INSTRUMENTAL VARIABLE ESTIMATION (II)

Joy Shi, Barbra Dickerman, Miguel Hernán
DEPARTMENTS OF EPIDEMIOLOGY AND BIOSTATISTICS

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Learning objectives
At the end of this lecture you will be able to

- Critique the validity of proposed instruments
- Describe the dangers of weak instruments
- Define additional assumptions required for IV estimation

☐ Key concepts

- Weak instrument
- Homogeneity vs. Monotonicity

Outline

1. IV estimation in randomized experiments
2. IV estimation in observational studies
3. Application to smoking cessation study
   - Standard IV estimator
   - Two-stage estimator
4. Limitations of IV estimation
5. Conclusions

From last time: Key variables

- Treatment A
  - quitting smoking between baseline and 1982 (1: yes, 0: no)
- Continuous outcome Y
  - Weight gain: weight in 1982 minus baseline weight (in kg)
- Proposed instrument Z
  - Price of a pack of cigarettes in 1982 (adjusted for inflation to 2008 US $) in the state in which the individual was born
  - Dichotomized as "Price higher than $1.50"
    - 1: yes, 0: no
- Y and Z available for **1476** participants
  - For simplicity, analysis restricted to them

Data analysis
Standard IV estimator

- $\hat{E}[Y|Z=1] - \hat{E}[Y|Z=0]$
  - $= 2.686 - 2.536 = 0.1503$
- $\hat{E}[A|Z=1] - \hat{E}[A|Z=0] =$
  - $0.2578 - 0.1951 = 0.0627$

- The IV estimate of $E[Y^{a=1}] - E[Y^{a=0}]$ is
  - 0.1503 / 0.0627 = 2.4 kg

"Price higher than $1.50"
seems an arbitrary choice

- What if we use $2.00 as the cutoff?
- New proposed instrument Z
  - Price of a pack of cigarettes in 1982 (adjusted for inflation to 2008 US $) in the state in which the individual was born
  - Dichotomized as "Price higher than $2.00"
    - 1: yes, 0: no

Revised data analysis
Standard IV estimator

- $\hat{E}[Y|Z=1] - \hat{E}[Y|Z=0]$
  - $= 3.4978 - 2.6213 = 0.8765$
- $\hat{E}[A|Z=1] - \hat{E}[A|Z=0] =$
  - $0.2549 - 0.2562 = -0.0013$

- The IV estimate of $E[Y^{a=1}] - E[Y^{a=0}]$ is
  - 0.8765 / −0.0013 = −674.23 kg

7

Wow.
What about other cutoffs?

- "Price higher than \$1.60"
  - 41.3 kg
- "Price higher than \$1.70"
  - −40.9 kg
- "Price higher than \$1.80"
  - −21.1 kg
- "Price higher than \$1.90"
  - −12.8 kg

Something is amiss here

See iv.R, lines 27-35

8

Four problems of IV estimation

1. The proposed instrument Z may not be an instrument
2. A weak instrument Z amplifies and creates bias
3. An instrument is not enough for IV estimation
4. IV estimation with time-varying treatments and with survival analyses is computationally challenging and intensive

9

Reminder: Z is an instrument if it meets the instrumental conditions

i. Z and treatment A are associated
   - relevance condition

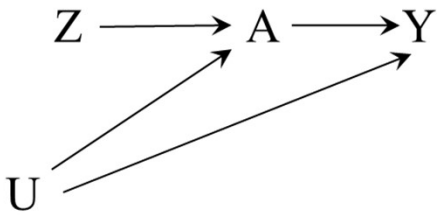
ii. Z affects the outcome Y only through its potential effect on treatment A
   - Exclusion restriction: no direct effect of Z on Y

iii. Z does not share causes with the outcome Y
   - no confounding for the effect of Z on Y

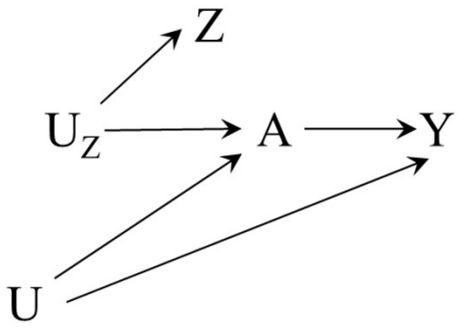
10

Instrument Z

Causal instrument



Surrogate instrument



11

Problem #1: The proposed instrument may not be an instrument

- If Z does not meet one or more instrumental conditions, the IV estimand is not the average causal effect of A on Y

$$\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}] \neq \frac{\mathrm{E}[Y|Z=1] - \mathrm{E}[Y|Z=0]}{\mathrm{E}[A|Z=1] - \mathrm{E}[A|Z=0]}$$

- Can we use the data to empirically verify whether the instrumental conditions hold?

