

# Causal Inference Methods and Case Studies

STAT24630

Jingshu Wang

# Lecture 10

Topic: Non-compliance in randomized experiments, instrumental variables

- Non-compliance in randomized experiment
  - Intention-to-treat effect
  - Principal stratification
  - The monotonicity and exclusion restriction assumptions
  - CATE estimand and the moment-based estimator
  - Connection with two-stage least square estimator
  - Weak instrument

# The Sommer-Zeger vitamin A supplement data

- In principle, 8 different possible values of the triple  $(Z_i, W_i^{\text{obs}}, Y_i^{\text{obs}})$
- Non-compliance:  $Z_i \neq W_i^{\text{obs}}$

Assignment $Z_i$	Vitamin Supplements $W_i^{\text{obs}}$	Survival $Y_i^{\text{obs}}$	Number of Units ( $N = 23,682$ )
0	0	0	74
0	0	1	11,514
1	0	0	34
1	0	1	2385
1	1	0	12
1	1	1	9663

# Three types of traditional analyses

Method	Estimate	Calculation	Row Comparison
ITT	0.0026	$= \frac{2385 + 9663}{12 + 9663 + 34 + 2385} - \frac{11514}{74 + 11514}$	3, 4, 5, & 6 vs. 1 & 2
As-treated	0.0065	$= \frac{9663}{12 + 9663} - \frac{11514 + 2385}{74 + 11514 + 34 + 2385}$	5 & 6 vs. 1, 2, 3, & 4
Per-protocol	0.0052	$= \frac{9663}{12 + 9663} - \frac{11514}{74 + 11514}$	5 & 6 vs. 1 & 2

Assignment $Z_i$	Vitamin Supplements $W_i^{\text{obs}}$	Survival $Y_i^{\text{obs}}$	Number of Units ( $N = 23,682$ )
0	0	0	74
0	0	1	11,514
1	0	0	34
1	0	1	2385
1	1	0	12
1	1	1	9663

Can we provide a better analysis?

# Setup of the framework

- Treatment assignment (randomized encouragement):  $Z_i \in \{0,1\}$
- Potential treatment variables:  $(W_i(0), W_i(1))$ 
  - $W_i(z) = 1$ : would receive the treatment if  $Z_i = z$
  - $W_i(z) = 0$ : would not receive the treatment if  $Z_i = z$
- Observed treatment received:  $W_i^{\text{obs}} = W_i(Z_i)$
- In the non-compliance setting, there are two “treatment”: assignment to treatment and receipt of treatment
- Potential outcomes:  $Y_i(z, w)$  potential outcome if unit is assigned to  $z$  and receive  $w$
- Observed outcome:  $Y_i^{\text{obs}} = Y_i(Z_i, W_i(Z_i))$
- We can also write the potential outcomes as  $Y_i(z) = Y_i(z, W_i(z))$

# Underlying assumptions

- No interference assumption for  $W_i(z)$  and  $Y_i(z, w)$
- Randomization of the treatment assignment
$$(Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1), W_i(0), W_i(1)) \perp Z_i$$
- We don't have
$$(Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1)) \perp W_i^{\text{obs}}$$
or
$$(Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1)) \perp W_i^{\text{obs}} | Z_i$$
We don't know why some units comply and some units don't
- Compliance can not be controlled by randomized experiment

# Intention-to-treat (ITT) effects

- ITT effect on the receipt of treatment level

$$\text{ITT}_{W,i} = W_i(1) - W_i(0) \quad \text{ITT}_W = \frac{1}{N} \sum_{i=1}^N \text{ITT}_{W,i} = \frac{1}{N} \sum_{i=1}^N (W_i(1) - W_i(0))$$

- ITT effect on the outcome of primary interest

$$\text{ITT}_{Y,i} = Y_i(1, W_i(1)) - Y_i(0, W_i(0))$$

$$\text{ITT}_Y = \frac{1}{N} \sum_{i=1}^N \text{ITT}_{Y,i} = \frac{1}{N} \sum_{i=1}^N (Y_i(1, W_i(1)) - Y_i(0, W_i(0)))$$

