

ABDUL LATIF JAMEEL
Poverty Action Lab

TRANSLATING RESEARCH INTO ACTION



ipa
INNOVATIONS FOR
POVERTY ACTION

Case Study 4: Deworming in Kenya

Addressing threats to experimental integrity



This case study is based on Edward Miguel and Michael Kremer, “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities,” *Econometrica* 72(1): 159-217, 2004.

J-PAL and IPA thank the authors for allowing us to use their study.

Key vocabulary

Phase-in design: a study design in which groups are individually phased into treatment over a period of time; groups that are scheduled to receive treatment later act as the comparison groups in earlier rounds.

Equivalence: when groups are identical on all baseline characteristics, both observable and unobservable. It is ensured by randomization, *in expectation*.

Attrition: the process of individuals dropping out of either the treatment or comparison group over the course of the study.

Attrition bias: a statistical bias that occurs when individuals systematically drop out of either the treatment or the comparison group for reasons related to the treatment.

Partial compliance: when individuals do not “comply” with their assignment (to treatment or comparison). Also termed “diffusion” or “contamination.”

Intention-to-treat analysis: the measured impact of a program comparing study (treatment versus comparison) groups, regardless of whether they actually received the treatment.

Externality: an indirect cost or benefit incurred by individuals who did not directly receive the treatment. Also known as a “spillover”.

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—hookworm, roundworm, whipworm, and schistosomiasis. Schools were phased in randomly.

Randomization ensures that the treatment and comparison groups are comparable at the beginning, but there can be external influences that can make them incomparable at the end of the program. Imagine we have a pile of seeds from five different plants. If we split this pile randomly into two bags, both bags should have the same composition of seeds. Suppose now that one of the bags gets punctured; the hole is small enough for only the smallest seed variety to pass through. What can we say about the composition of the two bags after this event? Are the two bags still comparable? This type of event can happen between initial randomization and the endline and can reintroduce selection bias; it diminishes the validity of the impact estimates and is a threat to the integrity of the experiment.

How can common threats to experimental integrity be managed?

Worms: a common problem with a cheap solution

Worm infections account for over 40 percent of the global tropical disease burden. Infections are common in areas with poor sanitation. More than 2 billion people are affected. Children, who typically have poorer sanitary habits, are particularly vulnerable: 400 million school-age children are chronically infected with intestinal worms.

Symptoms include listlessness, diarrhea, abdominal pain, and anemia. But worms affect more than the health of children. Heavy worm infections can impair children's physical and mental development, leading to poor attendance and performance in school.

Poor sanitation and personal hygiene habits facilitate transmission. Infected people excrete worm eggs in their feces and urine. In areas with poor sanitation, the eggs contaminate the soil or water. Other people are infected when they ingest contaminated food or soil (hookworm, whipworm, and roundworm), or when hatched worm larvae penetrate their skin upon contact with contaminated soil (hookworm) or fresh water (schistosome). 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). So treating a child not only reduces her own worm load; it may also reduce disease transmission—and so benefit the community at large.

Treatment kills worms in the body, but does not prevent reinfection. Oral medication that can kill 99 percent of worms in the body is available: albendazole or mebendazole for treating hookworm, roundworm, and whipworm infections; and praziquantel for treating schistosomiasis. These drugs are cheap and safe. A dose of albendazole or mebendazole costs less than 3 US cents while one dose of praziquantel costs less than 20 US cents. The drugs have very few and minor side effects.

Worms colonize the intestines and the urinary tract, but they do not reproduce in the body; their numbers build up only through repeated contact with contaminated soil or water. The World Health Organization (WHO) recommends presumptive school-based mass deworming in areas with high prevalence. Schools with hookworm, whipworm, and roundworm prevalence over 50 percent should be mass treated with albendazole every 6 months, and schools with schistosomiasis prevalence over 30 percent should be mass treated with praziquantel once a year.

The Primary School Deworming Program

International Child Support Africa (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 public

health nurses from the Ministry of Health, in the presence of health officers from ICS.

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

Evaluation design: the experiment as planned

Because of administrative and financial constraints, the PSDP could not be implemented in all schools immediately. Instead, the 75 schools were randomly divided into three groups of 25 schools and phased in over three years. Group 1 schools were treated starting in both 1998 and 1999, Group 2 schools in 1999, and Group 3 schools starting in 2001. Group 1 schools were the treatment group in 1998, while schools in Group 2 and Group 3 were the comparison. In 1999 Group 1 and Group 2 schools formed the treatment group and Group 3 schools the comparison.

TABLE 1

The planned experiment: the PSDP treatment timeline showing experimental groups in 1998 and 1999

	1998	1999	2000
Group 1	Treatment	Treatment	Treatment
Group 2	Comparison	Treatment	Treatment
Group 3	Comparison	Comparison	Treatment

For the purpose of the following questions, we will only look at results after the 1998 period.

Threats to integrity of the planned experiment

DISCUSSION TOPIC 1

Threats to experimental integrity

Randomization ensures that the groups are in expectation equivalent, and therefore comparable, at the beginning of the program. The impact is then estimated as the difference in the average outcome of the treatment group and the average outcome of the comparison group, both at the end of the program. To be able to say that the program caused the impact, you need to be able to say that the program was the only difference between the treatment and comparison groups over the course of the evaluation.

- What does it mean to say that the groups are equivalent at the start of the program?
- Can you check if the groups are equivalent at the beginning of the program? How?