12

Condition (i) is empirically verifiable

0

Recall, condition (i) states: Z is associated with treatment A

True 0%

False 0%

13

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

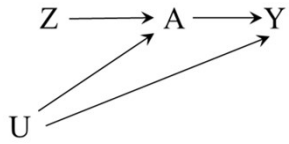
In our example

- The proposed instrument Z and the treatment A are weakly associated
- The probability of quitting smoking is
  - 25.8% for those with $Z = 1$
  - 19.5% for those with $Z = 0$
- The risk difference is ~6%
  - Weak association = Weak instrument (more later)

14

Conditions (ii) and (iii) may not hold
Consider the example of Mendelian randomization

- A: Alcohol intake
- Y: Coronary heart disease
- Z: Genetic variant



- Condition (ii) will not hold if the genetic variant has a direct effect on the outcome (direct arrow from Z to Y)
- Condition (iii) will not hold if the genetic variant is in linkage disequilibrium with another variant that affects the outcome (common cause of Z and Y)
- Can we verify whether conditions (ii)-(iii) hold?

15

Condition (ii) is empirically verifiable

0

Recall, condition (ii) states: Z affects the outcome Y only through treatment A

True 0%

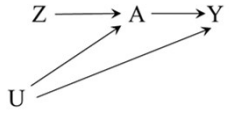
False 0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

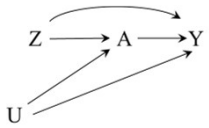
16

What if we check whether Z and Y are associated?

- Unconditionally
  - Z and Y will be associated if Z is an instrument,



  - but also if condition (ii) is violated



- Within levels of A
  - Z and Y will be associated if Z is an instrument
    - Because A is a collider
  - but also if condition (ii) is violated
    - Because Z has a direct effect on Y

17

Condition (iii) is empirically verifiable

0

Recall, condition (iii) states: Z does not share causes with the outcome Y

True 0%

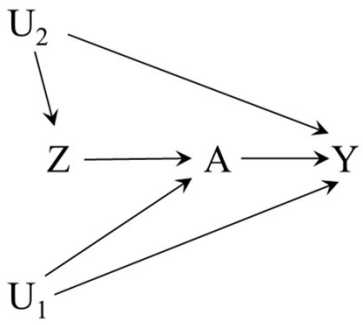
False 0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

18

Can confounding be disproven?

- Confounding for the proposed instrument is always possible in observational studies
  - same as for any other variable not under the investigators' control
- Confounding contributes to the numerator of the IV estimator and is inflated by the denominator
  - as if it were part of the effect of treatment



19

If we cannot verify conditions (ii)-(iii), then what can we do?

- Be cautious with our conclusions
  - Our use of Z may have actually introduced bias
- Attempt to empirically falsify conditions (ii)-(iii)
  - Sometimes the conditions can be falsified, which proves that Z is not an instrument
- Measure and adjust for the risk factors U_2
  - For condition (iii), we might feel more comfortable assuming that there is no unmeasured confounding for the proposed instrument within levels of the covariates, rather than unconditionally

20

Try to falsify conditions (ii) and (iii) by finding empirical evidence against them

- Various types of falsification tests
  - leveraging prior causal assumptions
  - assessing inequalities that can detect extreme violations
  - making use of effect modifiers
  - comparing estimates from several potential instruments
  - See review by Labrecque and Swanson (Cur. Epidemiol. Rep., 2018)
- Unfortunately, these tests may fail to reject a proposed instrument even if conditions (ii)-(iii) are violated

21

Condition on covariates to make condition (iii) more likely to hold

- This is where the two-stage estimator helps
  - Easier to add the covariates to the models than computing the standard IV estimator in each stratum of the covariates, and then pooling the estimates (under a homogeneity assumption)
- In our example, the two-stage estimator with covariates resulted in the smaller estimate with a larger 95% confidence interval
  - -1.0 kg; 95% CI (-57.7, 59.8) *See iv.R, lines 38-47*

