Instrumental Variables

What if we don't think we have measured all confounding covariates?
Is there any hope?

Road Map

- Motivating Example: The Effect of Military Service on Civilian Mortality:
- Natural experiment
- What can we/should we estimate: ITT vs.CACE
- IV Assumptions
- Two Estimation methods : Wald vs. 2SLS
- Identifiability
- What can go wrong with IV?
- An Applied Example Using Stata.

Motivating Example: The Effect of Military Service on Civilian Mortality/ Lifetime Earnings (Angrist et al)

- Policy question: Does active military service negatively impact future (civilian) outcomes such as health status or earnings?

  https://www.youtube.com/watch?v=XhLbysRh8XY
- Why might this be difficult to answer without bias?

Motivating Example: The Effect of Military Service on Civilian Mortality/ Lifetime Earnings (Angrist et al)

- Policy question: Does active military service negatively impact future (civilian) outcomes such as health status or earnings?
- Analysis problem: Those who self-select into the military tend to be different from those who don't in ways that might affect health outcomes
- What solutions might exist?

Motivating Example: The Effect of Military Service on Civilian Mortality/ Lifetime Earnings (Angrist et al)

- Policy question: Does active military service negatively impact future (civilian) outcomes such as health status or earnings?
- Analysis problem: Those who self-select into the military tend to be different from those who don't in ways that might affect health outcomes
- Natural experiment? Between 1970 and 1973 priority for the draft for the Vietnam war was *randomly assigned* in a lottery using dates of birth

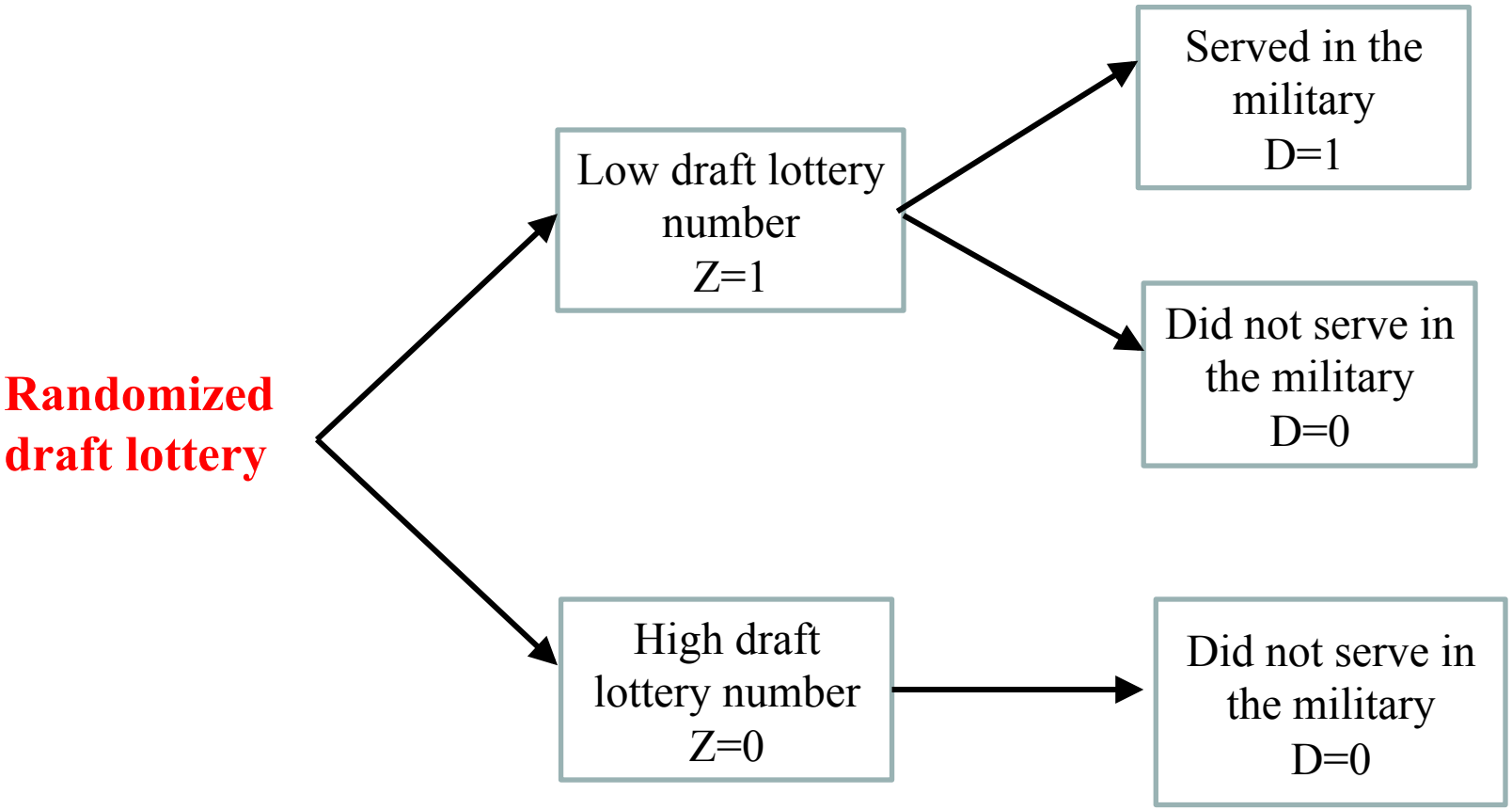
Instrumental Variables -- general

- IV is an approach to causal estimation that plays off the randomization of a variable related to the treatment, rather than the treatment itself (the randomized encouragement design is a special case)
- The catch is that you have to observe a special variable, called an "instrument" that (informally) has the following features:
  - It is related to the treatment
  - It is only related to the outcomes through its impact on the treatment

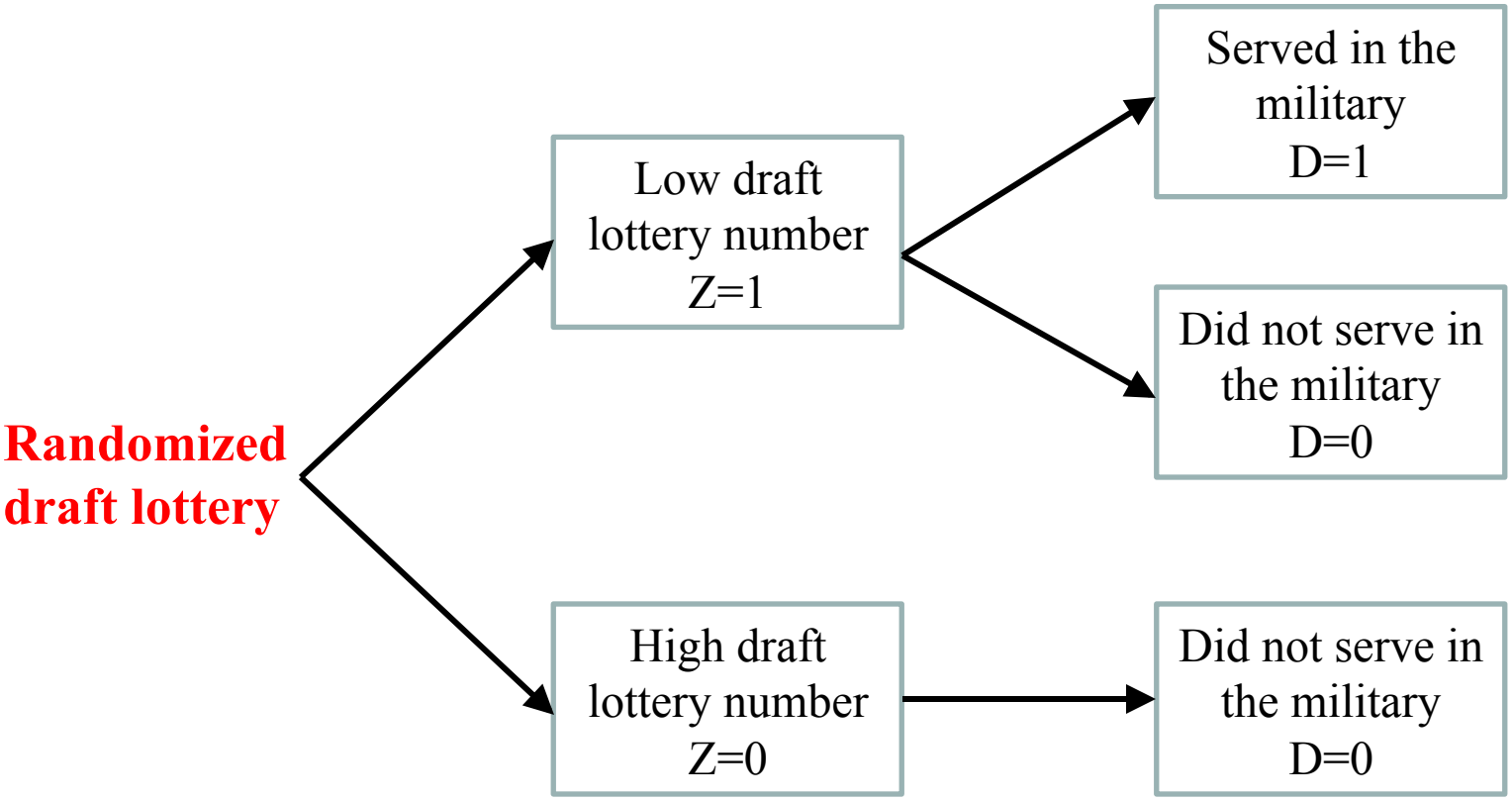
Instrumental Variables -- applied to this example

- The instrument here is generated by the Vietnam draft lottery. We will operationalize as
  - Z=1 lottery number low enough to be drafted
  - Z=0 lottery number high enough to avoid the draft
- The treatment, D, is "serving in Vietnam"
- The outcomes considered were subsequent (civilian) mortality or earnings

