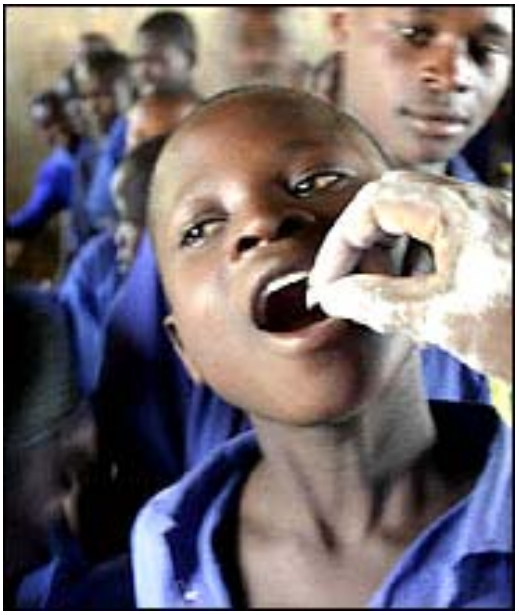
Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs
Spring 2009

ABDUL LATIF JAMEEL
Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

Massachusetts Institute of Technology



Case 4: Deworming in Kenya

Managing 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 thanks the authors for allowing us to use their paper

Between 1998 and 2001, the NGO International Child Support Africa implemented a school-based mass deworming program in 75 primary schools in western Kenya. The program treated the 30,000 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 it cannot ensure that they remain comparable at the end of the program. Nor can it ensure that people comply with the treatment they were assigned. Life also goes on after the randomization: other events besides the program happen between randomization and the end-line. These events can reintroduce selection bias; they diminish the validity of the impact estimates and are threats 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, still learning good sanitary habits, are particularly vulnerable: 400 million school-age children are chronically infected with intestinal worms.

Worms affect more than the health of children. Symptoms include listlessness, diarrhea, abdominal pain, and anemia. Beyond their effects on health and nutrition, heavy worm infections can impair children's physical and mental development and reduce their 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 re-infection. 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 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 Abdul Latif Jameel Poverty Action Lab
@MIT, Cambridge, MA 02130, USA | @IFMR, Chennai 600 008, India | @PSE, Paris 75014, France

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 3 groups of 25 schools, and phased-in over 3 years. Group 1 schools were treated starting in both 1998 and 1999, Group 2 schools in 1999, and Group 3 starting in 2001. Group 1 schools were the treatment group in 1998, while schools Group 2 and Group 3 were the comparison. In 1999 Group 1 and Group 2 schools were the treatment and Group 3 schools the comparison.

Figure 1: The planned experiment: the PSDP treatment timeline showing experimental groups in 1998 and 1999

	1998	1999	2001
Group 1	**Treatment**	**Treatment**	**Treatment**
Group 2	Comparison	**Treatment**	**Treatment**
Group 3	Comparison	Comparison	**Treatment**

Threats to integrity of the planned experiment

Discussion Topic 1: Threats to experimental integrity

Randomization ensures that the groups are equivalent, and therefore comparable, at the beginning of 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. 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.

1.	What does it mean to say that the groups are equivalent at the start of the program?
2.	Can you check if the groups are equivalent at the beginning of the program? How?
3.	What happen over the course of the evaluation to make the groups non-equivalent?
4.	How does non-equivalence at the end threaten the integrity of the experiment?
5.	You randomized, creating equivalent treatment and comparison groups. If the groups remain equivalent, what else can happen after randomization to threaten your ability to say the program was the only difference between the two groups?
6.	In Case 1, you learned about other methods to estimate program impact, such as simple difference, multiple regression, multiple regression with panel data, and matching. **a.** For each threat you just worked, say if and how the threat exists for each of these methods. **b.** Are the threats to experimental integrity unique to randomization?

Managing attrition—when the groups do not remain equivalent

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

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 get a worm load score of 3, medium worm infections a score of 2, and light infections a score of 1.

The program is school-based, so it is natural and cost-effective to collect data at the schools—the children are gathered in one place, so the enumerator does not have to travel to every child's home. The enumerator takes the measurements on all children in school on a randomly chosen day (the school authorities are not given prior warning).

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

Protocol compliance is 100 percent: all children who are in the treatment get treated and none of the children in the comparison are treated. Deworming at the beginning of the school year results in a worm load of 1 at the end of the year because of re-infection. 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.

