

CASE STUDY 4: DEWORMING IN KENYA

Understanding Threats to Experimental Integrity



This case study is based on Miguel, Edward, and Michael Kremer. 2004. "**Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.**" *Econometrica* 72 (1): 159-217. <https://doi.org/10.1111/j.1468-0262.2004.00481.x>.

J-PAL thanks the authors for allowing us to use their paper.

KEY VOCABULARY

Treatment assignment, Treatment status	An individual's treatment assignment is the group they were randomly assigned to: the treatment or comparison group. An individual's treatment status is what actually happened to them: were they treated or not?
Attrition	Attrition is an occurrence when individuals or groups leave the study. This can happen for many reasons: they move away from the study area, they no longer wish to participate, they are absent on the surveyors' attempt to survey them, and many more. What is key to note is that if a unit <i>attrits</i> , they do not appear in your data—regardless of their treatment status and their outcome. <i>Random</i> attrition is a concern because it reduces your sample size, which all else equal, makes it harder to detect differences between treatment and comparison groups. <i>Non-random</i> attrition, or when certain groups are more likely to attrit than others, is a larger concern, because it introduces selection bias (described below) in your study sample.
Balance	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.
Selection bias	<p>Selection bias is bias that occurs when the individuals who receive 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 do not balance (for example, those who took it up may be more motivated, or they may be weaker students). Randomization removes selection bias because it breaks the link between characteristics of the individual and their treatment status.</p> <p>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 - Surveyors can choose to only survey the closest houses
Attrition bias	<p>Attrition bias is a type of selection bias that occurs when people choose to 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 choose to leave may be correlated with the treatment. Suppose some of the treatment group finds your job training classes to be too difficult and leave the study. This could mean that workers who have higher levels of ability or motivation are more likely to receive the training, which would create bias in your results.
Compliance	<p>In many randomized evaluations, researchers randomize the assignment to treatment instead of the actual treatment (e.g., they randomly pick which group to offer vaccines instead of randomly administering vaccines). The study sample can be split into three distinct groups:</p> <ol style="list-style-type: none"> 1. Always-takers: This group of people will always take up the program, regardless of assignment status. 2. Never-takers: This group of people will never take up the program, regardless of assignment status. 3. 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. <p>When respondents do not comply with their treatment assignment, the study has <i>partial compliance</i>. 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 <i>non-compliers</i>, and the action of not complying with treatment status as <i>non-compliance</i>.</p> <p>Note: You may hear of a fourth group called defiers. These are individuals who always do the opposite of their treatment status (i.e., they will take up treatment only if they are assigned to the comparison group</p>

	and they will refuse treatment only if assigned to the treatment group). However, there are no defiers in this case study.
Intention-to-treat (ITT):	The ITT is a method for estimating the effect of the <i>program</i> where you compare the average outcomes of those <i>assigned to</i> the treatment group to the average outcomes of those <i>assigned to</i> the comparison group, regardless of whether individuals within those groups have actually received the treatment (also known as treatment status). The ITT measures the impact of delivering a program in the real world, where some people don't take up the program when they are supposed to, and others do take up the program when they are not supposed to.
Local Average Treatment Effect (LATE)	The LATE is a method for estimating the effect of the program <i>on those who complied with their treatment status</i> . The LATE divides the ITT by the difference in the proportion of 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.
Spillovers	Spillovers occur when one individual's action of taking up a treatment impacts another individual, regardless of that individual's assignment status. An illustrative example of spillovers are vaccines: If you are randomly assigned to be offered a vaccine—and <i>you choose to take it up</i> —you reduce the risk of others around you contracting the disease. It does not matter if the people around you are vaccinated or not—or even if they are in the study—the fact that you took up the treatment has impacted them.

LEARNING OBJECTIVE

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

Between 1998 and 2001, the NGO International Child Support Africa (ICS) implemented a school-based mass deworming program in 75 primary schools in western Kenya. The program treated the 45,000 pupils enrolled at these schools for worms and was evaluated by Michael Kremer and Ted Miguel. This case study draws from the evaluation but incorporates hypothetical examples not present in the paper.

