J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

Threats and Analysis

Lina Marliani
Executive Director
J-PAL Southeast Asia

Course Overview

1. What is Evaluation?
2. Outcomes, Impact, and Indicators
3. Why Randomize?
4. How to Randomize
5. Sampling and Sample Size
6. Threats and Analysis
7. Evaluation from Start to Finish
8. Evidence from Community-Driven Development, Health, and Education Programs
9. Using Evidence from Randomized Evaluations

Lecture Overview

- Attrition
- Spillovers
- Partial Compliance and Sample Selection Bias
- Intention to Treat & Treatment on Treated
- Choice of outcomes
- External validity

Lecture Overview

- Attrition
- Spillovers
- Partial Compliance and Sample Selection Bias
- Intention to Treat & Treatment on Treated
- Choice of outcomes
- External validity

Attrition

- Is it a problem if some of the people in the experiment vanish before you collect your data?
  - It is a problem if the type of people who disappear is correlated with the treatment.
- Why is it a problem?
- Why should we expect this to happen?

Attrition bias: an example

- The problem you want to address:
  - Some children don't come to school because they are too weak (undernourished)
- You start a school feeding program and want to do an evaluation
  - You have a treatment and a control group
- Weak, stunted children start going to school more if they live next to a treatment school
- First impact of your program: increased enrollment.
- In addition, you want to measure the impact on child's growth
  - Second outcome of interest: Weight of children
- You go to all the schools (treatment and control) and measure everyone who is in school on a given day
- Will the treatment-control difference in weight be over-stated or understated?

	Before Treatment			After Treament	
	T	C		T	C
	20	20		**22**	20
	25	25		**27**	25
	30	30		**32**	30
Ave.					
Difference			Difference		

	Before Treatment			After Treament	
	T	C		T	C
	20	20		**22**	20
	25	25		**27**	25
	30	30		**32**	30
Ave.	**25**	**25**		**27**	**25**
Difference		0	Difference		2

What if only children > 21 Kg come to school?

What if only children > 21 Kg come to school?

	Before Treatment			After Treament	
	T	C		T	C
	20	20		**22**	20
	25	25		**27**	25
	30	30		**32**	30

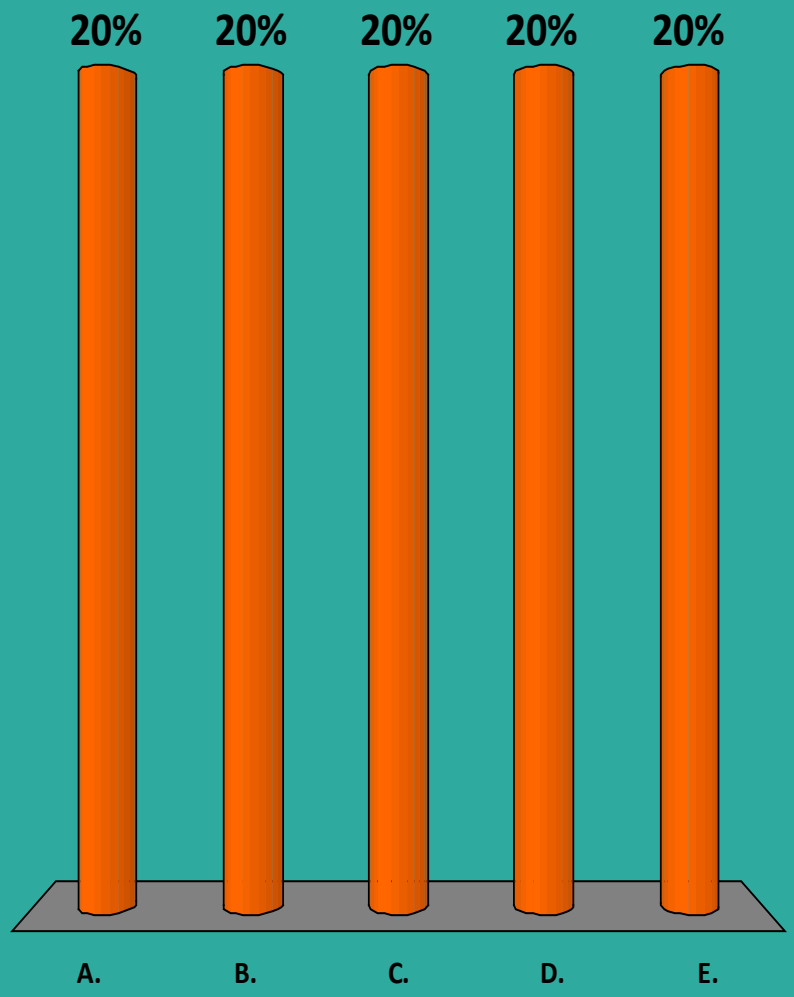
A. Will you underestimate the impact?

B. Will you overestimate the impact?

C. Neither

D. Ambiguous

E. Don't know



What if only children > 21 Kg come to school?

	Before Treatment			After Treament	
	T	C		T	C
	[absent]	[absent]		**22**	[absent]
	25	25		**27**	25
	30	30		**32**	30
Ave.	**27.5**	**27.5**		**27**	**27.5**
Difference		0	Difference		-0.5

When is attrition not a problem?

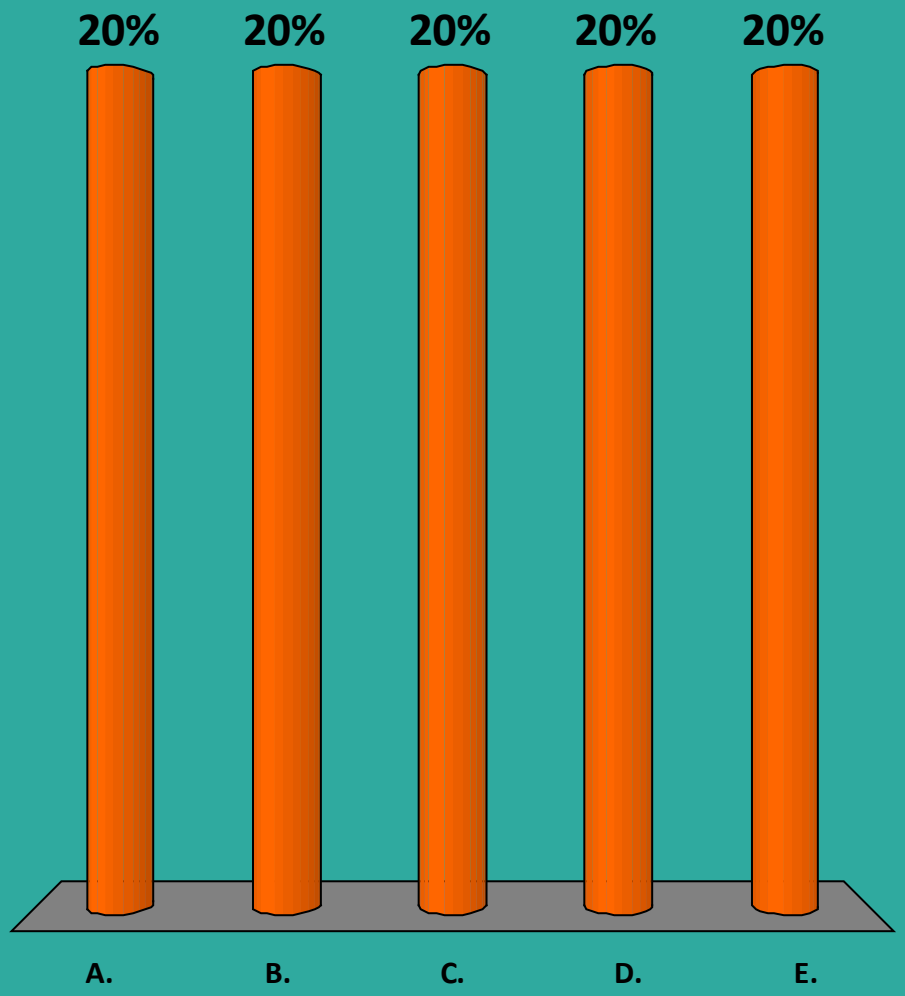
A. When it is less than 25% of the original sample