22

Problem #2: A weak instrument amplifies existing bias

- Weak instrument Z = small denominator of IV estimator
  - 0.0627 in our example
- Therefore biases in the numerator
  - violations of condition (ii): direct effect of Z
  - violations of condition (iii): unmeasured confounding for Z
- will be greatly exaggerated
  - in our example, any bias affecting the numerator of the IV estimator would be multiplied by approximately 15.9 (1/0.0627)

23

Problem #2: A weak instrument creates additional bias

- Suppose we randomly generate a variable Z and propose it as an instrument
- In an infinite population,
  - the Z-A association (denominator of IV estimand) will be exactly zero
  - the IV estimate will be undefined
- In a study with a finite sample
  - By chance, the Z-A association will be small but not exactly zero
  - the IV estimate will be very biased because the numerator is incorrectly inflated

24

Problem #2: A weak instrument creates additional bias

- In our smoking cessation example, the proposed instrument "Price higher than $1.50" behaves like a randomly generated variable
  - When we redefine it as price higher than $1.60, $1.70, $1.80, or $1.90, the IV estimate changes dramatically
  - Bias

25

Problems #1 and #2

- A proposed instrument Z that fails to be an instrument
  - because it violates the instrumental conditions
- or that is a weak instrument
  - because it barely meets condition (i)
- may result in much bias when using IV estimation
  - bias may be in a counterintuitive direction
- Given how much bias weak instruments may create, a strong proposed instrument that slightly violates conditions (ii)-(iii) may be preferable to a valid, but weaker, proposed instrument

26

Problem #3: An instrument is not enough

- Suppose we have a perfect instrument Z
  - It meets the 3 instrumental conditions
  - It is not weak
- Is it then true that the IV estimand equals the average causal effect of A on Y?

$$\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}] = \frac{\mathrm{E}[Y|Z=1] - \mathrm{E}[Y|Z=0]}{\mathrm{E}[A|Z=1] - \mathrm{E}[A|Z=0]} \quad \textbf{?}$$

- **No**
  - the above equality requires an additional fourth condition

27

Problem #3: An instrument is not enough

- Remember:
  - In randomized trials, the intention-to-treat effect in the numerator is inflated by the degree of non-adherence in the denominator
- But what if the intention-to-treat effect is further from the null than the per-protocol effect?
  - Then the IV estimand would incorrectly inflate (rather than deflate) the numerator

28

Problem #3: An instrument is not enough

- An instrument only allows to compute bounds for, not a point estimate of, the average causal effect of A on Y
- Bounds too wide, not very informative
  - For dichotomous outcomes
    - Natural bounds (Robins 1989)
    - Sharp bounds, possibly tighter but still wide (Balke and Pearl 1994)
  - For continuous outcomes
    - Natural bounds (Manski 1990)
- In addition, 95% confidence intervals around the bounds

29

Problem #3: An instrument is not enough

- We need an additional condition (iv)
  - Some sort of homogeneity
- There are several versions of the homogeneity condition (iv)
- Under any of them,

$$\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}] = \frac{\mathrm{E}[Y|Z=1] - \mathrm{E}[Y|Z=0]}{\mathrm{E}[A|Z=1] - \mathrm{E}[A|Z=0]}$$

30

Several versions of homogeneity condition (iv)