Randomization creates groups that on average are balanced at the start of the intervention. However, it should be noted that there can be external influences that make them unbalanced at the end of the program. Imagine we randomly assign villages into two groups: treatment and comparison. Survey staff travel to the treatment villages first, and by the time they reach comparison villages the adult men have migrated for seasonal work. What can we say about the composition of the two groups after this event? Are the groups of people still balanced? These types of changes can introduce bias, diminishing the validity of the impact estimates and threatening the integrity of the experiment.

This raises the question: What other threats to experimental integrity exist?

WORMS: A COMMON PROBLEM WITH A COST-EFFECTIVE SOLUTION

Worm infections are common in areas with poor access to sanitation. Children, who typically are less aware of the importance of good sanitary and personal hygiene habits, are particularly vulnerable: The World Health Organization (WHO) estimated that 272.2 million school-age children were at risk in 2017 (WHO, 2018).

Infected people excrete worm eggs in their feces and urine. In areas with poor sanitation, the eggs contaminate the soil or water. School-age children are more likely to spread worms because they have riskier hygiene practices (more likely to swim in contaminated water, more likely to not use the latrine, less likely to wash hands before eating). Thus, treating a child not only reduces her own worm load, it may also reduce disease transmission—this, in turn, benefits the community at large.

Treatment kills worms in the body (although it does not prevent reinfection). Oral medications that can kill 99 percent of worms in the body are available, inexpensive, and safe. A dose of one of the common medications costs between US\$0.02 and US\$0.20. The WHO recommends regular preemptive school-based mass deworming in areas with high prevalence.

THE PRIMARY SCHOOL DEWORMING PROGRAM

ICS implemented the Primary School Deworming Program (PSDP) in the Busia District in Western Kenya, a densely settled region with high worm prevalence. Treatment followed WHO guidelines: the medicine was administered by nurses from the Ministry of Health in the presence of health officers from ICS.

The PSDP was expected to affect health, nutrition, and education outcomes. To measure impact, ICS collected data on a series of outcomes, including prevalence of worm infection; worm loads (severity of worm infection); self-reported illness; and school participation rates and test scores.

THREATS TO THE INTEGRITY OF THE PLANNED EXPERIMENT

DISCUSSION TOPIC 1

Threats to experimental integrity (5 minutes)

Recall that randomization creates groups that on average are balanced at the start of the experiment.

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

UNDERSTANDING ATTRITION: WHEN THE GROUPS DO NOT REMAIN BALANCED

Attrition is an occurrence 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 group—reduces the sample size in the study. Barring any other changes to the study design, a smaller sample size 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. If a specific type of person leaves *only* the treatment or comparison group, it reduces the balance of the two groups and introduces selection bias into the estimate of the treatment effect.

DISCUSSION TOPIC 2

Understanding Attrition (25 minutes)

There are 45,000 children: 15,000 in treatment schools and 30,000 in comparison schools. After you randomize, you compare the treatment and comparison groups on several characteristics:

Table 1: Baseline Balance Test

(1)	(2)	(3)	(4)
Characteristic:	Treatment average	Comparison average	Difference (Treat. – Comp.)
Age	7.09 (1.23)	7.32 (0.99)	-0.23 (0.34)
Household income (Kenyan Shillings)	11,592.23 (4,230)	11,603.54 (4,265)	-11.31 (101.99)
Reading test score (out of 100)	45.39 (8.67)	44.96 (9.12)	-0.43** (0.12)
Male (=1 if male, =0 otherwise)	0.49 (0.24)	0.48 (0.23)	0.01 (0.10)
English (=1 if school taught in English, =0 otherwise)	0.85 (0.15)	0.82 (0.18)	0.03 (0.09)
Baseline worm load	1.728 (.98)	1.727 (.94)	0.001 (0.13)
Number of children within group	15,000	30,000	
Notes:	Standard errors in parenthesis. Stars indicate statistical significance: * =0.10, ** =.05. ***=0.01		