B. When it happens in the same proportion in both groups

C. When it is correlated with treatment assignment

D. All of the above

E. None of the above



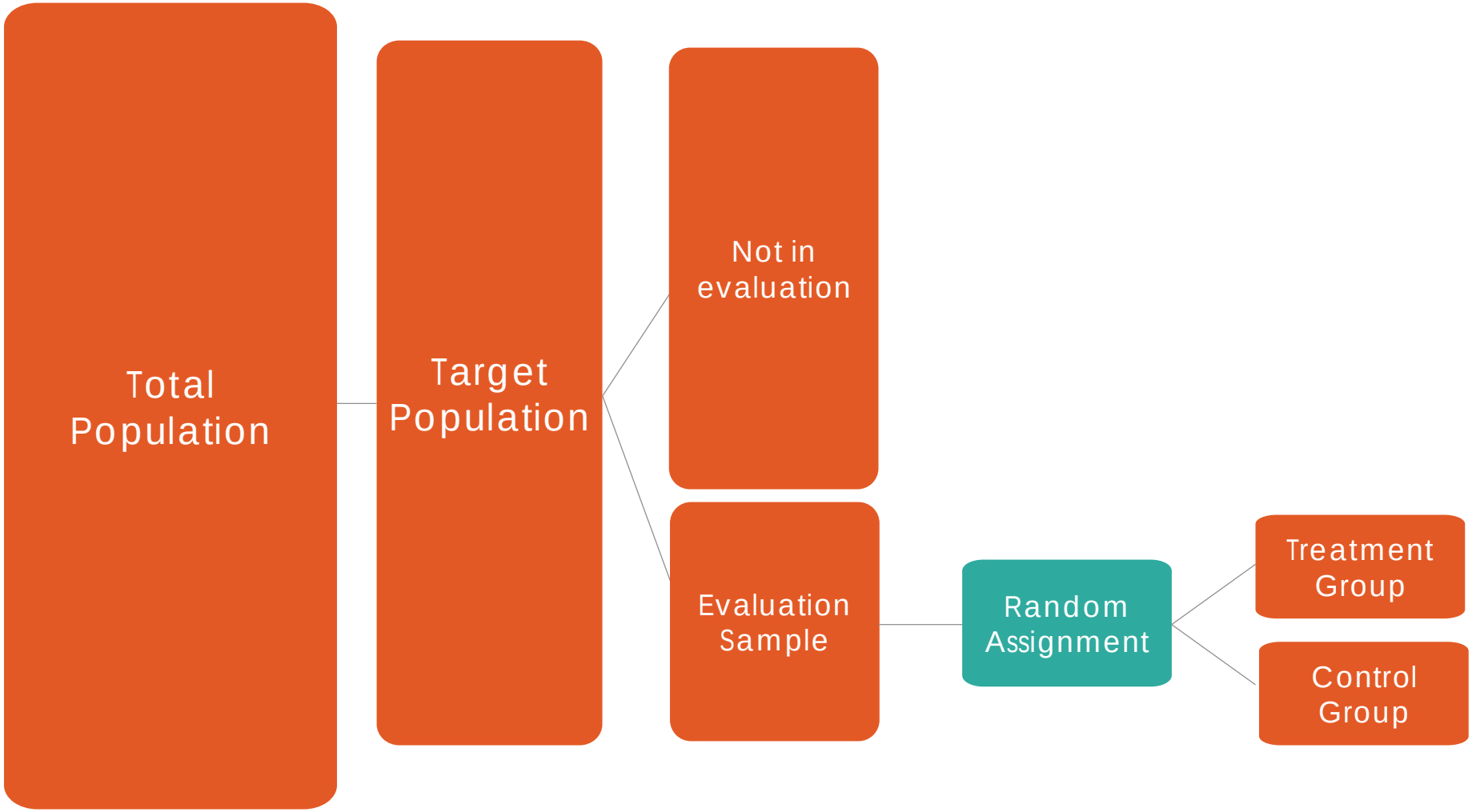
Attrition Bias

- Devote resources to tracking participants after they leave the program
- If there is still attrition, check that it is not different in treatment and control. Is that enough?
- Also check that it is not correlated with observables.
- Try to bound the extent of the bias
  - suppose everyone who dropped out from the treatment got the lowest score that anyone got; suppose everyone who dropped out of control got the highest score that anyone got…
  - Why does this help?

