University of BRISTOL

MRC Centre for Causal Analyses in Translational Epidemiology

Mendelian randomization:
using genetic data in epidemiological studies

Tom Palmer

MRC CAiTE Centre
School of Social and Community Medicine

10 February 2011, RSS Avon Local Group Meeting

Outline

Causal inference from observational data

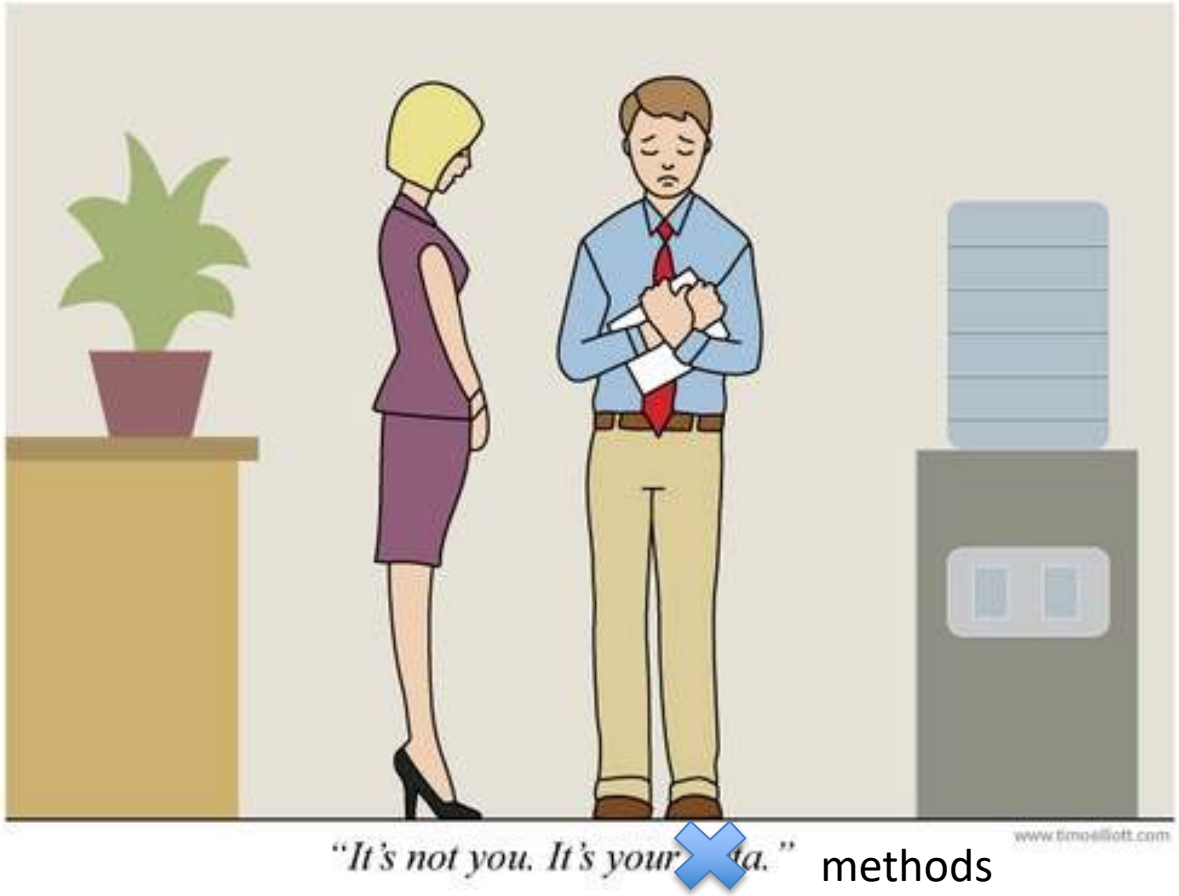
What is Mendelian randomization?

Instrument strength example

Application: effect of body mass index on Ischaemic Heart Disease

Summary

(Problem of) causal inference from observational data



methods

Disadvantages

Unmeasured confounding

Reverse causation

Advantages

Ethics & economics

Famous results:

Richard Doll –

Smoking & lung cancer

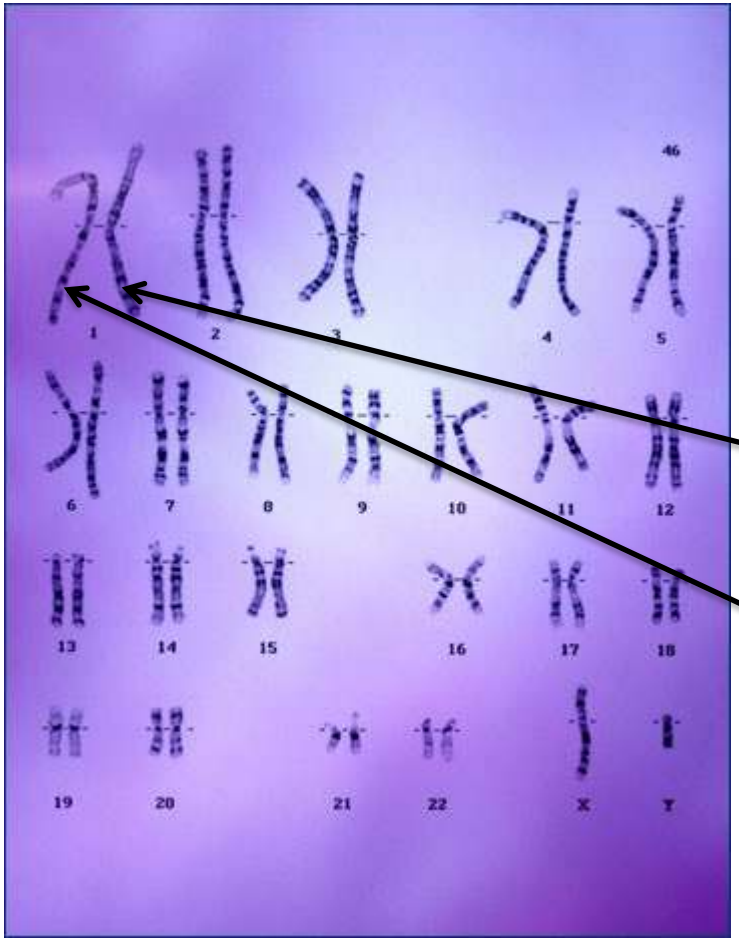
New methods:

propensity scores, marginal structural models, **instrumental variables, structural mean models**, measurement error, missing data approaches, structural equation models, ...

What is Mendelian randomization?

Genotypes as instrumental variables

What is a genotype?



Allele 1: 0, 1

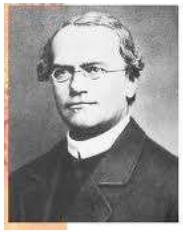
Allele 2: 0, 1

Genotype

0 + 0 = 0

0 + 1 = 1

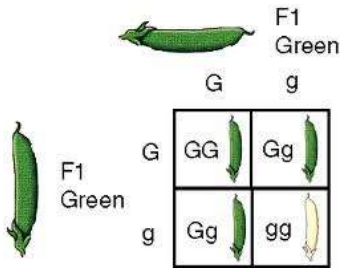
1 + 1 = 2



What is Mendelian randomization?

Genotypes as instrumental variables

What is an instrumental variable?



(ii)

Confounders

(i) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The Wellcome Trust Case Control Consortium*

Genotype → Exposure → Outcome

(iii)

2003

2004

Window Help

/ 22 74% trumental variable

© International Epidemiological Association 2003 Printed in Great Britain

International Journal of Epidemiology 2003;**32**:1–22
DOI: 10.1093/ije/dyg070

30TH THOMAS FRANCIS JR MEMORIAL LECTURE

'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*

George Davey Smith and Shah Ebrahim

Associations between modifiable exposures and disease seen in observational epidemiology are sometimes confounded and thus misleading, despite our best efforts to improve the design and analysis of studies. Mendelian randomization—the random assortment of genes from parents to offspring that occurs during gamete formation and conception—provides one method for assessing the causal nature of some environmental exposures. The association between a disease and a

sease is not
retations of
ymorphisms
Mendelian
ing by poly-
ymorphisms
orphisms for
ts of genetic
pportunities
y contribute
exposures.

Adobe Reader

Reader has finished searching the document. No matches were found.

OK

Genetic epidemiology—the theme of this issue of the *International Journal of Epidemiology*—is seen by many to be the only future for epidemiology, perhaps reflecting a growing awareness of the limitations of observational epidemiology[1] (Box 1). Genetic epidemiology is concerned with understanding heritable aspects of disease risk, individual susceptibility to disease, and ultimately with contributing to a comprehensive molecular understanding of pathogenesis. The massive investment and expansion of human genetics, if it is to return value for the common good, must be integrated into public health functions. The human genome epidemiology network (HuGE Net—http://www.cdc.gov/genetics/huge.htm) has been established to promote the use of genetic knowledge—in terms of genetic tests and services—for disease prevention and health promotion.[2,3] A broad taxonomy of human genome studies of public health relevance has been developed[4] (Box 2). In this issue of the *IJE*, we publish a paper by Miguel Porta,[5] who highlights the need for a more rational approach to genetic testing, given the likely low penetrance of many genes associated with cancers,[6] likening the role of the genome to a jazz score that is interpreted and developed through experience and context—and is seldom predictable. Such insights may well temper enthusiasm for genetic testing in populations.

University of Bristol, Department of Social Medicine, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK.

* 30th Thomas Francis Jr Memorial Lecture, to be delivered by George Davey Smith at the University of Michigan, School of Public Health, 6 March 2003.

However, in parallel to the approaches advocated by HuGE, genetic epidemiology can lead to a more robust understanding of environmental determinants of disease (e.g. dietary factors, occupational exposures, and health-related behaviours) relevant to whole populations (and not simply to genetically susceptible subpopulations).[7–10] This approach has recently been referred to as 'Mendelian randomization'.[11–15] Here we begin by briefly reviewing reasons for current concerns about aetiological findings generated by conventional observational epidemiology and then we outline the potential contribution (and limitations) of Mendelian randomization.

Observational epidemiology: yet more residually confounded associations of no causal significance?

Over the last decade several severe indictments of epidemiology have appeared, with the major thrust being that spurious non-replicable and non-causal findings are produced and sometimes widely disseminated.[16–20] The most salient examples come from situations in which observational epidemiological studies have highlighted an apparently substantial causal association that has later failed to be confirmed in large-scale randomized controlled trials (RCT). An important example of this is the contradictory set of findings regarding the association between the antioxidant vitamin β-carotene and smoking-related cancers.

