J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

Threats and Analysis

Course Overview

1. Why Evaluate
2. Theory of Change & Measurement
3. Why & When to Randomize
4. How to Randomize
5. Sample Size & Power
6. Ethical Considerations for Randomized Evaluations
7. Threats & Analysis
8. Randomized Evaluation from Start to Finish
9. Applying & Using Evidence
10. The Generalizability Framework

Introduction

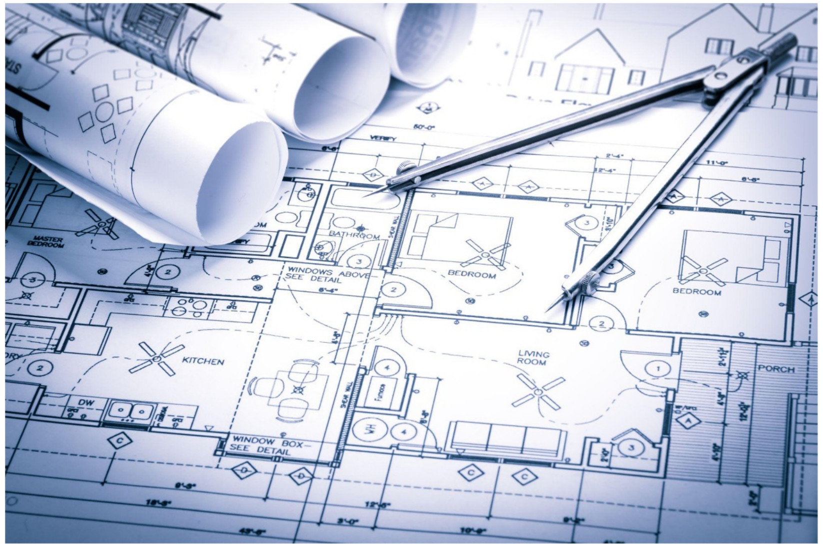


Photo credit: Shutterstock.com

During the **conception phase,** we design an evaluation that enables us to answer our research questions

Photo credit: Shutterstock.com

But the **implementation phase** of the evaluation is also extremely important: many things can go wrong

Learning Objectives

- Identify the main **threats to validity** that can arise while implementing an intervention and evaluation
  - Main focus on internal validity (whether the estimated impact reflects a causal relationship between the treatment and the outcome)
- Discuss strategies to mitigate these threats during the **implementation phase**
- Learn some strategies to account for threats during the **analysis phase**

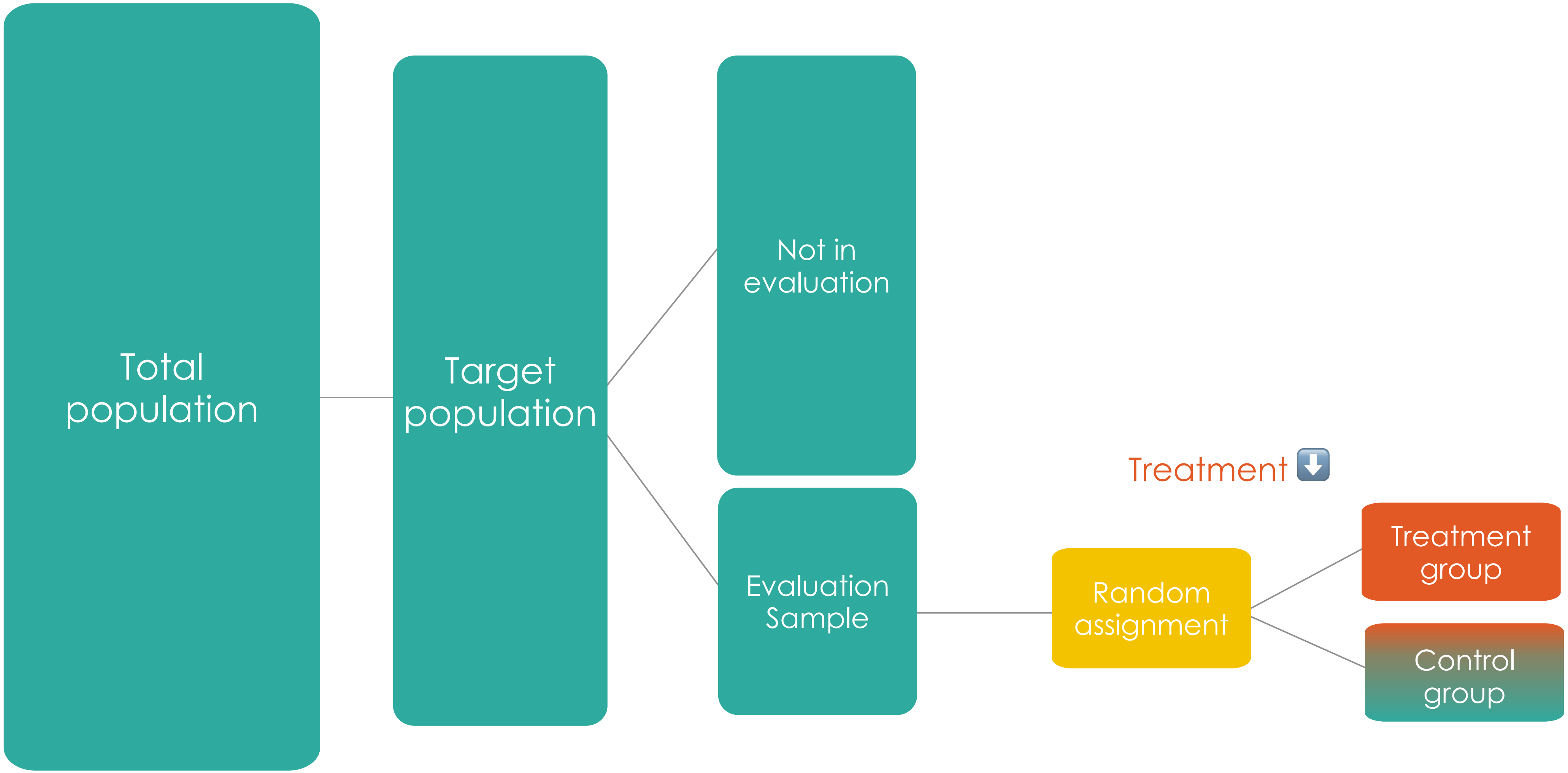
Lecture Overview

- Threats to validity
  - Spillovers
  - Attrition
  - Evaluation-driven Effects
  - Partial Compliance
- Generating impact estimates
  - Intention to Treat
  - Local Average Treatment Effect
  - Reporting Results

Lecture Overview

- Threats to validity
  - Spillovers
  - Attrition
  - Evaluation-driven Effects
  - Partial Compliance
- Generating impact estimates
  - Intention to Treat
  - Local Average Treatment Effect
  - Reporting Results

Reminder from Lecture 4: Spillovers



Spillovers

Spillovers occur when the outcomes of untreated units are indirectly affected by the treatment given to others.