Simple form of observed noncompliance: researcher view

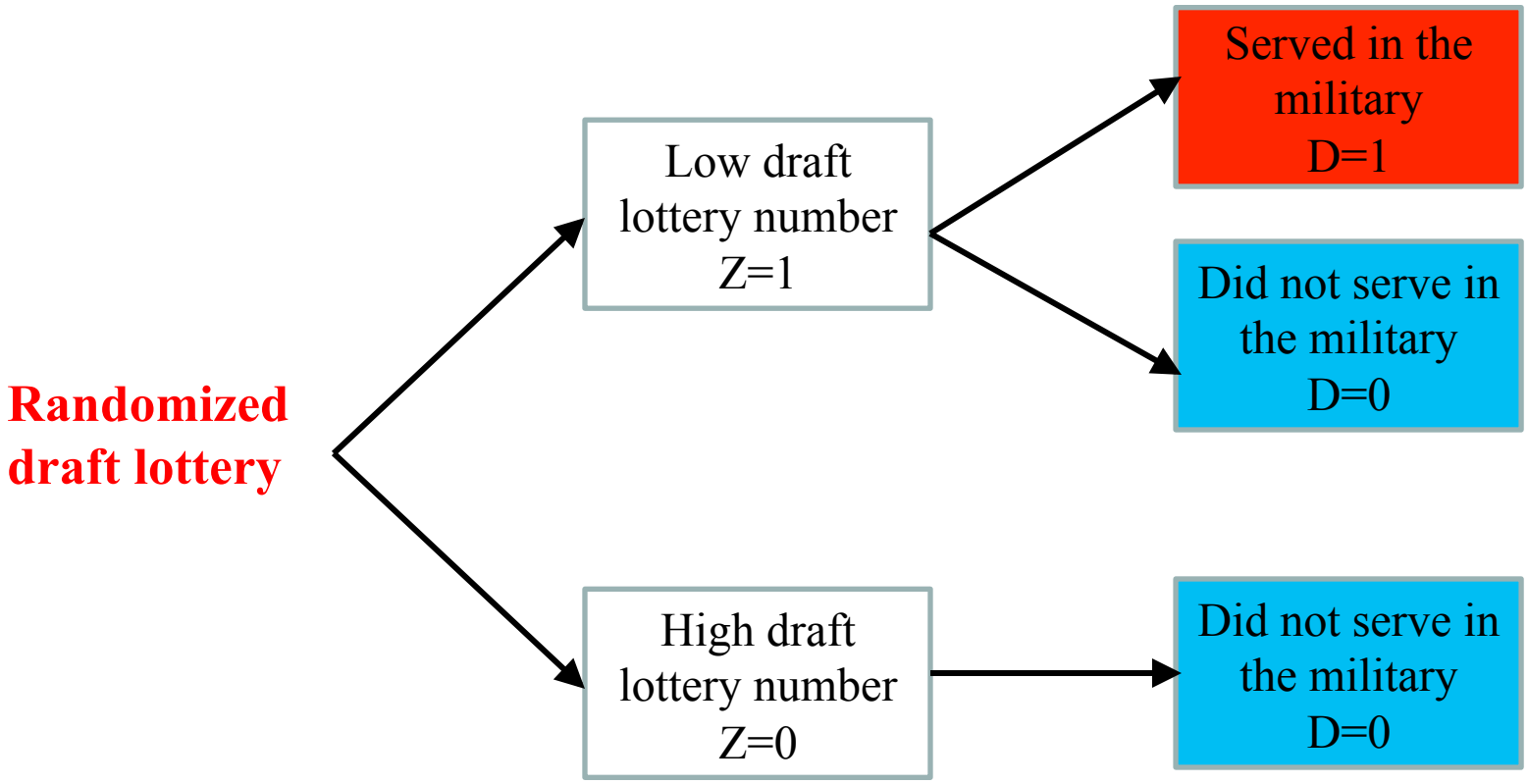


Simple form of observed noncompliance: researcher view



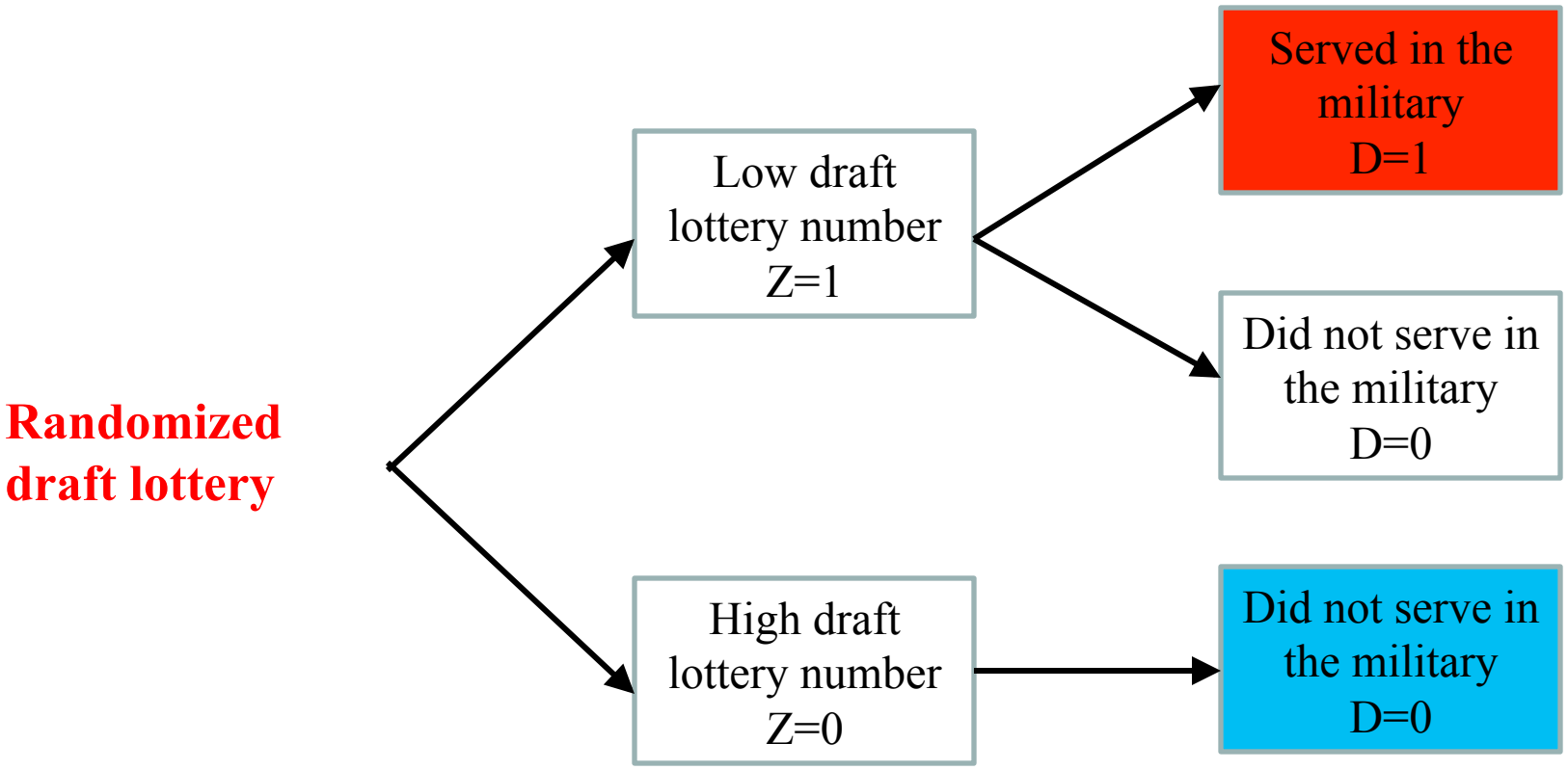
If we are interested in answering the question "What is the effect of serving in the military on subsequent health and earnings?" what types of comparisons should we make?

Naïve comparison 1: As Treated



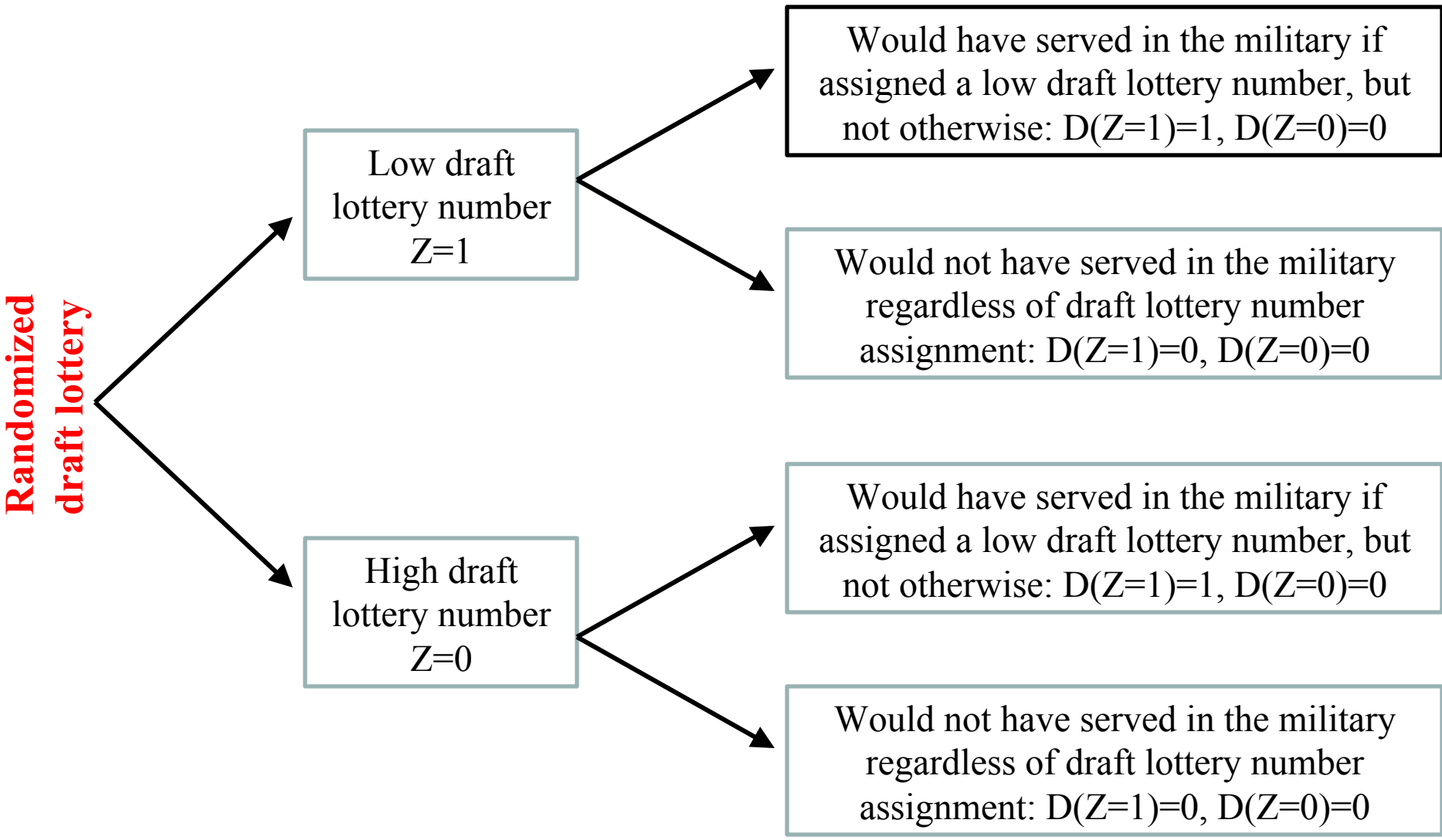
As-treated analyses make comparisons between those who received the treatment and those who did not.