		Pretest		***Posttest***	
	Worm Load	Treatment	Comparison	Treatment	Comparison
	3	5,000	5,000	0	**Dropped out**
	2	5,000	5,000	0	5,000
	1	5,000	5,000	15,000	5,000
	Total children tested at school	15,000	15,000	15,000	10,000

1.	**a.** What is the average pretest worm load for the treatment group? **b.** What is the average pretest worm load for the comparison group? **c.** Are the groups equivalent?
2.	**a.** What is the average posttest worm load for the treatment group? **b.** What is the average posttest worm load for the comparison group? **c.** What is the difference?
3.	**a.** Calculate the outcome differences at the beginning and at the end of the year? **b.** Is this outcome difference an accurate estimate of impact of the program? **c.** If it is not accurate does it overestimate or underestimate the impact?
4.	Because the treatment was treated, you expected there to be a difference between the groups at the end of the year. **a.** If this difference is an effect, what is the source of attrition bias, if any? **b.** How can you solve the problem to get a better estimate of program impact?
5.	**a.** What is the average posttest worm load for the comparison group if you also tested the 5,000 dropouts? **b.** Calculate the impact of the program. **c.** What is the size of the attrition bias?
6.	**a.** The PSPD also looked at school attendance rates and test scores. **b.** Would differential attrition bias either of these outcomes? **c.** Would the impact be underestimated or overestimated?
7.	In their song *A Day in the Life*, the Beatles sing, "And though the holes were rather small, they had to count them all." Why should your consider adopting *A Day in the Life* as your theme song when you are thinking about managing attrition?

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

Some people assigned to the treatment may in the end not actually get treated. For example, children assigned to after-school tutoring programs simply do not show up for tutoring. Some people assigned to the comparison may get access to the treatment, either from the program or from another provider. For example, children assigned to the after-school tutoring comparison group may get extra help from the teachers or get program materials and methods from their classmates. Either way, these people are not complying with their assignment in the planned experiment. This is called "partial compliance" or "diffusion" or, less benignly, "contamination." The effects are ubiquitous 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 manage threats arising from partial compliance?

Discussion Topic 3: Managing partial compliance

All the children from the poorest families don't have shoes and so they have worm loads of 3. Though their parents had not paid the school fees, the children were allowed to stay on in school during the year. Parental consent was required for treatment and to give consent, the parents had to come to the school and sign a consent form in the headmaster's office. Because they had not paid school fees, the poorest parents were reluctant to come to the school. So none of the children with worm loads of 3 were actually treated. Their worm loads scores remain 3 at the end of the year. No one assigned to comparison was treated. All the children in the sample at the beginning of the year were followed up, if not at school then at home.

	Pretest		***Posttest***	
Worm Load	Treatment	Comparison	Treatment	Comparison
3	5,000	5,000	5,000	5,000
2	5,000	5,000	0	5,000
1	5,000	5,000	10,000	5,000
Total children tested	15,000	15,000	15,000	15,000

1. **a.** Calculate the impact estimate based on the original assignments.
b. What does this "intention to treat" estimate measure?
c. This is an accurate measure of the effect of the program, but is it a good measure? What are the considerations? When is it useful? When is it not useful?

You are interested in learning the effect of treatment on those actually treated.

2. 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.
a. What is the impact using only the treated?
b. Is the advice sound? Why? Why not?

3. 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.
a. What is the impact estimate based on this strategy?
b. Is the advice sound? Why? Why not?

4. Another colleague suggests that you use the compliance rates, the proportion of people in each group that complied with the treatment assignment. You should divide the "intention to treat" estimate with the difference in the compliance rates.
a. What are the compliance rates in the treatment and comparison groups?
b. What is the impact estimate based on this strategy?
c. Is the advice sound? Why? Why not?

5. The program raised awareness of the worms, so some parent in the comparison bought the drugs and treated the children at home. Altogether 2,000 comparison children were treated.

What is the "treatment on the treated" impact estimate?

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 PSPD, randomization was at the school level.