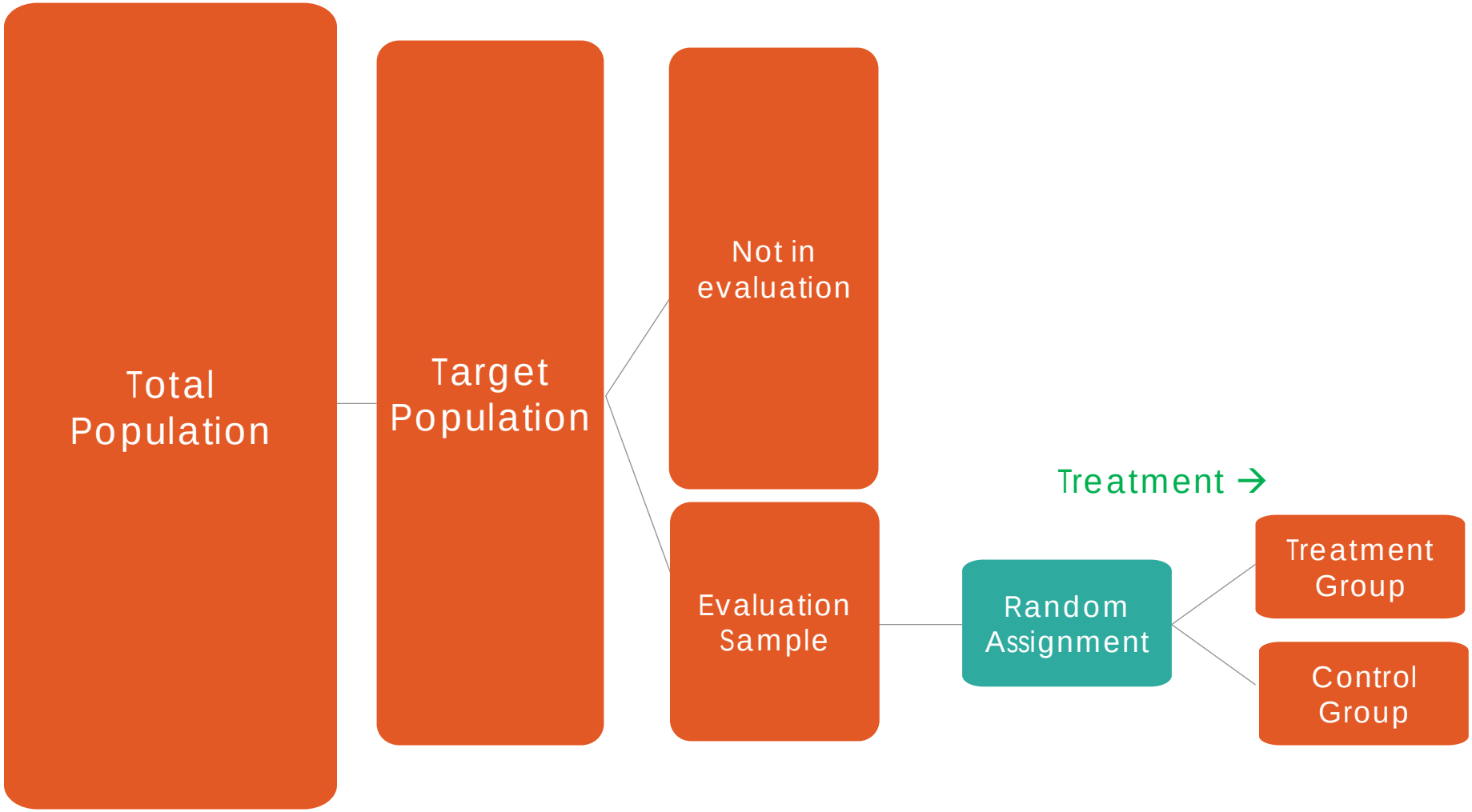
Lecture Overview

- Attrition
- Spillovers
- Partial Compliance and Sample Selection Bias
- Intention to Treat & Treatment on Treated
- Choice of outcomes
- External validity

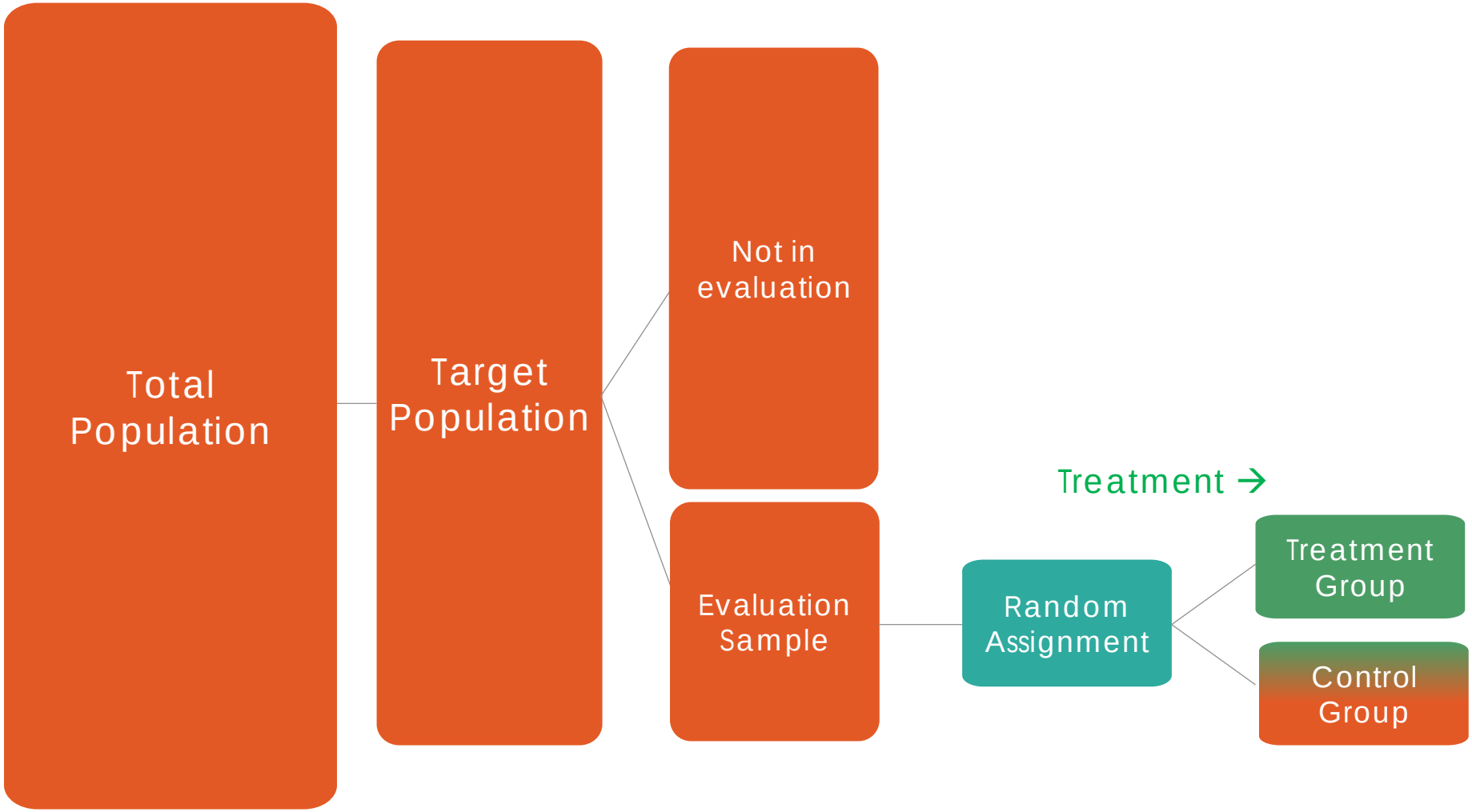
What else could go wrong?



Spillovers, contamination



Spillovers, contamination



Example: Vaccination for chicken pox

- Suppose you randomize chicken pox vaccinations within schools
  - Suppose that prevents the transmission of disease, what problems does this create for evaluation?
  - Suppose externalities are local? How can we measure total impact?