- Spillovers violate the key assumption that one unit's treatment assignment has no effect on the outcomes of other units
- Spillovers are not limited to subjects in the study sample, but can affect anyone who is not treated
- Common causes: geographic proximity, social networks
- Make it difficult or impossible to measure the impact of the program
  - Comparison group no longer serves as a valid estimate of the counterfactual

Spillovers - Outcomes

- Spillovers may not put a study in jeopardy if they are contained or measured, but are problematic if they affect the comparison group.
- Spillovers can be positive or negative.
  - Positive spillovers: the comparison group benefits from the treatment group.
  - Negative spillovers: the comparison group is harmed by the treatment group.
- Spillovers can cause impact to be underestimated or overestimated.
- Channels through which spillovers occur include physical, marketwide or general equilibrium, informational, and behavioral.

Physical Spillover

Example: A member of the treatment group receives the transfer and gives some of the money to friends or relatives who are assigned to the control group

Control group benefits from treatment group

Treatment

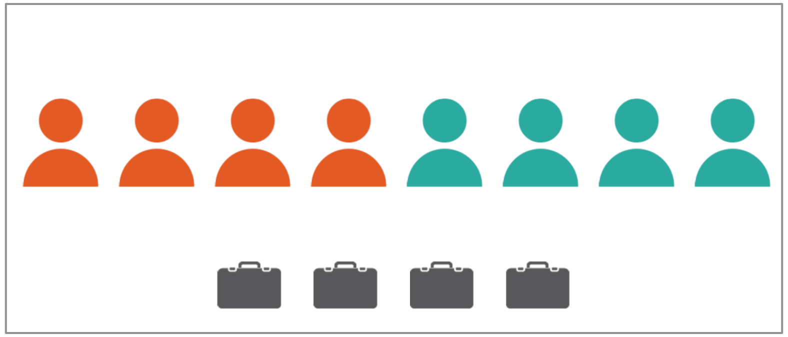
Treatment group

Control group

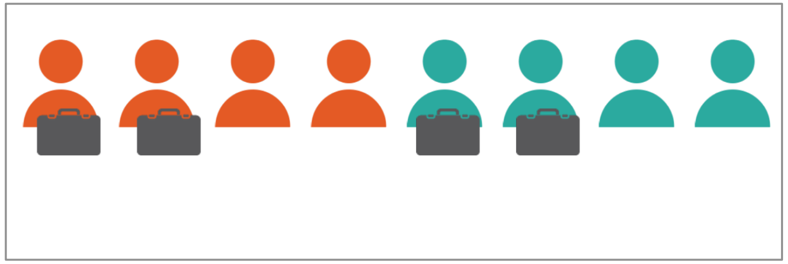
Marketwide/General Equilibrium Effects Spillover

Example: Displacement effects from job training programs

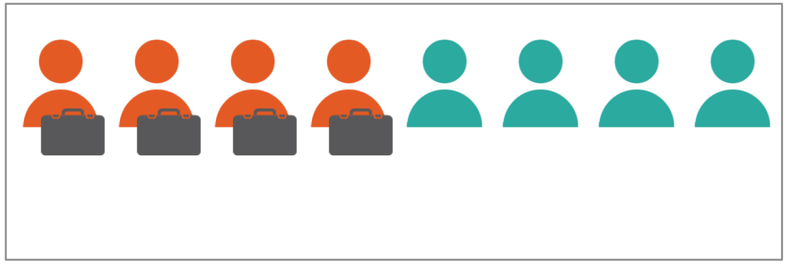
- Evaluations of job training programs traditionally compare employment outcomes between those who were trained (treatment) and those in the same area/population who were eligible but not trained (control)
- This does not take into account the possibility that the control group could be harmed if jobs are limited and treatment/control are in competition



Limited number of jobs available



Without intervention



With intervention (if displacement occurs)

