# Instrumental variables and Mendelian randomization

Arvid Sjölander

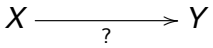
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet

Causal Inference for Epidemiological Research



## The aim of epidemiological research

- Most epidemiological research questions are centered around a particular **exposure** and a particular **outcome**
- Typically, we would like to know if the exposure has a causal effect on the outcome
  - hypothesis testing
- If there is a causal effect, we would also like to know its magnitude
  - effect estimation






## Randomized trials

- In randomized trials, the exposure-outcome association can be given a causal interpretation
- E.g. the statistical risk ratio

$$\frac{p(Y=1|X=1)}{p(Y=1|X=0)}$$

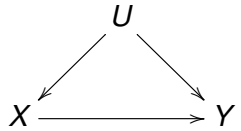
is equal to the causal risk ratio

$$\frac{p(Y_1=1)}{p(Y_0=1)}$$

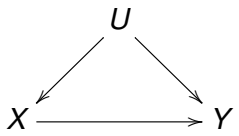


## Observational studies

- In observational studies, the exposure-outcome association is usually confounded and cannot be given a causal interpretation



## Confounder control




- Observed confounders can be controlled for, e.g. in a regression model
- But unobserved confounders may still bias the results



## Instrumental variables

- Instrumental variables (IV) methods can be used to test/estimate causal effects in the presence of unobserved confounding
- IV methods have a long history in econometrics
- They have more recently become popular in epidemiology
  - in particular through the Mendelian randomization (MR) studies, where the IV is a gene



## Outline





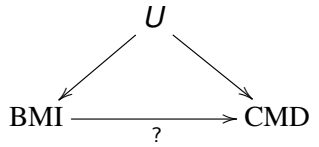
## Outline

## Motivating example

- Kivimäki et al. (2011) aimed to study if obesity causes common mental disorders (CMD)
- They carried out a large prospective study, where both BMI and CMD were measured on 4 occasions during follow-up
  - several analyses of the repeated measures were performed
  - here we only consider mean BMI/CMD across follow-up



## The study in a DAG




- A huge potential for confounding, e.g. by
  - age
  - sex
  - socioeconomic status (SES)
  - comorbid conditions
  - etc



## The FTO gene

- The study participants were genotyped on the fat mass and obesity (FTO) gene
- Variants in this gene have previously been associated with BMI
- This gene was used as an instrumental variable (IV) in the analysis



## The IV assumptions

- All IV analyses rely on three fundamental assumptions
- 1) The IV is associated with the exposure
- 2) The association between the IV and the outcome is unconfounded
- 3) The IV affects the outcome only through the exposure
- *Display these assumptions by presence/absence of arrows in DAGs*

# Solution



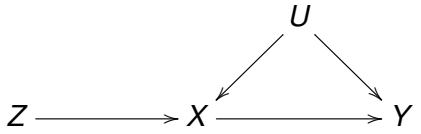
# Solution, cont'd



# Solution, cont'd



# The IV against the outcome

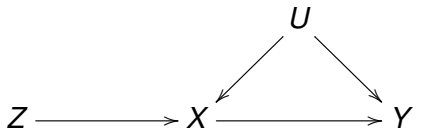



- Suppose that we observe a statistical association between the IV $Z$ and the outcome $Y$ in the DAG above
- *Is it then possible that the exposure $X$ has no causal effect on the outcome $Y$?*

# Solution



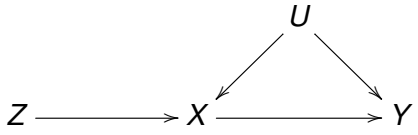
# Hypothesis testing with IVs




- And the other way around: **if $Z$ and $Y$ are associated, then $X$ has a causal effect on $Y$**
- This suggests a very simple strategy for hypothesis testing
  - check if the IV $Z$ is associated with the outcome $Y$
  - if it is, then conclude that the exposure $X$ has a causal effect on the outcome $Y$



# Generality of the approach




- This IV hypothesis test is completely general in that it allows for
  - arbitrary types of variables (binary, categorical, continuous, time-to-event etc)
  - arbitrary statistical methods for testing (non-parametric test, $\chi^2$-test, $t$-test, regression model etc)
  - arbitrary sampling schemes (cross-sectional, prospective, retrospective etc)
- ... as long as the three IV assumptions hold



# Results from the Kivimäki et al. (2011) study

**Table 2.** Associations of Fat Mass and Obesity-Associated (*FTO*) Genotype With Body Mass Index, Overweight, Obesity, and Common Mental Disorders From 4 Repeated Assessments in Men and Women, Whitehall II Study, 1985–2004