IJE vol.33 no.1 © International Epidemiological Association 2004; all rights reserved. *International Journal of Epidemiology* 2004;**33**:21–25
DOI: 10.1093/ije/dyh048

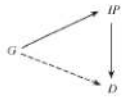
Commentary: The concept of 'Mendelian randomization'

Duncan C Thomas and David V Conti

This issue of the *International Journal of Epidemiology* reprints a seminal letter to the editor by Martijn Katan,[1] which appears to be the first description of the concept of 'Mendelian randomization.' In discussing the controversy over whether the association between low serum cholesterol and cancer is causal or might simply reflect an effect of the disease to lower cholesterol levels ('reverse causation') or confounding by diet or other factors, Katan proposed a test of causality by studying instead the relationship between cancer and a genetic determinant of serum cholesterol, the apolipoprotein A (*APOE*) gene. His rationale was that since alleles are allocated essentially at random, such an association would not be subject to either confounding or reverse causation. Thus, if a causal relationship between *APOE* and serum cholesterol were clearly established, then an association between *APOE* and cancer would provide indirect evidence for the causality of the association between serum cholesterol and cancer. Although Katan did not use the term 'Mendelian randomization', the concept has been attributed to him and subsequently developed by a number of other authors.[2–6] In particular, Davey Smith and Ebrahim[2]

Department of Preventive Medicine, University of Southern California, Los Angeles, CA 9089-9011, USA.

have shown how the magnitude of the estimated effects of a gene (G) on an intermediate phenotype (IP) and on disease (D) can be combined to yield an estimate of the causal effect of the intermediate phenotype on disease, as illustrated in the following figure:

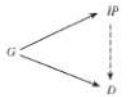


(where the dotted arrow from G to D represents the indirect association assumed to be mediated entirely through IP).

Use of instrumental variables in epidemiology

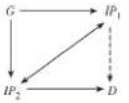
It may seem perverse to try to study the causality of a relationship between IP and D through the relationship of each with G, but there is merit in the idea. While its application to molecular epidemiology is novel, the idea is more than 70 years old, apparently first introduced into the econometrics literature by Wright[7] and later adopted into the statistical measurement error and causal inference literature under the rubric of 'instrumental variables'.[8–10] The basic idea is that if a causal pathway is correctly specified as in the above Figure (including certain additional assumptions discussed in the Appendix), then the causal effect of IP on D can be estimated by the ratio of the coefficients for the regression of D on G and of IP on G. (An exactly analogous argument applies in randomized controlled trials, where G would represent 'intent to treat' and IP the treatment actually received: although the IP–D association could be biased various ways, the G–D association is

connection induced by confounding by G:



This would be a case of a 'false-positive' inference—an incorrect conclusion that there is a causal connection between IP and D when in fact none exists. Of course, negative confounding could also lead to a false-negative conclusion—that there was no association between IP and D when there really is one.

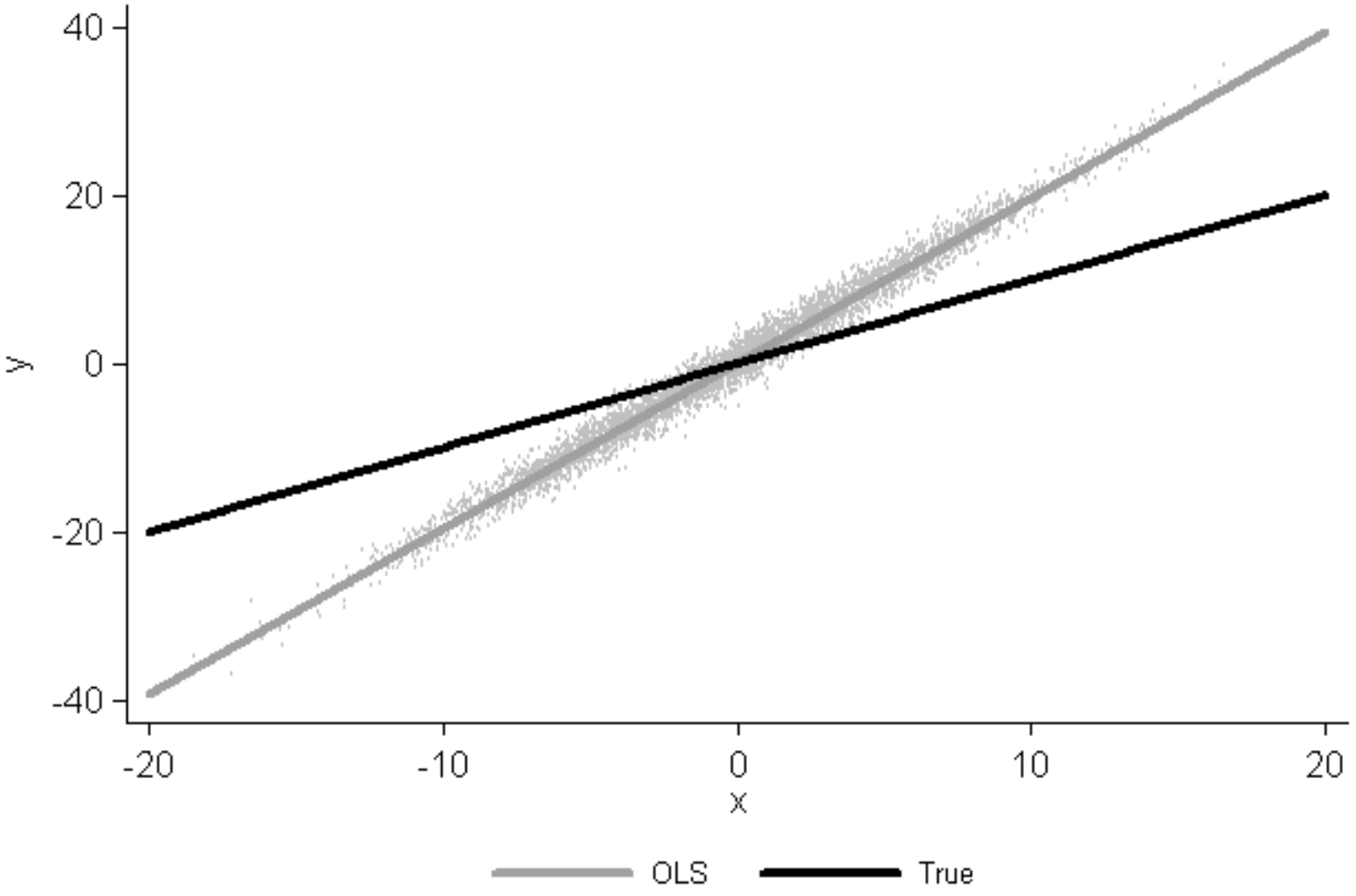
One way such a situation could come about is when a single gene has pleiotropic effects. Suppose, for argument sake, that the true causal picture were as follows:



where the solid arrows indicate causal connections and the dashed arrow indicates a non-causal association induced by the other associations.

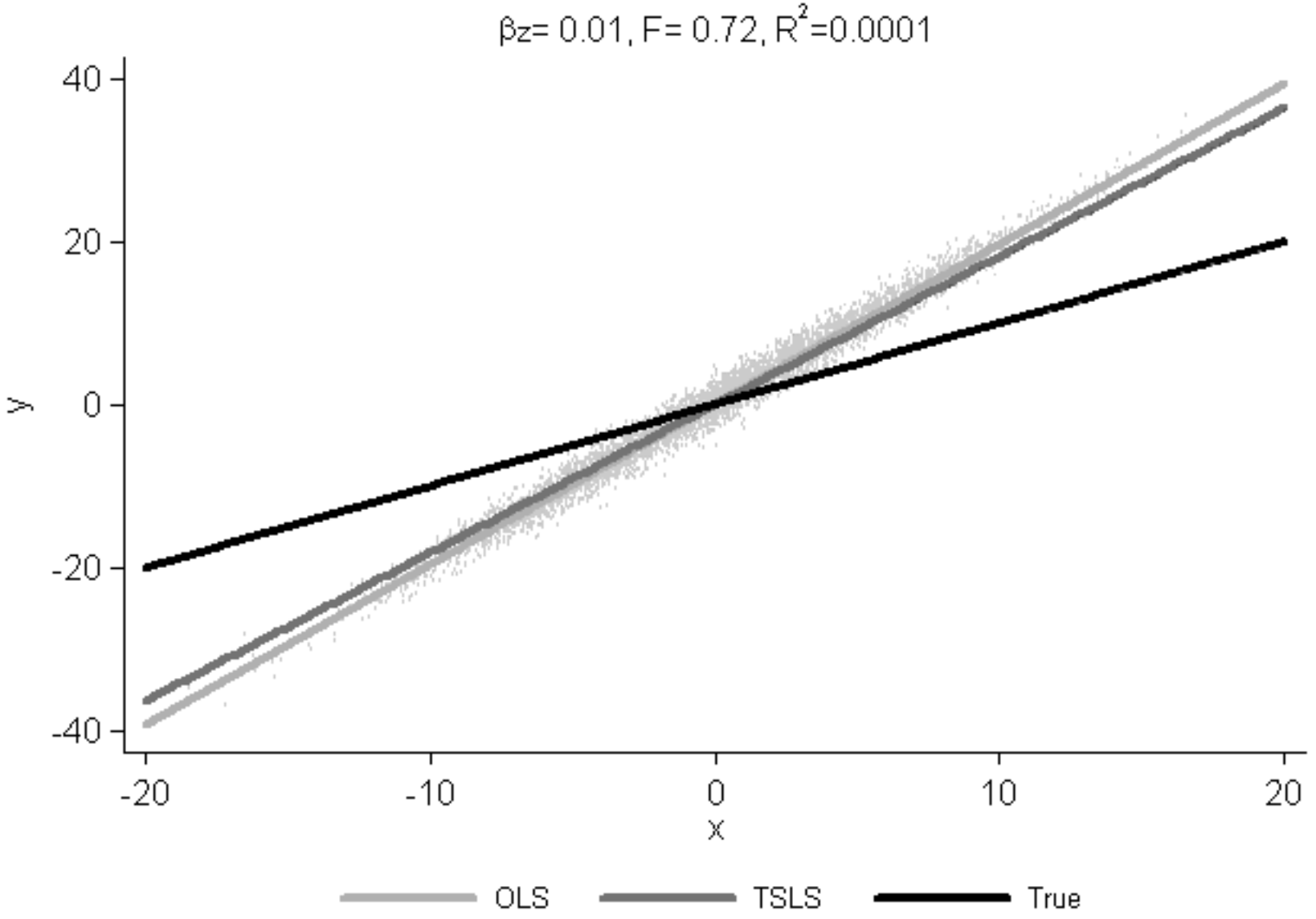
For example, Davey Smith and Ebrahim[2] provide an interesting discussion of the role of folate, homocysteine, and the methylenetetrahydrofolate reductase (*MTHFR*) gene in the aetiology of coronary heart disease (CHD) and neural tube defects (NTD). This is a very complex pathway, involving several feedback