Treatment group

Control group

Control group is harmed by treatment group

Appendix: More details

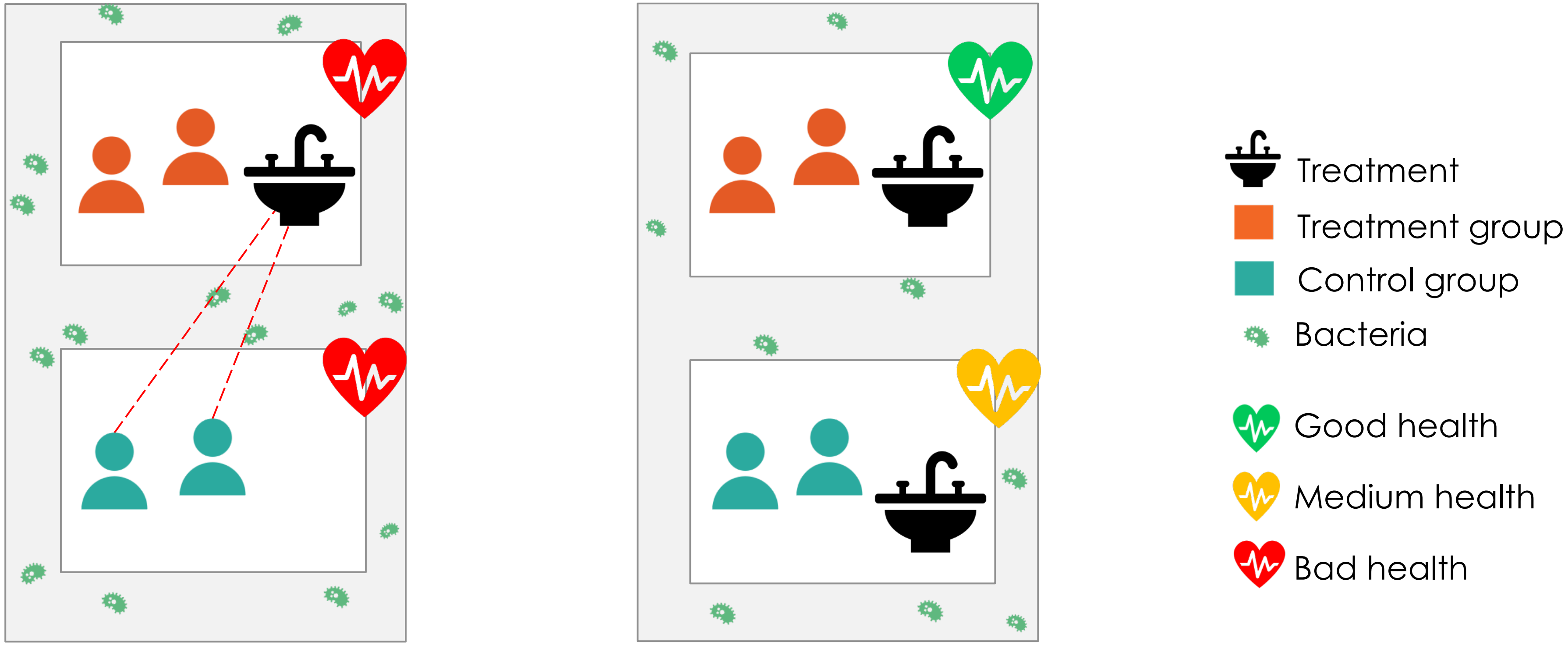
Behavioral/Informational Spillover

Example: Control group imitates neighbors' hygiene practices or learns about the health benefits

Level of randomization: household

Control group sees intervention

...and improves hygiene practices



Control group benefits from treatment group

What can be done about spillovers?

(1) Avoid spillovers

- Incorporate spatial buffers between treatment and control units
- Choose level of randomization wisely, and randomize at a higher level if concerned about spillovers

(2) Measure spillovers

- Build plans to collect data on spillovers into the experimental design
- Measure spillovers in the analysis phase

Thought exercise: Measuring informational spillovers

Imagine you are designing a randomized evaluation of a television program that features educational storylines about HIV/AIDs to understand the impact on viewers' knowledge, attitudes, and behaviors.

- **How could you design the evaluation to measure knowledge and behavior changes for viewers of the program—as well as the potentially positive informational spillovers to peers within their social networks?**

To learn more about the results of an HIV/AIDs edutainment intervention in Nigeria, **see appendix** and Banerjee, La Ferrara, and Orozco (2019), "The Entertaining Way to Behavioral Change: Fighting HIV with MTV."

Lecture Overview

- Threats to validity
  - Spillovers
  - Attrition
  - Evaluation-driven Effects
  - Partial Compliance
- Generating impact estimates
  - Intention to Treat
  - Local Average Treatment Effect
  - Reporting Results

Attrition

Attrition occurs when study group members leave the study and data on their outcomes cannot be collected.

Discussion question: Why is it a problem if some of the people in the experiment leave the study before you finish collecting your data? Why might we expect this to happen?

- It may be a problem depending on how much of the study sample we lose
- It is a problem if the type of people who disappear is correlated with the treatment
- Common drivers of attrition include mobility or migration, motivation, and mortality

Attrition Bias Example: A School Feeding Program

- Malnutrition can affect children's ability to consistently attend school.
- Imagine you start a school feeding program and want to do an evaluation.
  - You have a treatment and a control group
- Measure the program's effects on child growth (e.g., weight of children)
- You go to all the schools (treatment and control) and weigh everyone who is in school on a given day.

Photo credit: CatherineLProd, Shutterstock.com

Student weight in kilograms before and after a school feeding program					
	Before Treatment			**After Treatment**	
	T	C		T	C
	20	20		22	20
	25	25		27	25
	30	30		32	30
Avg.					
	Difference:			Difference:	

Student weight in kilograms before and after a school feeding program					
	Before Treatment			**After Treatment**	
	T	C		T	C
	20	20		22	20
	25	25		27	25
	30	30		32	30
Avg.	25	25		27	25
	Difference:	0		Difference:	2

What if Only Children > 21 Kg Come to School?

Student weight in kilograms before and after a school feeding program					
	Before Treatment			After Treatment	
	T	C		T	C
	[absent]	[absent]		22	[absent]
	25	25		27	25
	30	30		32	30
Avg.	27.5	27.5		27	27.5
	Difference:	0		Difference:	-0.5

What if Only Children > 21 Kg Come to School?

Student weight in kilograms before and after a school feeding program					
	Before Treatment			After Treatment	
	T	C		T	C
	[absent]	[absent]		22	[absent]
	25	25		27	25
	30	30		32	30
Avg.	27.5	27.5		27	27.5
	Difference:	0		Difference:	-0.5

A. You will underestimate the impact
B. You will overestimate the impact
C. Neither
D. Ambiguous
E. Don’t know

What can be done about attrition?

Implementation phase

- More intensive follow-up efforts with survey respondents
  - Account for follow-up costs in project planning and funding
  - For example: Follow-up visits and tracking of respondents who moved to neighboring areas

Analysis phase

- Use bounded estimates to mitigate the effects of attrition on impact estimates
  - Bounded estimates: take the percentage difference between treatment and comparison and drop the top percentile and bottom percentile from the group with less attrition to bound the estimates, creating worst case and best case scenarios

When is attrition NOT a problem?

A. When the attrition rates are similar in both treatment and control groups
B. When the estimated treatment effect is zero (among those who remain in the study)
C. When the true treatment effect is zero
D. None of the above

Lecture Overview

- Threats to validity
  - Spillovers
  - Attrition
  - Evaluation-driven Effects
  - Partial Compliance
- Generating impact estimates
  - Intention to Treat
  - Local Average Treatment Effect
  - Reporting Results

Evaluation-driven effects

Evaluation-driven effects occur when respondents change their behavior in response to the evaluation itself instead of the intervention.

