverse causality
- And in RCTs by:
  - Dealing with non-compliance

14

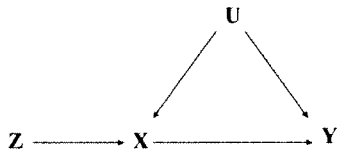
"Bottom line. All methods have assumptions
People forget that in the methods they're comfortable with

[But do people know the deep assumptions]

Definition

- An instrumental variable is a variable that is associated with an exposure of interest, but is NOT associated with the outcome of interest – except through its association with the exposure of interest

15



Z – IV

X – Exposure

Y – Outcome

U – Unmeasured confounders (incl. confounders measured but not modeled correctly)

Assumptions of IV analysis

- Z associated with X
- Z is independent of U
- Z is independent of Y given U and X (conditional for any level of U or of X)
- **IF** want to estimate association of X-Y with 95%CI need to know Z-X association & SE and Z-Y association & SE

17

Uses of IV in epidemiology

- Dealing with confounding
- Dealing with reverse causality
- Dealing with non-compliance

18

/

Dealing with confounding

- X-Y association confounded by unmeasured covariables but neither Z-X nor Z-Y are confounded

1. Z independent of U
2. Z is associated with X (magnitude of this association is known)
3. Z is independent of Y given U and X

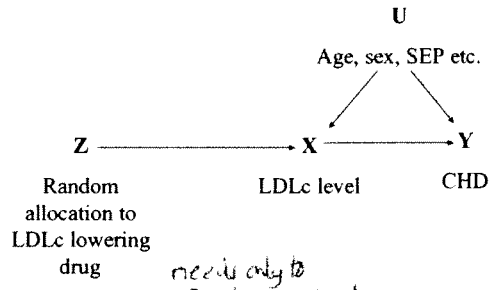
19

Dealing with confounding

- Randomisation in RCT
- Mendelian Randomisation (more this afternoon)
- 'Natural' experiments – Vietnam military service lottery, age at school entry policy

20

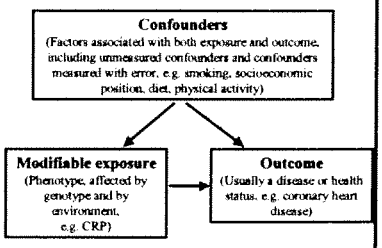
Randomisation as an IV to deal with confounding



21

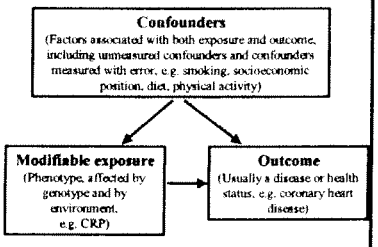
Genetic variants as IVs to deal with confounding

Conventional observational epidemiology



22

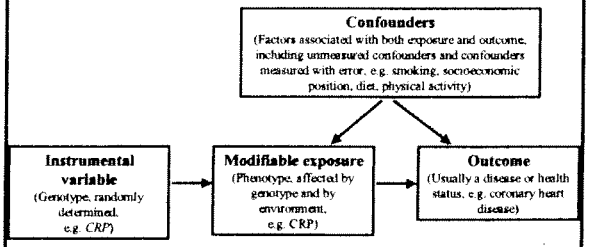
Conventional observational epidemiology



It is often impossible to exclude unmeasured or residual confounding as an explanation for observed exposure/outcome associations

