

RURAL SANITATION IN INDONESIA

Understanding Threats to Experimental Integrity



PATIRO facilitator shares maternal neonatal health scorecards with community activists.
Photo: Jessica Creighton

This case study is based on: Cameron, Lisa, Susan Olivia, and Manisha Shah. 2019. **“Scaling Up Sanitation: Evidence from an RCT in Indonesia”**. *Journal of Development Economics*, 138: 1-16.

J-PAL thanks the authors for allowing us to use their paper as a teaching tool.

KEY VOCABULARY

Treatment assignment, treatment status	<p>An individual's treatment assignment is the group they were randomly assigned to: were they assigned to the treatment group or the comparison group?</p> <p>Note that whether a unit/individual actually receives the treatment will depend on compliance with their treatment assignment.</p>
Balance	<p>Randomization creates two groups that on average look very similar. This can be tested by collecting some baseline demographic information—such as age, gender, years of education, income, etc.—and comparing the average value of these characteristics in the treatment group to the average value of them in the comparison groups. Even when randomization is done correctly, some of these average values will be different; however, this reflects differences that occur by chance. We say the comparison and treatment groups balance if they have similar average values for baseline characteristics</p>
Selection Bias	<p>Selection bias occurs when individuals who receive or opt into the program are systematically different from those who do not. Consider an elective after school tutoring program. Is it effective at raising children's exam scores? If we compare those who take up the tutoring program to those who don't, we will get a biased estimate of the effect of the tutoring program, because those who chose to take it up are likely different from those who don't. The two groups likely are not balanced (for example, those who took it up may be more motivated, or they may have lower grades). Randomization minimizes selection bias because it breaks the link between characteristics of the individual and their treatment status. Selection bias can occur in other ways in a randomized evaluation. For example:</p> <ul style="list-style-type: none">- Participants can choose to take up a treatment or refuse it- Participants can choose to leave the study (i.e., attrit/attrition)
Attrition Bias	<p>Attrition bias is a type of selection bias that occurs when people leave the study. This can bias the estimate of the treatment impact in two ways:</p> <ol style="list-style-type: none">1. It may be the case that people with certain characteristics (say, those with the highest levels of education) in both the treatment and comparison groups leave. This means your study population looks less like the general population. The treatment effect you estimate might not represent the true effect for the general population.2. The reasons people leave may be correlated with the treatment. Suppose the students who have the most resources at home who are in the treatment group improve their performance and test into elite private schools, leaving your study sample. Then comparing treatment and comparison groups at endline would underestimate the impact of the program, because the students with the highest grades are 'missing' from the treatment group.
Compliance	<p>When a unit's treatment assignment (assigned to treatment or comparison group) matches their treatment status (took up or did not take up the treatment), we say they have complied.</p> <p>Any study sample can be split into three distinct groups:</p> <ol style="list-style-type: none">1. Compliers: This group of people will follow their assignment status. If they are assigned to the treatment group, they will take up the treatment; if they are assigned to the control group they will not take up the program.

	<ol style="list-style-type: none"> 2. Always-takers: This group of people will always take up the program, regardless of assignment status. 3. Never-takers: This group of people will never take up the program, regardless of assignment status. <p>When respondents do not comply with their treatment assignment, the study has partial compliance. In the treatment group, the people who do not comply are never-takers, while in the comparison group, those who do not comply are always-takers. We collectively refer to those who do not comply as non-compliers, and the action of not complying with treatment status as non-compliance.</p> <p>Note that when there is <i>two-sided non-compliance</i> (i.e., non-compliance in both the treatment and comparison group), we have to make the monotonicity assumption, which states that assignment to treatment does not dissuade someone from taking up the treatment (in which case, we would classify them as “defiers”).</p>
<p>Intention-to-Treat (ITT)</p>	<p>The ITT is a method for estimating the effect of the program where you compare the average outcomes of those assigned to the treatment group to the average outcomes of those assigned to the comparison group, regardless of whether individuals within those groups have actually received the treatment. The ITT measures the impact of delivering a program in the real world, where some people don't take up the program when offered it, and others take up the program even when they are not expressly encouraged to do so.</p>
<p>Local Average Treatment Effect (LATE)</p>	<p>The LATE is a method for estimating the effect of the program on those who complied with their treatment status. The LATE divides the ITT by the difference in the proportion of the treatment group who took up the program and the proportion of the comparison group who took up the program. Recall that the ITT compares the average outcome of the treatment group to that of the comparison group. This means that under partial compliance, the average changes we measure in the treatment group will be diluted by changes in outcomes among those who did not take it up. Intuitively, you should think of the LATE as a way of adjusting the ITT to reflect that not all of those assigned to treatment were treated while some who were assigned to the comparison group were treated.</p>

LEARNING OBJECTIVES

To explore how common threats to experimental integrity can influence the effect of a program.