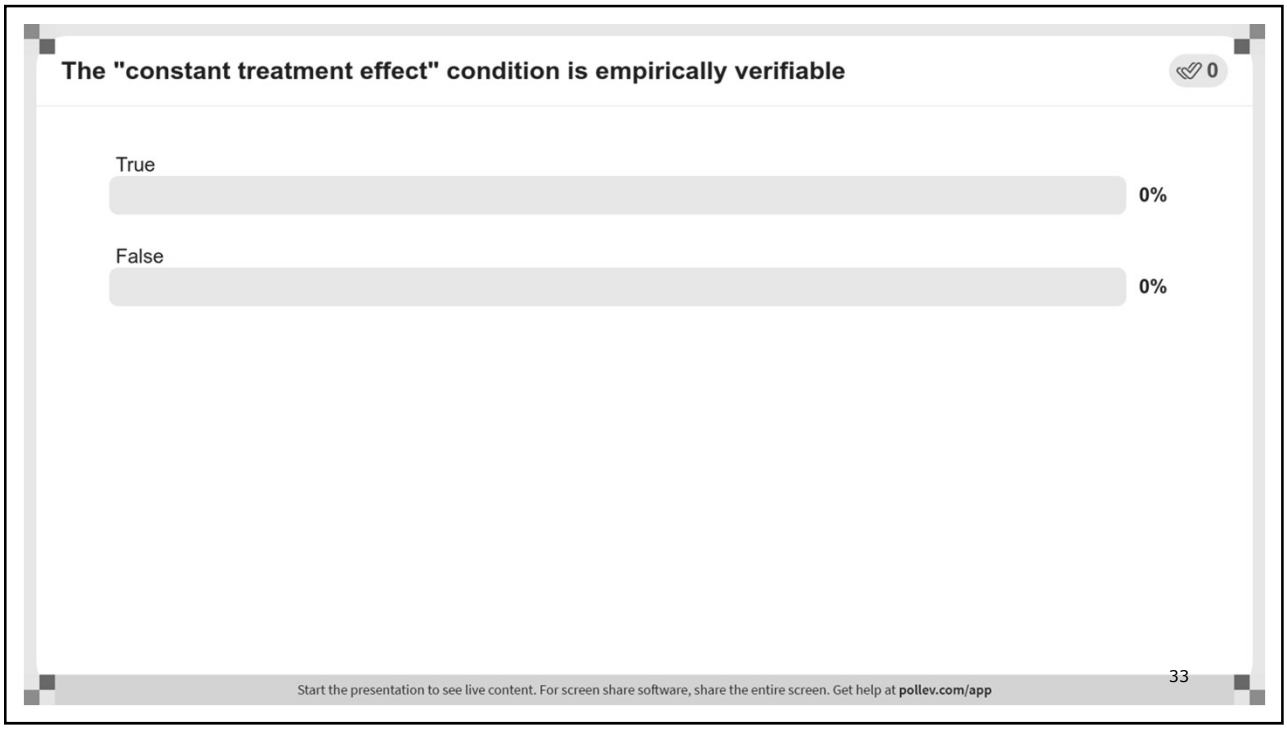
a. Treatment effect is constant for all individuals
b. Treatment effect is constant within levels of the instrument (separately in the treated and the untreated)
   - No additive "effect modification" by Z
c. Treatment effect is constant within levels of the unmeasured confounders
   - No additive "effect modification" by U
d. The Z-A association (additive scale) is constant across levels of the unmeasured confounders

31

a) Constant treatment effect

- The effect of A on Y is the same for every individual
  - Often implicit in two-stage estimators
  - Special case: sharp null hypothesis
- Biologically implausible for continuous outcomes and logically impossible for dichotomous outcomes
- Too strong an assumption
  - the extreme "no effect modification" assumption

32



b) No effect modification by Z

- No modification by Z (or U_Z) of the effect of A on Y
  - Among the treated and among the untreated

- If no effect modification on the
  - additive scale: $\mathrm{E}[Y^{a=1}]-\mathrm{E}[Y^{a=0}]$ is the standard IV estimand
  - multiplicative scale: $\mathrm{E}[Y^{a=1}]-\mathrm{E}[Y^{a=0}]$ is another IV estimand
    - See Theorem 4 in Hernán and Robins 2006

34

b) No additive effect modification by Z

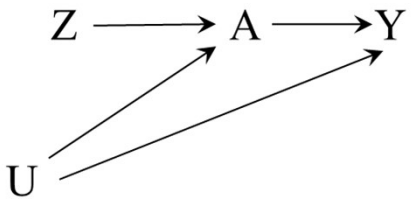
- The effect of A on Y on the risk difference scale is the same in treated individuals with Z=1 as in treated individuals with Z=0

  $$\mathrm{E}[Y^{a=1} - Y^{a=0} | A=1, Z=1] = \mathrm{E}[Y^{a=1} - Y^{a=0} | A=1, Z=0]$$