| Predictor | No. of Participants | BMI[a] | | Overweight[b] | | Obesity[b] | | Common Mental Disorders[c] | |
|---|---|---|---|---|---|---|---|---|---|
| | | β | 95% CI | β | 95% CI | β | 95% CI | β | 95% CI |
| *Men* | | | | | | | | | |
| No. of *FTO* adiposity alleles | | | | | | | | | |
| 0 | 1,046 | 0.00 | Referent | 0.00 | Referent | 0.00 | Referent | 0.00 | Referent |
| 1 | 1,442 | 0.283 | 0.039, 0.526 | 0.200 | 0.064, 0.335 | 0.018 | −0.054, 0.090 | 0.020 | −0.065, 0.105 |
| 2 | 493 | 0.935 | 0.608, 1.262 | 0.331 | 0.144, 0.519 | 0.192 | 0.094, 0.289 | 0.172 | 0.058, 0.286 |
| Per-allele increase | 2,981 | 0.433 | 0.275, 0.592 | 0.173 | 0.083, 0.262 | 0.081 | 0.034, 0.129 | 0.074 | 0.019, 0.129 |
| *P* for trend | | <0.0001 | | <0.0001 | | 0.001 | | 0.009 | |
| *Women* | | | | | | | | | |
| No. of *FTO* adiposity alleles | | | | | | | | | |
| 0 | 448 | 0.00 | Referent | 0.00 | Referent | 0.00 | Referent | 0.00 | Referent |
| 1 | 543 | −0.183 | −0.741, 0.376 | 0.015 | −0.196, 0.225 | −0.076 | −0.441, 0.289 | 0.073 | −0.074, 0.219 |
| 2 | 173 | 0.310 | −0.474, 1.094 | 0.145 | −0.155, 0.445 | 0.121 | −0.391, 0.632 | −0.011 | −0.217, 0.194 |
| Per-allele increase | 1,164 | 0.077 | −0.294, 0.449 | 0.058 | −0.083, 0.120 | 0.012 | −0.089, 0.114 | 0.012 | −0.085, 0.110 |
| *P* for trend | | 0.68 | | 0.42 | | 0.81 | | 0.80 | |

Abbreviations: BMI, body mass index; CI, confidence interval; *FTO*, fat mass and obesity-associated gene; GHQ, General Health Questionnaire.
[a] Mean BMI (weight (kg)/height (m)$^2$) across 4 repeated clinical examinations conducted over a 19-year follow-up period.
[b] Number of times (range, 0–4) a participant was found to be overweight (BMI 25.0–29.9) or obese (BMI ≥30) in 4 examinations conducted over a 19-year follow-up period. The analysis of overweight did not include obese participants and was therefore based on 2,473 men and 861 women.
[c] Number of times (range, 0–4) a participant was designated a GHQ "case" in 4 examinations conducted over a 19-year follow-up period. GHQ score was used as the measure of common mental disorders.

# Conclusion

- If we believe that the FTO gene is a valid IV, then there appears to be a causal effect of BMI on CMD for men, but possibly not for women
- *What could be the mechanisms behind such causal effect? Biological? Sociological?*



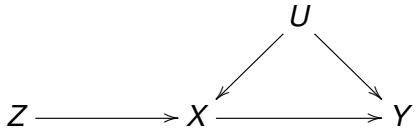
# Testability of the assumptions

- Like all statistical methods, IV methods rely on assumptions
- Ideally these assumptions should be tested and verified before the results are trusted
- In contrast to assumptions in 'ordinary' statistical models, the IV assumptions are to a large extent untestable



# Assumption 1

- The IV is associated with the exposure




- This assumption is easy to verify empirically (i.e. by data analysis)



# Results from the Kivimäki et al. (2011) study

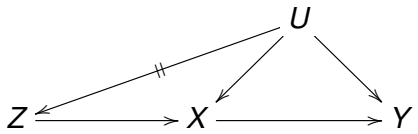
**Table 2.** Associations of Fat Mass and Obesity-Associated (*FTO*) Genotype With Body Mass Index, Overweight, Obesity, and Common Mental Disorders From 4 Repeated Assessments in Men and Women, Whitehall II Study, 1985–2004

| Predictor | No. of Participants | BMI[a] | | Overweight[b] | | Obesity[b] | | Common Mental Disorders[c] | |
|---|---|---|---|---|---|---|---|---|---|
| | | β | 95% CI | β | 95% CI | β | 95% CI | β | 95% CI |
| | | | | *Men* | | | | | |
| No. of *FTO* adiposity alleles | | | | | | | | | |
| 0 | 1,046 | 0.00 | Referent | 0.00 | Referent | 0.00 | Referent | 0.00 | Referent |
| 1 | 1,442 | 0.283 | 0.039, 0.526 | 0.200 | 0.064, 0.335 | 0.018 | −0.054, 0.090 | 0.020 | −0.065, 0.105 |
| 2 | 493 | 0.935 | 0.608, 1.262 | 0.331 | 0.144, 0.519 | 0.192 | 0.094, 0.289 | 0.172 | 0.058, 0.286 |
| Per-allele increase | 2,981 | 0.433 | 0.275, 0.592 | 0.173 | 0.083, 0.262 | 0.081 | 0.034, 0.129 | 0.074 | 0.019, 0.129 |
| *P* for trend | | <0.0001 | | <0.0001 | | 0.001 | | 0.009 | |
| | | | | *Women* | | | | | |
| No. of *FTO* adiposity alleles | | | | | | | | | |
| 0 | 448 | 0.00 | Referent | 0.00 | Referent | 0.00 | Referent | 0.00 | Referent |
| 1 | 543 | −0.183 | −0.741, 0.376 | 0.015 | −0.196, 0.225 | −0.076 | −0.441, 0.289 | 0.073 | −0.074, 0.219 |
| 2 | 173 | 0.310 | −0.474, 1.094 | 0.145 | −0.155, 0.445 | 0.121 | −0.391, 0.632 | −0.011 | −0.217, 0.194 |
| Per-allele increase | 1,164 | 0.077 | −0.294, 0.449 | 0.058 | −0.083, 0.120 | 0.012 | −0.089, 0.114 | 0.012 | −0.085, 0.110 |
| *P* for trend | | 0.68 | | 0.42 | | 0.81 | | 0.80 | |

Abbreviations: BMI, body mass index; CI, confidence interval; *FTO*, fat mass and obesity-associated gene; GHQ, General Health Questionnaire.
[a] Mean BMI (weight (kg)/height (m)$^2$) across 4 repeated clinical examinations conducted over a 19-year follow-up period.
[b] Number of times (range, 0–4) a participant was found to be overweight (BMI 25.0–29.9) or obese (BMI ≥30) in 4 examinations conducted over a 19-year follow-up period. The analysis of overweight did not include obese participants and was therefore based on 2,473 men and 861 women.
[c] Number of times (range, 0–4) a participant was designated a GHQ "case" in 4 examinations conducted over a 19-year follow-up period. GHQ score was used as the measure of common mental disorders.