Naïve comparison 2: Per protocol

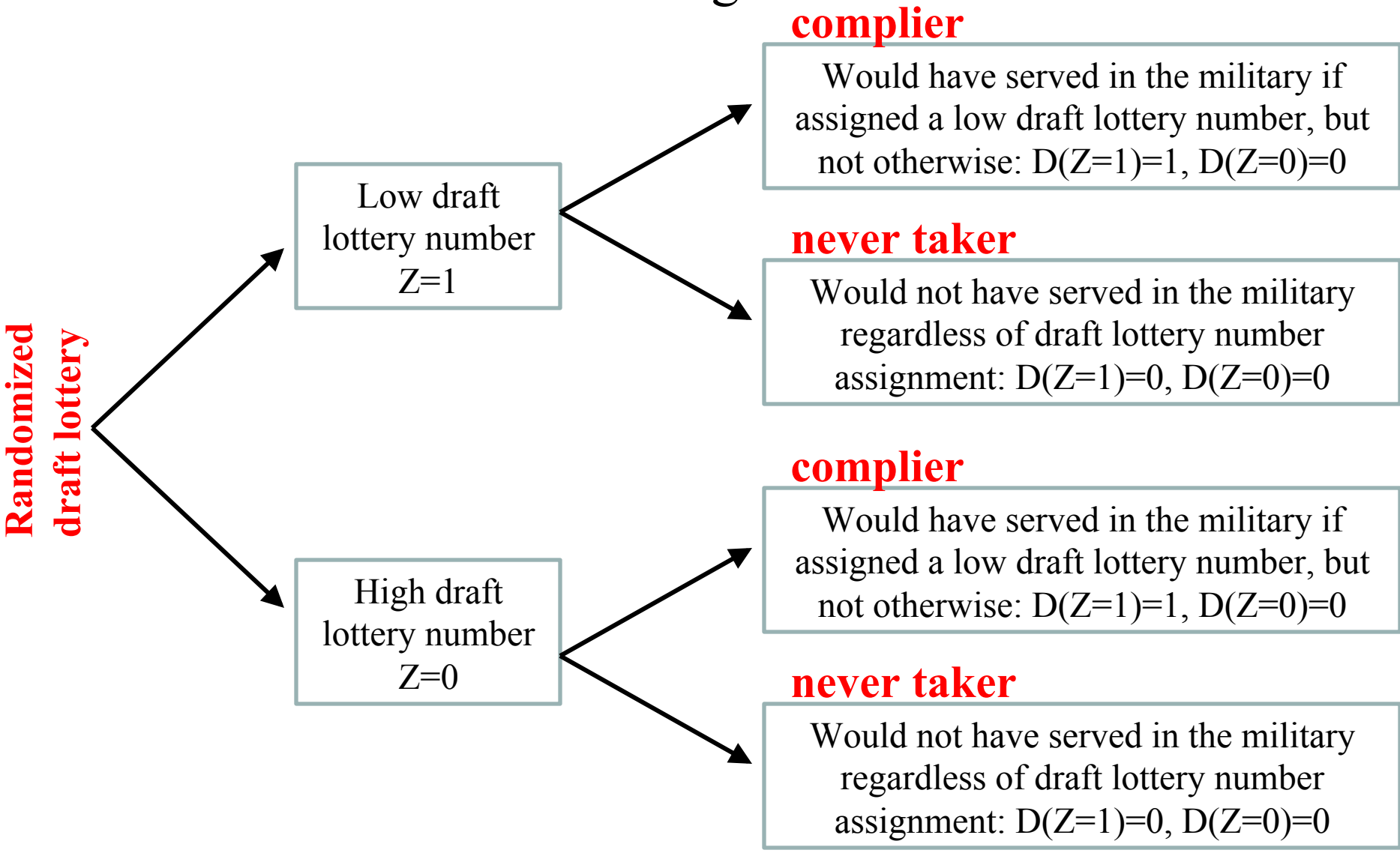


Per protocol analyses make comparisons between those who received the treatment who were assigned to receive treatment and those who did receive treatment among those who were not assigned to receive treatment. Thus observed non-compliers are ignored.

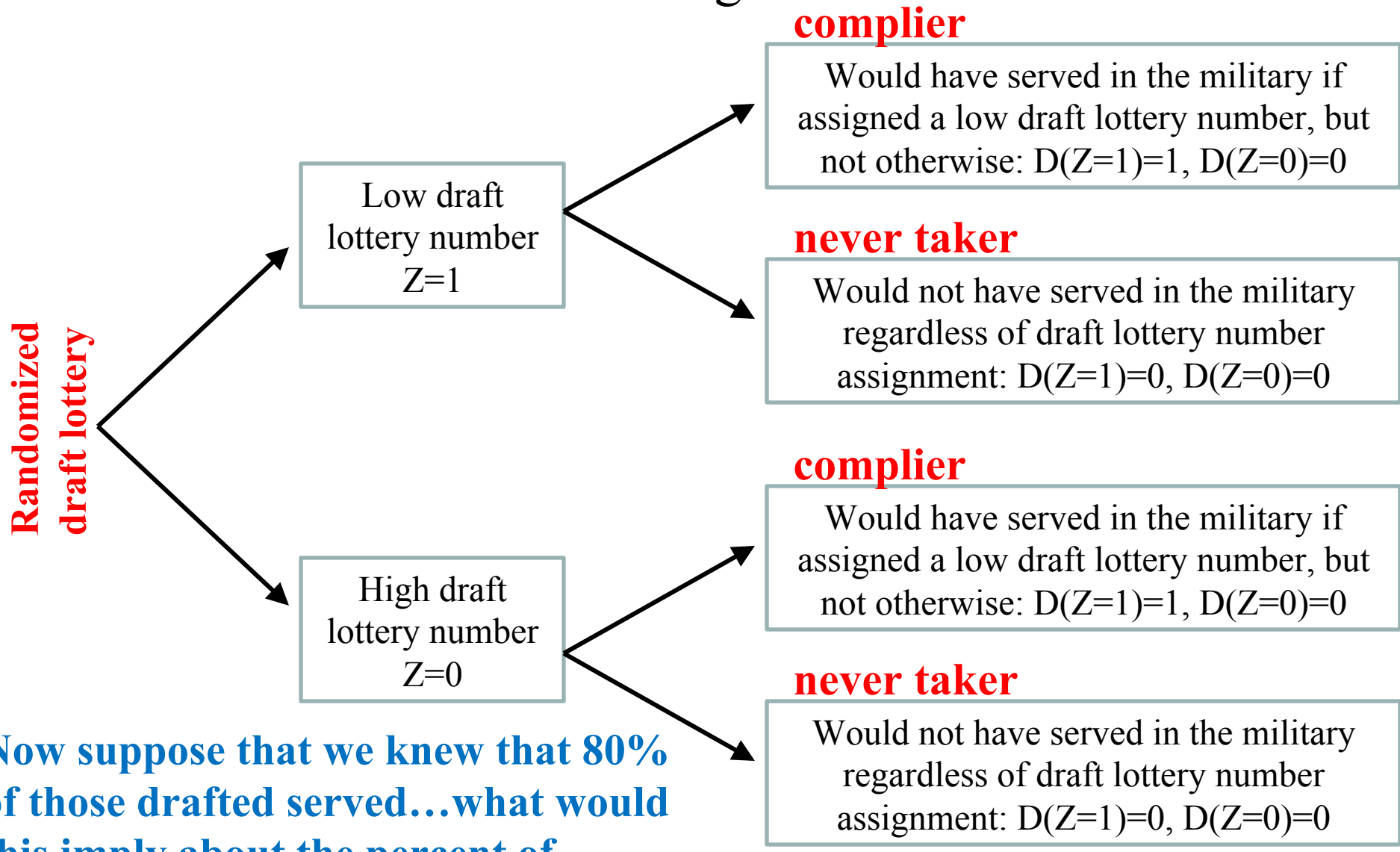
Consider a counterfactual definition of (non)compliance Viewed with "god-vision"



Consider a counterfactual definition of (non)compliance Viewed with "god-vision"