Externalities Within School

Without Externalities

School A	Treated?	Outcome
Pupil 1	Yes	no chicken pox
Pupil 2	No	chicken pox
Pupil 3	Yes	no chicken pox
Pupil 4	No	chicken pox
Pupil 5	Yes	no chicken pox
Pupil 6	No	chicken pox

Total in Treatment with chicken pox

Total in Control with chicken pox

Treament Effect

With Externalities

Suppose, because prevalence is lower, some children are not re-infected with chicken pox

School A	Treated?	Outcome
Pupil 1	Yes	no chicken pox
Pupil 2	No	no chicken pox
Pupil 3	Yes	no chicken pox
Pupil 4	No	chicken pox
Pupil 5	Yes	no chicken pox
Pupil 6	No	chicken pox

Total in Treatment with chicken pox

Total in Control with chicken pox

Treatment Effect

Externalities Within School

Without Externalities

School A	Treated?	Outcome
Pupil 1	Yes	no chicken pox
Pupil 2	No	chicken pox
Pupil 3	Yes	no chicken pox
Pupil 4	No	chicken pox
Pupil 5	Yes	no chicken pox
Pupil 6	No	chicken pox

Total in Treatment with chicken pox: 0%

Total in Control with chicken pox: 100%

Treament Effect: -100%

With Externalities

Suppose, because prevalence is lower, some children are not re-infected with chicken pox

School A	Treated?	Outcome
Pupil 1	Yes	no chicken pox
Pupil 2	No	no chicken pox
Pupil 3	Yes	no chicken pox
Pupil 4	No	chicken pox
Pupil 5	Yes	no chicken pox
Pupil 6	No	chicken pox

Total in Treatment with chicken pox: 0%

Total in Control with chicken pox: 67%

Treatment Effect: -67%

How to measure program impact in the presence of spillovers?

- Design the unit of randomization so that it encompasses the spillovers
- If we expect externalities that are all within school:
  - Randomization at the level of the school allows for estimation of the overall effect

Example: Price Information

- Providing farmers with spot and futures price information by mobile phone
- Should we expect spillovers?
- Randomize: individual or village level?
- Village level randomization
  - Less statistical power
  - “Purer control groups”
- Individual level randomization
  - More statistical power (if spillovers small)
  - Ability to measure spillovers

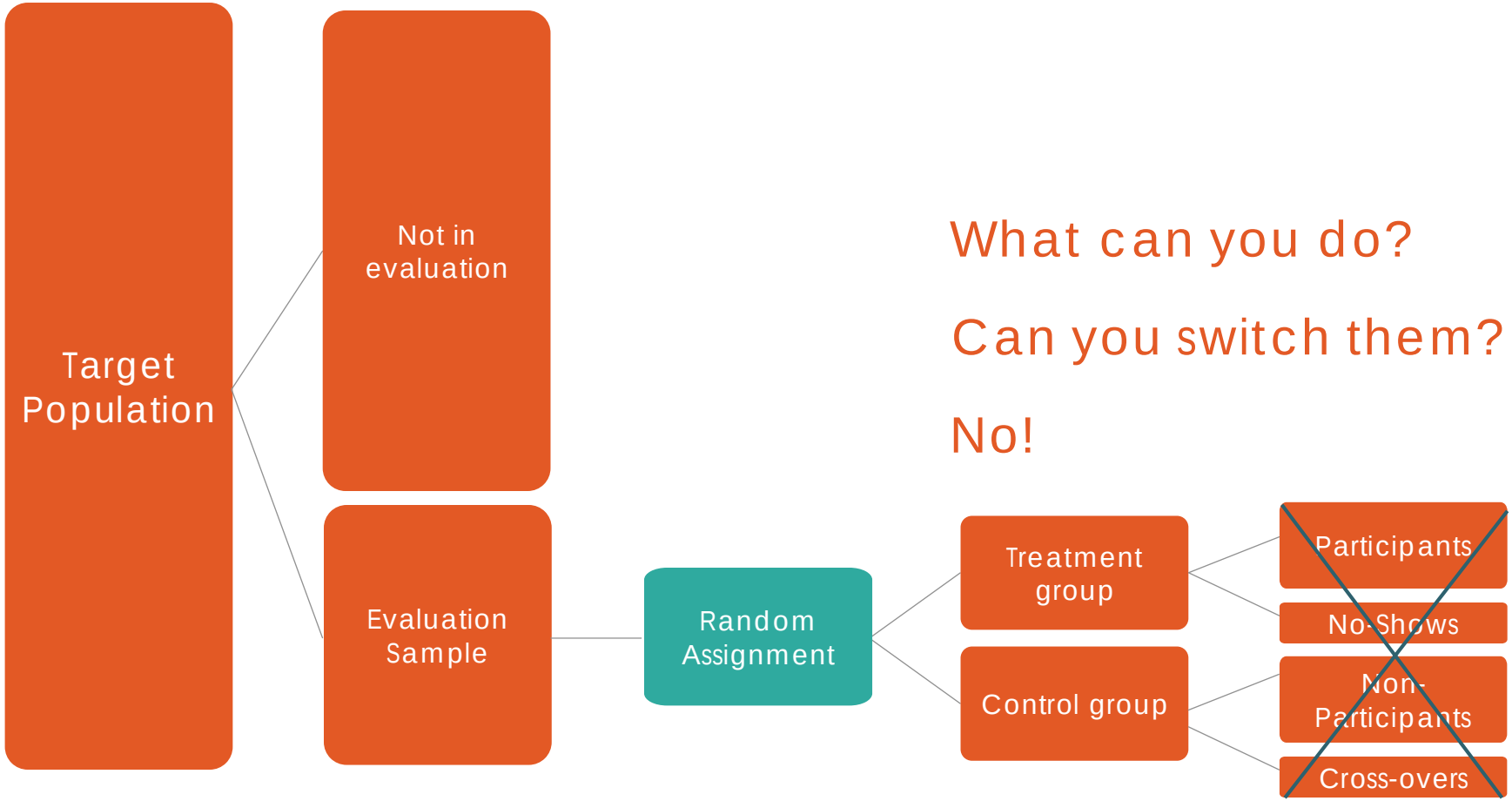
Example: Price Information

- Can we do both?
- Randomly assign villages into one of four groups, A, B, C, & D
- Group A Villages
  - SMS price information to all individuals with phones
- Group B Villages
  - SMS price information to randomly selected 75% of individuals with phones
- Group C Villages
  - SMS price information to randomly selected 25% of individuals with phones
- Group D Villages
  - No SMS price information

Lecture Overview

- Attrition
- Spillovers
- Partial Compliance and Sample Selection Bias
- Intention to Treat & Treatment on Treated
- Choice of outcomes
- External validity