SUBJECTS COVERED

Balance, attrition, selection bias, compliance, spillovers, intention-to-treat effect (ITT), local average treatment effect (LATE).

INTRODUCTION

As part of the Indonesian government's national strategy to improve environmental and health outcomes in rural areas and to address the dangers of fecal-borne illness linked to poor sanitation, the government of Indonesia rolled out the Community-Led Total Sanitation (CLTS) program.

Aimed to change community norms and create demand for sanitation, the CLTS program facilitates community discussions of the negative health consequences of existing sanitation practices. During the implementation, facilitators are sent to villages to initiate a community analysis of existing sanitation practices and a discussion of the negative health consequences of such practices. The community actively participates in the facilitated meeting and then forge their own plans to improve village sanitation with limited follow-up support and monitoring from facilitators. Discussions are held in public places and are open to all. CLTS facilitators aim to introduce a feeling of shame about open defecation during the discussions to strongly motivate behavior change and investment in sanitation facilities. In contrast to other approaches that have been used widely in the past in Indonesia and elsewhere, no funding for infrastructure or subsidies of any kind are provided.

The CLTS program was implemented across 29 rural districts in the province of East Java, Indonesia, intending to reach a total of 1.4 million people. In partnership with the Water and Sanitation Program of the World Bank, researchers conducted a randomized evaluation in eight of East Java's rural districts to study the impact of the large-scale CLTS program in Indonesia on sanitation practices, attitudes towards open defecation, and child health.

Random assignment was conducted at the village level, stratified by sub-district.¹ Within rural areas of East Java, 160 villages in eight districts (roughly 2,000 households) were randomly assigned either to receive the CLTS program (80 villages) or to a comparison group that did not receive CLTS (80 villages). A baseline survey was conducted in 2008 prior to implementation, followed by an endline survey two years after implementation in 2010 and 2011, and outcomes were measured at the household level.

This case study will take us through different threats to experimental integrity, including non-compliance, imbalance, and attrition, and spillovers, in the context of the evaluation of the CLTS program in Indonesia. It draws from the evaluation by Lisa Cameron, Susan Olivia, and Manisha Shah but incorporates hypothetical examples that did not occur in the actual study.

THREATS TO THE INTEGRITY OF THE PLANNED EXPERIMENT

DISCUSSION TOPIC 1: BALANCE BETWEEN GROUPS

Randomization creates groups that, on average, are “balanced,” meaning they look very similar in terms of their average age, gender composition, education levels, etc. However, even when randomization is done correctly, meaningful differences can occur by chance, especially when your sample size is small, and these differences can bias your results if not accounted for in your analysis. Moreover, as the experiment unfolds, external influences can reintroduce selection bias at the end of the program—people may migrate (potentially disproportionately in one group or another), we may find it harder to track and survey respondents in the comparison group, or people in the comparison group might be less willing to respond to an endline survey. These and other events can potentially reintroduce selection bias, diminishing the validity of the impact estimates, and are threats to the integrity of the experiment.

1. Can you check if the groups actually are balanced at the start of a program? How would you do so?

¹ In each district ten villages were randomly selected to participate in the impact evaluation as treatment villages and ten were randomly selected to act as control villages.

2. Can you check if the groups are balanced at the end of a program? How would this be different from checking in the beginning of the program?

DISCUSSION TOPIC 2: UNDERSTANDING ATTRITION

Attrition occurs when people drop out of the sample over the course of the experiment. Attrition is a concern for several reasons: First, attrition—whether in the treatment or comparison groups—reduces the sample size of the study, which makes it harder to detect the effect of the program.