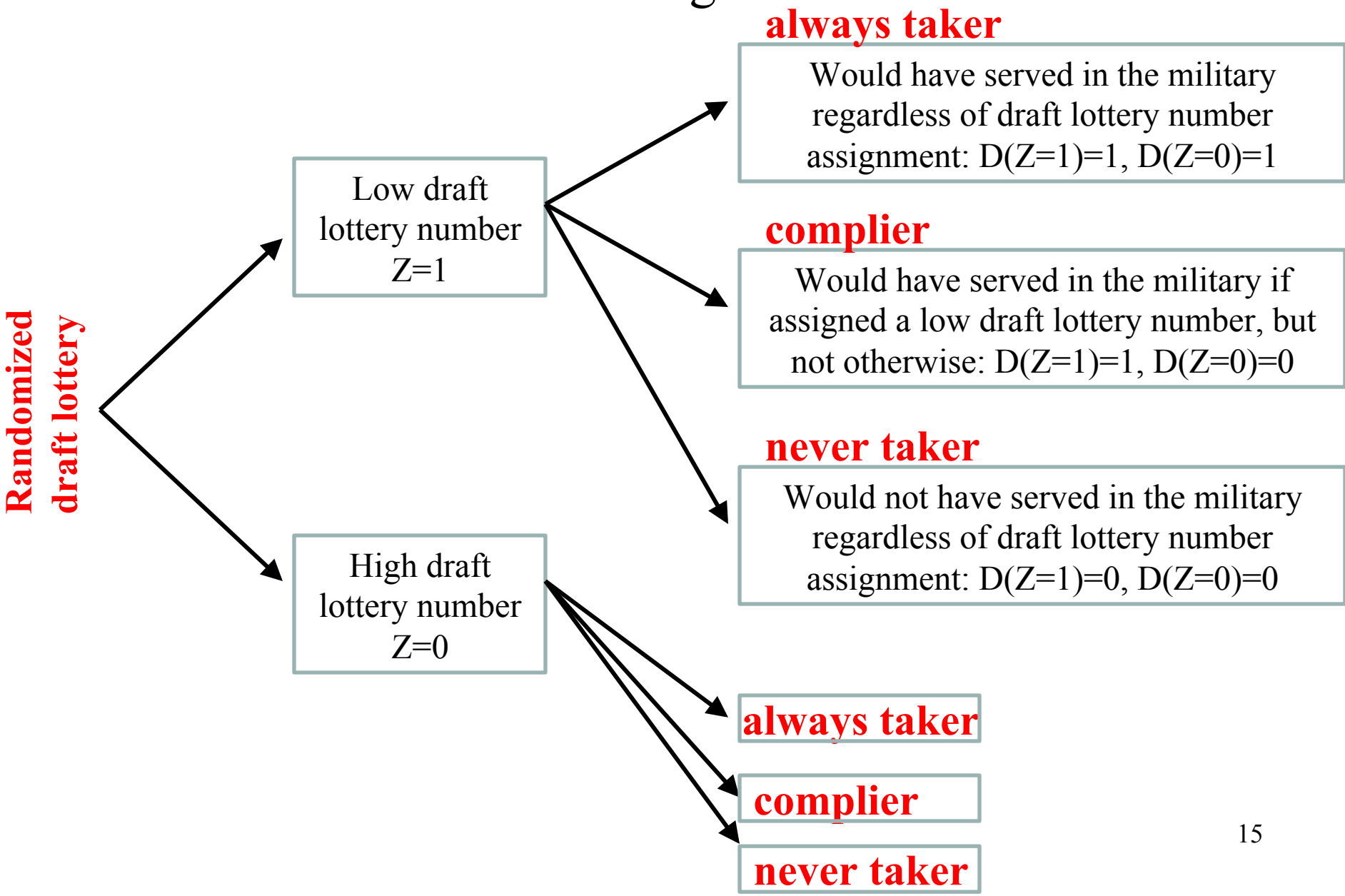


Consider a counterfactual definition of (non)compliance Viewed with "god-vision"

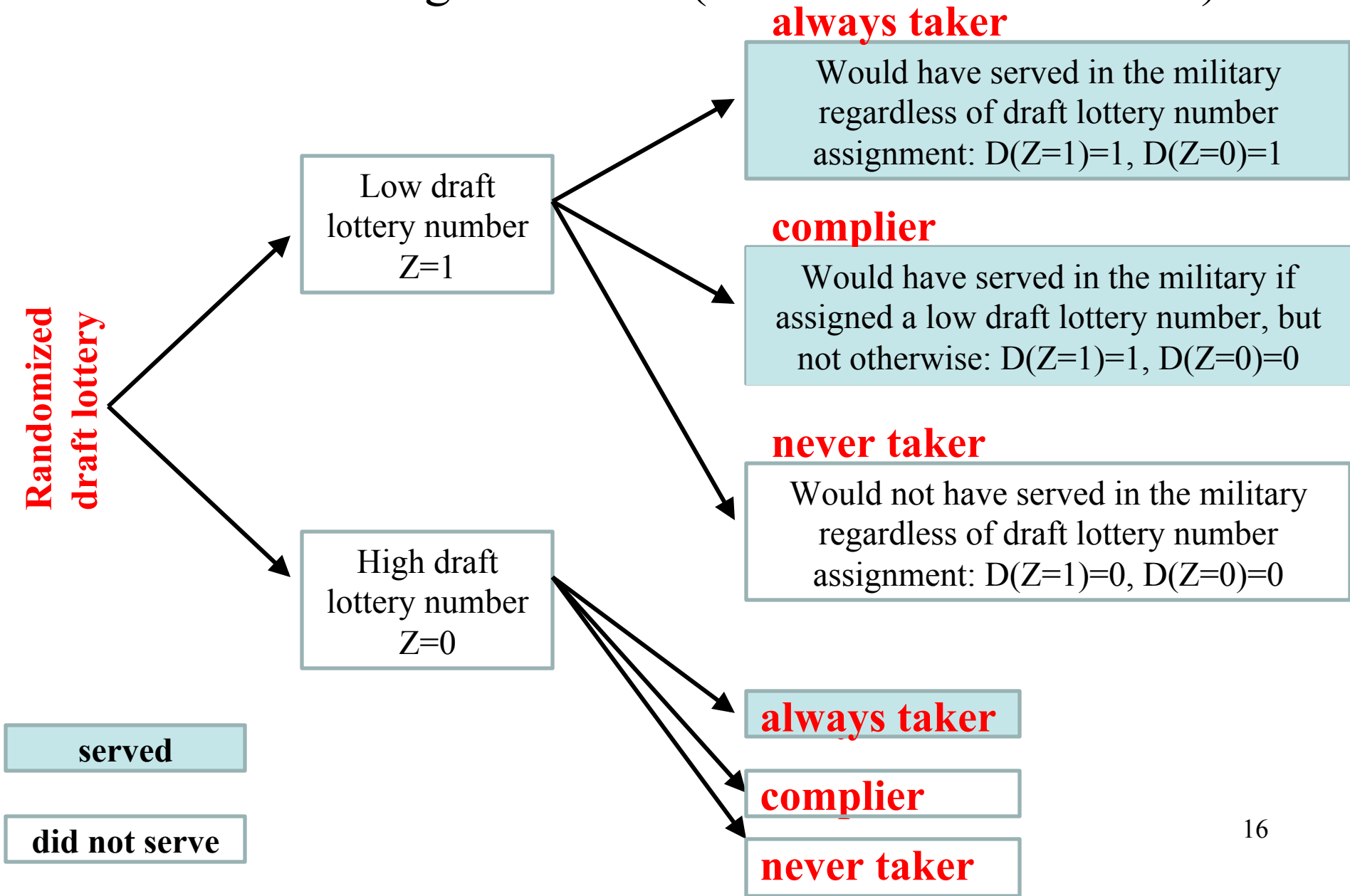


Now suppose that we knew that 80% of those drafted served…what would this imply about the percent of compliers overall?

A slightly more complicated form of (non)compliance Viewed with "god-vision"



A slightly more complicated form of (non)compliance Viewed with "god-vision" (researcher view in blue)



Notation (As before (and in AIR paper))

- Z denotes the randomly assigned *instrument*
- Y(0) and Y(1) stand for potential outcomes
- Y denotes the observed outcome
- D(0) ≡ D(Z=0) and D(1) ≡ D(Z=1) denote potential "program participation" under each assignment of the instrument
- D denotes observed participation behavior (often thought of as the *treatment*)
- C recategorizes D(0) and D(1) into 4 "types" of people
  - compliers: D(0)=0 and D(1)=1
  - never takers: D(0)=0 and D(1)=0
  - always takers: D(0)=1 and D(1)=1
  - defiers: D(0)=1 and D(1)=0

Let's think more about compliance potential outcomes

- D(1) and D(0), treatment participation if instrument is 1 or 0, respectively
- Y(Z,D(Z)) (this notation will simplify later)
- We can define C as "true compliance status":

Compliance	C	D(0)	D(1)
Complier	c	0	1
Never taker	n	0	0
Always taker	a	1	1
Defier	d	1	0

What is the relationship between the ITT and the causal effect for compliers?

What is the CACE: *complier average causal effect*?

It is a function of:

- ITT or intent-to-treat effect is the effect of the randomized variable, E[Y(Z=1) - Y(Z=0)] – here, the effect of having a lottery number that meant you could be drafted
- The probability of being a complier, E[D(Z=1) - D(Z=0)], which in our example is the probability that if you were drafted you went to Vietnam and if you weren't you didn't
- Relationship with CACE?

Form of the CACE estimand

$$IV_{wald} = \frac{ITT}{Pr(complier)} = \frac{E[Y(Z=1) - Y(Z=0)]}{E[D(Z=1) - D(Z=0)]}$$

Deriving the IV estimand

First a reminder about more general potential outcome notation

- Initially we will write out potential outcomes as a function of both the instrument, Z, and the treatment, D
- $Y(0,D(0)) \equiv Y(Z=0,D(Z=0))$, and

  $Y(1,D(1)) \equiv Y(Z=1,D(Z=1))$ and

Understanding the IV estimand as a ratio of two ITT effects

- Let's decompose the ITT effect:

$$E[Y(Z=1) - Y(Z=0)] = E[Y(1,D(1)) - Y(0,D(0))]$$

$$\begin{aligned}
(1)\quad & E[Y(1, D(1)=0) - Y(0, D(0)=0)] * Pr(D(0)=0 \,\&\, D(1)=0) + \\
(2)\quad & E[Y(1, D(1)=1) - Y(0, D(0)=1)] * Pr(D(0)=1 \,\&\, D(1)=1) + \\
(3)\quad & E[Y(1, D(1)=1) - Y(0, D(0)=0)] * Pr(D(0)=0 \,\&\, D(1)=1) + \\
(4)\quad & E[Y(1 ,D(1)=0) - Y(0, D(0)=1)] * Pr(D(0)=1 \,\&\, D(1)=0)
\end{aligned}$$

This is a weighted average of the average effects for each of the 4 types of people defined by combinations of D(0) and D(1)

Understanding the IV estimand as a ratio of two ITT effects

- Let's decompose the ITT effect:

$$E[Y(Z=1) - Y(Z=0)] = E[Y(1,D(1)) - Y(0,D(0))]$$

Effect for Complier — Pr(Complier)

(1) $E[Y(1, D(1)=0) - Y(0, D(0)=0)] * Pr(D(0)=0 \ \& \ D(1)=0) +$

Effect for Always Taker — Pr(Always taker)

(2) $E[Y(1, D(1)=1) - Y(0, D(0)=1)] * Pr(D(0)=1 \ \& \ D(1)=1) +$

Effect for Never Taker — Pr(Never taker)

(3) $E[Y(1, D(1)=1) - Y(0, D(0)=0)] * Pr(D(0)=0 \ \& \ D(1)=1) +$

Effect for Defier — Pr(defier)

(4) $E[Y(1 ,D(1)=0) - Y(0, D(0)=1)] * Pr(D(0)=1 \ \& \ D(1)=0)$

This is a weighted average of the average effects for each of the 4 types of people defined by combinations of D(0) and D(1)

Exclusion restriction

- The exclusion restriction says that if your treatment wouldn't be different even if your instrument assignment was different, then your outcome (or, more generally, the distribution of your outcome) also won't be different
- A colloquial way of phrasing this is that the instrument only affects the outcome through the treatment
- How would that impact the formula for the ITT effect?