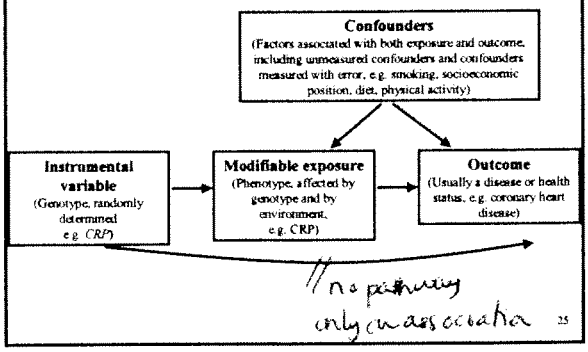
23

Genetic Variant / Mendelian randomisation approach



24

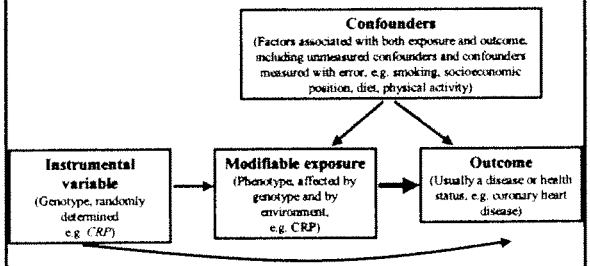
Mendelian randomisation approach



// no pathway
only an association

25

Mendelian randomisation approach



We use associations that are not affected by confounding to estimate the causal exposure-outcome association

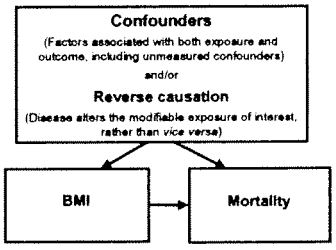
26

Dealing with reverse causality

- Back to the BMI example
  - Dealing with reverse causality
  - & Dealing with confounding

27

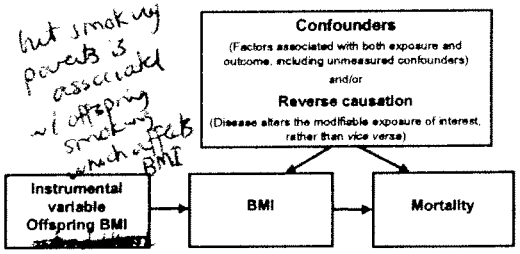
Conventional observational epidemiology



28

Instrumental variables approach

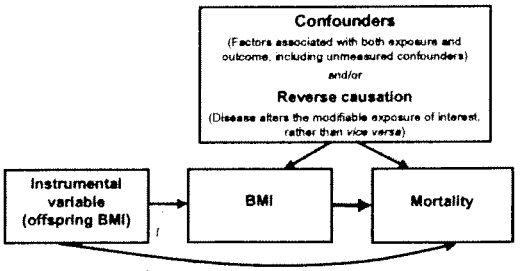
but smoking prevents is associated w/ offspring smoking which affects BMI



29

even if BMI is dropping pre-clinically, offspring BMI isn't

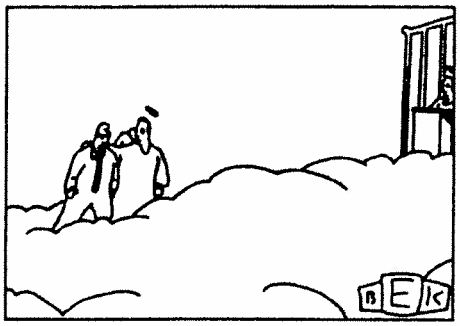
Instrumental variables approach



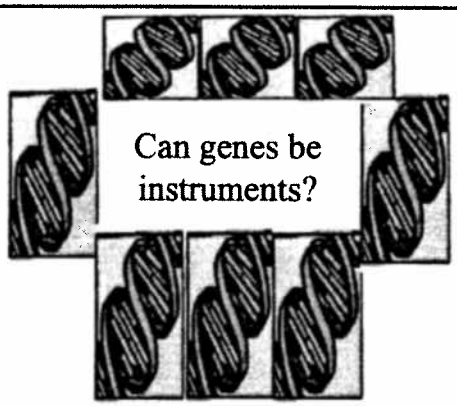
// Compare these associations

30

G. Davey Smith U. Bristol
July '07



"Well, so much for antioxidants."



Mendel on Mendelian randomization

Mendel in 1862

"the behaviour of each pair of differentiating characteristics in hybrid union is independent of the other differences between the two original plants, and, further, the hybrid produces just so many kinds of egg and pollen cells as there are possible constant combination forms"

(Sometimes called Mendel's second law – the law of independent assortment)

Gregor Mendel, 1865.