Managing attrition: when the groups do not remain equivalent

Attrition is when people drop out of the sample—both treatment and comparison groups—over the course of the experiment. One common example in clinical trials is when people die; so common indeed that attrition is sometimes called experimental mortality.

Worm Load	T	C	T	C
3	5,000	10,000	0	10,000
2	5,000	10,000	0	10,000
1	5,000	10,000	15,000	10,000
Total children tested at school	15,000	30,000	15,000	30,000
Average				

DISCUSSION TOPIC 2

Managing attrition

You are looking at the health effects of deworming. In particular you are looking at the worm load (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

There are 30,000 children: 15,000 in treatment schools and 15,000 in comparison schools. After you randomize, the treatment and comparison groups are equivalent, meaning children from each of the three worm load categories are equally represented in both groups.

Suppose protocol compliance is 100 percent: all children who are in the treatment get treated and none of the children in the comparison are treated. Children that were dewormed at the beginning of the school year (that is, children in the treatment group) end up with a worm load of 1 at the end of the year. The number of children in each worm-load category is shown for both the pretest and posttest.

1.

- a. At pretest, what is the average worm load for each group?
- b. At posttest, what is the average worm load for each group?
- c. What is the impact of the program?
- d. Do you need to know pretest values? Why or why not?

Suppose now that children who have a worm load of 3 only attend half the time and drop out of school if they are not treated. The number of children in each worm-load category is shown for both the pretest and posttest.

TABLE 2

	Pretest	Posttest

TABLE 3

Worm Load	Pretest		Posttest	
	T	C	T	C
3	5,000	10,000	0	Dropped out
2	5,000	10,000	0	10,000
1	5,000	10,000	15,000	10,000
Total children tested at school	15,000	30,000	15,000	20,000
Average				

2.
 - a. At posttest, what is the new average worm load for the comparison group?
 - b. What is the impact of the program?
 - c. Is this outcome difference an accurate estimate of the impact of the program? Why or why not?
 - d. If it is not accurate, does it overestimate or underestimate the impact?
 - e. How can we get a better estimate of the program's impact?

Besides worm load, the PSDP also looked at outcome measures such as school attendance rates and test scores.

3.
 - a. Would differential attrition (i.e., differences in dropouts between treatment and comparison groups) bias either of these outcomes? How?

- b. Would the impacts on these final outcome measures be underestimated or overestimated?

In Case Study 1, you learned about other methods to estimate program impact, such as pre-post, simple difference, difference-in-difference, and multivariate regression.

4.
 - a. Does the threat of attrition only present itself in randomized evaluations?

Managing partial compliance: when the treatment group does not actually get treated or the comparison group gets treated

Some people assigned to the treatment may in the end not actually get treated. In an after-school tutoring program, for example, some children assigned to receive tutoring may simply not show up for tutoring. Those assigned to the comparison group may obtain access to tutoring, either from the program or from another provider. Or comparison group children may get extra help from the teachers or acquire program materials and methods from their classmates. In any of these scenarios, people are not complying with their assignment in the planned experiment. This is called “partial compliance” or “diffusion” or, less benignly, “contamination.” In contrast to carefully controlled lab experiments, diffusion is a ubiquitous concern in social programs. After all, life goes on, people

will be people, and you have no control over what they decide to do over the course of the experiment. All you can do is plan your experiment and offer them treatments. How, then, can you deal with the complications that arise from partial compliance?

DISCUSSION TOPIC 3 Managing partial compliance

TABLE 4

	Pretest		Posttest	
	T	C	T	C
Worm Load				
3	5,000	10,000	5,000	10,000
2	5,000	10,000	0	10,000
1	5,000	10,000	10,000	10,000
Total children tested at school	15,000	30,000	15,000	30,000

1.
 - a. Calculate the impact estimate based on the original group assignments.
 - b. This is an unbiased measure of the effect of the program, but in what ways is it useful and in what ways is it not as useful?
 - c. Five of your colleagues are passing by your desk; they all agree that you should calculate the effect of the treatment using only the 10,000 children who were treated and compare them to the comparison group. Is this advice sound? Why or why not?

- d. Another colleague says that it's not a good idea to drop the untreated entirely; you should use them but consider them as part of the comparison. Is this advice sound? Why or why not?

Managing spillovers: when the comparison, itself untreated, benefits from the treatment being treated

People assigned to the control group may benefit indirectly from those receiving treatment. For example, a program that distributes insecticide-treated nets may reduce malaria transmission in the community, indirectly benefiting those who themselves do not sleep under a net. Such effects are called externalities or spillovers.

DISCUSSION TOPIC 4 Managing spillovers

In the deworming program, randomization was at the school level. However, while all boys at a given treatment school were treated, only girls younger than thirteen received the deworming pill. This was due to the fact that the WHO had not tested (and thus not yet approved) the deworming pill for pregnant women. Because it was difficult to determine which girls were at risk of getting pregnant, the program decided to not administer the medication to any girl thirteen or older. (Postscript: since the deworming evaluation was implemented, the WHO has approved the deworming medication for pregnant women.)

Thus, at a given treatment school, there was a distinct group of students that was never treated, but lived in very close proximity to a group that was treated.

Suppose protocol compliance is 100 percent: all boys and girls under thirteen in treatment schools get treated and all girls thirteen and over in treatment schools, as well as all children in comparison schools, do not get treated.

You can assume that due to proper randomization, the distribution of worm load across the three groups of students is equivalent between treatment and control schools prior to the intervention.

1.
 - a. If there are any spillovers, where would you expect them to show up?
 - b. Is it possible for you to capture these potential spillover effects? How?