# Assumption 2

- The association between the IV and the outcome is unconfounded




- This assumption is more difficult to verify, since some confounders are typically unobserved
  - this is the main reason for using IV methods
- However, if some confounders are observed, then we can at least verify that $Z$ is independent of these



# Results from the Kivimäki et al. (2011) study

**Table 3.** Associations of Fat Mass and Obesity-Associated (*FTO*) Genotype With Factors Potentially Confounding the Relation Between Obesity and Common Mental Disorders in Men and Women, Whitehall II Study, 1985–2004

| Baseline Characteristic | No. of *FTO* Adiposity Alleles | | | | | | | | | *P* Value |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | | | 1 | | | 2 | | | |
| | No. | % | Mean (SD) | No. | % | Mean (SD) | No. | % | Mean (SD) | |
| | | | | *Men* | | | | | | |
| Maximum no. of participants | 1,046 | | | 1,442 | | | 493 | | | |
| Age, years | | | 44.0 (6.1) | | | 43.7 (5.9) | | | 44.0 (5.7) | 0.75 |
| Marital status (married) | | 82.8 | | | 84.4 | | | 82.4 | | 0.45 |
| Socioeconomic position | | | | | | | | | | |
| High | | 39.6 | | | 41.5 | | | 44.4 | | 0.39 |
| Intermediate | | 54.8 | | | 53.7 | | | 51.3 | | |
| Low | | 5.6 | | | 4.9 | | | 4.3 | | |
| Chronic disease | | 8.3 | | | 10.8 | | | 9.7 | | 0.12 |
| Prevalent coronary heart disease | | 0.8 | | | 0.6 | | | 0.8 | | 0.75 |
| Prevalent diabetes | | 0.5 | | | 0.6 | | | 0.8 | | 0.72 |
| Use of psychotropic drugs | | 2.3 | | | 2.0 | | | 2.2 | | 0.88 |
| High alcohol consumption[a] | | 16.4 | | | 17.4 | | | 20.6 | | 0.13 |
| Physical inactivity | | 23.8 | | | 19.8 | | | 22.6 | | 0.051 |
| Current smoking | | 10.4 | | | 11.9 | | | 9.2 | | 0.20 |



# Results from the Kivimäki et al. (2011) study, cont'd

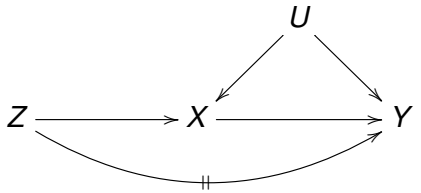
| | | | | |
|---|---|---|---|---|
| | | *Women* | | |
| Maximum no. of participants | 448 | 543 | 173 | |
| Age, years | 44.0 (6.1) | 44.5 (5.8) | 44.7 (5.9) | 0.13 |
| Marital status (married) | 60.5 | 57.3 | 61.9 | 0.44 |
| Socioeconomic position | | | | |
| High | 15.2 | 16.1 | 19.7 | 0.20 |
| Intermediate | 44.0 | 46.5 | 49.7 | |
| Low | 40.9 | 37.5 | 30.6 | |
| Chronic disease | 10.4 | 9.2 | 8.8 | 0.78 |
| Prevalent coronary heart disease | 0.4 | 0.4 | 0.0 | 0.69 |
| Prevalent diabetes | 0.5 | 0.4 | 0.0 | 0.69 |
| Use of psychotropic drugs | 4.0 | 4.4 | 4.1 | 0.94 |
| High alcohol consumption[b] | 3.0 | 3.5 | 5.8 | 0.24 |
| Physical inactivity | 41.3 | 38.0 | 39.3 | 0.59 |
| Current smoking | 13.3 | 17.3 | 17.1 | 0.20 |

Abbreviations: *FTO*, fat mass and obesity-associated gene; SD, standard deviation.
[a] >21 alcohol units per week.
[b] >14 alcohol units per week.



# Assumption 3

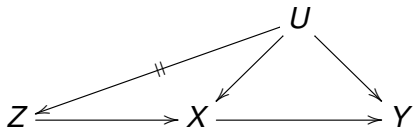
- The IV affects the outcome only through the exposure




- This assumption is impossible to verify, unless all confounders are completely measured
- *Why?*

## Assumption 2 in MR studies

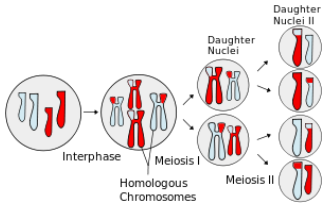
- The association between the IV and the the outcome is unconfounded




- If genes were randomized, then this assumption would be true
- *Are genes randomized?*



## Meiosis

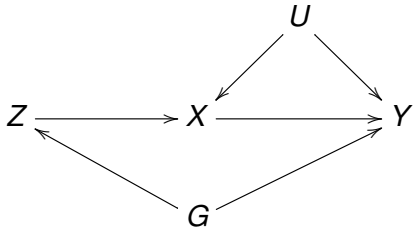



- In meiosis, the child randomly receives one allele from the father and one allele from the mother
- So in this sense, the child's alleles are randomized **conditional on the parents' genes**
  - thus the term 'Mendelian randomization'
- But in typical MR studies we have not measured the parents' genes, so the analysis cannot condition on these
- As a consequence, MR studies may suffer from problems due to parental genetic effects and population mixture