Non compliers

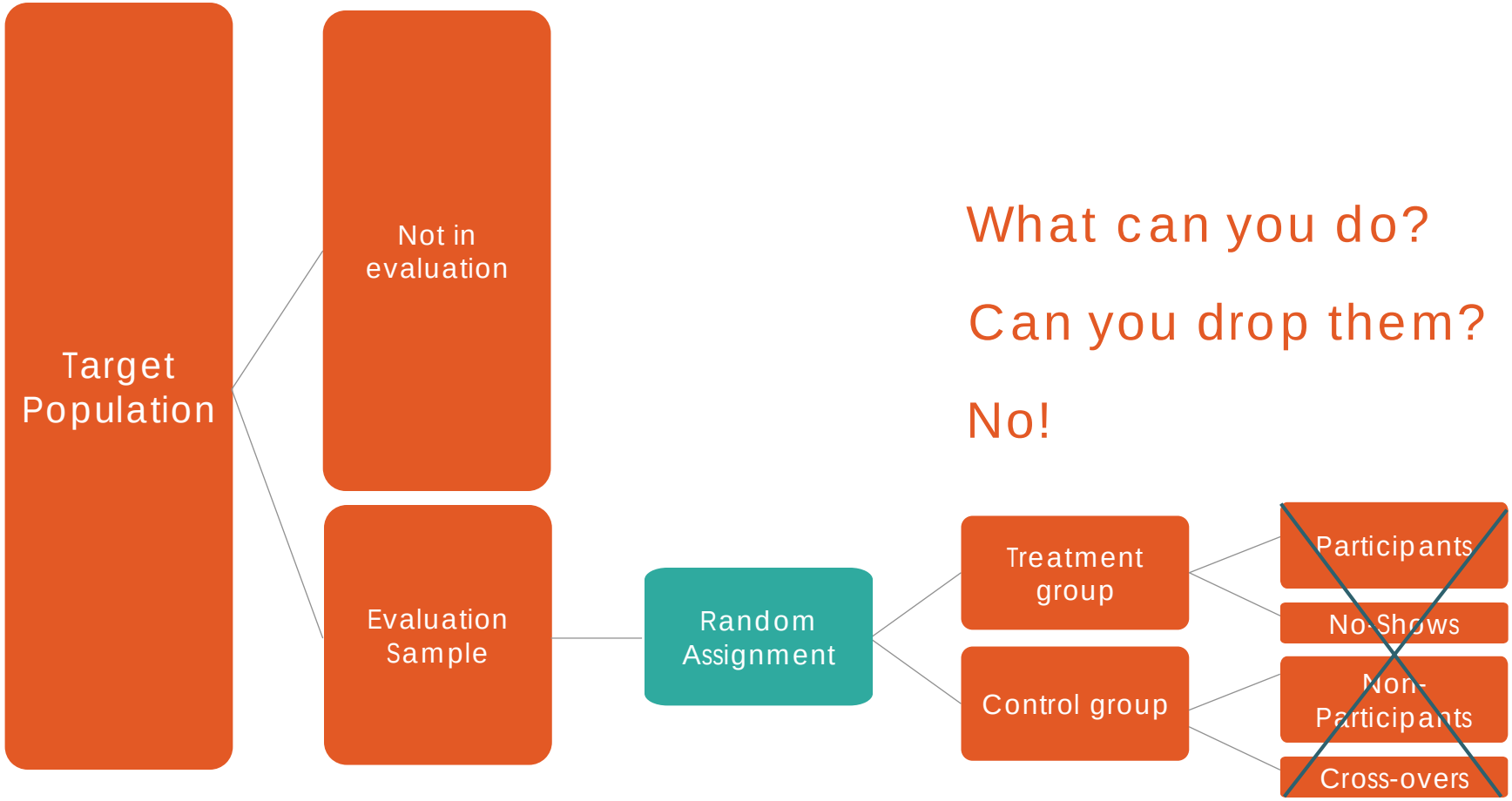


What can you do?

Can you switch them?

No!

Non compliers

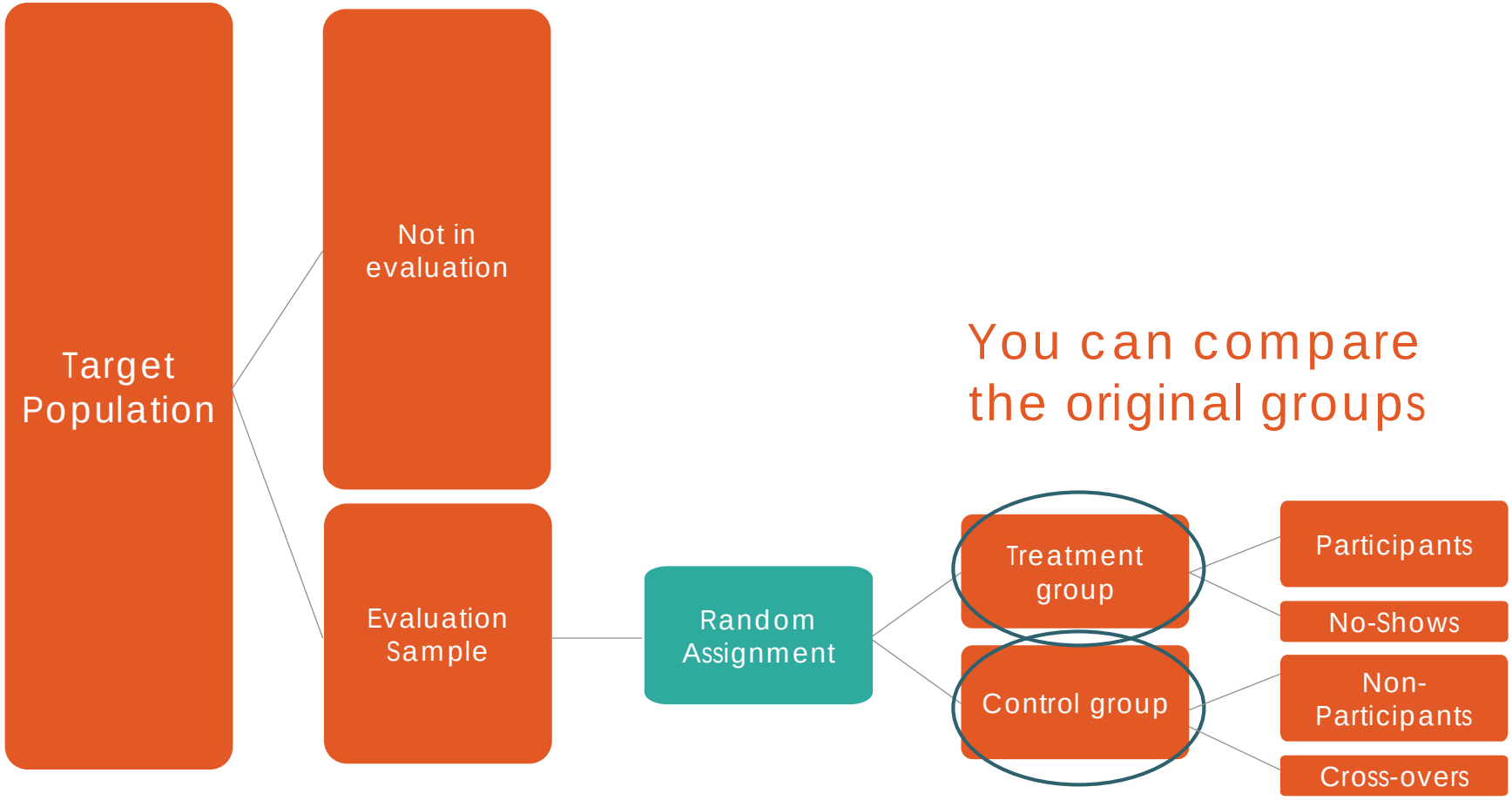


What can you do?

Can you drop them?

No!

Non compliers



You can compare the original groups

Sample selection bias

- Sample selection bias could arise if factors other than random assignment influence program allocation
  - Even if intended allocation of program was random, the actual allocation may not be

Sample selection bias

- Individuals assigned to comparison group could attempt to move into treatment group
  - School feeding program: parents could attempt to move their children from comparison school to treatment school
- Alternatively, individuals allocated to treatment group may not receive treatment
  - School feeding program: some students assigned to treatment schools bring and eat their own lunch anyway, or choose not to eat at all.