# Statistical analysis of ITT effects

- Statistical analyses of these effects follow exactly the same procedures as before

$$\widehat{\text{ITT}_W} = \bar{W}_1^{\text{obs}} - \bar{W}_0^{\text{obs}} \quad \widehat{\mathbb{V}}(\widehat{\text{ITT}_W}) = \frac{s_{W,0}^2}{N_0} + \frac{s_{W,1}^2}{N_1}$$

$$s_{W,z}^2 = \sum_{i: W_i^{\text{obs}}=z} \frac{(W_i^{\text{obs}} - \bar{W}_z^{\text{obs}})^2}{N_z - 1} = \frac{N_z}{N_z - 1} \bar{W}_z^{\text{obs}}(1 - \bar{W}_z^{\text{obs}})$$

$$\widehat{\text{ITT}_Y} = \bar{Y}_1^{\text{obs}} - \bar{Y}_0^{\text{obs}} \quad \widehat{\mathbb{V}}(\widehat{\text{ITT}_Y}) = \frac{s_{Y,1}^2}{N_1} + \frac{s_{Y,0}^2}{N_0}$$

- We can also use regression analyses
- Drawback is that it estimates 'programmatic effectiveness' instead of '**biologic efficacy**'

# Principal stratification

- Stratify individuals based on their compliance status
- Four principal strata

- Compliers (co)  $(W_i(0), W_i(1)) = (0,1)$
- Non-compliers (nc)  $\begin{cases} \text{Always - takers (at)} & (W_i(0), W_i(1)) = (1, 1) \\ \text{never - takers (nt)} & (W_i(0), W_i(1)) = (0, 0) \\ \text{Defiers (df)} & (W_i(0), W_i(1)) = (1, 0) \end{cases}$

$W_i(1)$		
	0	1
	0	nt
		co
$W_i(0)$		
1	df	at

# Principal stratification

- Principal stratification depends on latent states of units!!
- Can not decide which principal strata each unit belong to simply based on the observed data
  - **one-sided compliance:** control group can never receive the treatment, but treatment group may not follow the assignment

		Assignment $Z_i$	
		0	1
Receipt of treatment $W_i^{\text{obs}}$	0	nt/co	nt
	1	—	co

- In general

		$Z_i$	
		0	1
$W_i^{\text{obs}}$	0	nt/co	nt/df
	1	at/df	at/co

# ITT effect decomposition

- Denote the proportion of individuals that fall into each strata as  $\pi_c, \pi_a, \pi_n, \pi_d$ 
  - For one-sided compliance data,  $\pi_a = \pi_d = 0$
- Define the average ITT effect for each strata
  - For the treatment received  $\text{ITT}_{W,c}, \text{ITT}_{W,a}, \text{ITT}_{W,n}, \text{ITT}_{W,d}$   
 $\text{ITT}_{W,c} = 1, \text{ITT}_{W,a} = 0, \text{ITT}_{W,n} = 0, \text{ITT}_{W,d} = -1$

- For the primary outcome  $\text{ITT}_c, \text{ITT}_a, \text{ITT}_n, \text{ITT}_d$

- For the ITT effect on treatment received

$$\text{ITT}_W = \sum_{i=1}^N \text{ITT}_{W,i} = \pi_c \text{ITT}_{W,c} + \pi_a \text{ITT}_{W,a} + \pi_n \text{ITT}_{W,n} + \pi_d \text{ITT}_{W,d} = \pi_c - \pi_d$$

- For the ITT effect on primary outcome

$$\text{ITT}_Y = \sum_{i=1}^N \text{ITT}_{Y,i} = \pi_c \text{ITT}_c + \pi_a \text{ITT}_a + \pi_n \text{ITT}_n + \pi_d \text{ITT}_d$$

# Instrumental variables (IV)

Assumptions for  $Z_i$  being a valid IV:

- Randomization:  $Z_i \in \{0,1\}$  are randomized
- Monotonicity: no defiers  $\pi_d = 0$  or  $W_i(0) \leq W_i(1)$  for all  $i$
- Exclusion restriction: instrument affects the outcome only through treatment

$$Y_i(1, w) = Y_i(0, w)$$

- For always takers

$$\text{ITT}_{Y,i} = Y_i(1, W_i(1)) - Y_i(0, W_i(0)) = Y_i(1,1) - Y_i(0,1) = 0$$

so  $\text{ITT}_a = 0$

- For never takers

$$\text{ITT}_{Y,i} = Y_i(1, W_i(1)) - Y_i(0, W_i(0)) = Y_i(1,0) - Y_i(0,0) = 0$$

so  $\text{ITT}_n = 0$

- For compliers

$$\text{ITT}_{Y,i} = Y_i(1, W_i(1)) - Y_i(0, W_i(0)) = Y_i(1,1) - Y_i(0,0)$$

$\text{ITT}_c$  is the average ``biological efficacy'' of the treatment on compliers

- Relevance:  $\pi_c > 0$

# Instrumental variables

Assumptions of  $Z_i$  being a valid IV :

- Randomization:  $Z_i \in \{0,1\}$  are randomized
- Monotonicity: no defiers  $\pi_d = 0$  or  $W_i(0) \leq W_i(1)$  for all  $i$
- Exclusion restriction: instrument affects the outcome only through treatment

$$Y_i(1, w) = Y_i(0, w)$$

- Relevance:  $\pi_c > 0$
- Then  $\text{ITT}_W = \pi_c$  and  $\text{ITT}_Y = \pi_c \text{ITT}_c + \pi_a \text{ITT}_a + \pi_n \text{ITT}_n + \pi_d \text{ITT}_d = \pi_c \text{ITT}_c$
- IV estimand:  $\text{ITT}_c$  Complier average treatment effect (CATE)

$$\text{CATE} = \text{ITT}_c = \frac{\text{ITT}_Y}{\text{ITT}_W}$$

- We can identify  $\text{ITT}_Y$  and  $\text{ITT}_W$ , so  $\text{ITT}_c$  is also identifiable
- $\text{CATE} \neq \text{ATE}$  unless ATE for noncompliers equals CATE

# The monotonicity assumption

- **Monotonicity:** no defiers  $\pi_d = 0$  or  $W_i(0) \leq W_i(1)$  for all  $i$
- Defiers are individuals who never follow treatment assignment no matter what treatment assignment is
- For one-sided compliance data, monotonicity is always satisfied
- Check the monotonicity assumption in general:
  - $\text{ITT}_W = \pi_c - \pi_d > 0$  if  $\pi_d = 0$ , so if we can reject the null that  $\text{ITT}_W \geq 0$ , then monotonicity assumption must fail
  - Otherwise, the monotonicity assumption is not testable
  - Need to decide whether the monotonicity assumption is reasonable or not based on domain knowledge

# The exclusion restriction assumption

- **Exclusion restriction:** instrument affects the outcome only through treatment  
$$Y_i(1, w) = Y_i(0, w)$$
- Double-blinding in experiments guarantees exclusion restriction
- The assumption in general is not testable, and need subject-matter knowledge to judge
- The subject-matter knowledge needed is often more subtle than that required to evaluate SUTVA

# Moment-based IV estimator

- Causal estimand assuming a super population

$$\text{CATE} = \frac{\text{ITT}_Y}{\text{ITT}_W} = \frac{\mathbb{E}(Y_i(1) - Y_i(0))}{\mathbb{E}(W_i(1) - W_i(0))}$$

- Method-of-moment estimator:

$$\hat{\tau}^{iv} = \frac{\widehat{\text{ITT}}_Y}{\widehat{\text{ITT}}_W}$$

- How to estimate the variance of  $\hat{\tau}^{iv}$ ?