At population levels, if you take into account ancestry (say trade), there's no non-randomization

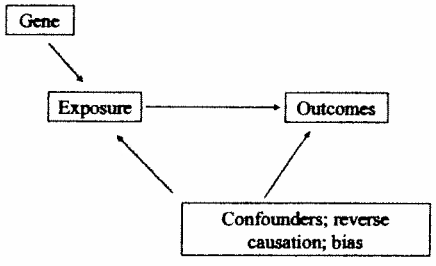
Compare the two sisters who share (on average) external conditions, but genes have not segregated

Mendelian randomization

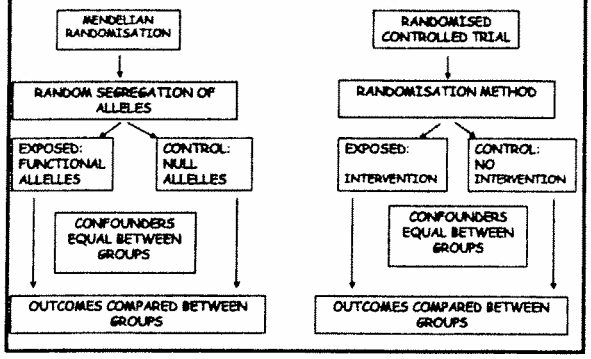
In a genetic association study the laws of Mendelian genetics imply that comparison of groups of individuals defined by genotype should only differ with respect to the locus under study (and closely related loci in linkage disequilibrium with the locus under study).

Genotypes can proxy for some modifiable environmental factors, and there should be no confounding of genotype by behavioural, socioeconomic or physiological factors (excepting those influenced by alleles at closely proximate loci or due to population stratification).

Mendelian randomization is an instrumental variables approach …



Medelian randomisation and RCTs



Categories of inference from Mendelian Randomization

- Exposure propensity
- Intermediate phenotypes
  - Single factors
  - In situations with correlated intermediate phenotypes
- Modifiers of environmental exposure
- Characterizing "difficult to measure" environmental exposures
- Intergenerational influences
- Indicator of the category of exposure of importance

Exposure Propensity: how does alcohol intake influence the risk of disease?

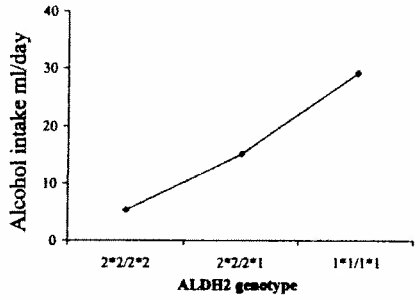
My Doctor said "Only 1 glass of alcohol a day". I can live with that.

Metabolism of alcohol

Ethanol ⇒ (ADH, CYP2E1) Acetaldehyde ⇒ (ALDH) Acetic acid

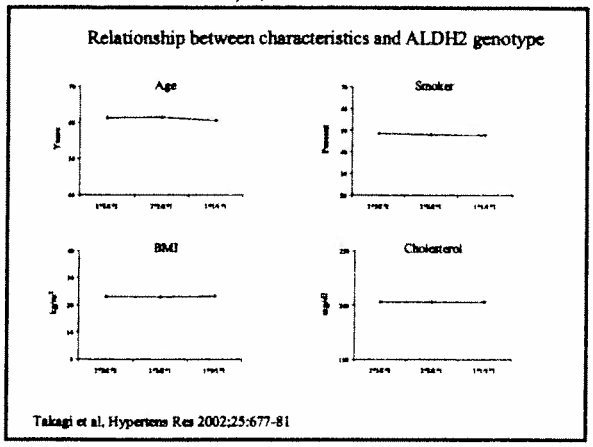
* Mainly occurs in the liver, but some activity is also present in the oral cavity and digestive tract