Lecture Overview

- Attrition
- Spillovers
- Partial Compliance and Sample Selection Bias
- Intention to Treat & Treatment on Treated
- Choice of outcomes
- External validity

ITT and ToT

- Vaccination campaign in villages
- Some people in treatment villages not treated
  - 78% of people assigned to receive treatment received some treatment
- What do you do?
  - Compare the beneficiaries and non-beneficiaries?
  - Why not?

Which groups can be compared ?

Treatment Group: Vaccination	Control Group
TREATED	NON-TREATED
NON-TREATED	

What is the difference between the 2 random groups?

Treatment Group	Control Group
1: treated – not infected 2: treated – not infected 3: treated – infected	5: non-treated – infected 6: non-treated – not infected 7: non-treated – infected 8: non-treated – infected
4: non-treated – infected	

Intention to Treat - ITT

Treatment Group:	50% infected
Control Group:	75% infected

- Y(T)= Average Outcome in Treatment Group
- Y(C)= Average Outcome in Control Group

$$ITT = Y(T) - Y(C)$$

- ITT = 50% - 75% = -25 percentage points

Intention to Treat (ITT)

- What does “intention to treat” measure?

  “What happened to the average child who is in a treated school in this population?”

- Is this difference the causal effect of the intervention?

When is ITT useful?

- May relate more to actual programs
- For example, we may not be interested in the medical effect of deworming treatment, but what would happen under an actual deworming program.
- If students often miss school and therefore don't get the deworming medicine, the intention to treat estimate may actually be most relevant.

School 1	Intention to Treat ?	Treated?	Observed Change in weight
Pupil 1	yes	**yes**	**4**
Pupil 2	yes	**yes**	**4**
Pupil 3	yes	**yes**	**4**
Pupil 4	yes	no	0
Pupil 5	yes	**yes**	**4**
Pupil 6	yes	no	2
Pupil 7	yes	no	0
Pupil 8	yes	**yes**	**6**
Pupil 9	yes	**yes**	**6**
Pupil 10	yes	no	0
Avg. Change among Treated A=			

School 2			
Pupil 1	no	**no**	**2**
Pupil 2	no	**no**	**1**
Pupil 3	no	yes	3
Pupil 4	no	**no**	**0**
Pupil 5	no	**no**	**0**
Pupil 6	no	yes	3
Pupil 7	no	**no**	**0**
Pupil 8	no	**no**	**0**
Pupil 9	no	**no**	**0**
Pupil 10	no	**no**	**0**
Avg. Change among Not-Treated B=			

School 1:
Avg. Change among Treated (A)

School 2:
Avg. Change among not-treated (B)

A-B

School 1	Intention to Treat ?	Treated?	Observed Change in weight
Pupil 1	yes	**yes**	**4**
Pupil 2	yes	**yes**	**4**
Pupil 3	yes	**yes**	**4**
Pupil 4	yes	no	0
Pupil 5	yes	**yes**	**4**
Pupil 6	yes	no	2
Pupil 7	yes	no	0
Pupil 8	yes	**yes**	**6**
Pupil 9	yes	**yes**	**6**
Pupil 10	yes	no	0
Avg. Change among Treated A=			3

School 2			
Pupil 1	no	**no**	**2**
Pupil 2	no	**no**	**1**
Pupil 3	no	yes	3
Pupil 4	no	**no**	**0**
Pupil 5	no	**no**	**0**
Pupil 6	no	yes	3
Pupil 7	no	**no**	**0**
Pupil 8	no	**no**	**0**
Pupil 9	no	**no**	**0**
Pupil 10	no	**no**	**0**
Avg. Change among Not-Treated B=			0.9

School 1:
Avg. Change among Treated 3 (A)
School 2:
Avg. Change among not-treated 0.9 (B)

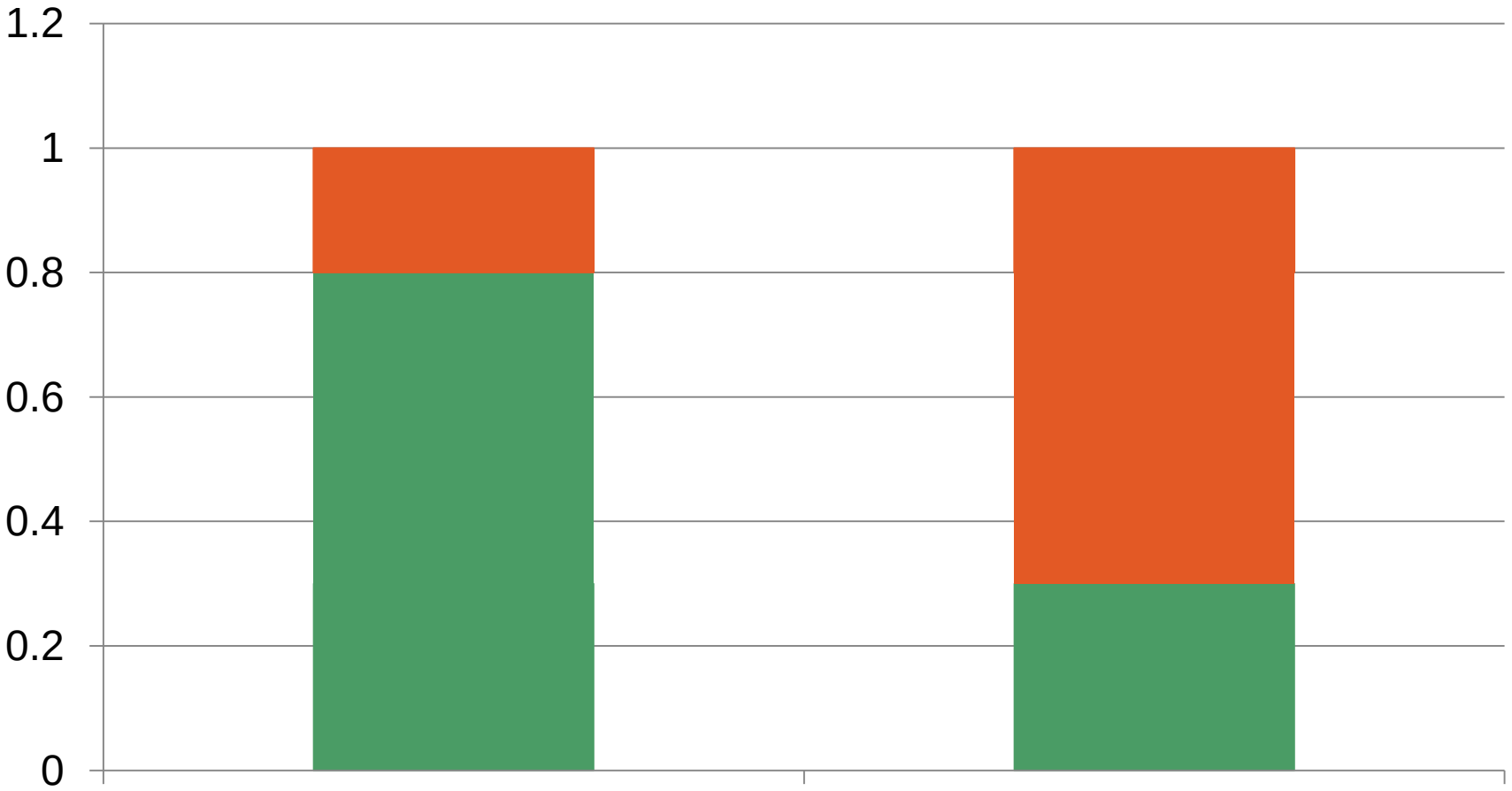
A-B 2.1

From ITT to effect of treatment on the treated (TOT)