Second, attrition can cause bias. This bias can arise when certain types of people leave the study (e.g., those who live furthest from the village center, those from the richest households, etc.). If a specific type of person leaves the study in *both* the treatment and comparison group, then the study sample looks less like the general population, meaning the results of the study are harder to generalize to the actual population.

More consequentially, if a specific type of person disproportionately leaves in either the treatment or comparison group, it reduces the balance of the two groups and introduces bias into the estimate of the treatment effect.

Suppose there are 2,000 households in our sample (1,000 in the treatment group and 1,000 in the comparison group). Suppose all of the households in the treatment group participate in the CLTS program, while none of the comparison group households do. The ownership of sanitation facilities for each group are shown for both baseline and endline in Table 1.

Table 1: Toilet ownership at baseline and endline				
Type of sanitation facilities	Baseline		Endline	
	Treatment	Comparison	Treatment	Comparison

Access to improved sanitation ²	490	460	640	460
Access to unimproved sanitation	110	140	150	140
No sanitation	400	400	210	400
Sample Size	1,000	1,000	1,000	1,000

1. Using the table above, calculate the following:
 - a. At baseline, what share of households have any sanitation facilities (either improved or unimproved) for each group?
 - b. At endline, what share of households have any sanitation facilities for each group?
 - c. What is the impact of the program on the ownership of any sanitation facilities?

Suppose now that in the treatment group, half of the households who have no sanitation facilities at the endline refused to respond to the endline survey. Sanitation facilities ownership for respondents in the endline sample in each group under this scenario is displayed in Table 2:

Table 2: Toilet ownership at baseline and endline with attrition in the comparison group				
Type of sanitation facilities	Baseline		Endline	
	Treatment	Comparison	Treatment	Comparison
Access to improved sanitation	490	460	640	460

² Improved sanitation facilities include: a) a flush toilet or latrine that flushes to a sewer, septic tank, or pit; b) a ventilated improved pit (VIP) latrine; c) a pit latrine with the pit well covered by a slab; or d) composting toilets. Shared and public toilets are considered “unimproved” regardless of their type.

Access to unimproved sanitation	110	140	150	140
No sanitation	400	400	105	400
Sample Size	1,000	1,000	895	1,000

2. Using the table above, calculate the following:
- What is the measured impact of the program on the ownership of any sanitation facilities?
 - Is this outcome difference between groups an accurate estimate of the impact of the program? Why or why not?
 - If it is not accurate, does it overestimate or underestimate the impact? By how much?
 - Does this threat of attrition only present itself in randomized evaluations?
 - Could you think of possible ways to mitigate attrition?
3. Suppose we have strong reason to believe that the true treatment effect on toilet construction is large, positive, and significant. How might the following scenarios influence our ability to accurately estimate the treatment effect?
- The top 20% of the sample in terms of household income—both comparison and treatment—leaves the study sample.

- b. In the treatment group, the households with a low initial level of community participation were not happy with the program as it did not provide subsidies/funding. They exit the study sample by refusing to answer the endline survey.

DISCUSSION TOPIC 3: PARTIAL COMPLIANCE

In the study of CLTS in Indonesia, random assignment determined which villages received the program. However, not all of the treatment group villages followed through by delivering the program.

The endline survey data indicates that 66 percent of the 80 treatment villages received the program and that 13.8 percent of the 80 control villages were exposed to the program.³

In this section, we will examine the consequences of partial compliance and how to prevent or minimize this. For the purposes of this discussion, we will focus primarily on compliance of villages with their treatment assignment and consider households’ treatment status to reflect their village’s treatment status.⁴

Table 3a: Toilet Ownership by Treatment Assignment		
Group	Toilet Ownership	Number of households
Treatment	79%	1,000
Comparison	60%	1,000

³ Cameron, Olivia, and Shah (2019) note that “Non-compliance was largely a result of district governments changing some of their target communities after the randomization plan had been agreed upon.”

⁴ Another consideration that intersects with compliance would be the reach of the CLTS program within treatment (and comparison) villages and how many households were exposed to the program directly.

1. Imagine you compare toilet ownership of those *assigned to the treatment group* to those *assigned to the comparison group*, regardless of the treatment status of the households within those groups. What is the impact of the treatment?