Common causes: salience of being evaluated, social pressure

These include observer-driven effects and enumerator effects

- Hawthorne effects: Behavior changes due to attention from the study or intervention
- Anticipation effects: Comparison group changes behavior because they expect to receive the treatment later (particular concern for phase-ins)
- Resentment / demoralization effects: Comparison group resents missing out on treatment and changes behavior
- Demand effects: Behavior changes due to perceptions of evaluator's objectives
- Survey effects: Being surveyed changes subsequent behavior

What can be done about evaluation-driven effects?

- Use a **different level** of randomization
- **Minimize salience** of evaluation as much as possible
  - Do not announce phase-in
    - Downside is that this can be useful to reduce attrition!
  - Make sure staff is impartial and treats both groups similarly
    - E.g., blind data collection staff to treatment arm
- Measure the evaluation-driven effects in a **subset** of the sample
  - Prime a subset of the sample by reminding them of the evaluation (Mummolo and Peterson 2019)
  - Supplement survey data with other measures of behavioral outcomes (Fearon, Humphreys, and Weinstein 2008)

Thought exercise: Feedback to teachers

Imagine you are designing a randomized evaluation of a program that provides feedback to teachers (based on students' testing performance) to help understand the impact on teacher effort and ultimately student learning outcomes. However, classroom observation and the presence of enumerators to measure teacher activity may drive teachers' behavior, rather than the treatment itself.

How could you disentangle program effects from potential Hawthorne effects?

Reminder: Hawthorne effects are behavior changes due to attention of the study or intervention.

To learn more about the results of a teacher feedback intervention in India, **see appendix** and Muralidharan and Sundararaman (2010), "The Impact of Diagnostic Feedback to Teachers on Student Learning."

Lecture Overview

- Threats to validity
  - Spillovers
  - Attrition
  - Evaluation-driven Effects
  - Partial Compliance
- Generating impact estimates
  - Intention to Treat
  - Local Average Treatment Effect
  - Reporting Results

Partial Compliance and Sample Selection Bias

Noncompliance occurs when a unit's treatment assignment (assigned to treatment or comparison group) does not match their treatment status (received or did not receive the treatment).

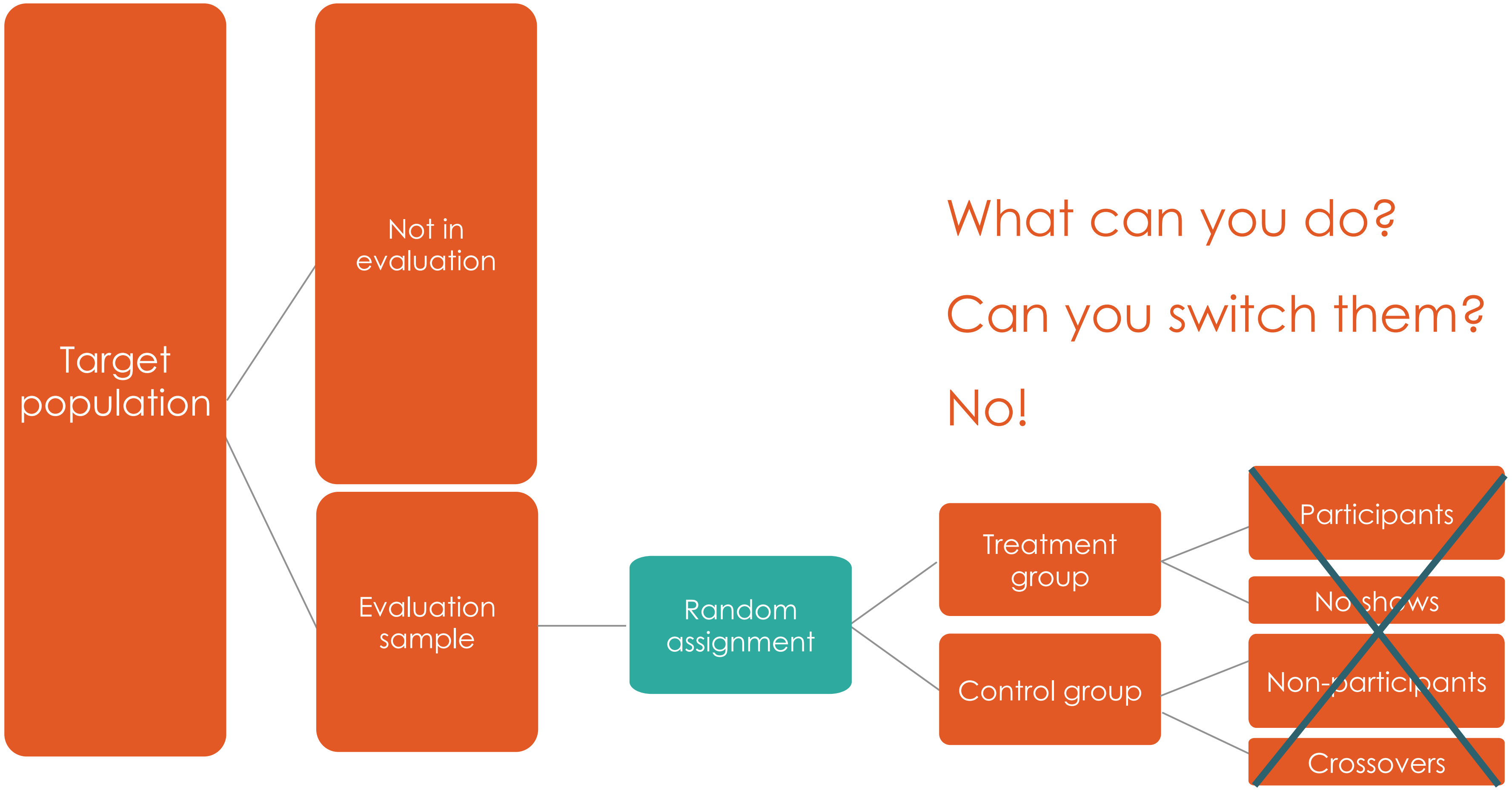
- ➔ Individuals assigned to treatment group may not receive the program
- ➔ Individuals assigned to the comparison group may access the treatment
- ➔ Can be due to project **implementers** or the **participants** themselves

When some of the participants are noncompliant, we say there is **partial compliance.**

Noncompliance can lead to **sample selection bias** and threaten internal validity if not properly accounted for in analysis.