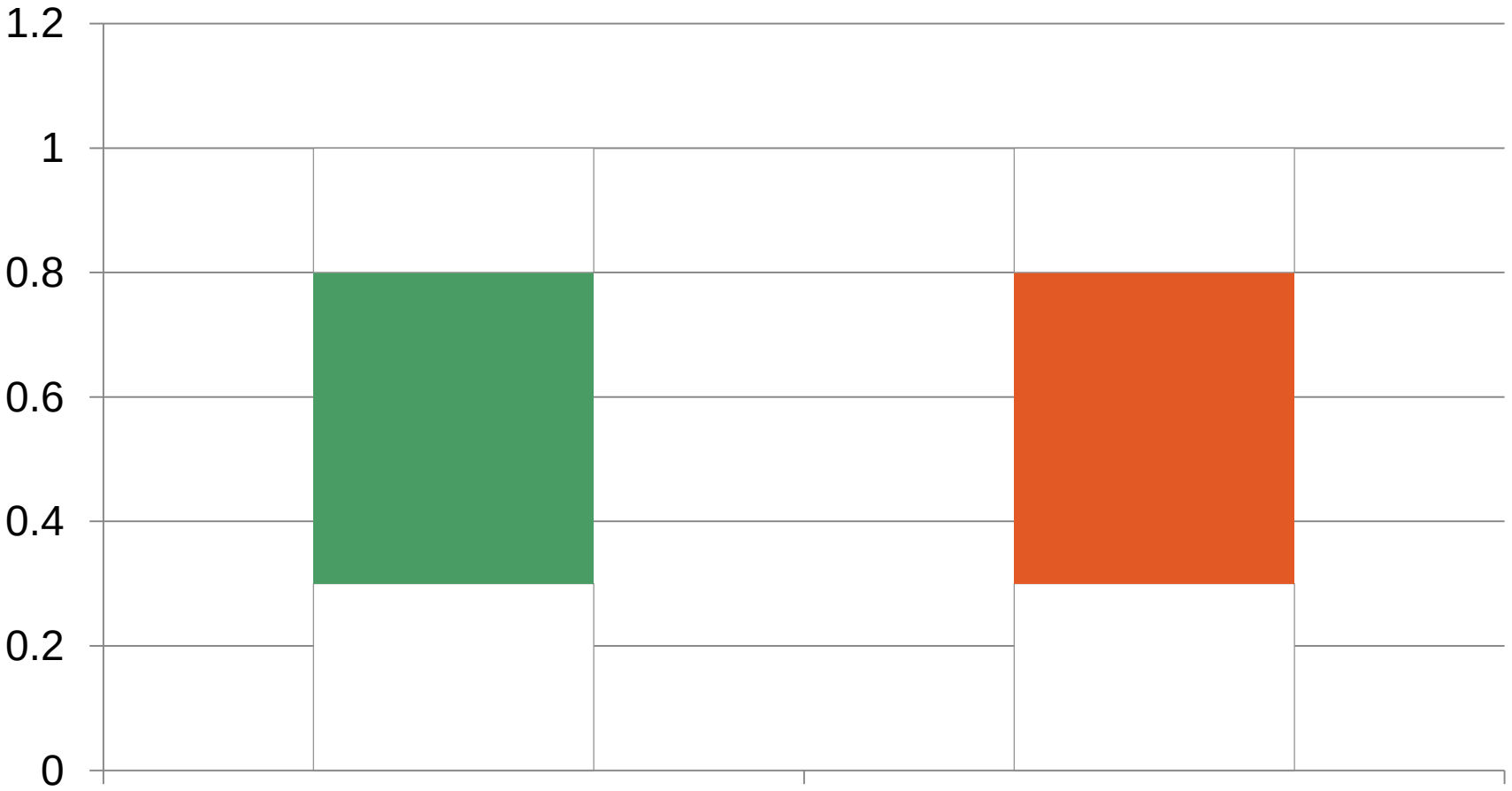
- The point is that if there is leakage across the groups, the comparison between those originally assigned to treatment and those originally assigned to control is smaller
- But the difference in the probability of getting treated is also smaller
- Formally this is done by "instrumenting" the probability of treatment by the original assignment

39

Estimating ToT from ITT: Wald



Interpreting ToT from ITT: Wald



Estimating TOT

- What values do we need?
- Y(T)
- Y(C)

- Prob[treated | T]
- Prob[treated | C]

Treatment on the treated (TOT)

- Starting from a simple regression model:
- $Y_i = a + B * S_i + e_i$
- [Angrist and Pischke, p. 67 show]:

$$B = \frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[s_i|z_i = 1] - E[s_i|z_i = 0]}$$

Treatment on the treated (TOT)

$$B = \frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[s_i|z_i = 1] - E[s_i|z_i = 0]}$$

$$\frac{Y(T) - Y(C)}{Prob[treated|T] - Prob[treated|C]}$$

TOT estimator

School 1	Intention to Treat ?	Treated?	Observed Change in weight
Pupil 1	yes	yes	4
Pupil 2	yes	yes	4
Pupil 3	yes	yes	4
Pupil 4	yes	no	0
Pupil 5	yes	yes	4
Pupil 6	yes	no	2
Pupil 7	yes	no	0
Pupil 8	yes	yes	6
Pupil 9	yes	yes	6
Pupil 10	yes	no	0
		Avg. Change Y(T)=	

School 2			
Pupil 1	no	no	2
Pupil 2	no	no	1
Pupil 3	no	yes	3
Pupil 4	no	no	0
Pupil 5	no	no	0
Pupil 6	no	yes	3
Pupil 7	no	no	0
Pupil 8	no	no	0
Pupil 9	no	no	0
Pupil 10	no	no	0
		Avg. Change Y(C) =	

A = Gain if Treated
B = Gain if not Treated

ToT Estimator: A-B

$$\textbf{A-B} = \frac{Y(T)-Y(C)}{Prob(Treated|T)-Prob(Treated|C)}$$

Y(T)
Y(C)
Prob(Treated|T)
Prob(Treated|C)

Y(T)-Y(C)
Prob(Treated|T)-Prob(Treated|C)

A-B

TOT estimator

School 1	Intention to Treat ?	Treated?	Observed Change in weight
Pupil 1	yes	yes	4
Pupil 2	yes	yes	4
Pupil 3	yes	yes	4
Pupil 4	yes	no	0
Pupil 5	yes	yes	4
Pupil 6	yes	no	2
Pupil 7	yes	no	0
Pupil 8	yes	yes	6
Pupil 9	yes	yes	6
Pupil 10	yes	no	0
		Avg. Change Y(T)=	3