## Parental genetic effects

- The parents' genes clearly affect the child's genes
- Suppose that the parents' genes also affect the child's outcome
  - e.g. by affecting how the parents behave to the child

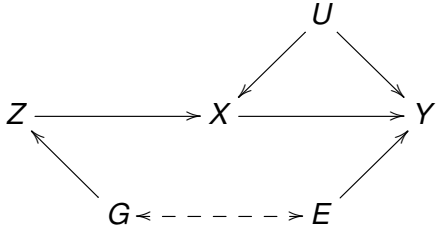



- Then assumption 2 would be violated



## Population mixture

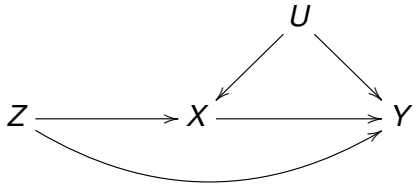
- Suppose that there is population mixture
  - i.e. the study population consists of different ethnic groups
- The parents' ethnicity may affect/be correlated with their genes
- Suppose that the parents' ethnicity also affects the child's outcome
  - e.g. due to social or cultural differences, some ethnic groups are more likely to develop the outcome than other




- Then assumption 2 would be violated

## Pleiotropic effects

- Even if genes were truly randomized, an additional concern could be pleiotropic effects
- In particular, suppose that the IV affects the outcome through mechanism that are unrelated to the exposure




- Then assumption 3 would be violated



## Outline



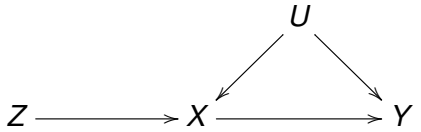


## A special application of IVs

- There is one special application of IVs, in which the IV assumptions are (almost) guaranteed to hold
- This is the randomized controlled trial (RCT) with non-compliance



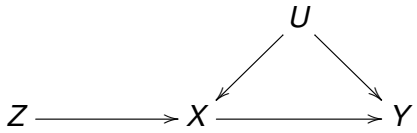
## The basic scenario




- Let $Z$ be 'assigned treatment level'
  - $Z = 1$ for 'assigned to treatment', $Z = 0$ for 'assigned to no treatment'
- Let $X$ be 'treatment level actually taken'
  - $X = 1$ for 'took the treatment', $X = 0$ for 'did not take the treatment'
- Simplifying assumption:
  - no blinding
  - 'treatment' vs 'no treatment', not 'treatment' vs 'placebo'

## Assumption 1

- The IV is associated with the exposure

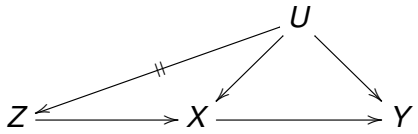



- This assumption is true, since the decision to take the treatment or not is most likely influenced by the assignment



## Assumption 2

- The association between the IV and the outcome is unconfounded

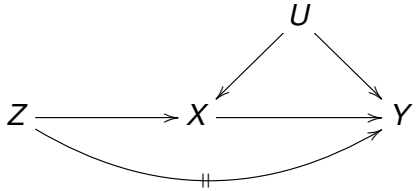



- This assumption is true, due to randomization of assigned treatment level



## Assumption 3)

- The IV affects the outcome only through the exposure




- In RCT's, a direct effect of $Z$ on $Y$ is possible in principle, but unlikely in practice
- *Why?*



## The ITT analysis

- One conventional method to analyze RCTs with non-compliance is the 'intention-to-treat' (ITT) analysis
- In this analysis, one tests for an association between 'assigned treatment level' ($Z$) and the outcome ($Y$)
- This is exactly the IV hypothesis test we discussed previously

# Outline





# Outline





# Motivating example

- Sommer and Zeger (1991) considered a RCT of vitamin A supplement in northern Sumatra
- In the treatment group, oral doses of vitamin A were administered at 3 and 6 months
- At 12 months, mortality was determined



# Results from the Sommer and Zeger (1991) study

Table I. Mortality rates in control and programme villages, months 4–12, stratified by compliance

| Study group | Complied | Children | Deaths | Mortality (per 1000) |
|---|---|---|---|---|
| Control | — | 11,588 | 74 | 6·4 |
| Treatment | — | 12,094 | 46 | 3·8 |
| | Yes | 9,675 | 12 | 1·2 |
| | No | 2,419 | 34 | 14·1 |

An intent-to-treat analysis was performed to estimate the programmatic effectiveness of the vitamin A distribution. The mortality in the two groups from month 4 (following completion of the first distribution cycle) to month 12 was compared. As shown in Table I, there were 46 deaths out of 12,094 children (0.38 per cent) in the villages randomized to vitamin A, as opposed to 74 out of 11,588 (0.64 per cent) in control villages. The effectiveness relative risk is estimated by 0·38/0·64 = 0·59 (95 per cent confidence interval: 0·41 to 0·86). Hence, the data suggest roughly a 40 per cent reduction in mortality associated with the vitamin A distribution programme.

- This the effect of 'treatment assignment' (the ITT effect), not the effect of treatment
- What is the treament effect?