Interpreting the IV estimand

$E[Y(Z=1) - Y(Z=0)] = E[Y(1,D(1)) - Y(0,D(0))]$ (always take

(1) $E[Y(1, D(1)=0) - Y(0, D(0)=0)] * Pr(D(0)=0 \& D(1)=0) +$

(2) $E[Y(1, D(1)=1) - Y(0, D(0)=1)] * Pr(D(0)=1 \& D(1)=1) +$

(3) $E[Y(1, D(1)=1) - Y(0, D(0)=0)] * Pr(D(0)=0 \& D(1)=1) +$

(4) $E[Y(1 ,D(1)=0) - Y(0, D(0)=1)] * Pr(D(0)=1 \& D(1)=0)$

- Recall we are interested in the effect of D on Y.

  (1) and (2) aren't going to give us any info about that
- What if we assume that $Y(1, D(1)=0) = Y(0, D(0)=0)$

  and $Y(1, D(1)=1) = Y(0, D(0)=1)$ (No always taker and never taker)

This assumption is called the exclusion restriction

Exclusion restriction

Think about what the exclusion restriction implies for never takers? Does this seem likely to be satisfied?

Think about what the exclusion restriction implies for always takers? Does this seem likely to be satisfied?

Monotonicity assumption

- What if we were to assume that there were no defiers?
- This is called the monotonicity assumption because it implies that $D(1) \geq D(0)$
- Recall the defiers are those who would take the treatment if assigned not to but would not take the treatment if assigned to take it: $D(0) = 1, D(1) = 0$

Interpreting the IV estimand

E[Y(Z=1) - Y(Z=0)] = E[Y(1,D(1)) - Y(0,D(0))]

(1) E[Y(1, D(1)=0) - Y(0, D(0)=0)] * Pr(D(0)=0 & D(1)=0) +

(2) E[Y(1, D(1)=1) - Y(0, D(0)=1)] * Pr(D(0)=1 & D(1)=1) +

(3) E[Y(1, D(1)=1) - Y(0, D(0)=0)] * Pr(D(0)=0 & D(1)=1) +

(4) E[Y(1 ,D(1)=0) - Y(0, D(0)=1)] * Pr(D(0)=1 & D(1)=0)

- What are the implications of assuming that there are no defiers?

Interpreting the IV estimand

$$E[Y(Z=1) - Y(Z=0)] = E[Y(1,D(1)) - Y(0,D(0))]$$

(1) $E[Y(1, D(1)=0) - Y(0, D(0)=0)] * Pr(D(0)=0 \& D(1)=0) +$

(2) $E[Y(1, D(1)=1) - Y(0, D(0)=1)] * Pr(D(0)=1 \& D(1)=1) +$

(3) $E[Y(1, D(1)=1) - Y(0, D(0)=0)] * Pr(D(0)=0 \& D(1)=1) +$

(4) $E[Y(1 ,D(1)=0) - Y(0, D(0)=1)] * Pr(D(0)=1 \& D(1)=0)$

- Line 4 pertains to "defiers" (recall, those for whom D(0)=1,D(1)=0)
- If we assume that no defiers exist that is the same as saying that Pr(D(0)=1 & D(1)=0)=0

This assumption is referred to as monotonicity

Monotonicity in our example

How would we interpret monotonicity in the context of our example?

Is it likely to be satisfied?

Interpreting the IV estimand

$$E[Y(Z=1) - Y(Z=0)] = E[Y(1,D(1)) - Y(0,D(0))]$$

(1) $E[Y(1, D(1)=0) - Y(0, D(0)=0)] * Pr(D(0)=0 \ \& \ D(1)=0) +$

(2) $E[Y(1, D(1)=1) - Y(0, D(0)=1)] * Pr(D(0)=1 \ \& \ D(1)=1) +$

(3) $E[Y(1, D(1)=1) - Y(0, D(0)=0)] * Pr(D(0)=0 \ \& \ D(1)=1) +$

(4) $E[Y(1 ,D(1)=0) - Y(0, D(0)=1)] * Pr(D(0)=1 \ \& \ D(1)=0)$

- $E[Y(Z=1) - Y(Z=0)] =$
  $E[Y(1, D(1)=1) - Y(0, D(0)=0)] * Pr(D(0)=0 \ \& \ D(1)=1)$

$$E[Y(1, D(1)=1) - Y(0, D(0)=0)] = \frac{E[Y(1) - Y(0)]}{Pr(D(0)=0 \ \& \ D(1)=1)}$$

Denominator of the IV estimand

- To better understand the next assumption let's consider the following equivalency:

  $\Pr(D(1)=1 \;\&\; D(0)=0) = \Pr(C=\text{complier}) = E[D(Z=1) - D(Z=0)]$

- Why does this make sense?

Denominator of the IV estimand

- To better understand the next assumption let's consider the following equivalency:

  $\Pr(D(1)=1 \ \& \ D(0)=0) = \Pr(C=\text{complier}) = E[D(Z=1) - D(Z=0)]$
- Why does this make sense?
- First, we know that, *if monotonicity holds*,

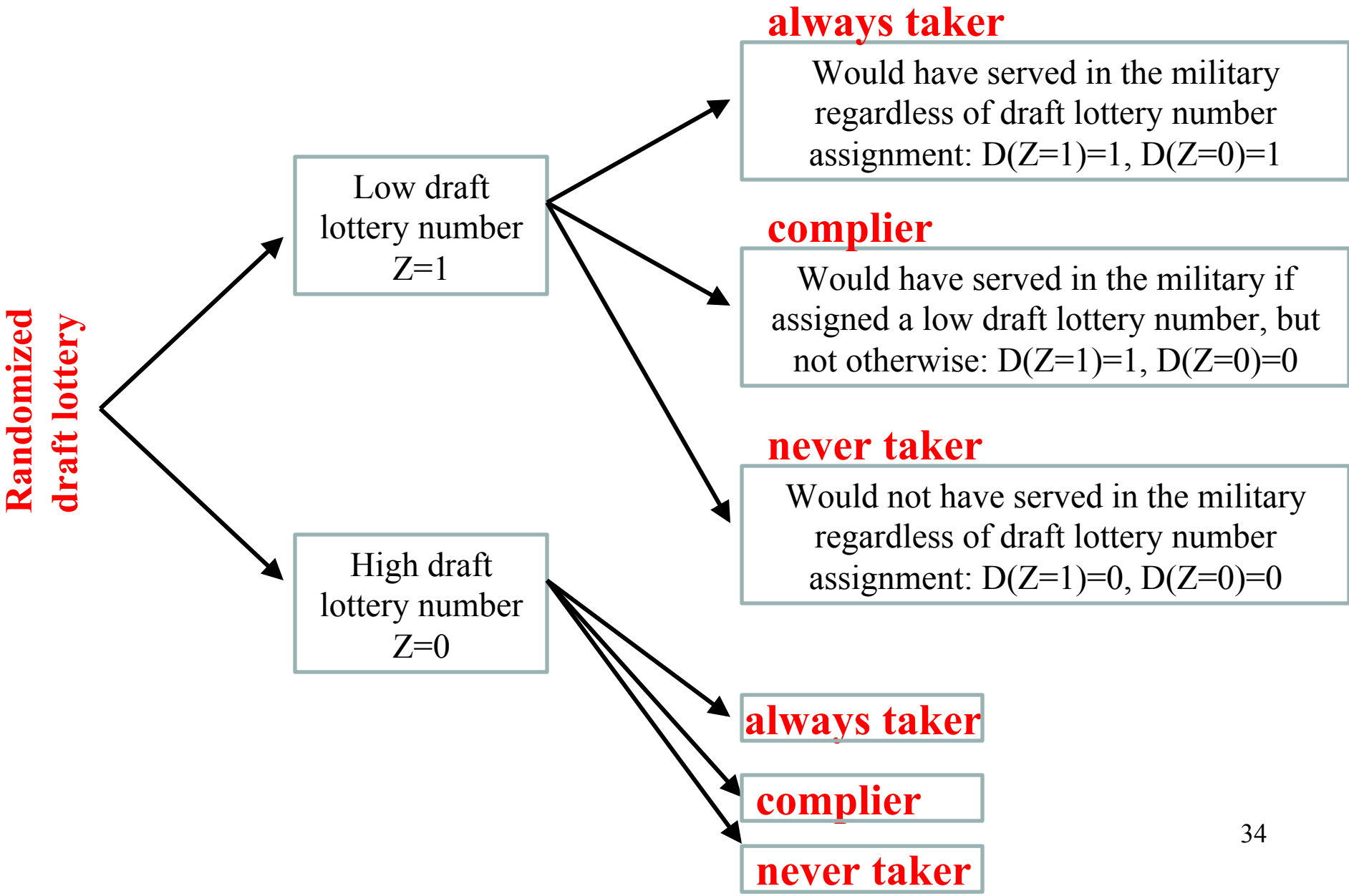
  $E[D(Z=1)] = \Pr(C = \text{complier} \cup C=\text{always taker})$

  $= \Pr(C = \text{complier}) + \Pr(C=\text{always taker})$

  and

  $E[D(Z=0)] = \Pr(C=\text{always taker})$
- Therefore, $E[D(Z=1) - D(Z=0)] = \Pr(C = \text{complier})$

Visually…..