School 2			
Pupil 1	no	no	2
Pupil 2	no	no	1
Pupil 3	no	yes	3
Pupil 4	no	no	0
Pupil 5	no	no	0
Pupil 6	no	yes	3
Pupil 7	no	no	0
Pupil 8	no	no	0
Pupil 9	no	no	0
Pupil 10	no	no	0
		Avg. Change Y(C) =	0.9

A = Gain if Treated
B = Gain if not Treated

ToT Estimator: A-B

$$\textbf{A-B} = \frac{Y(T)-Y(C)}{Prob(Treated|T)-Prob(Treated|C)}$$

Y(T)	3
Y(C)	0.9
Prob(Treated|T)	60%
Prob(Treated|C)	20%
Y(T)-Y(C)	2.1
Prob(Treated|T)-Prob(Treated|C)	40%
A-B	5.25

Generalizing the ToT Approach: Instrumental Variables

1. First stage regression:

$$T_{Actual} = \alpha_0 + \alpha_1 T_1 + \alpha_i X_i + e$$

2. Predict treatment status using estimated coefficients

$$\hat{T}_{predicted} = \hat{a}_0 + \hat{a}_1 T_1 + \hat{a}_i X_i$$

3. Regress outcome variable on predicted treatment status

$$Y_i = \beta_0 + \beta_1 \hat{T}_{predicted} + \beta_X X_i + \varepsilon$$

4. $\hat{\beta}_1$ gives treatment effect

Requirements for Instrumental Variables

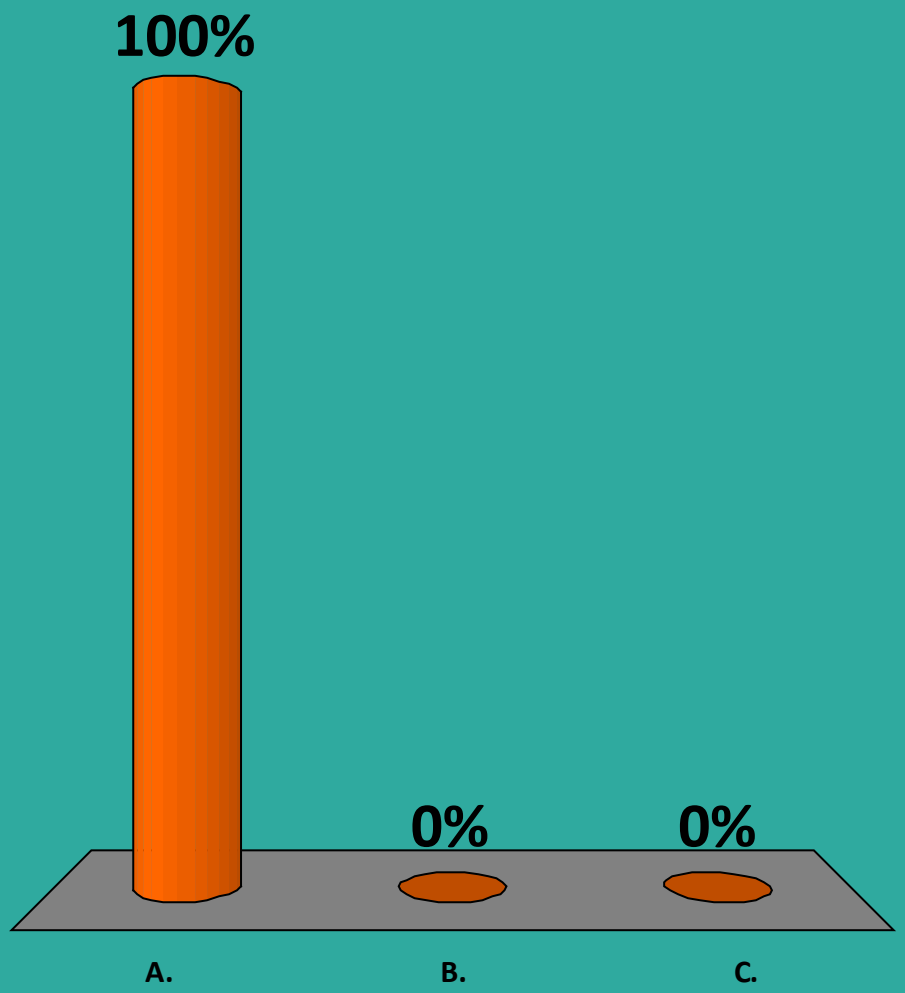
- First stage
  - Your experiment (or instrument) meaningfully affects probability of treatment
- Exclusion restriction
  - Your experiment (or instrument) does not affect outcomes through another channel

The ITT estimate will always be smaller (e.g., closer to zero) than the ToT estimate

A. True

B. False

C. Don't Know



TOT not always appropriate…

- Example: send 50% of MIT staff a letter warning of flu season, encourage them to get vaccines
- Suppose 50% in treatment, 0% in control get vaccines
- Suppose incidence of flu in treated group drops 35% relative to control group
- Is (.35) / (.5 – 0) = 70% the correct estimate?
- What effect might letter alone have?

Lecture Overview

- Attrition
- Spillovers
- Partial Compliance and Sample Selection Bias
- Intention to Treat & Treatment on Treated
- Choice of outcomes
- External validity

Multiple outcomes

- Can we look at various outcomes?
- The more outcomes you look at, the higher the chance you find at least one significantly affected by the program
  - Pre-specify outcomes of interest
  - Report results on all measured outcomes, even null results
  - Correct statistical tests (Bonferroni)

Covariates

- Why include covariates?
  - May explain variation, improve statistical power
- Why not include covariates?
  - Appearances of “specification searching”
- What to control for?
  - If stratified randomization: add strata fixed effects
  - Other covariates

Rule: Report both “raw” differences and regression-adjusted results

Lecture Overview

- Attrition
- Spillovers
- Partial Compliance and Sample Selection Bias
- Intention to Treat & Treatment on Treated
- Choice of outcomes
- External validity

Threat to external validity:

Behavioral responses to evaluations

Generalizability of results

Threat to external validity: Behavioral responses to evaluations

- One limitation of evaluations is that the evaluation itself may cause the treatment or comparison group to change its behavior
  - Treatment group behavior changes: Hawthorne effect
  - Comparison group behavior changes: John Henry effect
- Minimize salience of evaluation as much as possible
- Consider including controls who are measured at end-line only

Generalizability of results

- Depend on three factors:
  - Program Implementation: can it be replicated at a large (national) scale?
  - Study Sample: is it representative?
  - Sensitivity of results: would a similar, but slightly different program, have same impact?

Conclusion

- There are many threats to the internal and external validity of randomized evaluations…
- …as are there for every other type of study
- Randomized trials:
  - Facilitate simple and transparent analysis
    - Provide few "degrees of freedom" in data analysis (this is a good thing)
  - Allow clear tests of validity of experiment

Further resources

- Using Randomization in Development Economics Research: A Toolkit (Duflo, Glennerster, Kremer)
- Mostly Harmless Econometrics (Angrist and Pischke)
- Identification and Estimation of Local Average Treatment Effects (Imbens and Angrist, Econometrica, 1994).