Instrument strength example



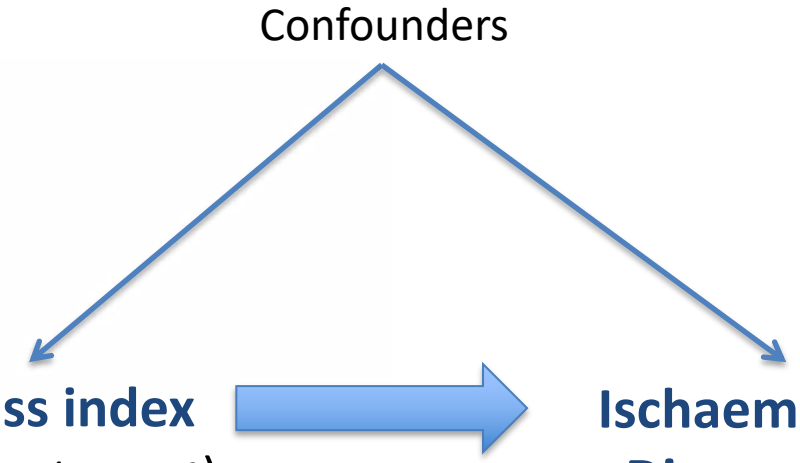
Unmeasured confounder: OLS biased

Instrument strength example



IV estimates consistent rather than unbiased

Application

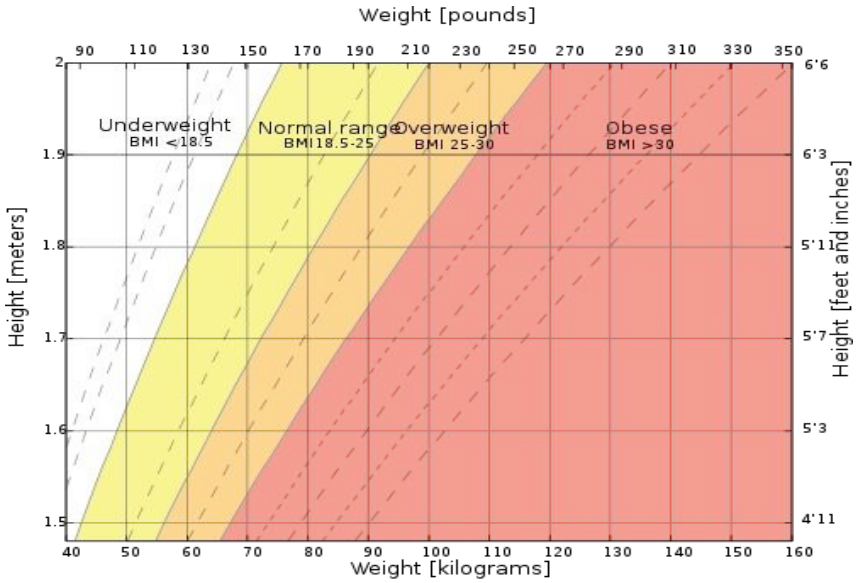


FTO
(rs9939609) *chr 16*
MC4R
(rs17782313) *chr 18*
TMEM18
(rs6548238) *chr 2*

body mass index
(BMI=weight/height^2)
[adiposity]

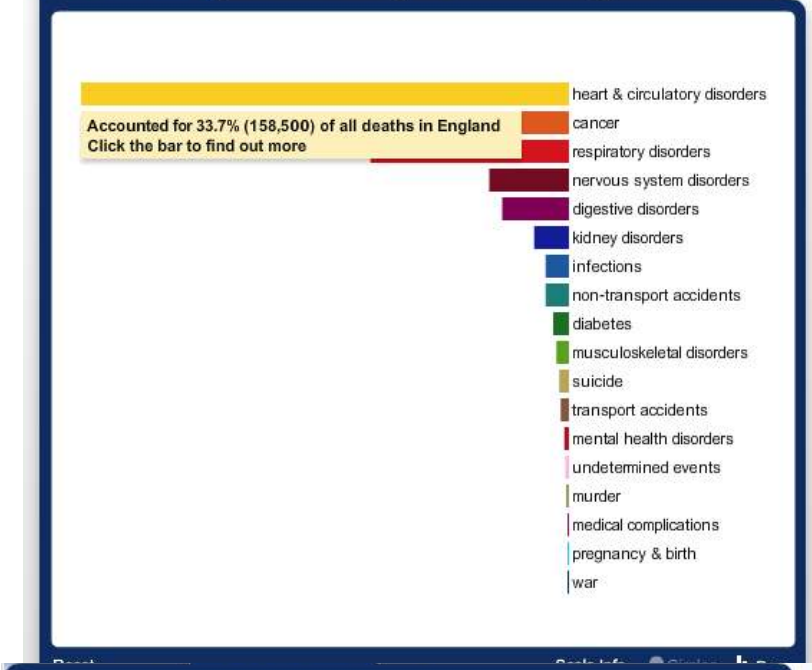
Ischaemic Heart Disease (IHD)