- ➔ **Selection bias** occurs when individuals who receive or opt into the program are systematically different from those who do not.

Non-compliers

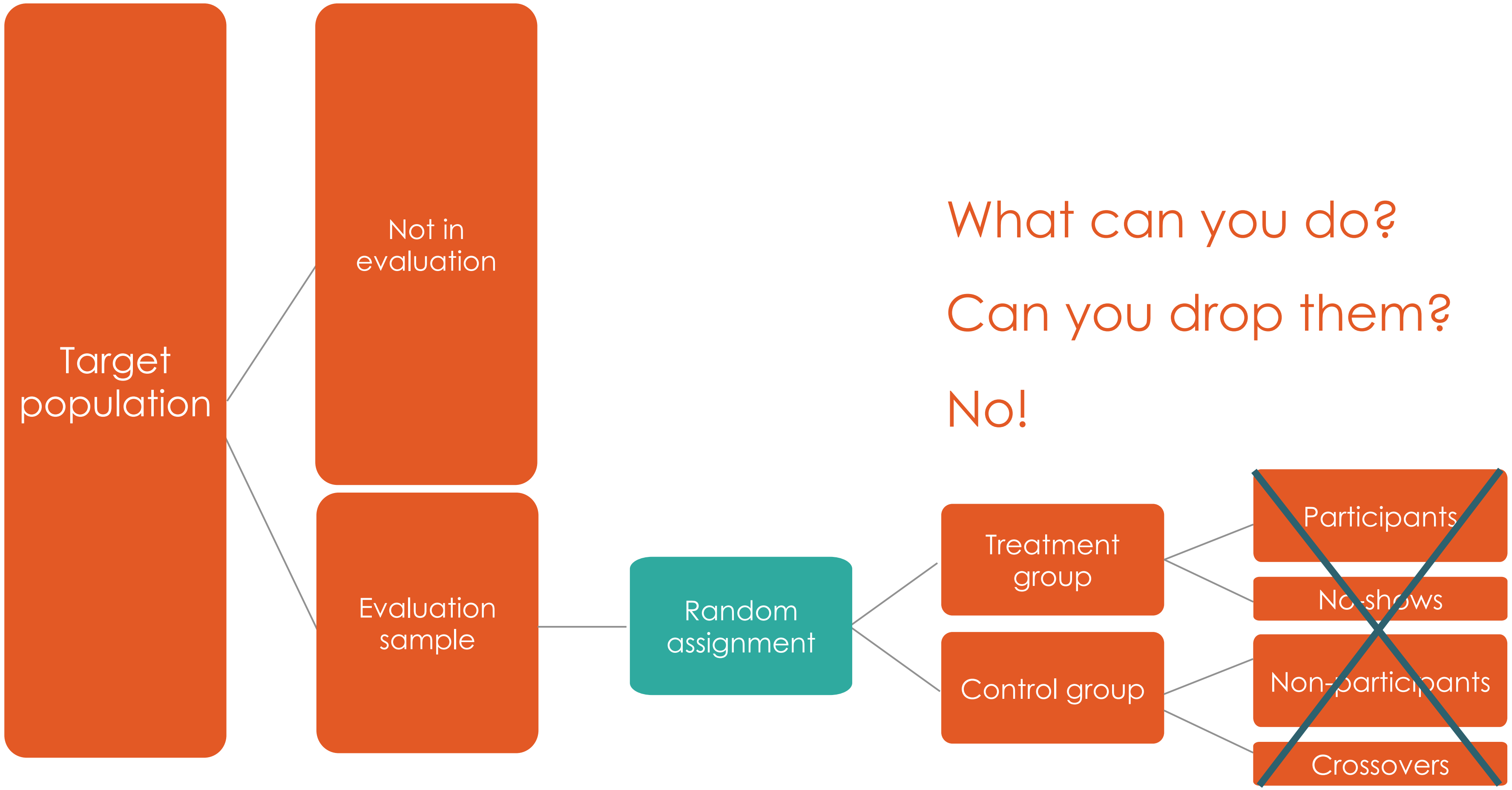


What can you do?

Can you switch them?

No!