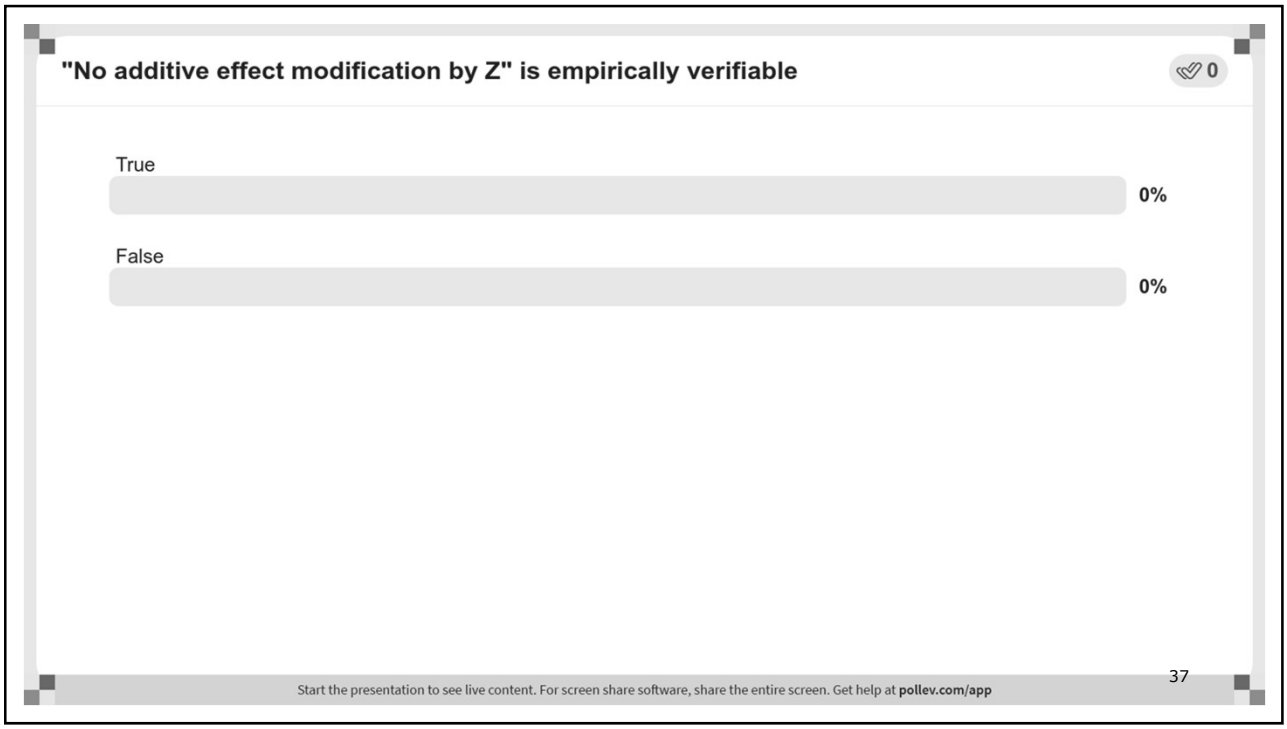
  - and similarly among untreated individuals

Example
Genetic variants

- A: Alcohol intake (1: heavy drinking, 0: mild/no drinking)
- Y: Coronary heart disease
- Z: Genetic variants associated with alcohol metabolism, e.g., ALDH2 polymorphisms in Asian populations



Effect modification in the treated if the risk difference for the effect of alcohol intake A on coronary heart disease Y was modified by genetic variants Z



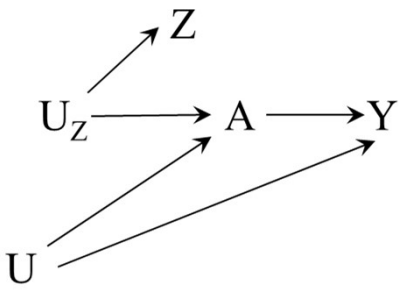
c) No additive "effect modification" by U

- Suppose the average causal effect of A on Y is not the same across levels of an unmeasured confounder
  - Then U is an additive effect modifier
- More intuitive to think about no additive effect modification by treatment-outcome confounders rather than by a proposed instrument
  - e.g., consider the use of physician's preference as a proposed instrument—it's unclear how to use subject-matter knowledge to assess whether physician's preference could modify the effect of, for example, chemotherapy on mortality
  - Even less intuitive if we have a surrogate instrument

38

Example
Preference (surrogate)

- Lung cancer patients
  - A: Type of chemotherapy
  - Y: 3-year mortality
  - U_Z: Physician's preference for type of chemotherapy
  - Z: Proportion of patients recently treated by that physician who received A=1
  - U: Past response to treatment



Effect modification in the treated if the risk difference for the effect of chemotherapy A on mortality Y was modified by past response to treatment U

39

d) The Z-A association (additive scale) is constant across levels of the confounders

- $\mathrm{E}[A|Z=1,U]-\mathrm{E}[A|Z=0,U]=\mathrm{E}[A|Z=1]-\mathrm{E}[A|Z=0]$
  - Wang and Tchetgen-Tchetgen (2018)

- More generally:
  - Any modifiers for the effect of A on Y are independent of any modifiers for the association between Z and A

40

In practice, any of these versions of

- homogeneity condition (iv) may be questionable
- What then? Is IV estimation doomed?
- Any ideas? Anyone?
- Help!