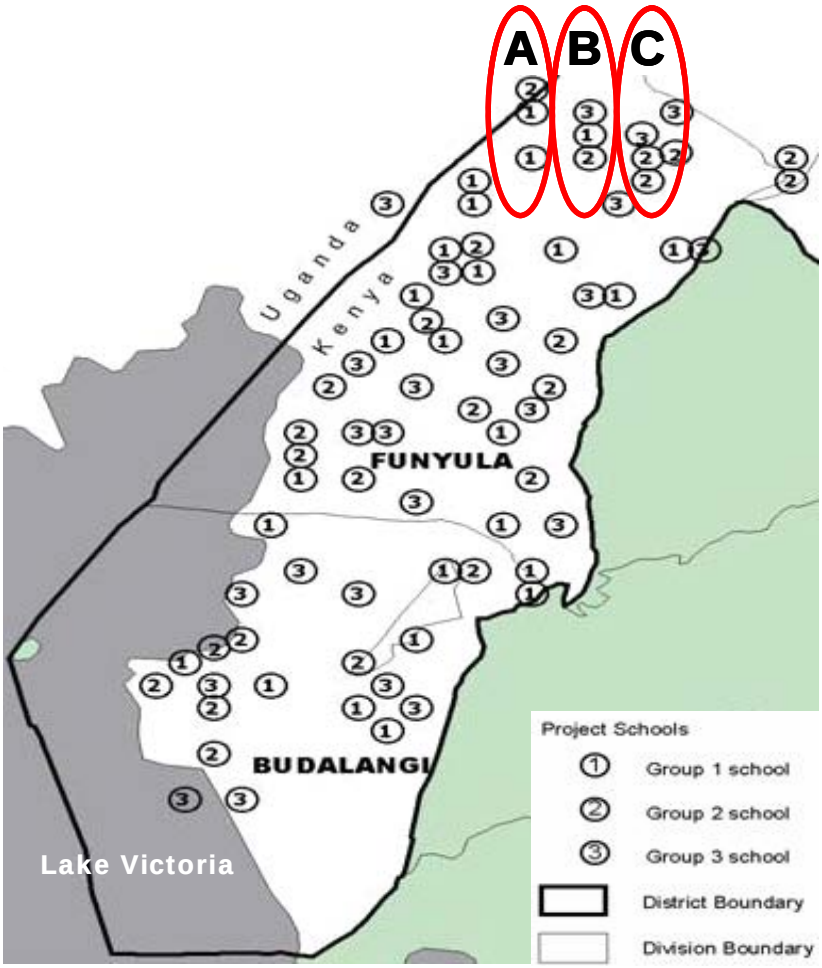
People in the evaluation areas lived on farms close together. Clusters of farms can be divided into areas of 3km radius. Three such areas—A, B, and C—are shown in the diagram below.* Farms are closed enough for children from neighboring farms to play with one another. Families also had a choice of primary schools.

There are three schools in area A, three in area B, and five in area C. It was common for children from neighboring farms, or even siblings, to go to different schools. Some of the schools in each cluster were treatment, others were control. Group 1 schools were the treatment in year 1, and group 2 and 3 were the comparison.

Each school has 100 children. Protocol compliance is 100 percent: all the children in treatment get treated and all the children in comparison do not get treated.

1.	You estimate impact by comparing average worm loads at treatment and comparison schools. Would this estimate be an underestimate or overestimate of the impact?
2.	**a.** The treatment density is the proportion of treated to untreated in a given grouping of people. **b.** What is the treatment density at the treatment schools in year 1? **c.** What is the treatment density of comparison schools? **d.** What are the treatment densities in areas A, B, and C in year 1? **e.** What are the treatment densities in areas A, B, and C in year 2 and year 3?
3.	**a.** If there are any spillovers, where would you expect them to come from? **b.** Is it possible for you to capture spillover effects within the schools? **c.** If you don't expect to be able to capture the spillover effect, what would you need to be able to capture them? **d.** Is it possible for you capture cross-school spillovers?
4.	Rank the areas A, B, and C in terms of the amount of treatment spillover effects expected in years 1, 2, and 3.
5.	**a.** If you had randomized at the individual level, what could you have done to capture interpersonal spillover? **b.** If you had randomized at the school level what can you do to capture cross-school spillovers? **c.** What general strategy does this suggest?

Discussion Topic 4: Managing spillovers



* The GPS locations were collected before May 2000, when the U.S. was still downgrading international GPS accuracy. Readings may only be accurate to within several hundred meters. So one Group 3 school appears to be in Uganda, but it's actually on the Kenyan side of the border. The school that appears to be in Lake Victoria is actually on a very small island.

References:

Crompton, D.W.T. 1999. "How Much Helminthiasis Is There in the World?" Journal of Parasitology 85: 397 – 403.

Kremer, Michael and Edward Miguel. 2007. "The Illusion of Sustainability," Quarterly Journal of Economics 122(3)

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," Econometrica 72(1): 159-217.

Shadish, William R, Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin Company

World Bank. 2003. "School Deworming at a Glance," Public Health at a Glance Series. http://www.worldbank.org/hnp

WHO. 1999. "The World Health Report 1999," World Health Organization, Geneva.

WHO. 2004. "Action Against Worms" Partners for Parasite Control Newsletter, Issue #1, January 2004, www.who.int/wormcontrol/en/action_against_worms.pdf