1. Are there any characteristics for which the treatment and control groups are different? If so, which ones? Which differences are statistically significant?
2. For the differences that are statistically significant, how economically meaningful are these differences? Are you worried that they indicate the groups are not balanced?
3. After treatment is delivered, you replicate table 1 with endline statistics. Suppose the wealthiest families in the study had decided to move to larger cities and don't appear in your data. Which of the above statistics would this change? Would you expect this change to be the same in the treatment and comparison groups? How could this bias your treatment effect estimate?
4. Now suppose that only wealthy families in the comparison group had moved to larger cities, as they both had the means to leave the area and were concerned about school quality (because their children's schools did not receive the treatment). Wealthy families in the treatment group did not

move. Which of the above statistics would this attrition change? How could this bias your treatment effect estimate?

UNDERSTANDING PARTIAL COMPLIANCE:

Recall that in this study, schools were randomly selected to receive an *offer* of deworming medication, then decided whether to take up the treatment. Likewise, individual children in schools that took up the treatment could decide whether to take the medication. Thus, it is necessary to distinguish between children's treatment *assignment* (whether they attended a school, which was assigned to receive the deworming medication) and treatment *status* (whether they took deworming medication).

From these two different measures, we can classify study participants into three groups: compliers, never-takers, and always-takers. Compliers are those who follow the study design; they take the deworming medication when their school is randomly selected to receive it, and do not take the deworming medication when their school is randomly selected not to receive it. Never-takers are those in the treatment group who did not (and never would) take the deworming medication. Always-takers are those in the comparison group who took the deworming medication despite being in a school assigned not to receive it.

In this section, your group will examine the ways partial compliance can change a study.

DISCUSSION TOPIC 3

Partial compliance (10 minutes)

1. Consider three cases of non-compliance. In each case, discuss whether this partial compliance would bias the treatment effect. If so, how?
 - a. Some families whose child attended a comparison school send the child to live with family members whose children attend treatment schools in order to get their child into a treatment school.
 - b. Some children in treatment schools distrust the NGO workers who distribute the medication and refuse to take it.
 - c. Some principals of comparison schools hear their colleagues in treatment schools discussing the deworming medication and decide to implement the program on their own.

DISCUSSION TOPIC 4

Treatment effect with partial compliance (15 minutes)

Next, we will walk through a hypothetical example with partial compliance. The study looks at the health effects of deworming. In particular, researchers are interested in worm loads (the severity of worm infection). Worm loads are scaled as follows:

Heavy worm infections = score of 3

Medium worm infections = score of 2

Light infections = score of 1

No worms = score of 0

We want to know if there was a treatment effect—did the deworming program work? To test this, we can use a similar method as testing for balance: compare the average endline worm loads of each group to see if there is a difference. This difference can be attributed to the treatment instead of differences in the groups precisely *because* of the randomization. As shown in table 1, the two groups were balanced and have similar baseline worm loads. Since their treatment assignment was random, we can attribute their endline differences to the treatment.

Table 2: Endline Worm Loads

Worm Load	Treatment schools	Comparison schools
3	876	5,236
2	5,488	11,423
1	6,604	13,341
0 (no worms)	2032	0
Total children tested at school	15,000	30,000
Average:		

1. What is the average worm load of each group at endline?
2. What is the estimated impact of the program, or the ITT?
3. When we calculate the ITT, who is being compared to whom? Are we using their treatment *assignment* or treatment *status*?

Table 3: Treatment Assignment vs. Treatment Status

Take-up status:	Assignment status:		Total:
	Treatment	Comparison	
Took the medication	12,000	0	12,000
Did not take medication	3,000	30,000	33,000
Total:	15,000	30,000	45,000

- Suppose we took a closer look at the data from Table 2 and were able to determine who actually had taken the medication, shown in Table 3. Determine who did not comply with their treatment status. What percentage of each group took up the treatment? What percentage of each group complied with their treatment assignment?
- Use your estimate of the ITT from question 2 to estimate the LATE, as follows:

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

- Is the LATE bigger or smaller than the ITT? Does that surprise you? Why would the LATE (the effect of the deworming *medication*) be different from the ITT (the effect of the school-based deworming *program*)?
- Can we compare the 12,000 treatment school children who took up the program to the 30,000 comparison school children who did not take up the program? Why or why not?
- Can we compare the 12,000 children who took up the program to the 33,000 children who did not take up the program? Why or why not?