41

An alternative to
the homogeneity condition (iv)

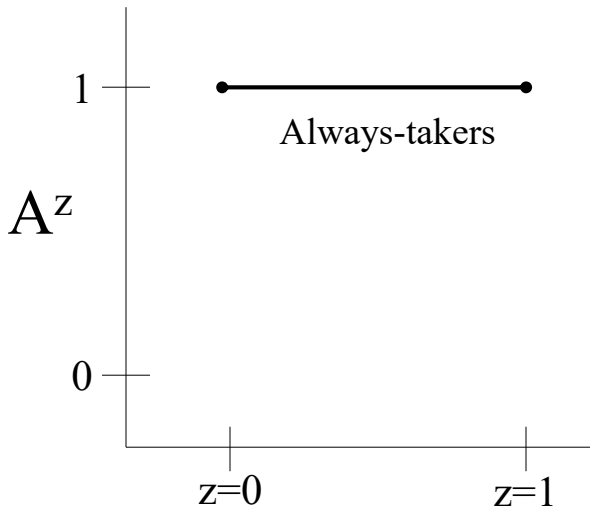
- Monotonicity condition (iv)
  - For all individuals, the level of treatment A that an individual would take if given a level of the instrument Z is a monotone increasing function of the level of Z
- Let us see what this means in a randomized trial

42

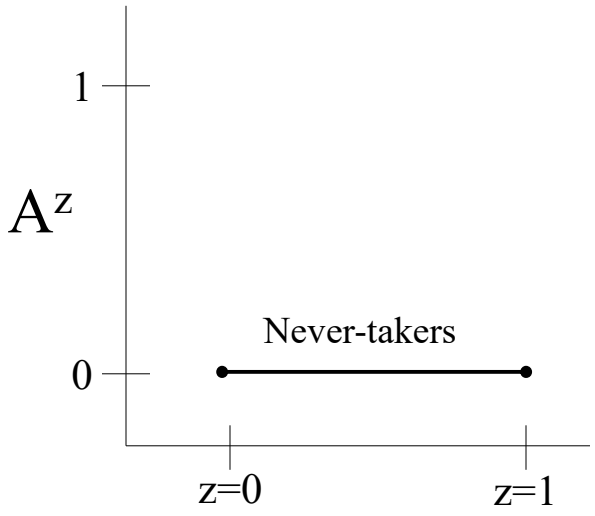
Monotonicity
Randomized trial

- Proposed instrument Z is "randomized assignment" to aspirin
  - A causal instrument
- 4 types of people:
  - Always-takers: $A^{z=0}=1, A^{z=1}=1$
  - Never-takers: $A^{z=0}=0, A^{z=1}=0$
  - Compliers: $A^{z=0}=0, A^{z=1}=1$
  - Defiers: $A^{z=0}=1, A^{z=1}=0$
- Monotonicity means that there are no defiers