Non-compliers

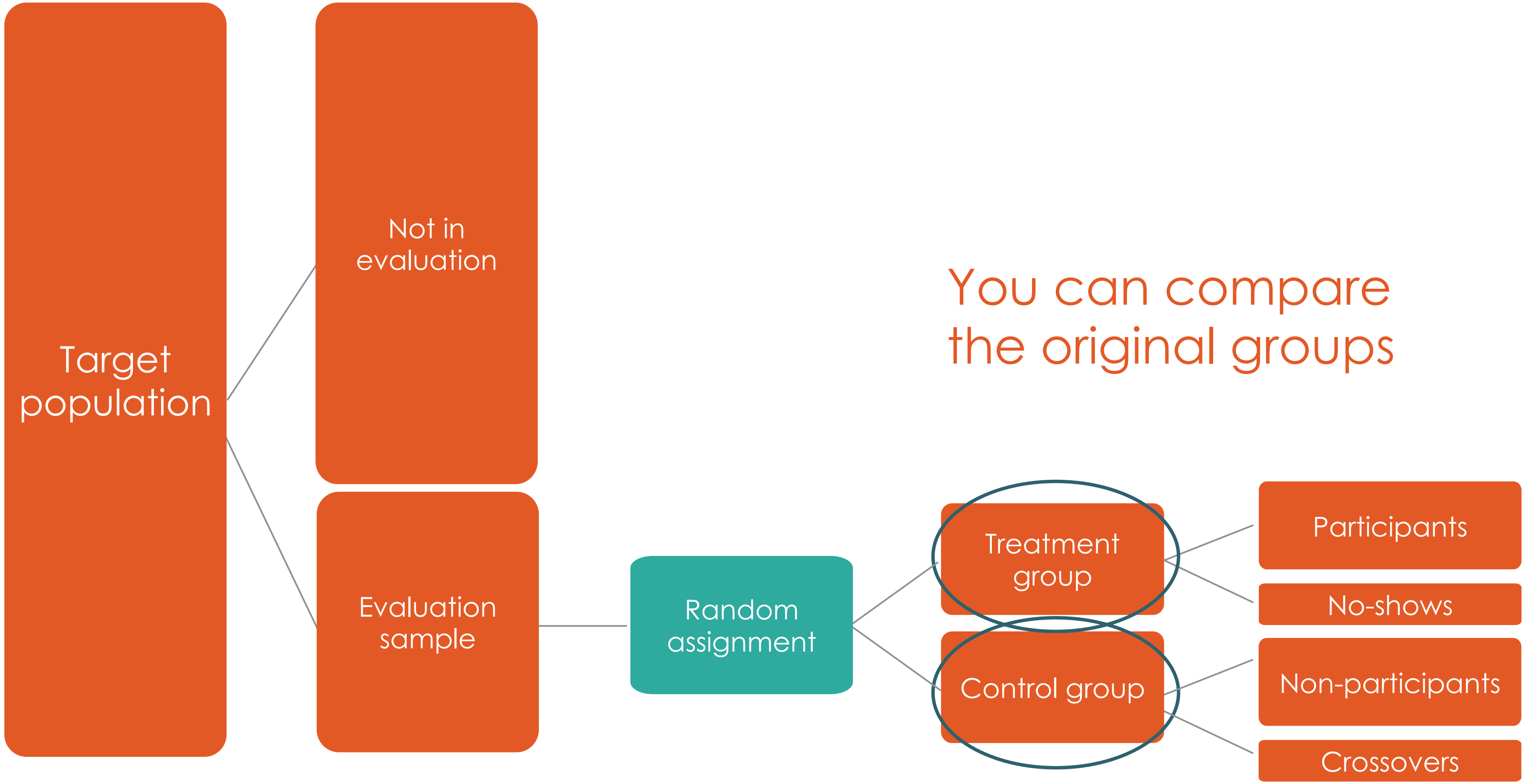


What can you do?

Can you drop them?

No!

Non-compliers



You can compare the original groups

Based on what we just discussed, our treatment group for analysis is…

A. Individuals assigned to treatment who were *actually* treated

B. All individuals who were *actually* treated

C. Individuals assigned to treatment, regardless of whether or not they were treated

D. Don't know

Example: Measuring Take-Up

Fazzio et al. (2021) study the impact of an alternative to government-run primary schools in isolated rural areas in Guinea-Bissau:

- The intervention provides 4 years of primary education classes
- Randomized at the village level: comparison villages continue with existing school options, and treatment villages receive the intervention

Do enrolled children in intervention villages attend classes?

Attendance level	Percent of students in treatment villages
Mean attendance	**85.72%**
Attend 0% of classes	9.27%
Attend >0 to 25% of classes	1.24%
Attend >25 to 50% of classes	2.32%
Attend >50 to 75% of classes	2.01%
Attend >75% to 100% of classes	85.16%

Example: Measuring Take-Up

Discussion question: What steps would you take in the design or implementation phases of the program to maximize take-up of the intervention?

Attendance level	Percent of students in treatment villages
Mean attendance	**85.72%**
Attend 0% of classes	9.27%
Attend >0 to 25% of classes	1.24%
Attend >25 to 50% of classes	2.32%
Attend >50 to 75% of classes	2.01%
Attend >75% to 100% of classes	85.16%

What can be done about noncompliance?

Implementation phase

Prevent noncompliance during design or implementation phase, for example by making take-up of the program easy, incentivizing take-up, or randomizing at a higher level

=> cannot always be done

Monitor it during implementation phase

=> important to be aware that it happens

Analysis phase

Interpret it during analysis phase

=> see next section

Lecture Overview

- Threats to validity
  - Spillovers
  - Attrition
  - Evaluation-driven Effects
  - Partial Compliance
- Generating impact estimates
  - Intention to Treat
  - Local Average Treatment Effect
  - Reporting Results

After-school supplementary lessons program

Photo: Students and teacher in the Gambia | Alex Eble

- Let's take the example of an after-school lessons program
- Some villages receive the program, some don't (random assignment)
- But enrolled children in treatment villages do not always attend the classes
  - and some children in control villages find a way to attend anyways!

Intention to Treat (ITT)

- Easiest way to deal with partial compliance - calculate the Intent to Treat (ITT):
  - The difference between the average outcome of the group that was randomly assigned to treatment and the group that was randomly assigned to comparison, *regardless of whether they actually received the treatment.*

ITT = (avg. outcome in group assigned to treatment) - (avg. outcome in group assigned to control)

- What does "intention to treat" measure?

"What happened to the average child in a treated village in this population?"

- Is this difference the causal effect of the intervention?

Village 1: Treatment

	Assigned to treatment	Treated	Change in reading score (in points)
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

Village 2: Control

	Assigned to treatment	Treated	Change in reading score (in points)
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	1
Pupil 10	No	No	0

Village 1: Treatment

	Intention to treat?	Treated	Change in reading score (in points)
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

Village 2: Control

	Assigned to treatment	Treated	Change in reading score (in points)
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	1
Pupil 10	No	No	0

Mean change in reading score:

Treated pupils in Village 1	4.67
Non-treated pupils in Village 2	0.5
Difference	4.17

Effect of treatment on reading scores?

NOT CORRECT!

Village 1: Treatment

	Intention to treat?	Treated	Change in reading score (in points)
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

Village 2: Control

	Assigned to treatment	Treated	Change in reading score (in points)
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	1
Pupil 10	No	No	0

The *Intent to Treat* estimate:	
Mean in village 1 :	3.0
Mean in village 2 :	1.0
Difference:	2.0

Village 1: Treatment

	Intention to treat?	Treated	Change in reading score (in points)
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

Village 2: Control

	Assigned to treatment	Treated	Change in reading score (in points)
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	1
Pupil 10	No	No	0

The *Intent to Treat:*	
Mean in village 1 :	3.0
Mean in village 2 :	1.0
Difference:	2.0

Treatment Probability:	
Fraction treated in village 1:	0.6
Fraction treated in village 2:	0.2
Difference:	0.4

Lecture Overview

- Threats to validity
  - Spillovers
  - Attrition
  - Evaluation-driven Effects
  - Partial Compliance
- Generating impact estimates
  - Intention to Treat
  - Local Average Treatment Effect
  - Reporting Results

Local Average Treatment Effect (LATE)

- In general, the Local Average Treatment Effect (LATE) is:

$$\textbf{LATE} = \frac{\textbf{ITT}}{\textbf{(take-up in treatment group) - (take-up in control group)}}$$

- What does the LATE estimate?

