CASE STUDY 4: TECHNOSERVE COFFEE IN RWANDA

Addressing threats to experimental integrity

This case study is based on a current study by Esther Duflo and Tavneet Suri.

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

KEY VOCABULARY

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

Equivalence: groups are identical on all baseline characteristics, both observable and unobservable. Ensured by randomization.

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

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

Partial Compliance: individuals do not "comply" with their assignment (to treatment or comparison). Also termed "diffusion" or "contamination."

Intention to Treat: the measured impact of a program comparing study (treatment versus control) 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 termed "spillover."

INTRODUCTION

In 2010, the Technoserve (TNS) Coffee Initiative partnered with J-PAL researchers to conduct a randomized evaluation on their coffee agronomy-training program in Nyarubaka sector in southern Rwanda. Technoserve carried out their regular recruitment sign-up processes across all 27 villages in the sector and registered 1600 coffee farmers who were interested in attending the monthly training modules. The study design for the evaluation then required that this pool of farmers be split into treatment and control groups, meaning those who would participate in the training, and those who wouldn't (for now—they would be trained in later phases). The trainings in Nyarubaka included 800 coffee farmers, randomly selected from the pool of 1600.

Randomization ensures that the treatment and comparison groups are equivalent at the beginning, mitigating concern for selection bias. But it cannot ensure that they remain comparable until the end of the program. Nor can it ensure that people comply with the treatment, or even the non-treatment, that they were assigned. Life also goes on after the randomization: other events besides the program happen between initial randomization and the end-line data collection. 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?

EVALUATION DESIGN — THE EXPERIMENT AS PLANNED

As previously mentioned, the agronomy training evaluation consisted of 1600 farmers, half of which attended monthly training sessions, and the other half did not.

In addition, there was a census done of the entire sector to show us which households were coffee farmers and which ones were not. The census showed that there were 5400 households in Nyarubaka - 2400 non-coffee farming households and 3000 coffee farming households (1600 of which were already in our sample).

Each month a Technoserve farmer trainer would gather the farmers assigned to his/her group and conduct a training module on farming practices (e.g. weeding, pruning, bookkeeping, etc). The farmers were taught the best practices by using a practice plot so they could see and do exactly what the instructor was explaining.

To think about:

How can we be certain that the control group farmers did not attend the training too? What can be done to reduce this risk?

Since we have a census for Nyarubaka, how might this be helpful in at least controlling for or documenting any spillovers? (think about what can be done at the trainings themselves).

What type of data might you need/want to try to control for any spillovers in this case?

What were other forms or opportunities for agronomy training in the area?

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 THE PROGRAM. THE IMPACT IS THEN ESTIMATED AS THE DIFFERENCE BETWEEN 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.

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. Other than the program's direct and indirect impacts, what can happen over the course of the evaluation (after conducting the random assignment) to make the groups non-equivalent?

4. How does non-equivalence at the end threaten the integrity of the experiment?

5. In the Technoserve agronomy training example, why is it useful to randomly select from the farmers who signed up for the Technoserve training program, rather than amongst all the coffee farmers in the sector?

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 in clinical trials is when people die; so common indeed that attrition is sometimes called experimental mortality.

Discussion Topic 2
Managing Attrition

You are looking at how much farmers adopt the recommendations and techniques from the agronomy trainings. Using a stylized example, let's divide adoption of the techniques as follows:

Full adoption = score of 2

Partial adoption = score of 1

No adoption = score of 0

Let's assume that there are 1800 farmers: 900 treatment farmers who receive the training and 900 comparison farmers who do not receive the training. After you randomize and collect some baseline data, you determine that the treatment and comparison groups are equivalent, meaning farmers from each of the three categories are equally represented in both groups.

Suppose protocol compliance is 100 percent: all farmers who are in the treatment go to the training and none of the farmers in the comparison attend the training. Let's Farmers who attend all agronomy trainings end up with full adoption, scoring a 2. Let's assume that there was a drought during this period, and those who adopted best-practices managed to protect their crops against damage. However, the farmers who have adoption level 0 see most of their crops perish, and members of the household enter the migrant labor market to generate additional income. The number of farmers in each treatment group, and each adoption category is shown for both the pre-adoption and post-adoption.

	in sample		

1. At program end, what is the average adoption for the treatment group?

a. At program end, what is the average adoption for the comparison group?

b. What is the difference?

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?

2. Besides level of adoption, the Technoserve agronomy training evaluation also looked at outcome measures such as yields and farm labor. In the Technoserve agronomy evaluation, identify some other causes for attrition in the Tretment group and the Control groups? What can be done to mitigate these?

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

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

3. You may know of other research designs to measure impact, such as the non-experimental or quasi-experimental methodologies (eg. Pre-post difference-in-difference, regression discontinuity, instrumental variables (IV), etc)

a. Is the threat of attrition unique to randomized evaluations?

TABLE 1

Adoption Level	Pre-adoption		Post-adoption	
	T	C	T	C
0	300	300	0	Dropped out
1	300	300	0	300
2	300	300	900	300
Total farmers	900	900	900	600

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. In an after-school tutoring program, for example, some children assigned to receive tutoring may simply not show up for tutoring. And the others assigned to the comparison may obtain access to the treatment, 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 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 deal with the complications that arise from partial compliance?

Discussion Topic 3
Managing partial compliance

Suppose that farmers who have adoption level 0 are too risk averse to adopt the techniques they learn at the training. Famers believe that there is no way for them to adopt the techniques that are described in early trainings and stop attending. Consequently, none of the treatment farmers with adoption level 0 increased their adoption and remained at level 0 at the end of the program. No one assigned to comparison had attended the trainings. All the farmers in the sample at the beginning of the program were followed up.

TABLE 2

Adoption Level	Pre-adoption		Post-adoption	
	T	C	T	C
0	300	300	300	300
1	300	300	0	300
2	300	300	600	300
Total # farmer in the sample	900	900	900	900

1. Calculate the impact estimate based on the original group assignments.

   a. Is this an unbiased measure of the effect of the program?

   b. In what ways is it useful and in what ways is it not as useful?

   You are interested in learning the effect of treatment on those actually treated ("treatment on the treated" (TOT) estimate).

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 farmers who attended the training.

3. Is this advice sound? Why or why not?

4. Another colleague says that it's not a good idea to drop the farmers who stopped attending the trainings entirely; you should use them but consider them as part of the control group.

5. Is this advice sound? Why or why not?

6. Another colleague suggests that you use the compliance rates, the proportion of people 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 the treatment ratios (i.e. proportions of each experimental group that received the treatment).

7. 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 Technoserve agronomy training evaluation, randomization was at the farmer level, meaning that while one farmer might have been selected to be in the training, his neighbor didn't have the same fortunes during the randomization process.

Depending on the evaluation and the nature of the program, it might be more challenging to prevent spillovers of agronomic knowledge between friends, than it is for delivering hard tangible objects in farmers' hands, like a weighing scale or calendar to maintain harvest records.

1. How do you imagine spillovers might occur in agronomy training?

2. What types of mechanisms can you think of that could be used to reduce or manage spillovers?

Discussion Topic 5
Measuring spillovers

1. Can you think of ways to design the experiment explicitly to measure the spillovers of the agronomy training?