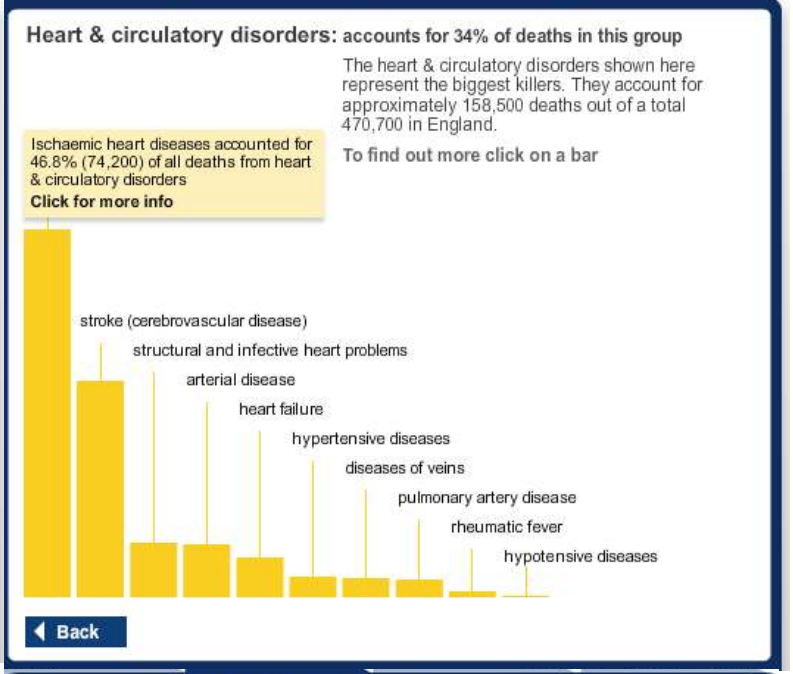
[reduced blood supply to heart, angina,
heart attack]

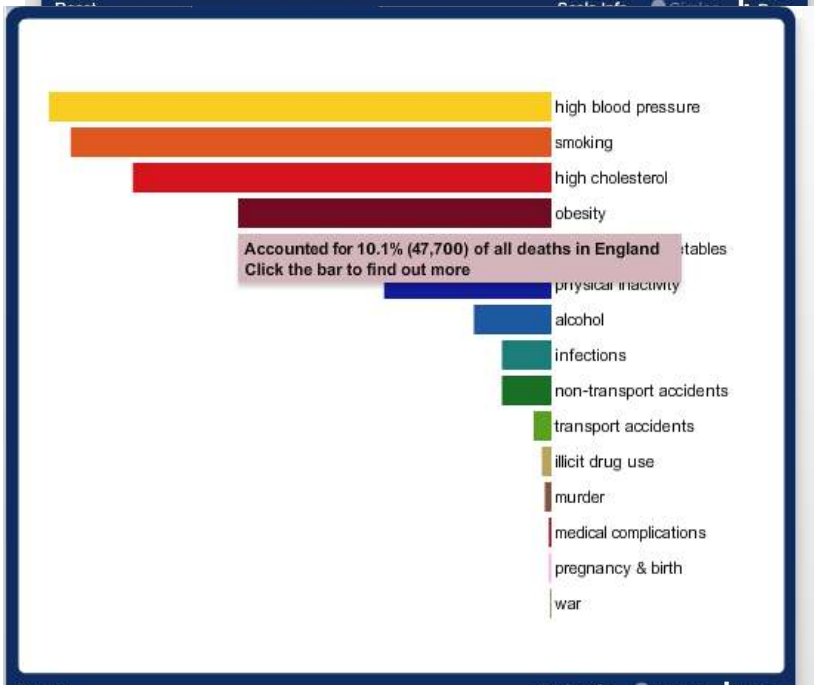


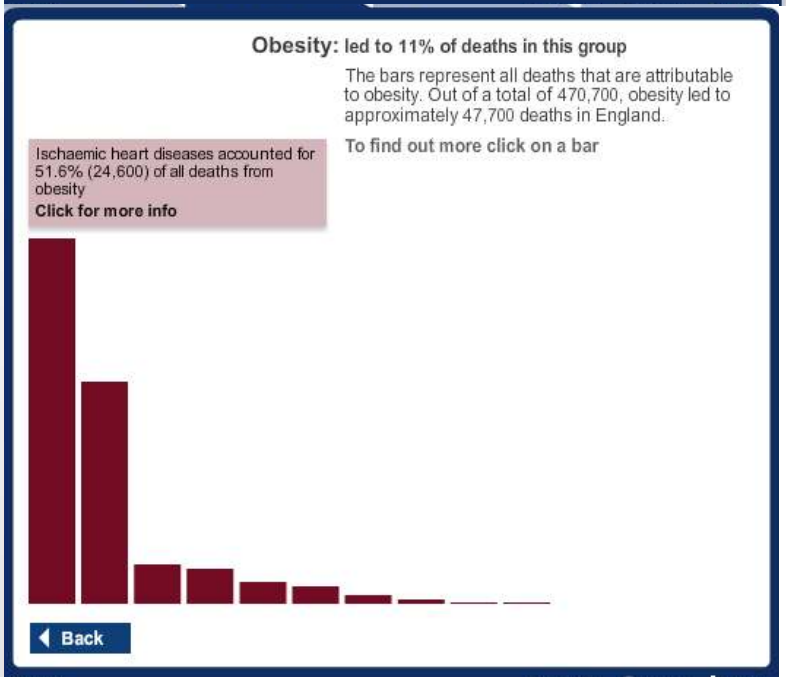
3 large studies in Copenhagen

- Copenhagen General Population Study(CGPS)
  N=54613, 3780 IHD events
- Copenhagen City Heart Study (CCHS)
  N=10474, 2006 IHD events
- Copenhagen Ischaemic Heart Disease Study (CIHDS) *case-control, BMI not measured*
  N=10540, 5270 IHD events