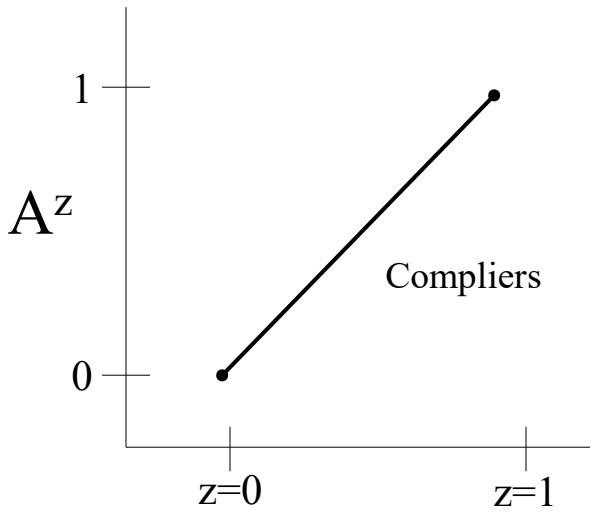
43



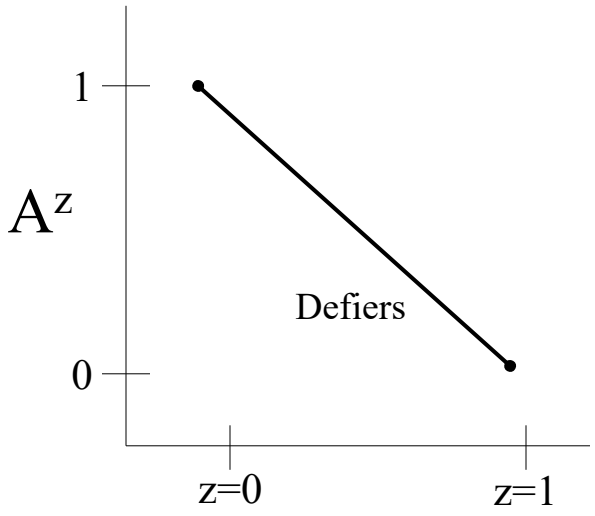
44



45



46



47

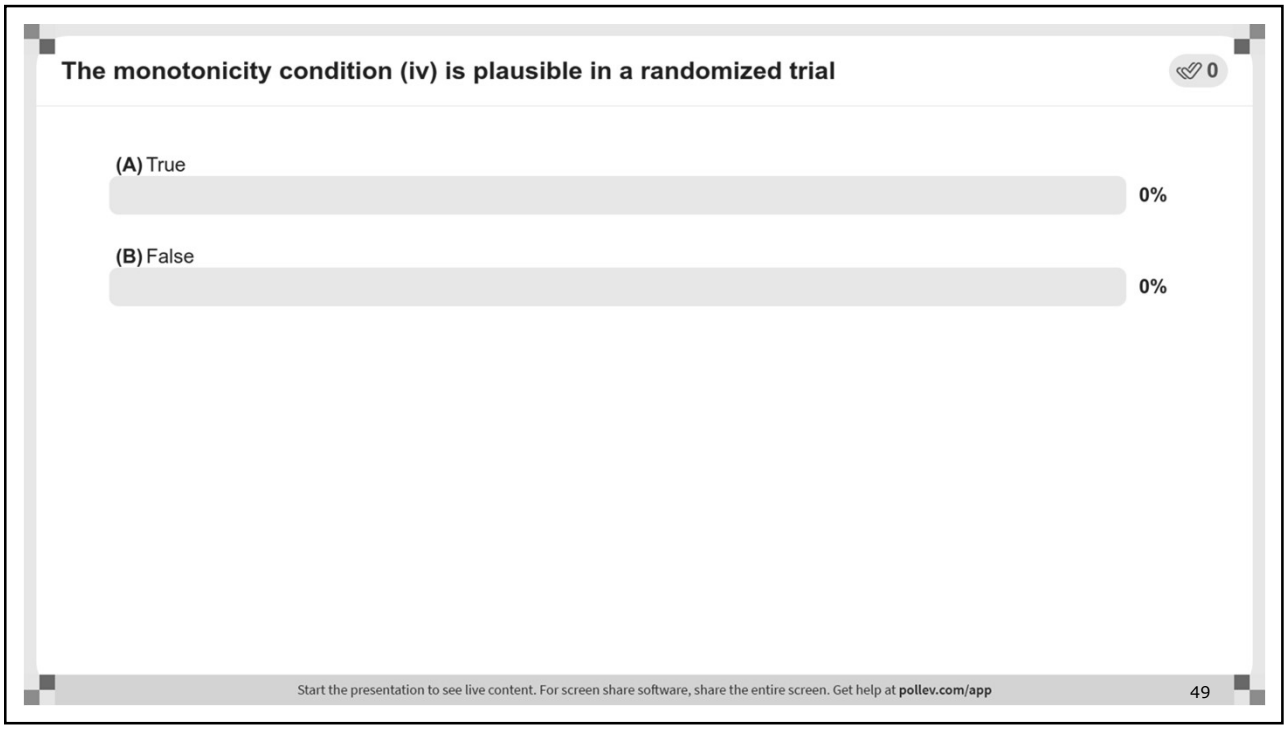
The monotonicity condition (iv) is empirically verifiable in a randomized trial 0

True 0%

False 0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

48



Under monotonicity, the standard IV estimand equals an average causal effect of treatment

- Not the average causal effect of A on Y in the population
- But rather the average causal effect of A on Y in the "compliers"
  - Local average treatment effect (LATE)
  - Complier average causal effect (CACE)
  - A result obtained by Imbens and Angrist (1995)
    - Proof in *Causal Inference: What If* Technical Point 16.7

50

Under monotonicity, how many compliers are in the randomized trial?