Randomized draft lottery
Low draft lottery number Z=1
High draft lottery number Z=0
always taker
Would have served in the military regardless of draft lottery number assignment: D(Z=1)=1, D(Z=0)=1
complier
Would have served in the military if assigned a low draft lottery number, but not otherwise: D(Z=1)=1, D(Z=0)=0
never taker
Would not have served in the military regardless of draft lottery number assignment: D(Z=1)=0, D(Z=0)=0
always taker
complier
never taker

Interpreting the IV estimand

$$E[Y(Z=1) - Y(Z=0)] = E[Y(1,D(1)) - Y(0,D(0))]$$

(1) $E[Y(1, D(1)=0) - Y(0, D(0)=0)] * Pr(D(0)=0 \ \& \ D(1)=0) +$

(2) $E[Y(1, D(1)=1) - Y(0, D(0)=1)] * Pr(D(0)=1 \ \& \ D(1)=1) +$

(3) $E[Y(1, D(1)=1) - Y(0, D(0)=0)] * Pr(D(0)=0 \ \& \ D(1)=1) +$

(4) $E[Y(1 ,D(1)=0) - Y(0, D(0)=1)] * Pr(D(0)=1 \ \& \ D(1)=0)$

- $E[Y(Z=1) - Y(Z=0)] =$
  $E[Y(1, D(1)=1) - Y(0, D(0)=0)] * Pr(D(0)=0 \ \& \ D(1)=1)$

$$E[Y(1, D(1)=1) - Y(0, D(0)=0)] = \frac{E[Y(1) - Y(0)]}{E[D(Z=1) - D(Z=0)]}$$

Non-zero correlation between instrument and treatment

- An instrument isn't useful if it doesn't actually predict the treatment
- Why is that important for our formula?

Interpreting the IV estimand

$$E[Y(Z=1) - Y(Z=0)] = E[Y(1,D(1)) - Y(0,D(0))]$$

(1) $E[Y(1, D(1)=0) - Y(0, D(0)=0)] * Pr(D(0)=0 \;\&\; D(1)=0) +$

(2) $E[Y(1, D(1)=1) - Y(0, D(0)=1)] * Pr(D(0)=1 \;\&\; D(1)=1) +$

(3) $E[Y(1, D(1)=1) - Y(0, D(0)=0)] * Pr(D(0)=0 \;\&\; D(1)=1) +$

(4) $E[Y(1 ,D(1)=0) - Y(0, D(0)=1)] * Pr(D(0)=1 \;\&\; D(1)=0)$

- $E[Y(Z=1) - Y(Z=0)] =$
  $E[Y(1, D(1)=1) - Y(0, D(0)=0)] * Pr(D(0)=0 \;\&\; D(1)=1)$

$$E[Y(1, D(1)=1) - Y(0, D(0)=0)] = \frac{E[Y(1) - Y(0)]}{E[D(Z=1) - D(Z=0)]}$$

Non-zero correlation assumption: this can't be equal to zero

Interpreting the IV estimand

Rearranging terms we get:

$$E[Y(1, D(1)=1) - Y(0, D(0)=0)] = \frac{E[Y(Z=1) - Y(Z=0)]}{E[D(Z=1) - D(Z=0)]}$$

But there is one more simplification possible.

Another consequence of exclusion is that we can write

$E[Y(1, D(1)=1)]$ as $E[Y(D=1)]$ and

$E[Y(0, D(0)=1)]$ as $E[Y(D=0)]$, leaving:

$$E[Y(D=1) - Y(D=0) \mid D(0)=0, D(1)=1] = \frac{E[Y(Z=1) - Y(Z=0)]}{E[D(Z=1) - D(Z=0)]}$$

Interpreting the IV estimand

$$E[Y(D=1) - Y(D=0) \mid D(0)=0, D(1)=1] = \frac{E[Y(Z=1) - Y(Z=0)]}{E[D(Z=1) - D(Z=0)]}$$

In words

$$\text{CACE} = \frac{\text{ITT of Z on Y}}{\text{ITT of Z on D}}$$

Each of these effects on the right is easy to estimate if we can make what assumption???

Ignorability!

- Let's assume Z, our instrument is randomly assigned: $Y(1,D(1)), Y(0,D(0)) \perp Z$ (ignorability)
- Then we know we can unbiasedly estimate both parts of the IV estimand

$$E[Y(Z=1) - Y(Z=0)]$$

$$E[D(Z=1) - D(Z=0)]$$

Ignorability for our example

Does ignorability seem likely to be satisfied in our example?

https://www.youtube.com/watch?v=25cmARm05qs

https://www.nytimes.com/1970/01/04/archives/statisticians-charge-draft-lottery-was-not-random.html

http://web.mit.edu/berinsky/www/files/draftlottery.pdf

Review of Assumptions

- Ignorability of the instrument (i.e. it is randomized or conditionally randomized)
- Exclusion
- Monotonicity (no defiers)
- Non-zero correlation between instrument and treatment
- SUTVA

Comparison of assumptions under IV and pscore approaches

Instrumental Variables

- Ignorability of the instrument (i.e. it is randomized or conditionally randomized)
- Non-zero correlation between instrument and treatment
- Monotonicity (no defiers)
- Exclusion
- SUTVA

Propensity scores

- Ignorability of the treatment assignment
- Balance
- Overlap
- SUTVA

IV in a regression/econometric framework

For binary dependent variable

$Y_i = \beta_0 + \tau D_i + \varepsilon_i$

$D_i = \alpha_0 + \alpha_1 Z_i + \nu_i$

The assumptions are stated as:

1) Z_i is uncorrelated with both ε_i and ν_i and

2) $\text{cov}(Z_i, D_i) \neq 0$

Note that the exclusion restriction is represented in this formulation by the combination of the requirement $E[Z_i \, \varepsilon_i] = 0$ and the fact that Z_i is omitted from the first stage equation

Estimation

Estimation

The two most common estimation strategies when the instrument is binary are

- The Wald estimator
- Two stage least squares (TSLS)

Wald estimator

- The Wald estimator is a ratio estimator that separately estimates the numerator and denominator in the IV estimand

$$\text{CACE} = \frac{E[Y(1)-Y(0)]}{E[D(1)-D(0)]} = \frac{ITT_Y}{ITT_Z}$$

- Since Z is randomized the numerator can be estimated by estimating E[Y|Z=1] – E[Y|Z=0] using $\overline{Y}_1 - \overline{Y}_0$
- The same reasoning applies to the denominator which can be estimated as the difference in the percent observed to "comply" in each group $\overline{D}_1 - \overline{D}_0$
- Equivalently, the Wald estimate can be extended simply by taking the ratio of the coefficient for Z in E[Y|Z,X] and the coefficient for Z in E[D|Z,X] (see next slide)

Estimation for Wald estimator

- Thus if

  $E[Y \mid Z] = \mu_0 + \mu_1 Z$

  $E[D \mid Z] = \alpha_0 + \alpha_1 Z$
- then the estimate of CACE can be found by taking the ratio of the causal effect of Z on Y with the causal effect of Z on D,

  $$\tau_{CACE} = \mu_1 / \alpha_1$$
- Can this be extended to incorporate covariates?
- Absolutely! We can also estimate those parameters in the following regression specifications

  $$E[Y \mid Z] = \mu_0 + \mu_1 Z + \omega_1 X_1 + \ldots + \omega_k X_k$$

  $$E[D \mid Z] = \alpha_0 + \alpha_1 Z + \varphi_1 X_1 + \ldots + \varphi_k X_k$$
- Why might we want to condition on covariates as well?

Two-stage least squares

- A more versatile way to calculate IV estimates is via two-stage least squares (TSLS, 2SLS)
- The basic idea is to
  - Regress *D* on *Z* (first stage): $E[D \mid Z] = \alpha_0 + \alpha_1 Z$
  - Use this (first-stage) model to make predictions for $\hat{D}$,
  - Then regress *Y* on $\hat{D}$, (second stage): $E[Y \mid D] = \beta_0 + \tau_{CACE} \hat{D}$
  - The coefficient on $\hat{D}$ is τ_{CACE}, the IV estimand
  - The standard error on that coefficient would be inappropriate however because it doesn't account for the fact that the equations are correlated (*ivreg* will fix for you)
- Again we can also condition on covariates in these models
- These models can also be extended for use with a continuous instrument or treatment variable

Extension of assumptions to include covariates

- Assumption 2: Random Assignment
  $Y(0,D(0)), Y(1,D(1)) \perp Z \mid X$

- Assumption 3: Exclusion Restriction
  $p(Y(1) \mid C = n, X) = p(Y(0) \mid C = n, X)$
  and
  $p(Y(1) \mid C = a, X) = p(Y(0) \mid C = a, X)$

Standard errors

- The standard error for the IV estimate, 1, can be found by taking the estimate of the s.e. for $\hat{\beta}_1$ and multiplying it by the following ratio: $\frac{\text{MSE}_2}{\text{MSE}_1}$
- Here MSE_1 is just the mean squared error from the second regression above and MSE_2 is equal to the mean squared residuals formed by $Y_i - \hat{Y}_i$ where $\hat{Y}_i$ is the fit obtained by multiplying the coefficients from the 2nd stage model by the associated X variables and D (note D not D)

$$Y_i = \beta_0 + \tau' D_i + \gamma_1' X_{i1,} + \ldots + \gamma_k' X_{ik} + \varepsilon_i$$

Intuition for magnitude of IV versus ITT standard errors

- We expect the IV standard errors to be larger than the ITT standard errors
- Why?

Intuition for magnitude of IV versus ITT standard errors

- We expect the IV standard errors to be larger than the ITT standard errors
- Why? Because our "effective" sample size is getting smaller. For instance, in the simple a binary instrument, binary treatment scenario you can think of the LATE estimate as being based only on compliers who are just a subset of the entire dataset. Therefore the standard errors are inflated to reflect this smaller sample size.