## The IV assumptions are not enough for estimation

- We have seen that we can **test** for a causal exposure effect, assuming that the IV assumptions hold
- Balke and Pearl (1997) showed that the IV assumptions are not enough to **estimate** the causal exposure effect
- However, they showed that the causal effect can sometimes be bounded
  - i.e. that it may be possible to provide a range of possible values, given the observed data



## Notation

- Suppose that $Z$, $X$, and $Y$ are binary
- Let $p(Y_1 = 1)$ be the counterfactual probability of the outcome, had everybody been exposed
- Let $p(Y_0 = 1)$ be the counterfactual probability of the outcome, had everybody been unexposed
- Let $\psi$ be the causal risk difference

$$\psi = p(Y_1 = 1) - p(Y_0 = 1)$$

- Balke and Pearl (1997) showed that $\psi$ is not estimable even if the IV assumptions hold



## Bounds

- Define $p_{yx.z} = p(Y = y, X = x|Z = z)$
- Balke and Pearl (1997) showed that

$$\max \left\{ \begin{array}{c} p_{00.0} + p_{11.1} - 1 \\ p_{00.1} + p_{11.1} - 1 \\ p_{11.0} + p_{00.1} - 1 \\ p_{00.0} + p_{11.0} - 1 \\ 2p_{00.0} + p_{11.0} + p_{10.1} + p_{11.1} - 2 \\ p_{00.0} + 2p_{11.0} + p_{00.1} + p_{01.1} - 2 \\ p_{10.0} + p_{11.0} + 2p_{00.1} + p_{11.1} - 2 \\ p_{00.0} + p_{01.0} + p_{00.1} + 2p_{11.1} - 2 \end{array} \right\} \leq \psi \leq \min \left\{ \begin{array}{c} 1 - p_{10.0} - p_{01.1} \\ 1 - p_{01.0} - p_{10.1} \\ 1 - p_{01.0} - p_{10.0} \\ 1 - p_{01.1} - p_{10.1} \\ 2 - 2p_{01.0} - p_{10.0} - p_{10.1} - p_{11.1} \\ 2 - p_{01.0} - 2p_{10.0} - p_{00.1} - p_{01.1} \\ 2 - p_{10.0} - p_{11.0} - 2p_{01.1} - p_{10.1} \\ 2 - p_{00.0} - p_{01.0} - p_{01.1} - 2p_{10.1} \end{array} \right\}$$

- (Note: the right inequality sign is turned the wrong way in Balke and Pearl (1997))
- All components in the lower and upper bounds are estimable



## The Sommer and Zeger (1991) study revisited

Table I. Mortality rates in control and programme villages, months 4–12, stratified by compliance

| Study group | Complied | Children | Deaths | Mortality (per 1000) |
|---|---|---|---|---|
| Control | — | 11,588 | 74 | 6·4 |
| Treatment | — | 12,094 | 46 | 3·8 |
| | Yes | 9,675 | 12 | 1·2 |
| | No | 2,419 | 34 | 14·1 |

- $Z = 1$ for 'randomized to vitamine A', $X = 1$ for 'took vitamine A', $Y = 1$ for 'survived'
- To calculate bounds for the causal risk difference we need $p_{yx.z}$ for all combinations of $y$, $x$ and $z$

$p_{00.0} = 74/11588 = 0.0064$ $p_{00.1} = 34/12094 = 0.0028$

$p_{01.0} = 0$ $p_{01.1} = 12/12094 = 0.0010$

$p_{10.0} = (11588 - 74)/11588 = 0.9936$ $p_{10.1} = (2419 - 34)/12094 = 0.1972$

$p_{11.0} = 0$ $p_{11.1} = (9675 - 12)/12094 = 0.7990$

## Bounds in the Sommer and Zeger (1991) study

$$-0.1946 \leq \psi \leq 0.0054$$

- The bounds include the value 0, thus it is possible that there is no causal effect of vitamin A
- But we observed an ITT effect! Doesn't the presence of an ITT effect prove that there is a causal treatment effect?
- The presence of an ITT effect implies that there is a causal effect **for some subjects in the population**
- But the average (population) causal effect may still be 0, if there is heterogeneity in treatment response
  - i.e. the treatment is harmful for some subjects but not for other subjects (Balke and Pearl, 1997)



## Bounds in the Sommer and Zeger (1991) study, cont'd

$$-0.1946 \leq \psi \leq 0.0054$$

- The bounds include negative values, thus it is possible that causal effect of vitamin A is negative
  - even though the ITT effect is positive
- Again, this could happen if there is a strong heterogeneity in treatment response



## Bounds vs confidence intervals

$$-0.1946 \leq \psi \leq 0.0054$$

- The bounds should not be confused with a confidence interval
- A confidence interval quantifies the uncertainty due to sampling variability
  - decreases with sample size
- The bounds quantify the uncertainty due to unmeasured confounding
  - does not decrease with sample size
- In practice; estimate the lower and upper bound, and compute confidence intervals for the estimates



## Limitations of the bounds

- The bounds show how much (or little!) we can say about the causal effect, without further assumptions than the IV assumptions
- But sometimes additional assumptions are reasonable, such as (approximate) homogeneity in treatment response
- Also, the bounds are difficult to apply for non-binary variables

# Outline





# The causal linear model

$$E(Y_x|U) = \psi x + f(U)$$

- $f(U)$ is an unspecified (linear or non-linear) function of $U$



# Interpretation of $\psi$

- $\psi$ is the conditional causal effect of $X$ on $Y$, given $U$
- It measures the increase in the mean of $Y$, when $X$ is increased with 1 unit, for those with a given value of $U$

$$E(Y_{x+1}|U) - E(Y_x|U) = \{\psi(x+1) + f(U)\} - \{\psi x + f(U)\} = \psi$$



# Alternative interpretation of $\psi$

- Because the mean difference is collapsible, $\psi$ is also the marginal causal effect of $X$ on $Y$
- It measures the increase in the mean of $Y$, when $X$ is increased with 1 unit, for the whole population

$$\begin{aligned} & E(Y_{x+1}) - E(Y_x) \\ &= E\{E(Y_{x+1}|U)\} - E\{E(Y_x|U)\} \\ &= E\{E(Y_{x+1}|U) - E(Y_x|U)\} \\ &= \psi \end{aligned}$$