- Estimates  $\widehat{\text{ITT}}_Y$  and  $\widehat{\text{ITT}}_W$  are correlated because they use the same dataset
- We can approximate the variance of  $\hat{\tau}^{iv}$  when  $N$  is large (from delta method):

$$\mathbb{V}(\hat{\tau}^{iv}) \approx \frac{1}{\text{ITT}_W^4} \{ \text{ITT}_W^2 \mathbb{V}(\widehat{\text{ITT}}_Y) + \widehat{\text{ITT}}_Y^2 \mathbb{V}(\widehat{\text{ITT}}_W) - 2\text{ITT}_Y \text{ITT}_W \text{Cov}(\widehat{\text{ITT}}_W, \widehat{\text{ITT}}_Y) \}$$

- Plug-in estimator of  $\mathbb{V}(\hat{\tau}^{iv})$ :

$$\widehat{\mathbb{V}}(\hat{\tau}^{iv}) \approx \frac{1}{\widehat{\text{ITT}}_W^4} \{ \widehat{\text{ITT}}_W^2 \widehat{\mathbb{V}}(\widehat{\text{ITT}}_Y) + \widehat{\text{ITT}}_Y^2 \widehat{\mathbb{V}}(\widehat{\text{ITT}}_W) - 2\widehat{\text{ITT}}_Y \widehat{\text{ITT}}_W \widehat{\text{Cov}}(\widehat{\text{ITT}}_W, \widehat{\text{ITT}}_Y) \}$$

# Estimate the covariance

- The covariance between  $\widehat{\text{ITT}}_Y$  and  $\widehat{\text{ITT}}_W$ :

$$\begin{aligned}\text{Cov}(\widehat{\text{ITT}}_W, \widehat{\text{ITT}}_Y) &= \text{Cov}(\bar{W}_1^{\text{obs}} - \bar{W}_0^{\text{obs}}, \bar{Y}_1^{\text{obs}} - \bar{Y}_0^{\text{obs}}) \\ &= \frac{\text{Cov}(Y_i(1), W_i(1))}{N_1} + \frac{\text{Cov}(Y_i(0), W_i(0))}{N_0}\end{aligned}$$

- To estimate the covariance  $\text{Cov}(Y_i(z), W_i(z))$  for  $z = 0, 1$ :

$$\widehat{\text{Cov}}(Y_i(z), W_i(z)) = \frac{1}{N_z - 1} \sum_{i:Z_i=z} (W_i^{\text{obs}} - \bar{W}_z^{\text{obs}})(Y_i^{\text{obs}} - \bar{Y}_z^{\text{obs}})$$

- So, the plug-in estimator is

$$\widehat{\text{Cov}}(\widehat{\text{ITT}}_W, \widehat{\text{ITT}}_Y) = \sum_{z=0}^1 \frac{\sum_{i:Z_i=z} (W_i^{\text{obs}} - \bar{W}_z^{\text{obs}})(Y_i^{\text{obs}} - \bar{Y}_z^{\text{obs}})}{N_z(N_z - 1)}$$

- 95% confidence interval of CATE:  $[\hat{\tau}^{iv} - 1.96\sqrt{\hat{V}(\hat{\tau}^{iv})}, \hat{\tau}^{iv} + 1.96\sqrt{\hat{V}(\hat{\tau}^{iv})}]$

# Simplification for one-sided compliance data

As  $W_i(0) \equiv 0$ , we have

- $\widehat{\text{ITT}}_W = \bar{W}_1^{\text{obs}} - \bar{W}_0^{\text{obs}} = \bar{W}_1^{\text{obs}}$
- $\widehat{\mathbb{V}}(\widehat{\text{ITT}}_W) = \frac{s_{W,1}^2}{N_1} = \frac{\bar{W}_1^{\text{obs}}(1-\bar{W}_1^{\text{obs}})}{(N_1-1)}$  as  $s_{W,0}^2 = 0$
- $\widehat{\text{Cov}}(\widehat{\text{ITT}}_W, \widehat{\text{ITT}}_Y) = \frac{\sum_{i:Z_i=1} (W_i^{\text{obs}} - \bar{W}_1^{\text{obs}})(Y_i^{\text{obs}} - \bar{Y}_1^{\text{obs}})}{N_1(N_1-1)}$

# Result in Sommer-Zeger Vitamin Supplement data

ITT Estimates:

- $N_1 = 12 + 9663 + 34 + 2385 = 12094, N_0 = 74 + 11514 = 11588$
- $\widehat{\text{ITT}}_W = \bar{W}_1^{\text{obs}} = \frac{12+9663}{N_1} = 0.8, \widehat{\text{V}}(\widehat{\text{ITT}}_W) = \frac{\bar{W}_1^{\text{obs}}(1-\bar{W}_1^{\text{obs}})}{N_1-1} = \frac{0.2*0.8}{12093} = 0.0036^2$
- $\widehat{\text{ITT}}_Y = \frac{2385+9663}{N_1} - \frac{11514}{N_0} = 0.0026, \widehat{\text{V}}(\widehat{\text{ITT}}_Y) = \sum_{z=0}^1 \frac{\bar{Y}_z^{\text{obs}}(1-\bar{Y}_z^{\text{obs}})}{N_z-1} = 0.0009^2$
- 95% CI of  $\widehat{\text{ITT}}_Y$ : (0.0008, 0.0044)

CATE estimate:

- $\widehat{\tau}^{iv} = \frac{0.0026}{0.8} = 0.0032$
- $\widehat{\text{Cov}}(\widehat{\text{ITT}}_W, \widehat{\text{ITT}}_Y) = -0.0000017$  (correlation -0.05)
- $\widehat{\text{V}}(\widehat{\tau}^{iv}) = 0.0012^2$

Assignment $Z_i$	Vitamin Supplements $W_i^{\text{obs}}$	Survival $Y_i^{\text{obs}}$	Number of Units ( $N = 23,682$ )
0	0	0	74
0	0	1	11,514
1	0	0	34
1	0	1	2385
1	1	0	12
1	1	1	9663

- 95% CI of CATE: (0.0010, 0.0055)
- The as-protocol or as-treated estimates are possibly biased up

# Two-stage least square (2SLS) estimator

- Conventionally in econometrics, researchers use a two-stage least square approach for CATE
- The two-stage least square estimator is **equivalent** to  $\hat{\tau}^{iv}$
- Two-stage least square
  - Stage 1: regress  $W_i^{\text{obs}}$  on  $Z_i$  : the coefficient of  $Z_i$  is  $\text{ITT}_W$  (regression with no covariate)  
the fitted coefficient on  $Z_i$  is  $\widehat{\text{ITT}}_W$
  - Stage 1: regress  $Y_i^{\text{obs}}$  on  $Z_i$  : the coefficient of  $Z_i$  is  $\text{ITT}_Y$  (regression with no covariate)  
the fitted coefficient on  $Z_i$  is  $\widehat{\text{ITT}}_Y$
  - Take the ratio of estimated coefficients, which is exactly  $\hat{\tau}^{iv}$
- We can generalize 2SLS to incorporate covariates when estimating  $\text{ITT}_W$  and  $\text{ITT}_Y$

# The Angrist draft lottery data

## Background

- Policy makers are interested in whether veterans are adequately compensated for their service.
- Angrist (1991) aims to measure the long-term labor market consequences of military service during the Vietnam era
- Question: estimate the causal effect of serving in the military during the Vietnam War on earnings
- We can not directly compare veterans and non-veterans, as they can be systematically different in unobserved ways, even after adjusting for differences in observed covariates
- Serving in the military or not during the Vietnam War could not be randomized directly, but the military draft lottery of the Vietnam War was randomized
- This is called **a natural experiment**

# The Angrist draft lottery data

## Randomization

- For each birth year of birth cohort 1950-1952, a random ordering of the 365 days was constructed, a cutoff number was pre-determined, young men of that birth year who had a birth date with order before the cutoff “won” the lottery
- Randomization of birth date, instead of the individuals
- Theoretically, each date should be a unit, but in the book example, we treat each individual as a unit and consider the experiment as a completely randomized experiment (it’s actually a stratified cluster randomized experiment).  
Consequence is that we will tend to under-estimate the uncertainty of the causal estimator.