Instrumental variables and Stata

An applied example using Stata

Flu shot experiment

- Researchers were interested in the effect of flu shots on subsequent health measures
- However, it is logistically/ethically infeasible to directly randomize flu shots to individuals
- Instead, the researchers randomized doctors to either encourage their patients to receive flu shots or not. This is called a randomized encouragement design.
- This is similar to examples of randomized experiments with noncompliance

Flu shot example: data

- N=2893
- Y, Outcome: experienced a flu-related hospitalization within 3 years after randomization
- Z, Instrument: encouragement to get a flu shot
- D, Treatment: getting a flu shot
- X, Pre-treatment variables
  - COPD: Chronic Obstructive Pulmonary Disease
  - Age

  (original study by McDonald et al. 1992 MD Computing)

Assumptions

- Ignorability?

Assumptions

- Ignorability? The instrument, encouragement, was randomly assigned
- Exclusion?

Assumptions

- Ignorability? The instrument, encouragement, was randomly assigned
- Exclusion? Any chance that being encouraged would affect health status even if it didn't change behavior regarding flu shot receipt?

Assumptions

- Ignorability? The instrument, encouragement, was randomly assigned
- Exclusion? Any chance that being encouraged would affect health status even if it didn't change behavior regarding flu shot receipt?
- Monotonicity?

Assumptions

- Ignorability? The instrument, encouragement, was randomly assigned
- Exclusion? Any chance that being encouraged would affect health status even if it didn't change behavior regarding flu shot receipt?
- Monotonicity? Are there people people who would have gotten a flu shot if not encouraged but would not have gotten if they were encouraged?

Assumptions

- Ignorability? The instrument, encouragement, was randomly assigned
- Exclusion? Any chance that being encouraged would affect health status even if it didn't change behavior regarding flu shot receipt?
- Monotonicity? Are there people people who would have gotten a flu shot if not encouraged but would not have gotten if they were encouraged?
- Non-zero correlation between instrument and treatment?

Assumptions

- Ignorability? The instrument, encouragement, was randomly assigned
- Exclusion? Any chance that being encouraged would affect health status even if it didn't change behavior regarding flu shot receipt?
- Monotonicity? Are there people people who would have gotten a flu shot if not encouraged but would not have gotten if they were encouraged?
- Non-zero correlation between instrument and treatment?

  The instrument has an effect on flu shot receipt

Assumptions

- Ignorability? The instrument, encouragement, was randomly assigned
- Exclusion? Any chance that being encouraged would affect health status even if it didn't change behavior regarding flu shot receipt?
- Monotonicity? Are there people people who would have gotten a flu shot if not encouraged but would not have gotten if they were encouraged?
- Non-zero correlation between instrument and treatment?

  The instrument has an effect on flu shot receipt
- SUTVA?

Assumptions

- Ignorability? The instrument, encouragement, was randomly assigned
- Exclusion? Any chance that being encouraged would affect health status even if it didn't change behavior regarding flu shot receipt?
- Monotonicity? Are there people people who would have gotten a flu shot if not encouraged but would not have gotten if they were encouraged?
- Non-zero correlation between instrument and treatment?

  The instrument has an effect on flu shot receipt
- SUTVA? Plausible to believe that one person's treatment assignment would not affect another's outcome (families make decisions together?)

Analysis in Stata

As treated analysis

- As treated analysis compares those who received a flu shot with those who did not

```
. regress wcxho79 fluy2

      Source |       SS       df       MS              Number of obs =    2901
-------------+------------------------------           F(  1,  2899) =    0.02
       Model |  .001468216     1  .001468216           Prob > F      =  0.8910
    Residual |  226.797567  2899  .078233034           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared = -0.0003
       Total |  226.799035  2900  .078206564           Root MSE      =   .2797

------------------------------------------------------------------------------
     wcxho79 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       fluy2 |  -.0016439   .0119997    -0.14   0.891    -.0251727    .0218849
       _cons |    .085898   .0059947    14.33   0.000     .0741438    .0976523
------------------------------------------------------------------------------
```

Per protocol analysis

Per protocol analysis compares individuals across randomized groups who actually "did as they were told" – in this case the encouraged who took their flu shot -- versus the not encouraged -- who did not take their flu shot

```
. regress wcxho79 grp if (fluy2==1 & grp==1) | (fluy2==0 &grp==0)

      Source |       SS       df       MS              Number of obs =    1602
-------------+------------------------------           F(  1,  1600) =    1.79
       Model |   .13561045     1   .13561045           Prob > F      =  0.1807
    Residual |  120.987985  1600  .075617491           R-squared     =  0.0011
-------------+------------------------------           Adj R-squared =  0.0005
       Total |  121.123596  1601  .075654963           Root MSE      =  .27499

------------------------------------------------------------------------------
     wcxho79 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         grp |  -.0203759   .0152153    -1.34   0.181      -.05022    .0094682
       _cons |   .0882096   .0081266    10.85   0.000     .0722697    .1041495
------------------------------------------------------------------------------
```

Flu shot example: causal estimands

- The ITT effect of encouragement on receiving the flu shot

  E[D(1)-D(0)]
- The ITT effect of encouragement on flu-related hospitalizations

  E[Y(1)-Y(0)]
- The causal effect of receiving a flu shot on flu-related hospitalizations for compliers

  E[Y(1)-Y(0)]/ E[D(1)-D(0)]

Stata: ITT effect on outcome

```
. regress  wcxho79 grp

      Source |       SS       df       MS              Number of obs =    2901
-------------+------------------------------           F(  1,  2899) =    1.87
       Model |  .146135937     1  .146135937            Prob > F      =  0.1717
    Residual |  226.652899  2899  .078183132            R-squared     =  0.0006
-------------+------------------------------           Adj R-squared =  0.0003
       Total |  226.799035  2900  .078206564            Root MSE      =  .27961

------------------------------------------------------------------------------
     wcxho79 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         grp |     -.0142   .0103864    -1.37   0.172    -.0345655    .0061655
       _cons |   .0927762   .0074411    12.47   0.000     .0781858    .1073667
------------------------------------------------------------------------------
```

Stata: ITT effect on flu shot receipt

```
. regress  fluy2 grp

      Source |       SS       df       MS              Number of obs =    2901
-------------+------------------------------           F(  1,  2899) =   54.70
       Model |  10.0611709     1  10.0611709            Prob > F      =  0.0000
    Residual |   533.25079  2899  .183943012           R-squared     =  0.0185
-------------+------------------------------           Adj R-squared =  0.0182
       Total |  543.311961  2900  .187348952           Root MSE      =  .42889

------------------------------------------------------------------------------
       fluy2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         grp |   .1178239   .0159313     7.40   0.000     .0865861    .1490617
       _cons |   .1890935   .0114136    16.57   0.000     .1667138    .2114732
------------------------------------------------------------------------------
```

Simple IV estimate

- Is the just the ratio of the two ITT effects:

```
. display -.0142/.1178
-.12054329
```

- The intuition using a simple example is the following. Suppose we have the mean of 100 numbers but 88 percent of them (only 12 percent are compliers with a non-zero treatment effect). If we're really interested in the mean for just the non-zero numbers then the mean for all 100 of them can be seen as just dividing the sum by an incorrect divisor (100 instead of 12), so to get the correct divisor you just divide by this ratio (12/100).

$$\hat{\tau}_{ITT} = \frac{0+0+\ldots+0+a+\ldots+a}{100}$$

$$\hat{\tau}_{IV} = \frac{a+\ldots+a}{12} = \frac{0+0+\ldots+0+a+\ldots+a}{100} \times \frac{100}{12} = \frac{\hat{\tau}_{ITT}}{12/100}$$

$\tau_{CACE} = \tau_{CACE}$ in this slide

Simple IV estimate

. ivreg wcxho79 (fluy2=grp)

```
Instrumental variables (2SLS) regression
      Source |       SS       df       MS              Number of obs =    2901
-------------+------------------------------           F(  1,  2899) =    1.81
       Model | -7.67618731     1 -7.67618731           Prob > F      =  0.1790
    Residual |  234.475222  2899  .080881415           R-squared     =       .
-------------+------------------------------           Adj R-squared =       .
       Total |  226.799035  2900  .078206564           Root MSE      =   .2844

------------------------------------------------------------------------------
     wcxho79 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       fluy2 |  -.1205186   .0896603    -1.34   0.179     -.296323    .0552857
       _cons |   .1155655    .022991     5.03   0.000     .0704852    .1606458
------------------------------------------------------------------------------
Instrumented:  fluy2
Instruments:   grp
------------------------------------------------------------------------------
```

Interpretation

The effect of receiving a flu shot (versus not receiving it) for those who will actually take it when encouraged and won't if not encouraged (compliers) is a 12 percentage point reduction in the rate of subsequent flu-related hospitalizations.

With additional covariance adjustment…

```
. ivreg wcxho79 copd age (fluy2=grp copd age)

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =    2893
-------------+------------------------------           F(  3,  2889) =    3.46
       Model | -5.64506232     3 -1.88168744           Prob > F      =  0.0158
    Residual |  229.896704  2889  .079576568           R-squared     =       .
-------------+------------------------------           Adj R-squared =       .
       Total |  224.251642  2892  .077542062           Root MSE      =  .28209

------------------------------------------------------------------------------
     wcxho79 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       fluy2 |  -.1108732   .0899298    -1.23   0.218    -.2872062    .0654598
        copd |   .0399087   .0141511     2.82   0.005     .0121613     .067656
         age |  -.0000801   .0004982    -0.16   0.872     -.001057    .0008967
       _cons |   .1062615   .0286472     3.71   0.000     .0500906    .1624324
------------------------------------------------------------------------------
Instrumented:  fluy2
Instruments:   copd age grp
------------------------------------------------------------------------------
```

Postscript regarding assumptions

- More complicated analyses (see Hirano et al. 2000) reveal that the exclusion restriction does not appear to be satisfied
- The effect of encouragement for always takers and never takers appears to be about the same as the effect on compliers
- The original article which only calculated the ITT attributed the observed difference in means (though not statistically significant) to the flu-shot receipt. This is an important example of a case when both the ITT effect and the simple IV estimate can be misleading!