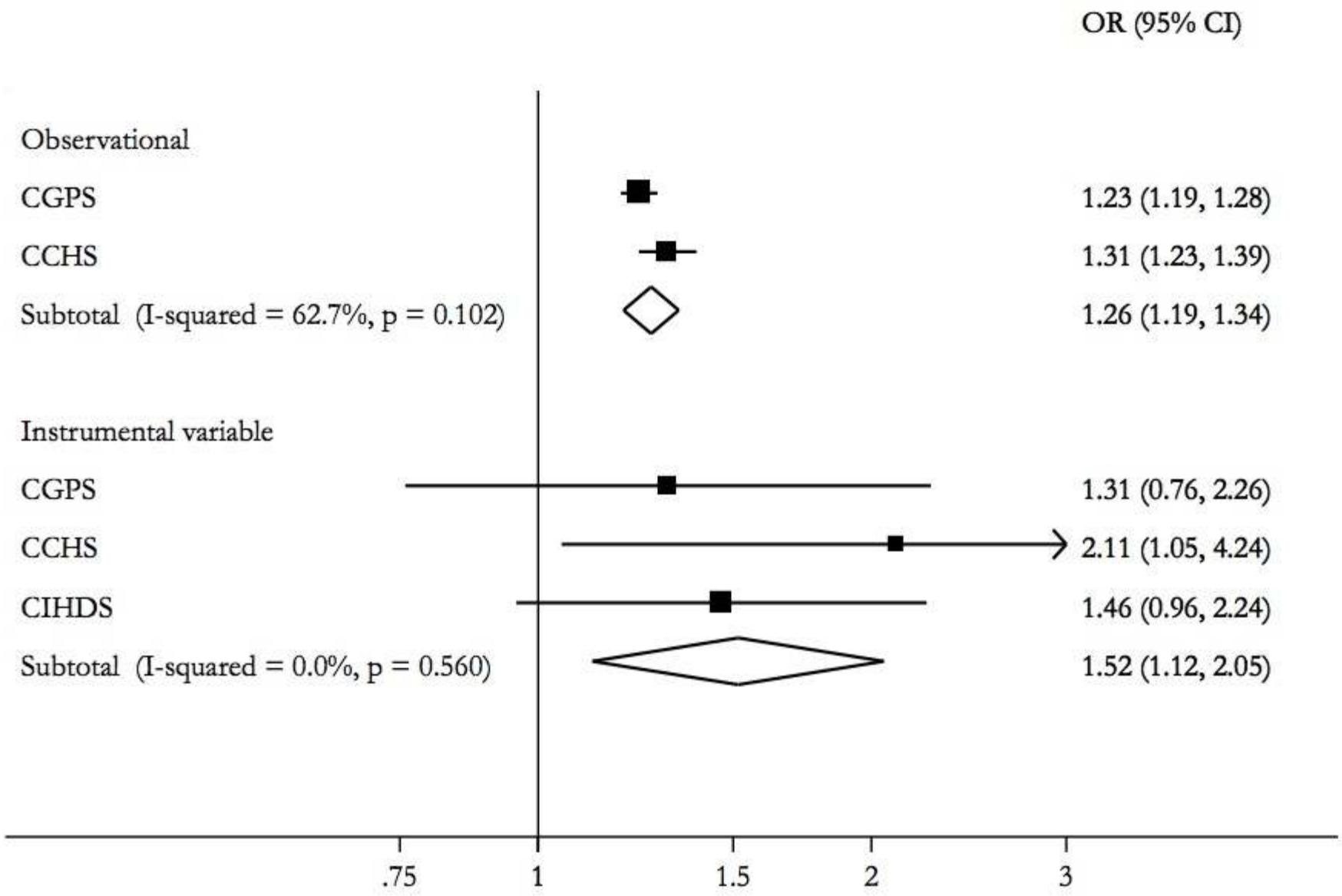






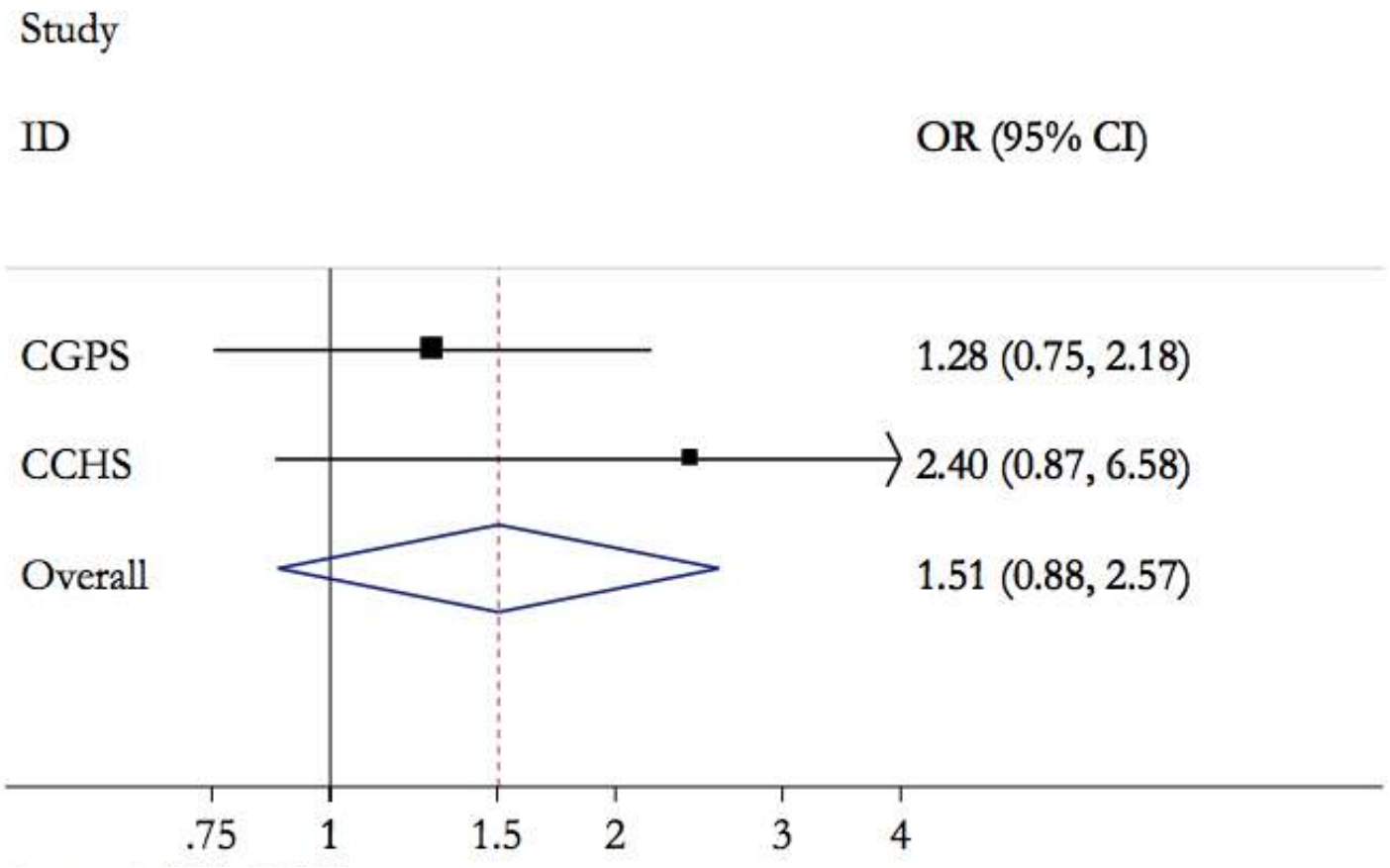
NHS Atlas of Risk (deaths registered in England in 2007)

Observational and instrumental variable estimates



OR for a 4 kg/m² increase in BMI

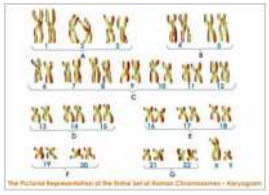
Logistic structural mean model IV estimates

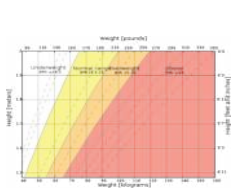