# Assumptions in the causal linear model

$$E(Y_x|U) = \psi x + f(U)$$

- Linear effect of $X$ on $Y$
- The effect of $X$ is the same across levels of $U$
  - no effect modification by $U$
  - aka no interaction between $X$ and $U$ on the linear scale
  - aka homogeneity in treatment response



# TS-estimation

$$E(Y_x|U) = \psi x + f(U)$$

- A convenient way to estimate $\psi$ is to use 'two-stage' (TS) estimation
- First stage: fit a 'working' linear regression model for $X$ on $Z$
- Second stage: fit a 'working' linear regression model for $Y$ on $Z$
- Estimate $\psi$ as the ratio of the two regression slopes
- TS-estimation gives an unbiased estimate of $\psi$ **regardless of whether the working models are correct or not**
  - provided that the IV assumptions hold, and the causal linear model is correct
  - doesn't matter whether we assume linear, dominant or recessive effect of genetic alleles in working models



# Results from the Kivimäki et al. (2011) study

**Table 2.** Associations of Fat Mass and Obesity-Associated (*FTO*) Genotype With Body Mass Index, Overweight, Obesity, and Common Mental Disorders From 4 Repeated Assessments in Men and Women, Whitehall II Study, 1985–2004

| Predictor | No. of Participants | BMI[a] | | Overweight[b] | | Obesity[b] | | Common Mental Disorders[c] | |
|---|---|---|---|---|---|---|---|---|---|
| | | β | 95% CI | β | 95% CI | β | 95% CI | β | 95% CI |
| *Men* | | | | | | | | | |
| No. of *FTO* adiposity alleles | | | | | | | | | |
| 0 | 1,046 | 0.00 | Referent | 0.00 | Referent | 0.00 | Referent | 0.00 | Referent |
| 1 | 1,442 | 0.283 | 0.039, 0.526 | 0.200 | 0.064, 0.335 | 0.018 | −0.054, 0.090 | 0.020 | −0.065, 0.105 |
| 2 | 493 | 0.935 | 0.608, 1.262 | 0.331 | 0.144, 0.519 | 0.192 | 0.094, 0.289 | 0.172 | 0.058, 0.286 |
| Per-allele increase | 2,981 | 0.433 | 0.275, 0.592 | 0.173 | 0.083, 0.262 | 0.081 | 0.034, 0.129 | 0.074 | 0.019, 0.129 |
| *P* for trend | | <0.0001 | | <0.0001 | | 0.001 | | 0.009 | |

$$\hat{\psi}_{men} = 0.074/0.433 = 0.17$$



# Results from the Kivimäki et al. (2011) study, cont'd

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Women* | | | | | | | | | |
| No. of *FTO* adiposity alleles | | | | | | | | | |
| 0 | 448 | 0.00 | Referent | 0.00 | Referent | 0.00 | Referent | 0.00 | Referent |
| 1 | 543 | −0.183 | −0.741, 0.376 | 0.015 | −0.196, 0.225 | −0.076 | −0.441, 0.289 | 0.073 | −0.074, 0.219 |
| 2 | 173 | 0.310 | −0.474, 1.094 | 0.145 | −0.155, 0.445 | 0.121 | −0.391, 0.632 | −0.011 | −0.217, 0.194 |
| Per-allele increase | 1,164 | 0.077 | −0.294, 0.449 | 0.058 | −0.083, 0.120 | 0.012 | −0.089, 0.114 | 0.012 | −0.085, 0.110 |
| *P* for trend | | 0.68 | | 0.42 | | 0.81 | | 0.80 | |

Abbreviations: BMI, body mass index; CI, confidence interval; *FTO*, fat mass and obesity-associated gene; GHQ, General Health Questionnaire.
[a] Mean BMI (weight (kg)/height (m)$^2$) across 4 repeated clinical examinations conducted over a 19-year follow-up period.
[b] Number of times (range, 0–4) a participant was found to be overweight (BMI 25.0–29.9) or obese (BMI ≥30) in 4 examinations conducted over a 19-year follow-up period. The analysis of overweight did not include obese participants and was therefore based on 2,473 men and 861 women.
[c] Number of times (range, 0–4) a participant was designated a GHQ "case" in 4 examinations conducted over a 19-year follow-up period. GHQ score was used as the measure of common mental disorders.

$$\hat{\psi}_{women} = 0.012/0.077 = 0.16$$

## Special case: binary variables

- When $Z$, $X$ and $Y$ are binary, $\psi$ is the risk difference

$$p(Y_1 = 1|U) - p(Y_0 = 1|U) = p(Y_1 = 1) - p(Y_0 = 1)$$

and the TS estimate can be written as

$$\frac{p(Y = 1|Z = 1) - p(Y = 1|Z = 0)}{p(X = 1|Z = 1) - p(X = 1|Z = 0)}$$

- The numerator is the ITT effect, as a risk difference
- The denominator is equal to 1 if 'compliance' is 100%



## TS-estimation for the Sommer and Zeger (1991) study

Table I. Mortality rates in control and programme villages, months 4–12, stratified by compliance

| Study group | Complied | Children | Deaths | Mortality (per 1000) |
|---|---|---|---|---|
| Control | — | 11,588 | 74 | 6·4 |
| Treatment | — | 12,094 | 46 | 3·8 |
| | Yes | 9,675 | 12 | 1·2 |
| | No | 2,419 | 34 | 14·1 |