- % compliers + % always-takers + % never-takers =100%
  - Because there are no defiers
- % of always-takers and never-takers is known
  - % always-takers = $\Pr[A=1|Z=0]$
  - % never-takers = $\Pr[A=0|Z=1]$
- Therefore the proportion of compliers is
  - $1 - \Pr[A=0|Z=1] - \Pr[A=1|Z=0] = \Pr[A=1|Z=1] - \Pr[A=1|Z=0]$
    - See IV_2_Technical_Points document for further detail

51

Monotonicity
Observational study

- Proposed instrument Z is not "randomized assignment"
  - Instrument may not be causal
- Still 4 types of people:
  - $A^{z=0}=1, A^{z=1}=1$
  - $A^{z=0}=0, A^{z=1}=0$
  - $A^{z=0}=0, A^{z=1}=1$
  - $A^{z=0}=1, A^{z=1}=0$
- Monotonicity holds if no ($A^{z=0}=1, A^{z=1}=0$) individuals
  - We use the term "defiers" even though there is no treatment assignment to be defied

52

Monotonicity: Observational study of chemotherapy and lung cancer

- Proposed instrument Z is "physician's preference"
- If only 2 doctors in the study: doctor $Z=1$ prefers $A=1$, and doctor $Z=0$ prefers $A=0$
  - Monotonicity if no patient would receive $A=0$ if treated by doctor $Z=1$ and $A=1$ if treated by doctor $Z=0$
- If many doctors in the study:
  - Monotonicity if no patient would receive $A=0$ if treated by any of the doctors $Z=1$ (those who prefer $A=1$) and $A=1$ if treated by any of the doctors $Z=0$ (those who prefer $A=0$)
- Plausible?

53

Who are the "compliers"?

- In the randomized trial example, the subset of the population who would comply with whichever treatment is assigned to them
- In the preference example, the subset of the population who would be treated
  - with $A=1$ by all doctors who prefer $A=1$
  - with $A=0$ by all doctors who prefer $A=0$

54

Issues to keep in mind when obtaining IV estimates under monotonicity (iv)

- The "compliers" cannot be identified
  - We are estimating an effect in an unknown subset of the population
- We can calculate the % of compliers, what is the relevance of the estimate if % of compliers is small?
  - This will happen when the instrument is weak
- When using a proposed instrument Z that is a proxy for a causal instrument U_Z
  - If U_Z is dichotomous, then the IV estimand based on Z still estimates the effect in the compliers but monotonicity is now defined with respect to U_Z
  - If U_Z is continuous, then the IV estimand based on Z estimates a hard-to-interpret weighted average of effects in many subgroups
    - Hernán and Robins 2006

55

Problem #4: Time-varying treatments / Survival analysis

- Most epidemiologic treatments are time-varying
  - And many are survival analyses
- The standard IV estimator estimates the effect of a non time-varying treatment using a non time-varying instrument
  - And not considering a failure-time outcome
- How to generalize to time-varying treatments and instruments and to survival analyses?
  - G-estimation of structural models

56

Outline

1. IV estimation in randomized experiments
2. IV estimation in observational studies
3. Application to smoking cessation study
   - Standard IV estimator
   - Two-stage estimator
4. Limitations of IV estimation
5. Conclusions

Conclusions (I)

- IV methods require FOUR assumptions
- The first 3 (instrumental) assumptions are nonnegotiable
  - association between instrument and treatment
  - no direct effect of the instrument on outcome
  - no unmeasured confounding (conditional exchangeability) for the instrument
- Wide variety of fourth assumptions
  - no effect modification by instrument, monotonicity, etc…
  - Different assumptions lead to different causal estimands

Conclusions (II)

- IV methods replace one unverifiable assumption
  - no unmeasured confounding (conditional exchangeability) for the treatment

by other unverifiable assumptions

  - no unmeasured confounding (conditional exchangeability) for the instrument
  - no direct effect of the instrument
  - homogeneity, monotonicity, etc…
- The fundamental problem of causal inference is not solved but simply shifted
  - Need to decide which assumptions are more likely to hold in each particular example

59

Conclusions (III)

- IV methods are underutilized
  - Especially in randomized trials
- But users need to be aware of the limitations of IV methods
  - Otherwise, conflicts between IV and non-IV estimates may be counterintuitive
- If large differences between IV and non-IV adjusted estimates AND you believe the IV estimates are correct, then you believe the ratio of unmeasured to measured confounding is large

60

Progress report

1. Introduction to modeling
2. Stratified analysis:
   - outcome regression
   - propensity scores
3. Standardization
4. Inverse probability weighting
   - Marginal structural models
5. Instrumental variable estimation
6. G-estimation