  The effect of the program on those who complied with their treatment status

- Note: Effects on those people who didn't take it up might have been quite different
- Very similar: "Treatment on the Treated" (TOT)

Local Average Treatment Effect (LATE)

- The intuitive idea:
  - Let's say the ITT effect of after-school lessons is a 3 point test score difference between treatment and control villages.
  - But only 50% of the children in the treatment villages actually went to the classes (let's assume no children in control schools got the classes).

- If the effect of 50% take-up is to increase scores by 3 points, then we can say that if everyone were to take the classes, the effect would be:

$$\text{LATE} = \frac{\text{ITT}}{\text{(take-up in treatment group) - (take-up in control group)}}$$

$$\text{LATE} = \frac{3}{(0.5) - (0)} = 3 * 2 = 6 \text{ points}$$

Village 1: Treatment

	Assigned to treatment	Treated	Change in reading score (in points)
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

Village 2: Control

	Assigned to treatment	Treated	Change in reading score (in points)
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	1
Pupil 10	No	No	0

The *Intent to Treat:*	
Mean in school 1 :	3.0
Mean in school 2 :	1.0
Difference:	2.0

Treatment Probability:	
Fraction treated in school 1:	0.6
Fraction treated in school 2:	0.2
Difference:	0.4

Local Average Treatment Effect:

$$2/.4 = 5$$

ITT vs LATE

If obtaining the estimate is easy, why not always use LATE?

- In order to estimate LATE we need data on compliance
- ITT may be the policy-relevant parameter of interest
  - 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.

ITT / LATE: Conclusions

- Both ITT and LATE can provide valuable information to decision-makers.
- LATE gives the effect of the intervention for those who comply with their assignment to treatment or control.
- ITT gives the overall effect of the intervention, admitting that noncompliance can happen (which is inherent to any policy).

Lecture Overview

- Introduction to threats:
  - Spillovers
  - Attrition
  - Behavioral Responses to Evaluations
  - Partial Compliance and Sample Selection Bias
- Generating impact estimates
  - Intention to Treat
  - Local Average Treatment Effect
  - Reporting Results

Reporting Results

Reporting bias occurs when the decision on whether and how to report impact estimates depends on the direction and significance of the estimate.

How researchers report their results can also threaten validity.

Potential sources of reporting bias:

- **Specification searching:** trying different analyses to find one that is statistically significant
  - The more outcomes and adjustments to covariates you look at, the higher the chance you find at least one significant effect.
- **File drawer problem:** significant results are more likely to be published

What can be done about reporting bias?

- Pre-specify outcomes of interest
  - Pre-analysis plans are becoming more common, and pre-specified analyses may be given more weight
  - Differentiate between pre-specified and exploratory analysis
- Report raw differences between treatment and control as well as regression estimates (adjusted based on covariates)
- Report all results, not just the most impressive or significant ones
- Share data and code along with research papers

Discussion question: What can we learn from a paper reporting no significant impact of the intervention on the outcome of interest?

Conclusions

- **Internal validity is a strength of well-designed randomized evaluations…**

  …so everything undermining it must be carefully considered
- **The design phase and project planning are important…**

  …but so is the ability to face challenges during implementation phase
- **Consider which threats are likely factors for a given evaluation…**

  …and plan to mitigate and monitor attrition, spillovers, partial compliance, and evaluation-driven effects
- **ITT and LATE are methods for generating impact estimates that can teach us different things about how a program works in context.**

Further Resources

- "Using Randomization in Development Economics Research: A Toolkit" (Duflo, Glennerster, and Kremer 2006)
- *Mostly Harmless Econometrics* (Angrist and Pischke 2008)
- "Identification and Estimation of Local Average Treatment Effects" (Imbens and Angrist 1994)
- *Impact Evaluation in Practice*, Chapter 9 (Gertler et al. 2016)
- "On Minimizing the Risk of Bias in Randomized Controlled Trials in Economics" (Eble, Boone, and Elbourne 2016)
- Duru, Maya, and Sarah Kopper. "Data analysis." J-PAL Research Resources. Accessed May 2023.

References

Banerjee, Abhijit, Eliana La Ferrara, and Victor H. Orozco-Olvera. 2019. "The entertaining way to behavioral change: Fighting HIV with MTV." No. w26096. National Bureau of Economic Research. https://www.nber.org/papers/w26096

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora. 2013. "Do labor market policies have displacement effects? Evidence from a clustered randomized experiment." The quarterly journal of economics 128, no. 2: 531-580. https://doi.org/10.1093/qje/qjt001

Duru, Maya, and Sarah Kopper. "Data analysis." J-PAL Research Resources. Accessed May 2023.

Eble, Alex, Peter Boone, and Diana Elbourne. "On minimizing the risk of bias in randomized controlled trials in economics." The World Bank Economic Review 31, no. 3 (2017): 687-707. https://doi.org/10.1093/wber/lhw034

Eble, Alex, Chris Frost, Alpha Camara, Baboucarr Bouy, Momodou Bah, Maitri Sivaraman, Pei-Tseng Jenny Hsieh et al. 2021. "How much can we remedy very low learning levels in rural parts of low-income countries? Impact and generalizability of a multi-pronged para-teacher intervention from a cluster-randomized trial in the Gambia." *Journal of Development Economics* 148 (2021): 102539. https://doi.org/10.1016/j.jdeveco.2020.102539

Fazzio, Ila, Alex Eble, Robin L. Lumsdaine, Peter Boone, Baboucarr Bouy, Pei-Tseng Jenny Hsieh, Chitra Jayanty, Simon Johnson, and Ana Filipa Silva. "Large learning gains in pockets of extreme poverty: Experimental evidence from Guinea Bissau." Journal of Public Economics 199 (2021): 104385. https://doi.org/10.1016/j.jpubeco.2021.104385

Fearon, James D., Macartan Humphreys, and Jeremy M. Weinstein. "Can development aid contribute to social cohesion after civil war? Evidence from a field experiment in post-conflict Liberia." American Economic Review 99, no. 2 (2009): 287-291. https://www.aeaweb.org/articles?id=10.1257/aer.99.2.287

J-PAL. "The Impact of Diagnostic Feedback for Teachers on Student Learning in India" J-PAL Evaluation Summary. 2010.