- $Z = 1$ for 'randomized to vitamine A', $X = 1$ for 'took vitamine A', $Y = 1$ for 'survived'

$$\begin{aligned} p(Y = 1|Z = 0) &= (11588 - 74)/11588 \\ &= 0.9936 \\ p(Y = 1|Z = 1) &= (12094 - 12 - 34)/12094 \\ &= 0.9962 \end{aligned}$$

$$\begin{aligned} p(X = 1|Z = 0) &= 0 \\ p(X = 1|Z = 1) &= 9675/12094 \\ &= 0.8000 \end{aligned}$$

$$\hat{\psi} = \frac{p(Y = 1|Z = 1) - p(Y = 1|Z = 0)}{p(X = 1|Z = 1) - p(X = 1|Z = 0)} = \frac{0.9962 - 0.9936}{0.8000 - 0} = 0.0033$$



## The bounds vs TS-estimation

- Bounds:

$$-0.1946 \leq \psi \leq 0.0054$$

- TS-estimation:

$$\hat{\psi} = 0.0033$$

- Clearly, the TS estimate relies heavily on the assumption of homogeneity in treatment response



## Alternative formulation of TS-estimation

- First stage: fit a 'working' linear regression model for $X$ on $Z$
- Create predictions $\hat{X}$ from the first stage model
- Second stage: fit a 'working' linear regression model for $Y$ on $\hat{X}$
- The two formulations give identical estimates when all models are linear
- Advantage of the 'ratio of slope' formulation: can be used when only regression slopes are available
- Advantage of the 'prediction' formulation: can be generalized to allow for exposure-covariate interactions

## Limitations of TS-estimation

- TS-estimation has the advantage of being simple
- However, for non-linear (e.g. logistic) models it gives biased estimates (Vansteelandt et al., 2011)
  - at best the bias is small, but the bias can be substantial
- Therefore, we will briefly consider a more general estimation technique called 'G-estimation'



## G-estimation

- With G-estimation, we obtain an estimate of $\psi$ by solving the equation

$$\sum_{i=1}^{n} d(Z_i)(Y_i - \psi X_i) = 0$$

where $d(Z_i)$ is an arbitrary function of $Z_i$ with mean 0



## Intuition behind G-estimation

$$\sum_{i=1}^{n} d(Z_i)(Y_i - \psi X_i) = 0$$

- The term $Y - \psi X$ can be thought of as a prediction of the counterfactual outcome $Y_0$
  - i.e. the outcome that we would have observed, had the subject counterfactually received exposure level $X = 0$
- Under the IV assumptions, $Y_0$ and $Z$ are independent in the population
- The solution to the equation above is the value of $\psi$ for which $Y_0$ and $Z$ are independent in the sample



## Analytic solution

- The equation for G-estimation has an analytic solution

$$\hat{\psi} = \frac{\sum_{i=1}^{n} d(Z_i) Y_i}{\sum_{i=1}^{n} d(Z_i) X_i}$$

## Relation between TS-estimation and G-estimation

- G-estimation:

$$\sum_{i=1}^{n} d(Z_i)(Y_i - \psi X_i) = 0$$

- Let $\bar{Z}$ be the sample mean of $Z$
- TS-estimation is a special case of G-estimation with $d(Z_i) = Z_i - \bar{Z}$:

$$\sum_{i=1}^{n} (Z_i - \bar{Z})(Y_i - \psi X_i) = 0$$

- Robins (1994) derived the most efficient choice of $d(Z_i)$, which is more complicated



## Example

| $i$ | $Z$ | $X$ | $Y$ |
|---|---|---|---|
| 1 | 0 | 25 | 4 |
| 2 | 0 | 26 | 4 |
| 3 | 0 | 27 | 5 |
| 4 | 1 | 26 | 4 |
| 5 | 1 | 27 | 5 |
| 6 | 1 | 29 | 3 |
| 7 | 2 | 25 | 5 |
| 8 | 2 | 27 | 4 |
| 9 | 2 | 29 | 5 |

$$\hat{\psi} = \frac{\sum_{i=1}^{n} d(Z_i) Y_i}{\sum_{i=1}^{n} d(Z_i) X_i}$$

- $Z$ = No. of FTO adiposity alleles, $X$ = BMI, $Y$ = CMD score
- *Use G-estimation to estimate $\psi$ for the data above, with $d(Z_i) = Z_i - \bar{Z}$. Interpret the result*



## Solution



## TS-estimation for the example

- First stage: linear regression model for $X$ on $Z$:

```
> fit1 <- lm(formula=X~Z)
> summary(fit1)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.2778     0.7982  32.921 6.18e-09 ***
Z             0.5000     0.6183   0.809    0.445
```

- Second stage: linear regression model for $Y$ on $Z$:

```
> fit2 <- lm(formula=Y~Z)
> summary(fit2)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.1667     0.3900  10.683 1.38e-05 ***
Z             0.1667     0.3021   0.552    0.598
```

- Ratio of the two regression slopes:

$$\hat{\psi} = 0.1667/0.5 = 0.3333$$

## TS-estimation for the example

- First stage: linear regression model for $X$ on $Z$:

```
. regress X Z
------------------------------------------------------------------------------
           X |      Coef.   Std. Err.      t    P>|t|     [95%
-------------+----------------------------------------------------------------
           Z |          .5    .618284     0.81   0.445    -.962
       _cons |   26.27778   .7982012    32.92   0.000     24.39
------------------------------------------------------------------------------
```

- Second stage: linear regression model for $Y$ on $Z$:

```
. regress Y Z
------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95%
-------------+----------------------------------------------------------------
           Z |   .1666667    .302109     0.55   0.598    -.5477
       _cons |   4.166667    .390021    10.68   0.000     3.244
------------------------------------------------------------------------------
```