Relationship between alcohol intake and ALDH2 genotype

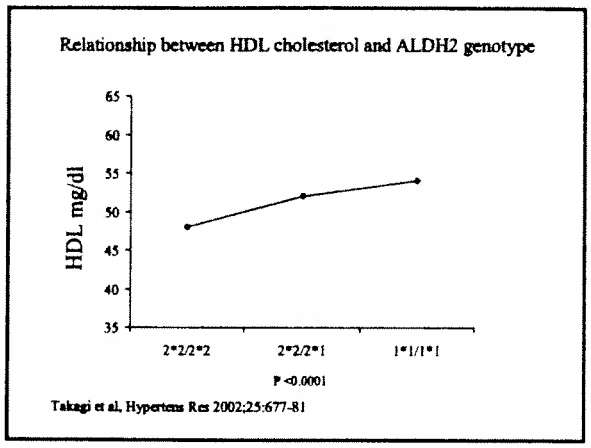


Takagi et al, Hypertens Res 2002;25:677-81

results of B.P. vs genotype
(with control of women not drinking)
(↑ dose response homozyg vs heterozyg)



[Interesting b/c drinking less might well lead to not hanging around smoking]

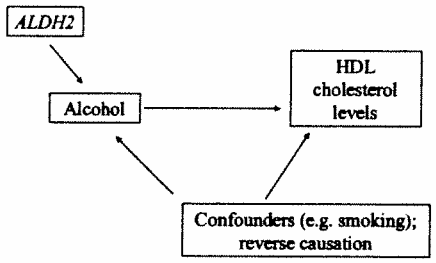


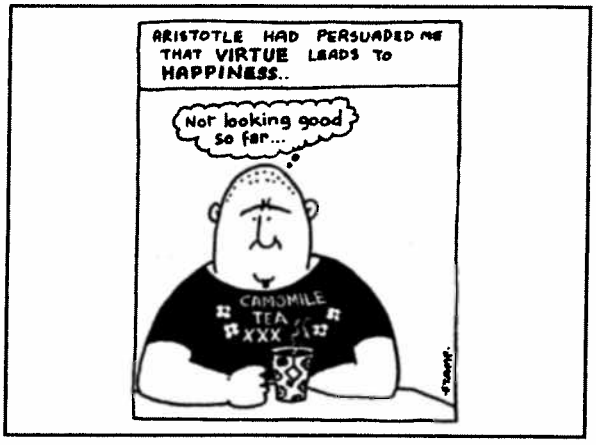
alcohol increases cholesterol
(b/c I.V. " ")

ALDH2 2*2/2*2 associations

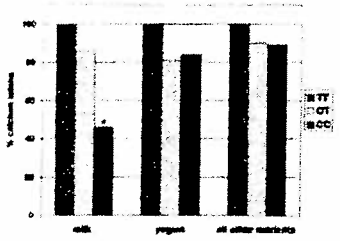
- Decreased risk of cirrhosis
- Decreased risk of oesophageal cancer
- Increased risk of MI – statistically explained by HDL levels

Mendelian randomisation is an instrumental variable approach …



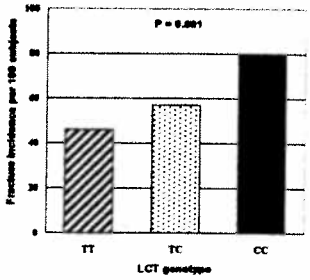


Calcium intake per day from milk, yogurt, and all other nutrients according to LCT genotypes



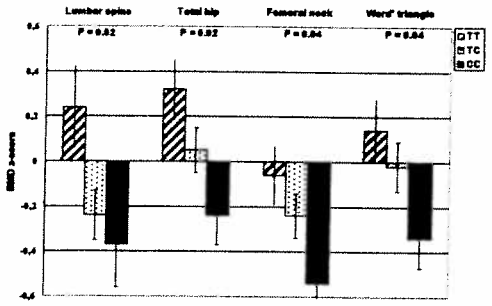
Obermayer-Pietsch et al, JBMR 2004;19:42-7