Mummolo, Jonathan, and Erik Peterson. 2019. "Demand effects in survey experiments: An empirical assessment." American Political Science Review 113, no. 2: 517-529. https://doi.org/10.1017/S0003055418000837

Muralidharan, Karthik, and Venkatesh Sundararaman. "The impact of diagnostic feedback to teachers on student learning: Experimental evidence from India." The Economic Journal 120, no. 546 (2010): F187-F203. https://doi.org/10.1111/j.1468-0297.2010.02373.x

Reuse and citation

To reference this lecture, please cite as:

J-PAL. "Lecture: Threats & Analysis." Abdul Latif Jameel Poverty Action Lab. 2023. Cambridge, MA

Appendix

Navigating challenges to RCT design

ESP WORKBOOK

1. Are you concerned about spillovers: i.e., that the treatment might indirectly affect participants assigned to the control group?

|Type answers here.

2. Are you concerned about non-compliance: i.e., when respondents do not comply with their treatment assignment?

|Type answers here.

3. Are you concerned about attrition bias: i.e., bias that arises when the probability of dropping out is related to treatment assignment?

|Type answers here.

4. Given the challenges you identified, how might you mitigate these challenges to the research design?

|Type answers here.

Measuring Marketwide/General Equilibrium Effects

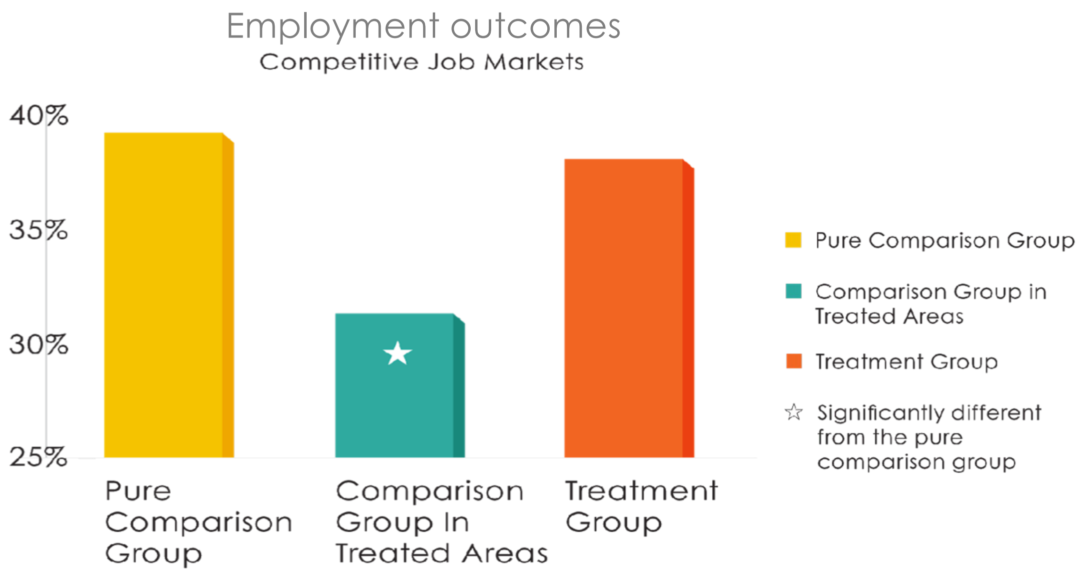
Example: Displacement effects from job training programs

- Crépon et al. (2012) evaluates the impact of a job placement program on unemployed populations across 235 labor markets in France
- Labor markets are randomly assigned to one of the following interventions:

- Study measures employment outcomes on treated groups AND control groups in treated areas

Next slide

General Equilibrium Effect: Untreated Job Seekers in Program Areas are Harmed by Treatment



Misleading comparison: versus

Ignoring the spillover effect, the study would have found that investing 100,000 euros into the job training program causes 9.7 people to find jobs within 8 months.

Better comparison: ver

Comparing the treatment group to a pure control group provides a better sense of the treatment effect. However, this still fails to account for the spillover.

Measuring the externality: versus

People living in areas with the job program that are not in the program have a harder time finding a job than people outside of those areas.

Total treatment effect: ver -

versus

When considering the spillover, the treatment is found to have no effect.

→ Return to Marketwide spillovers

MTV Shuga in in Nigeria

Example: Measuring informational spillovers

Study design: Banerjee et al. (2019) look at the effectiveness of edutainment programming on HIV/AIDs on knowledge, attitudes, and behaviors. Communities were randomly assigned to either:

- **Control group:** Placebo screening of TV series without educational message
- **Treatment group:** Screenings of MTV Shuga drama featuring educational storylines about HIV/AIDs (with additional random variation in social messages and who viewers watched with)
  - In addition to collecting data on study participants, researchers sampled a subset of their friends who did not attend screenings to measure informational spillovers

Photo: World Bank

Banerjee, La Ferrara, and Orozco (2019), "The Entertaining Way to Behavioral Change: Fighting HIV with MTV."
J-PAL Evaluation Summary: "MTV Shuga: Changing social norms and behaviors with entertainment education in Nigeria"

→ Return to measuring informational spillovers

Diagnostic tests and feedback to teachers in India

Example: Understanding Hawthorne effects

Study design: Muralidharan and Sundararaman (2010) assess whether providing low-stakes feedback to teachers leads to improved student learning outcomes. Schools were randomly assigned to either:

- **Control group:** No baseline test

  → Data collected during one unannounced visit

- **Treatment group:** Baseline test with results delivered to teachers one year later

  → Data collected during six unannounced visits over the course of the school year

  - Teachers in treatment schools seem to perform better on measures of teacher effort and activity, but no difference in student test scores across treatment and control at the end of the year, suggesting teachers only changed behavior while being observed

Photo: Robin Hayashi | J-PAL

Muralidharan and Sundararaman (2010), "[The Impact of Diagnostic Feedback to Teachers on Student Learning](#)."
J-PAL Evaluation Summary: "[The Impact of Diagnostic Feedback for Teachers on Student Learning in India](#)"

→ Return to evaluation-driven effects