- Ratio of slopes: $\hat{\psi} = 0.1667/0.5 = 0.3333$



## TS-estimation vs G-estimation

- For linear models, both TS-estimation and G-estimation gives unbiased estimates
- For non-linear (e.g. logistic) models, TS-estimation gives biased estimates
- G-estimation can be used in non-linear (e.g. logistic) models to obtain unbiased estimates
- This is beyond the scope of the course; we refer to Vansteelandt et al. (2011)

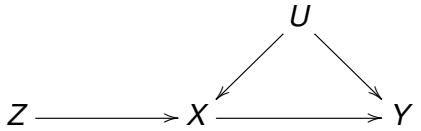


## Outline





## Case-control studies: testing



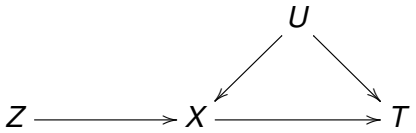

- To test for a causal effect of $X$ on $Y$ we test if $Z$ and $Y$ are associated
- This is straight-forward under case-control sampling as well
  - e.g. by a logistic regression model

## Case-control studies: estimation

- Can use TS-estimation with a logistic regression model for the outcome (second stage)
  - gives (more or less) biased estimates, due to non-linear model
- G-estimation gives unbiased estimates
- See Vansteelandt et al. (2011) and Bowden and Vansteelandt (2011)



## Time-to-event outcomes: testing




- To test for a causal effect of $X$ on $T$ we test if $Z$ and $T$ are associated
- This is straight-forward
  - e.g. by a Cox proportional hazards (PH) model

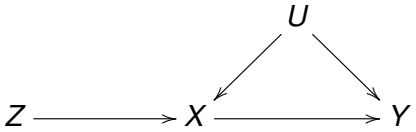


## Time-to-event outcomes: estimation

- Can use TS-estimation with a Cox PH model for the outcome (second stage)
  - but gives (more or less) biased estimates, due to non-linear model (Tchetgen Tchetgen et al., 2015)
- G-estimation gives unbiased estimates
- See Martinussen et al. (In press.)



## Weak instruments




- Assumption 1: the IV is associated with the exposure
- In practice, the association may be more or less strong
- When the association is weak, we say that the IV is a weak instrument
- In large samples, it doesn't matter that the IV is a weak instrument; it still produces valid inference
- In small samples, weak instruments create two problems:
  - low power for hypothesis testing
  - bias for estimation

## Allele scores

- In MR studies, the weak-instrument problem can sometimes be avoided by combining several genes/alleles into one 'allele score'
- The exposure is often more strongly associated with the allele score than with each separate allele
- It is important to consider the IV assumptions for each separate allele
  - if the IV assumptions hold for each separate allele, then they also hold for the allele score
  - if the IV assumptions are violated for at least one separate allele, then they are also violated for the allele score
- Burgess and Thompson (2013) provide an overview of the method



## Statistical uncertainty

- All estimators that we have discussed are asymptotically normal
- This means that we can assess statistical uncertainty with the usual 95% CI $\text{estimate} \pm 1.96 \times \text{standard error}$
- To obtain the standard error we can
  - stack the estimating equations and apply the sandwich formula, or
  - bootstrap

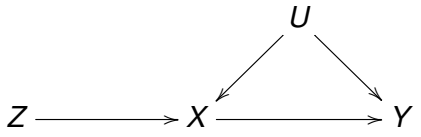


## Software

- TS-estimation for linear models:
  - `ivreg` in `AER` package in `R`
  - `ivregress` in `Stata`
- TS-estimation for log-linear models:
  - `ivpoisson` in `Stata`
- TS-estimation and G-estimation in linear, log-linear, logistic and Cox PH models:
  - `ivtools` package in `R`



## Summary




- Instrumental variables is a useful technique to deal with unmeasured confounding
- Hypothesis testing is straight-forward, whereas effect estimation requires more care
- All IV methods rely on the fundamental IV assumptions
- Some IV methods rely on additional (parametric assumptions) a well
- In real applications (e.g. Mendelian randomization studies), try to verify all assumptions to the extent possible

# References


Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association* **92,** 1171–1176.

Bonet, B. (2001). Instrumentality tests revisited. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pages 48–55. Morgan Kaufmann Publishers Inc.

Bowden, J. and Vansteelandt, S. (2011). Mendelian randomization analysis of case-control data using structural mean models. *Statistics in Medicine* **30,** 678–694.

Burgess, S. and Thompson, S. (2013). Use of allele scores as instrumental variables for mendelian randomization. *International Journal of Epidemiology* **42,** 1134–1144.

Kivimäki, M., Jokela, M., Hamer, M., Geddes, J., Ebmeier, K., Kumari, M., Singh-Manoux, A., Hingorani, A., and Batty, D. (2011). Examining overweight and obesity as risk factors for common mental disorders using fat mass and obesity-associated FTO genotype-instrumented analysis. The Whitehall II Study, 1985–2004. *American Journal of Epidemiology* **173,** 421–429.

Martinussen T., Sorensen D.N., and Vansteelandt S. (In press). Instrumental variables estimation under a structural Cox model. *Biostatistics*.

Robins, J. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics - Theory and Methods* **23,** 2379–2412.

Sommer, A. and Zeger, S. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine* **10,** 45–52.

Tchetgen Tchetgen, E., Walter, S., Vansteelandt, S., Martinussen, T., and Glymour, M. (2015). Instrumental variable estimation in a survival context. *Epidemiology* .

Vansteelandt, S., Bowden, J., Babanezhad, M., and Goetghebeur, E. (2011). On instrumental variables estimation of causal odds ratios. *Statistical Science* **26,** 403–422.