Non-vertebral fracture incidence per 100 subjects in postmenopausal women according to LCT genotypes



Obermayer-Pietsch et al, JBMR 2004;19:42-7

Bone Mineral Density according to LCT genotypes



Obermayer-Pietsch et al, JBMR 2004;19:42-7

Intermediate phenotypes: cholesterol and CHD

APOB mutation, cholesterol and CHD

Arg 3500 Gln carriers have cholesterol levels 2.6 mmol/l higher than non-carriers

Genotype unrelated to triglycerides, fibrinogen, glucose, BMI, W/H ratio etc etc

Tybjæg-Hansen et al NEJM 1998;338:1577-1584

APOB mutation, cholesterol and CHD

Arg 3500 Gln carriers CHD risk:
OR 7.0 (95% CI 2.2-22)

Central estimate higher than predicted by trials BUT not attenuated by measurement error and reflects life-time differences in cholesterol levels

Tybjæg-Hansen et al NEJM 1998;338:1577-1584

Implications of triangulation of cholesterol, *APOB* genotype and CHD risk

Lowering cholesterol will reduce CHD risk in the whole population

NOT

Screening for *APOB* mutations to detect high CHD risk

* Genotype effect size or PAR is NOT the issue! *

Environmentally modifiable factor

C-reactive protein (CRP) and cardiovascular disease and diabetes risk – cause, coincidence or confounded association?

CRP causes insulin resistance?

"Subclinical chronic low-grade inflammation might be an important player in the pathogenesis of insulin resistance and type II diabetes and CRP promotes atherosclerotic processes and endothelial cell inflammation … Inflammatory signalling pathways need to be explored in greater detail, as possible drug targets"

Sjoholm A, Nystrom T. Lancet 2005;365:610-2

CRP and "incident" diabetes in the Nurses' Health Study

- OR extreme CRP quintiles 7.1 (4.8-10.4)
- Strong associations of CRP with obesity, waist circumference, lifestyle
- Adjusted OR extreme CRP quintiles 4.4 (2.8-6.8)
- Interpreted as providing evidence of a causal role for CRP

Hu et al, Diabetes 2004;53:693-700

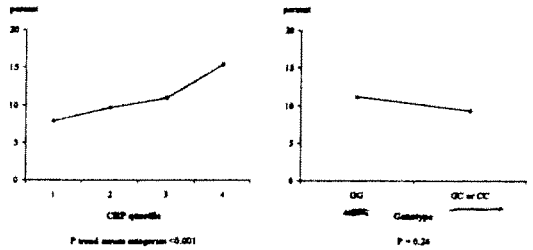
CRP (mg/l) by 1059 G/C variant

GG	GC or CC
1.81 (1.74-1.89)	1.39 (1.26-1.54)