## Relevance and two-sided non-compliance:

- Drafted individuals were required to prepare to serve in the military if fit for the service
- To serve the military, drafted individuals need to pass medical tests and have achieved minimum education level
- Individuals who were not draft eligible also can volunteer to serve in the military

# The Angrist draft lottery data

	Non-Veterans ( $N_c = 6,675$ )				Veterans ( $N_t = 2,030$ )			
	Min	Max	Mean	(S.D.)	Min	Max	Mean	(S.D.)
Draft eligible	0	1	0.24	(0.43)	0	1	0.40	(0.49)
Yearly earnings (in \$1,000's)	0	62.8	11.8	(11.5)	0	50.7	11.7	(11.8)
Earnings positive	0	1	0.88	(0.32)	0	1	0.91	(0.29)
Year of birth	50	52	51.1	(0.8)	50	52	50.9	(0.8)

## Check assumptions

- **Monotonicity:** appears to be a reasonable assumption
  - The lottery numbers impose restrictions on individuals' behaviors.
  - Monotonicity means that no one responds to these restrictions by serving only if they are not required to do so
  - It is possible that there are some individuals who would be willing to volunteer if they are not drafted but would resist the draft if required, but it must be a very small fraction and are likely ignorable

# The Angrist draft lottery data

	Non-Veterans ( $N_c = 6,675$ )				Veterans ( $N_t = 2,030$ )			
	Min	Max	Mean	(S.D.)	Min	Max	Mean	(S.D.)
Draft eligible	0	1	0.24	(0.43)	0	1	0.40	(0.49)
Yearly earnings (in \$1,000's)	0	62.8	11.8	(11.5)	0	50.7	11.7	(11.8)
Earnings positive	0	1	0.88	(0.32)	0	1	0.91	(0.29)
Year of birth	50	52	51.1	(0.8)	50	52	50.9	(0.8)

## Check assumptions

- **Exclusion restriction:** may be questionable
  - Consider the never-takers
  - Some never-takers are due to medical exemptions or exemptions due to their education or career choices. For them, the lottery numbers would likely not affect their future behaviors and the outcome
  - Some never-takers did have exemptions but changed their plan (enter graduate school or move to Canada) if they had a low draft number to avoid serving in the military. For them, exclusion restriction can be violated.

# Analysis results

ITT Estimates:

- $\widehat{\text{ITT}}_W = 0.1460, \widehat{\text{V}}(\widehat{\text{ITT}}_W) = 0.0108^2$
- $\widehat{\text{ITT}}_Y = -0.2129, \widehat{\text{V}}(\widehat{\text{ITT}}_W) = \sum_{z=0}^1 \frac{\bar{Y}_z^{\text{obs}}(1-\bar{Y}_z^{\text{obs}})}{N_z(N_z-1)} = 0.1980^2$
- 95% CI of  $\text{ITT}_Y$ :  $(-0.6010, 0.1752)$

If we are willing to assume monotonicity and exclusion restriction

CATE estimate:

- $\widehat{\tau}^{iv} = \frac{-0.2129}{0.1460} = -1.46$
- $\widehat{\text{V}}(\widehat{\tau}^{iv}) = 1.36^2$
- 95% CI of CATE:  $(-4.13, 1.2)$

# Weak instrument

- The instrumental variable is a weak instrument if the compliance probability ( $\pi_c$  or  $\widehat{\text{ITT}}_W$ ) is small
- Problems using weak instrument
  - $\hat{\tau}^{iv} = \frac{\widehat{\text{ITT}}_Y}{\widehat{\text{ITT}}_W}$ : the ratio is very unstable. If  $\widehat{\text{ITT}}_W$  is close to 0, then a small error (perturbation) in  $\widehat{\text{ITT}}_W$  can lead to a large error in  $\hat{\tau}^{iv}$
  - If the exclusion restriction assumption is violated, the bias in our estimator assuming exclusion restriction is inversely proportional to  $\pi_c$
- How to identify weak instrument?
  - In the first stage linear regression model  $W_i^{\text{obs}} = \alpha + \pi_c W_i + \varepsilon_i$ , calculate the F-statistics to test whether  $\pi_c = 0$
  - A rule of thumb is to check whether the F-statistics is larger than 10 or not.
  - F-statistics smaller than 10 indicates a weak instrument