Identifiability

Identifiability

- Broadly speaking, identifiability refers to whether or not the data contain sufficient information for unique estimation of a given parameter or set of parameters in a particular model.
- For example, in our IV model, the causal parameter is not identified without making the assumption of the exclusion restriction.

Multiple Instruments, Multiple Treatments, and Identification

Multiple Treatments

1. Sometimes we expect that the causal mechanism works through more than one "treatment". For example in the IHDP randomized experiment discussed last week, children in the treatment group received two primary services: access to very high-quality child care and access to home visits from a trained professional.
2. If we exclude one of these treatments from the model we run the risk that the exclusion restriction is not satisfied.
3. So can we include both?

For simplicity we will refer to τ_{CACE} simply as τ in the following slides..

Identifying a simple model using exclusion restriction

- Our model, more generally written (before imposing the exclusion restriction), and ignoring the error, is

$$E[Y \mid D, Z] = \beta_0 + \tau D + \beta_2 Z$$

$$E[D \mid Z] = \alpha_0 + \alpha_1 Z$$

  but 1 and 2 can't be estimated directly because of selection bias.

- However, we can estimate the causal effect of D on Y , β_1, by substituting the equation for D into the equation for Y

$$\begin{aligned}[Y \mid D, Z] &= \beta_0 + \tau D + \beta_2 Z \\ &= \beta_0 + \tau(\alpha_0 + \alpha_1 Z) + \beta_2 Z \\ &= \beta_0 + \tau\alpha_0 + (\tau\alpha_1 + \beta_2)Z \\ &= \mu_0 + \mu_1 Z\end{aligned}$$

Identifying a simple model using exclusion restriction (cont)

- To find β_1 we can solve the following equation

  $\mu_1 = \tau\alpha_1 + \beta_2$, which we can rearrange to get

  $\tau = (\mu_1 - \beta_2)/\alpha_1$
- But this is one equation with two unknowns, τ and β_2 (we can get an unbiased estimate of α_1. The exclusion restriction, however, lets us set β_2 to 0.
- Now we just have $\tau = \mu_1/\alpha_1$.

Identification with more than one treatment variable

- Suppose, however, that we have more than one treatment variable (e.g. in the IHDP). Assuming exclusion from the beginning:

$$E[Y \mid D_1, D_2] = \beta_0 + \tau_1 D_1 + \tau_2 D_2$$
$$E[D_1 \mid Z] = \alpha_0 + \alpha_1 Z$$
$$E[D_2 \mid Z] = \gamma_0 + \gamma_1 Z$$

and the goal is to estimate both β_1 and β_2. These can't be estimated directly from the 1st equation due to selection bias. However, if we substitute the equations for D1 and D2 into the equation for Y

$$\begin{aligned} E[Y \mid D_1, D_1] &= \beta_0 + \tau_1 D_1 + \tau_2 D_2 \\ &= \beta_0 + \tau_1(\alpha_0 + \alpha_1 Z) + \tau_2(\gamma_0 + \gamma_1 Z) \\ &= \beta_0 + \tau_1\alpha_0 + \tau_2\gamma_0 + (\tau_1\alpha_1 + \tau_2\gamma_1)Z \\ &= \mu_0 + \mu_1 Z \end{aligned}$$

Identification with more than one treatment variable (cont)

- So now to find τ_1 and τ_2 we need to solve the following equation

$$\mu_1 = \tau_1\alpha_1 + \tau_2\gamma_1.$$

- Now we can get unbiased estimates of μ_1, α_1, and γ_1 but that leaves us with just one equation and two unknowns. Therefore the parameters are not identifiable!

Identifying with more than one treatment variable and more than one instrument

- Suppose, however, that we have more than one treatment variable and more than one instrument. Assuming exclusion from the beginning:

$$E[Y \mid D_1, D_2] = \beta_0 + \tau_1 D_1 + \tau_2 D_2$$
$$E[D_1 \mid Z_1, Z_2] = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2$$
$$E[D_2 \mid Z_1, Z_2] = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2$$

  and the goal is to estimate both β_1 and β_2. These can't be estimated directly from the 1st equation due to selection bias.
- However, if we substitute the equations for D1 and D2 into the equation for Y

$$\begin{aligned} E[Y \mid D_1, D_2] &= \beta_0 + \tau_1 D_1 + \tau_2 D_2 \\ &= \beta_0 + \tau_1(\alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2) + \tau_2(\gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2) \\ &= \beta_0 + \tau_1\alpha_0 + \tau_2\gamma_0 + (\tau_1\alpha_1 + \tau_2\gamma_1)Z_1 + (\tau_1\alpha_2 + \tau_2\gamma_2)Z_2 \\ &= \mu_0 + \mu_1 Z_1 + \mu_2 Z_2 \end{aligned}$$

Identifying with more than one treatment variable and more than one instrument (cont)

- So now to find τ_1 and τ_2 we need to solve the following

$$\mu_1 = \tau_1 \alpha_1 + \tau_2 \gamma_1.$$

$$\mu_2 = \tau_1 \alpha_2 + \tau_2 \gamma_2.$$

- Since we can get unbiased estimates of μ_1, α_1, α_2, γ_1, and γ_2 that leaves us with two equations and two unknowns.
- Therefore the parameters are identifiable!
- The general rule then is that you always need to have at least as many instruments as treatment variables.
- "just-identified" means that # instruments = # treatment vars
- "over-identified" means that # instruments > # treatment vars

(over-identified models can have their own problems as we shall see in the slides ahead about weak instruments)

What can go wrong??

Violations of exclusion

- What happens if exclusion is violated?
- Then the ITT effect equals CACE plus a bias term that can be expressed as

$$\mathrm{E}[Y(1,d) - Y(0,d) \mid C = nc] \frac{\Pr(C = nc)}{\Pr(C \neq nc)}$$

(where d is the actual participation… 0 for never takers and 1 for always takers)

- Therefore the amount of bias depends on both the effect of the instrument on the noncomplier and the odds that someone is a noncomplier
- Thus the weaker the instrument the more trouble you can get into!

Similar argument in econometric speak

$$\text{plim}\ \hat{\tau}_{OLS} = \tau + \frac{\sigma_{V_0}}{\sigma_W} \cdot Corr(W, V_0)$$

$$\text{plim}\ \hat{\tau}_{IV} = \tau + \frac{\sigma_{V_0}}{\sigma_W} \cdot \frac{Corr(Z, V_0)}{Corr(Z, W)}$$

Violations of monotonicity

- If the monotonicity assumption is violated the IV estimand is equal to CACE plus a bias term that can be expressed as

$$-\lambda\{\mathrm{E}[Y(1)-Y(0)\,|\,C=d]-\mathrm{E}[Y(1)-Y(0)\,|\,C=c]\}$$

where

$$\lambda=\frac{\Pr(C=d)}{\Pr(C=c)-\Pr(C=d)}$$

- This bias is bigger when the proportion of defiers is big relative to the proportion of compliers (and is 0 when this proportion is 0).
- This bias is also bigger when the average causal effects of D on Y for the compliers and the defiers. If these effects are equal then violation of monotonicity yields no bias.

Violations of ignorability

1. Violations of ignorability will create a biased estimate of the ITT effect (the numerator of our IV estimand). The weaker our instrument, the more sensitive our results will be to such bias because it

$$\frac{E[Y(Z=1) - Y(Z=0)]}{E[D(Z=1) - D(Z=0)]}$$

   will be amplified by a small denominator

2. If ignorability is violated the denominator will also be wrong

Weak instruments: econometric perspective

1. Even if all of these IV assumptions hold we can get biased estimates of the treatment effect and standard errors with standard methods when we have "weak instruments" (that is when there isn't a sufficiently strong relationship between the instruments and the treatment, see Staiger and Stock 1997).
2. Diagnosis: Perform an F-test in the first stage regression (regression of treatment on instruments) for all the instruments being statistically significantly different from 0 (Bound and Jaeger paper gives suggested F-statistics for various numbers of instruments, Stock and Yogo (2005) extend this)
3. In just-identified models (i.e., # instruments = # treatments), estimates are approximately unbiased even with a weak instrument. These will always be a safer bet.

Weak Instruments, more thoughts

- When the instrument is weak we will only be able to make inferences about a small proportion of the sample who may or may not be representative of the population we care about
- When the instrument is weak it may make sense to focus on the ITT estimate instead since (if ignorability holds) this is unbiased and proportional to the treatment effect for compliers (assuming exclusion).
- As we have seen, when the instrument is weak our analyses are more vulnerable to deviations from the other modeling assumptions

Other Potential Pitfalls

1. In a situation where the randomization assumption (ignorability) or exclusion restriction rely upon conditioning on covariates, it is extremely important that the full set of covariates gets included.
2. It is important to remember that the exclusion restriction and randomization are separate assumptions. *Randomization is not sufficient for exclusion.*

How can we check if randomization is satisfied?

We can't directly. We can rule some things out:

Look for an association between the instrument(s) and personal characteristics that should not be affected by the instrument (anything that doesn't change or was determined before the treatment). For instance, if the instrument is binary can produce a balance table.

Can we check if exclusion is satisfied?

We can't directly. But we may be able to indirectly:

1. If possible, look for an association between the instrument and post-instrument outcomes that could be considered to be alternate paths but occur before the treatment or would not be expected to be affected by the treatment.
2. Look for an association between the instrument(s) and outcome(s) in samples where there is no reason for such a relationship (e.g. in the draft lottery example they used people born in 1953 for whom the lottery took place. The war ended before these people actually were drafted).

Thinking about IV assumptions in empirical research

Consider the following examples

Determine which variable is the instrument, treatment, and outcome. Then discuss the plausibility of the assumptions.

1. Weather
   Supply of fish
   Price of fish
2. Crime rates
   Electoral cycles
   Number of police employed
3. Receiving a phone call encouraging you to vote
   Voting in a given election
   Assignment to receive a phone call encouraging you to vote in an upcoming election
4. Labor market participation
   Sex ratio of 1st two children
   Having a third child
5. Attendance at charter school
   Test scores
   Lottery for admission into a charter school
6. Assignment of judges to juvenile cases
   High school completion
   Incarceration