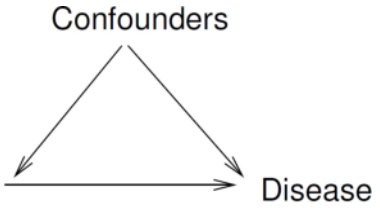
OR for a 4 kg/m² increase in BMI

Summary

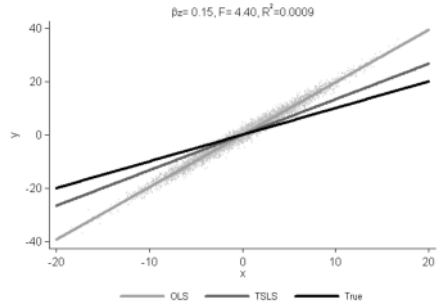
Mendelian randomization



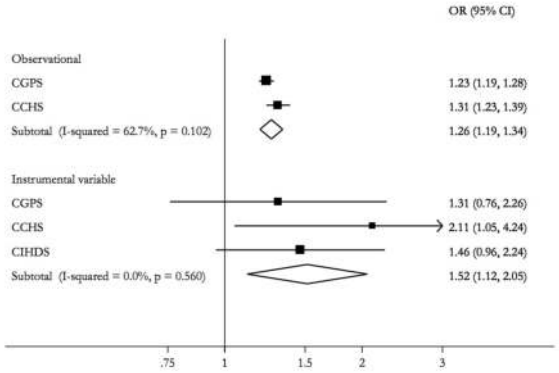




Instrument strength example



Application: effect of BMI on IHD



Bristol expertise: Maths, CMPO/Economics, School of Social and Community Medicine