UNDERSTANDING SPILLOVERS: WHEN GROUPS THAT ARE NOT DIRECTLY TREATED BENEFIT FROM THE TREATMENT GROUP BEING TREATED

Spillovers occur when one individual's action of taking up a treatment impacts another unit, regardless of that unit's assignment status. Spillovers are tricky to measure—they can often occur in people outside the study design, who you don't survey, or can occur in the comparison group, which reduces the measured treatment effect. Spillovers are not inherently good or bad, but they change the way we think of a program's effectiveness.

DISCUSSION TOPIC 4

Spillovers (25 minutes)

1. Consider the families of students in both the treatment and comparison group. Before the study, which group of families are less likely to have higher worm loads: families of schools assigned to the treatment group or families of students assigned to the comparison group? Using the information about worms presented in this guide, which group of families is less likely to have higher worm loads after the intervention is delivered to treatment schools?
2. You anecdotally hear from several treatment families that they have been experiencing fewer worm-related illnesses. You did not collect data on these families, but you hypothesize that treatment families have lower worm loads than comparison families.
 - a. Do you expect this difference to be larger or smaller than the difference in worm loads between children in treatment and comparison schools? Why?
 - b. Had you collected this data, how might you calculate the total impact of school-based deworming on treatment communities, given this information?
3. Spillovers can cause you to over or underestimate the effect of a program. Using the above example, if you were concerned that spillovers were obscuring some impact of the program, how could you make sure you capture this impact? At what stage of the project should you map out potential spillovers?

(OPTIONAL) EXPLORING THE COST-EFFECTIVENESS OF SCHOOL-BASED DEWORMING

Calculating the cost-effectiveness of a program—for instance, dollars spent per additional day of student attendance at school—can offer insights into which programs are likely to provide the greatest value for money in given situations. Cost-effectiveness analysis (CEA) summarizes complex programs in terms of a simple ratio of costs to impacts, allowing us to use a common measure to compare different programs evaluated in different countries in different years. To calculate the cost-effectiveness (CE) ratio, you need two pieces of data: an estimate of the program’s impact on specific outcomes and the cost of implementing the program.

$$CE\ ratio = \frac{\textit{Total cost of implementing program}}{\textit{Total impact of program on specific outcome}}$$

School-based deworming treatments have been shown to be cost effective (Morgan 2018). This section will explore some of the reasons why.

DISCUSSION TOPIC 5

Cost-Effectiveness (10 minutes)

1. Why do you think school-based deworming is cost-effective? Are there certain characteristics of this program that reward treating units larger than the individual? If so, what?

2. How would the cost-effectiveness of the program change if some types of worms became resistant to medication?

REFERENCES

- Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Morgan, Grace. 2018. "School-Based Deworming: A Low-Cost Intervention Achieving High-Impact Benefits." Accessed June 3, 2019. <https://www.evidenceaction.org/blog-full/2017-deworming-cost-effectiveness>.
- WHO (World Health Organization). 2018. *Weekly Epidemiological Record* 93 (50): 681–692. <https://www.who.int/wer/2018/wer9350/en/>.

REUSE AND CITATIONS

To request permission to access the accompanying teachers' guide, please email training@povertyactionlab.org. To reference this case study, please cite as:

J-PAL. "Deworming in Kenya: Threats and Analysis." Abdul Latif Jameel Poverty Action Lab. 2019. Cambridge, MA

To reference the original study by Miguel and Kremer, please cite as:

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica* 72 (1): 159-217. <https://doi.org/10.1111/j.1468-0262.2004.00481.x>.



J-PAL, 2019

This case study is made available under a Creative Commons Attribution 4.0 License (international): <https://creativecommons.org/licenses/by/4.0/>