

GUIDE 4: COUNSELING THE UNEMPLOYED

Addressing threats to experimental integrity



This case study is based on “Sample Attrition Bias in Randomized Experiments: A Tale of Two Surveys” By Luc Behaghel, Bruno Crépon, Marc Gurgand, Thomas Le Barbanchon, Paris-Jourdan Sciences Economiques, Version 1, 2011

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

DESCRIPTION

In France, there are two large government agencies that provide services to the unemployed. Unédic provides unemployment benefits to the unemployed (such as money to cover basic necessities) and ANPE (Agence Nationale Pour l'Emploi) provides counseling and job-placement services. In 2007, Unédic joined hands with the private sector to supply intensive counseling and job placement services to the unemployed community that it served. Simultaneously, the ANPE started its own program of providing intensive counseling and monitoring services. Both were large-scale efforts, each serving roughly 40,000 jobseekers.

A randomized evaluation was designed to test the relative effectiveness of each program. Randomization ensures that the treatment and comparison groups are comparable at the beginning; however it cannot ensure that they remain comparable until 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 initial randomization and the endline. These events can reintroduce selection bias; they diminish the validity of the impact estimates and are threats to the integrity of the experiment.

LEARNING OBJECTIVE

- To explore how common threats to experimental integrity can be managed
- Subjects covered Threats to experimental integrity, equivalence, comparability, compliance, spillovers or externalities, behavioral responses, intention to treat, treatment on the treated

GENERAL GUIDANCE

The section below, “Threats to experimental integrity”, is not included in the case study itself. You can use it as a reference throughout the case.

Additionally, this case involves more arithmetic than other cases—and while it is simple, be prepared to explain the reasoning behind each of the calculations in the Discussion Topics.

Finally, Discussion Topics 3 and 4 cover concepts (such as ITT, TOT, spillovers) that may require more in-depth explanation. If it is helpful, use an example of crossover from comparison to treatment for DT3. Likewise, a simplified example of spillovers is provided after DT4.

THREATS TO EXPERIMENTAL INTEGRITY

There are three main types of threats. This case covers types 1 and 2, but we include 3 in case you are asked about it.

1. When the groups do not remain equivalent—attrition bias

Estimates may become biased if people select themselves in or out of either of the groups—join or drop out—over the course of the experiment, and their reasons for doing so are systematically related to the treatment. While this can be seen as a program effect, it makes it more difficult to interpret any differences in outcomes. In a sense treatment correlated attrition reintroduces selection bias. The experimental groups comprise different people at the end; they are no longer equivalent and the planned comparison may no longer be valid.

2. When the planned experimental contrast is diminished—partial compliance, alternative services, and spillovers

The planned difference in treatment rates between the groups can disappear if people assigned to the treatment group are not actually treated or if people assigned to the comparison group do in fact get treated, directly or indirectly.

Some people assigned to the treatment may in the end not get treated. For example, children assigned to an

after-school tutoring program may simply not show up for tutoring. This is called partial compliance.

Some people assigned to the comparison may access program services or else get equivalent services 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. If this happens systematically, the treatment contrast between the groups begins to disappear and the impact comparison begins to become invalid. This is sometimes called contamination or, more benignly, diffusion of treatment to control.

Then people assigned to comparison may benefit indirectly from the treatment getting treated. So, 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.

3. When there are behavioral responses to the evaluation, not the treatment itself, responses that would not exist in the absence of the evaluation

When a program is being evaluated, participants may change their behavior because they are under observation; that is, they may respond to the program in ways they wouldn't if the program was not being evaluated. In such cases the impact estimates may capture not only the effects of the treatment but also the effects of the evaluation of the treatment.

People assigned to the comparison may start to compete with people in the treatment group. So, for example, in a program using contract teachers (treatment), the regular teachers (comparison) may work extra hard, harder than normal, during the

course of the experiment so as not to be outdone by the contract teachers. And once the experiment is over, they may revert to their normal level of effort. Competition makes the outcomes of the comparison higher than normal, biasing any impact downwards. These effects are sometimes called John Henry Effects.¹

People assigned to the treatment group may also change their behavior. For example, they may react positively to the novelty of the treatment. So when a school receives new inputs, morale goes up and students and the teachers temporarily perform better. Then the novelty wears off and performance drops. Or else the innovation is disruptive. Students and teachers struggle with a new way of learning and teaching and temporarily perform worse. Either way if the evaluation period coincides with the adjustment period, impact estimates would also capture the effects of the novelty or disruption. Such effects are sometimes called Hawthorne Effects.²

Discussion Topic 1

Threats to experimental integrity

(15 minutes)

1. What does it mean to say that the groups are equivalent at the start of the program?

Answer

It means they are composed of individuals that on average have comparable characteristics.

2. Can you check if the groups are equivalent at the beginning of the program? How?

Answer

Yes, compare the means of the groups on the characteristics that are important. Same as checking if "randomization was successful"

1. These effects are called John Henry effects after an American steel driver of the late nineteenth century, who worked in Virginia laying railway track with hammers. When steam drills were introduced, threatening to make steel drivers redundant, John Henry is said to have challenged the steam engine to a drilling competition, telling his captain "A man ain't nothing but a man. Before I am bitten by that steam drill, I will die with this hammer in my hand." He won the competition, but died

"with the hammer in his hand" from overexertion. His story survives in American folk music.

2. In a study carried out at Western Electric Company's Hawthorne, USA, site in the 1930s, it was thought that workers responded to being under observation by increasing productivity. This interpretation has since been challenged but the name survives.