Davey Smith et al, ATVB 2005;25:1051-6

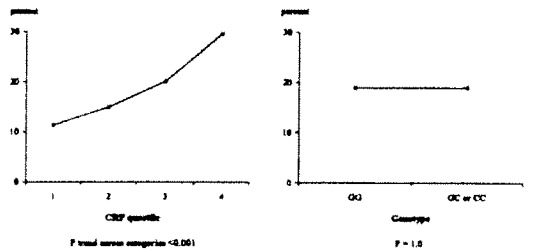
Confounders

Current smoker by CRP levels and *CRP* genotype



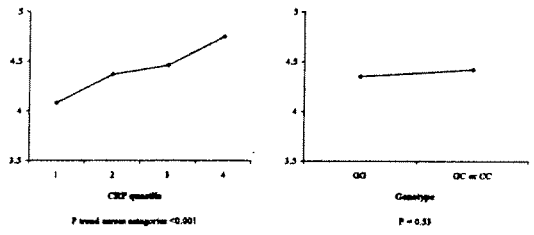
Davey Smith et al, ATVB 2005;25:1051-6

Physically inactive by CRP levels and *CRP* genotype

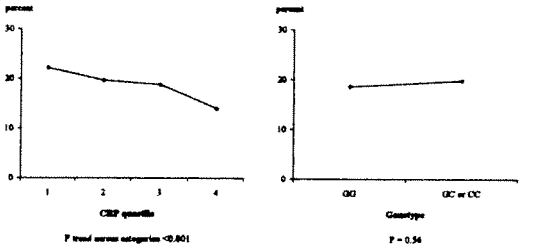


Davey Smith et al, ATVB 2005;25:1051-6

Lifecourse socioeconomic position score by CRP levels and *CRP* genotype

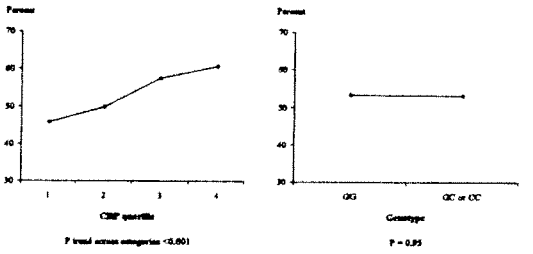


Moderate alcohol consumption by CRP levels and *CRP* genotype

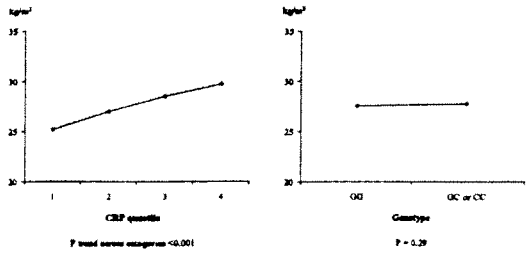


Outcomes

Hypertension by CRP levels and *CRP* genotype

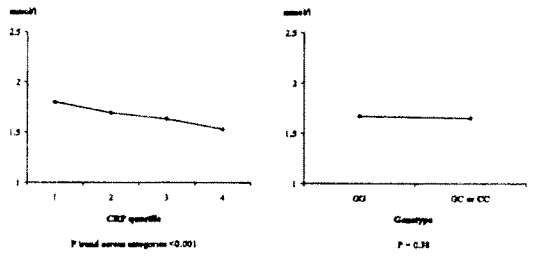


BMI by CRP levels and *CRP* genotype



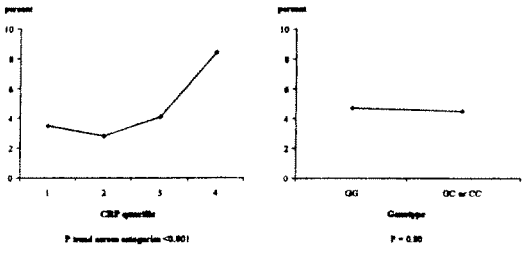
Davey Smith et al ATVB, 2005

HDLc by CRP levels and *CRP* genotype



Davey Smith et al ATVB, 2005

Doctor diagnosed diabetes by CRP levels and *CRP* genotype



Davey Smith et al ATVB, 2005

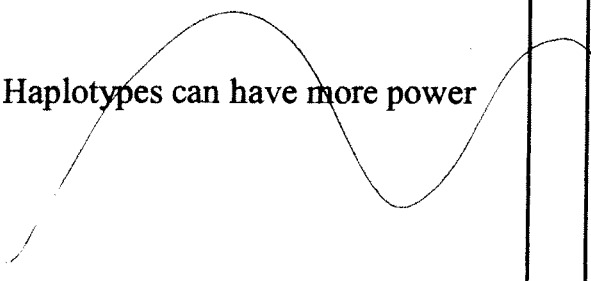
But – sample sizes need to be huge!

10,000 cases and 10,000 controls required to rule out a CRP-CHD top/bottom tertile OR of 1.5 at 80% power

(Davey Smith et al, QJM 2004;97:163-6)

but getting cheaper

Haplotypes can have more power



Triangulation with observational data : odds ratio for CHD for a doubling in C-reactive protein with and without full adjustment