2. Through the endline survey data, the researchers could identify which villages actually implemented the CLTS program, and found that 13.8% of the comparison group villages received the CLTS program. Meanwhile, 34% of the treatment group villages did not carry out the CLTS program. Thus, some of the comparison households participated in community discussions of the negative health consequences of existing sanitation practices regardless of their treatment assignment:

Table 3b: Treatment Assignment vs. Treatment Status for Households in the Sample			
	Treatment assignment:		
Treatment status:	Treatment	Comparison	Total:
Village exposed to the CLTS program	660	138	798
Village not exposed to the CLTS program	340	862	1,202
Total:	1,000	1,000	2,000

- a. Some of your colleagues are passing by your desk; they all agree that you should calculate the effect of the treatment by comparing the 660 households in villages who were assigned to and received the CLTS program to the 862 households in villages who were not assigned to and did not attend the program. Is this advice sound? Why or why not?

- b. Another colleague says you should compare the 798 households who participated in the program to the 1,202 households who did not participate in the program. Is this advice sound? Why or why not?
- c. Another colleague suggests that you use the compliance rates, the proportion of households in each group that did or did not comply with their treatment assignment. You should divide the “intention to treat” estimate by the difference in treatment ratios (i.e. proportions of each experimental group that received the treatment). Is this advice sound? Why or why not?
3. Using information from questions 1 and 2, calculate the percentage of the comparison group who complied with their randomized assignment. Calculate the percentage of each group who were treated.
- a. Use your estimate of the ITT from question 1 to estimate the LATE, as follows:

$$LATE = \frac{ITT}{\% \text{ of treatment group who took up treatment} - \% \text{ of comparison group who took up treatment}}$$

- b. In the LATE estimate, which two groups are we comparing to each other?

4. Is the LATE bigger or smaller than the ITT? Does that surprise you? Why would the LATE be different from the ITT?

DISCUSSION TOPIC 4: SPILLOVERS

Spillovers occur when an individual's treatment status has an impact on other individuals, in either the treatment or comparison groups. For example, when a mother vaccinates her child, that action also affects the health of her neighbor's children, because they will now be slightly less likely to get sick (even if they are already immunized, this is still true to some degree).

In randomized evaluations, spillovers pose a challenge because they can affect individuals in the comparison group. In case of immunizations, spillovers can make children in the comparison group *healthier* than they otherwise would be, leading us to underestimate the program's true effect.

1. In the case of the CLTS program, can you think of positive spillovers? Describe how they could happen.
2. Can you think now of negative spillovers? Describe how they could happen.
3. What are the two main strategies that a research team can use regarding spillovers? At what stage of the project should they be conceived and implemented?

REFERENCES AND FURTHER READING

- Cameron, Lisa, Susan Olivia, and Manisha Shah. 2019. “Scaling Up Sanitation: Evidence from an RCT in Indonesia”. *Journal of Development Economics*, 138: 1-16.
- Cameron, Lisa, Susan Olivia, and Manisha Shah. 2010. “Scaling Up Rural Sanitation: Findings from the Impact Evaluation Baseline Survey in Indonesia”. Water and Sanitation Program Working Paper. The World Bank.
- Cameron, Lisa, Susan Olivia, and Manisha Shah. 2013. “Impact Evaluation of a Large Scale Rural Sanitation Project in Indonesia”. World Bank Policy Research Working Paper (6360).
- Cameron, Lisa, Susan Olivia, and Manisha Shah. 2015. “Initial Conditions Matter: Social Capital and Participatory Development”. IZA Discussion Paper No. 9563.
- Gertler, Paul, Manisha Shah, Maria Laura Alzua, Lisa Cameron, Sebastian Martinez, and Sumeet Patil. 2015. “How Does Health Promotion Work? Evidence from the Dirty Business of Eliminating Open Defecation”. NBER Working Paper No. 20997.

REUSE AND CITATIONS

To request permission to reuse this case study or access the accompanying teachers’ guide, please email training@povertyactionlab.org. Please do not reuse without permission. To reference this case study, please cite as:

J-PAL. “Case Study: Rural Sanitation in Indonesia: Threats and Analysis.” Abdul Latif Jameel Poverty Action Lab. 2022. Cambridge, MA.