3. What can happen over the course of the evaluation (after conducting the random assignment) to make the groups non-equivalent?

Answer

First, people in the treatment group are treated, which would hopefully lead to some divergence in outcomes. This is what our experiment is trying to test.

Second, people can select in and out of the experimental groups, both treatment and comparison—due to attrition or partial compliance. This selection is problematic because often it happens for reasons directly related to the primary outcome(s), e.g. low-achieving students in comparison group drop out of school because they are falling behind, whereas low-achieving students in treatment group stay in school because they receive the intervention, which is intended to improve their learning.

Third, spillovers, or contamination, could also impact the comparison group indirectly. For example, the intervention could produce benefits or costs for the comparison group even though they do not directly receive the intervention (see point 2 of “Threats to experimental integrity” above). Such spillovers could produce non-equivalence between the treatment and comparison groups.

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

In the second example, the comparison is no longer based on the same people; the averages at the end are not taken on the same people as at the beginning; in a sense it reintroduces selection bias. In the third example, the control group no longer represents the counterfactual.

MANAGING ATTRITION—WHEN THE GROUPS DO NOT REMAIN EQUIVALENT

Attrition is when people join or drop out of the sample—treatment and/or comparison groups—over

the course of the experiment. One common example in clinical trials is when people die before the final outcomes are measured, earning it the term, “experimental mortality”.

Discussion Topic 2

Managing Attrition (25 minutes)

1. a. At pretest, what is the average employment outcome for each group?

	Public Intensive	Private Intensive	Control
Outcome	3	3	3

- b. At posttest, what is the average employment outcome for each group?

	Public Intensive	Private Intensive	Control
Outcome	1.875	1.5	2.25

- c. What is the impact of the program?

	Public Intensive	Private Intensive
Outcome	-0.375	-0.75

2. a. What is the impact of the program?

	Public Intensive	Private Intensive
Outcome	-0.125	-0.5

- b. Is this outcome difference an accurate estimate of impact of the program? Why or why not?

NO, it is not an accurate estimate because it omits the drop-outs. The jobseekers who dropped out were worse off than the average jobseeker, and this reason for dropping out (because they had no employment) is correlated with the treatment.

3. If it is not an accurate, does it overestimate or underestimate the impact?

Underestimates by 0.25

4. How can we get a better estimate of the program's impact?

Follow up the whole lot of them (Intention to treat—take the average on the same people at the beginning and at the end—compare the averages of treatment and comparison based on the original assignments).

5. In Case 2, you learned about other methods to estimate program impact, such as pre-post, simple difference, differences in differences, and multivariate regression.
6. Does the threat of attrition only present itself in randomized evaluations?

Answer

This question is supposed to show that the threats are general to all methods of estimating impact. It is to make sure that people don't leave thinking that these threats exist only if you are doing a randomized evaluation.

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 do 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?

Suppose 10,000 of the 40,000 jobseekers who were offered each treatment were not interested in receiving job counseling because they were intrinsically demotivated. Since, the 10,000 jobseekers who did not take-up the program were also not motivated to look for a job in the first place, they remained unemployed at the end of the year.

Discussion Topic 3

Managing partial compliance

(25 minutes)

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

Answer

	Public Intensive	Public Intensive
Posttest	2.125	2
Control	2.25	2.25
Impact	-0.125	-0.25

2. 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?

Answer

This estimate provides a measure of the effect of the program as a whole, not accounting for the fact that not everyone complied with the planned intervention protocol. This is referred to as the “intention to treat” (ITT) estimate.

Ultimately, it depends what you want to learn about. ITT may relate more to how programs are actually implemented on the ground. For example, we may not be interested in the added effect of intensive counseling, but what would happen under an actual intensive counseling program.

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

3. Five of your colleagues are passing by your desk; they all agree that you should calculate the effect of the treatment using only the 0,000 jobseekers who were treated.

- a. Is this advice sound? Why or why not?

Answer

This advice is not sound. The question that must be asked is, how are the jobseekers you exclude different from the average child? In this case they have below average (worse) employment outcomes and excluding them introduces attrition and selection bias, thereby producing non-equivalence between the treatment and comparison groups.

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

Answer

This advice is also not sound. It does not stick to the original assignments; the suggested manipulation reintroduces selection bias, by re-categorizing jobseekers with worse employment outcomes from the treatment group into the comparison group. Again, this produces non-equivalence between the two groups.

5. 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).

- a. Is this advice sound? Why or why not?

Answer

This is sound advice. You are still using the original assignments from your randomization, so unlike the first two methods, you are not reintroducing selection bias.

MANAGING SPILLOVERS—WHEN THE COMPARISON, ITSELF UNTREATED, BENEFITS FROM/ GETS HARMED BY THE TREATMENT BEING TREATED

This section is just meant for a discussion about spillovers and how they can render the treatment and control groups non-equivalent and thus bias our estimates. The following discussion is included in the case study:

People assigned to the control group may benefit or get harmed indirectly by those receiving treatment. In Case 3, how to randomize, we were concerned about such spillovers in the job-placement program when we chose the level of randomization. Specifically, we were concerned that the because of counseling, those in the treatment group were taking opportunities away from individuals in the control group. Alternatively, we could imagine a situation in which spillovers are positive. Increased employment in the treatment

group could improve the local economy, making it easier for control group jobseekers to find jobs. Or perhaps jobseekers in the control group had contacts in the treatment group and were now better connected to potential employers. In any of these cases, the control group would no longer represent the counterfactual—the state of the world had the program not been implemented.