

CASE STUDY 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.

KEY VOCABULARY

Counterfactual: what would have happened to the participants of a program had they not participated. The counterfactual cannot be observed from the **treatment group**; can only be inferred from the comparison group.

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

Attrition: individuals selected for a study drop out of the treatment or comparison group over the course of the study, before the final outcomes are measured.

Attrition Bias: statistical bias, which occurs when both: (a) individuals drop out of either the treatment or the comparison group, and (b) who drops out is correlated with which group they are assigned.

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

Intention to Treat: measured impact of a program that compares outcomes from all individuals assigned to the treatment group to those assigned to the control group (regardless of whether they actually availed of the treatment).

Treatment on the Treated: the estimated impact of a program on participants who participated in the program solely because they were assigned to the treatment group. (Requires some assumptions.)

Externality: an indirect cost or benefit incurred by individuals who did not directly receive the treatment. Also termed "spillover."

INTRODUCTION

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.

How can common threats to experimental integrity be managed?

PUBLIC COUNSELING VS. PRIVATE COUNSELING

Since the 1980s, the French labor market has been characterized by high unemployment rates and persistent long-term unemployment. At the beginning of the two programs, unemployment was decreasing but still 8.4 % of the labor force was unemployed; among them, about 30% had been unemployed for at least a year.

In the last case study we saw that supplemental private counseling was provided to some individuals whereas the rest went through the regular track of public counseling. The evaluators of that program found that more intensive private counseling did indeed make individuals better able to find employment that lasted at least 6 months, compared to those who received the status quo public counseling. However, it isn't clear whether we were measuring the impact of: (1) a more intensive program versus a less intensive program, or (2) the relative effectiveness of the public sector versus private sector, or (3) some combination of the two. In this new experiment, we are able to directly compare the relative effectiveness of the intense public program versus the intense private program.

The intensive counseling program mandated by Unédic was called the “private scheme” since it was provided by private agencies, while the intensive program offered by the public unemployment agency, ANPE, was referred to as the “public scheme”. Both programs had different content depending on the region they were offered in but the basic structure was the same. In both schemes there was a more intensive follow-up, with at least a weekly contact (email, phone) and a monthly face-to-face meeting between the jobseeker and his personal counselor. The regular public track (the control group), on the other hand, only required one contact every month.

EVALUATION DESIGN

The individuals were randomized during their first interview at the ANPE, after they were deemed eligible for the different programs. Consequently, they were

either assigned to treatment 1 (public scheme), treatment 2 (private scheme) or the control group (regular track).

Upon randomization, the jobseekers were informed about the track they were offered. If they were assigned to one of the two intensive counseling schemes, they were contacted by a counselor from the ANPE or by one of the private firms, respectively. The jobseekers were free to refuse the more intensive tracks, in which case they would have to enroll in the regular track. The take up of the program was far from complete and posed significant challenges in evaluating the effectiveness of the two programs.

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 in the average outcomes 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, we 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?

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

You are looking at the employment outcomes of intensive counseling for jobseekers. Employment outcomes are scaled as follows:

- Unemployment = score of 3
- Temporary employment = score of 2
- Permanent employment = score of 1

There are 120,000 jobseekers: 40,000 enrolled in the public scheme, 40,000 enrolled in the private scheme and 40,000 in the regular track. After you randomize, the treatment and comparison groups are equivalent, meaning jobseekers from each of the three categories are equally represented in all groups.

Suppose all jobseekers who are in the treatment receive intensive counseling and none of the jobseekers in the comparison receive do. Further suppose that all jobseekers that find either temporary or permanent employment. Finally, suppose that some jobseekers in the regular track find employment and others remain unemployed at the end of the year. The employment outcomes for jobseekers in each group are shown for both the pretest and posttest.

TABLE 1

Outcome	Prefest			Posttest		
	Public	Private	Control	Public	Private	Control
3	40,000	40,000	40,000	5,000	5,000	20,000
2	-	-	-	25,000	10,000	10,000
1	-	-	-	10,000	25,000	10,000
Total	40,000	40,000	40,000	40,000	40,000	40,000

1. Using the table above, calculate the following:
 - a. At pretest, what is the average employment outcome for each group?
 - b. At posttest, what is the average employment outcome for each group?
 - c. What is the impact of the program?

Suppose now that in the regular track, half of the jobseekers who remain unemployed and half of those who are temporarily unemployed at the end of the year feel ashamed and refuse to respond to the survey. The employment outcomes for jobseekers in each group are shown for both the pretest and posttest.

TABLE 2

Outcome	Pretest			Posttest		
	Public	Private	Control	Public	Private	Control
3	40,000	40,000	40,000	5,000	5,000	10,000
2	-	-	-	25,000	10,000	5,000
1	-	-	-	10,000	25,000	10,000
Total	40,000	40,000	40,000	40,000	40,000	25,000

2. Using the table above, calculate the following:
 - a. What is the impact of the program?

Is this outcome difference an accurate estimate of the impact of the program? Why or why not?
3. If it is not accurate, does it overestimate or underestimate the impact?
4. How can we get a better estimate of the program's impact?
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?

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. Those randomly assigned to a more intensive job counseling program may choose not to enroll. Or those assigned to the control group may decide they want intensive counseling anyway and pay out of pocket to get this service from a private agency. 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 we have no control over what they decide to do over the course of the experiment. All we can do is plan your experiment and offer them treatments. How, then, can we deal with the complications that arise from partial compliance?

Discussion Topic 3

Selection Bias Due to Incomplete Take-up

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.

TABLE 3

Outcome	Prefest			Postfest		
	Public	Private	Control	Public	Private	Control
3	40,000	40,000	40,000	15,000	15,000	20,000
2	-	-	-	15,000	10,000	10,000
1	-	-	-	10,000	15,000	10,000
Total	40,000	40,000	40,000	40,000	40,000	40,000

1. Calculate the impact estimate based on the original group assignment.
2. This is one potential method of evaluating the impact of the program. In what ways is it useful and in what ways is it not useful?

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 20,000 jobseekers who were treated.
 - a. Is this advice sound? Why or why not?
4. Another colleague says that it is 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?
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 treatment ratios (i.e. proportions of each experimental group that received the treatment).
 - a. Is this advice sound? Why or why not?

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

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 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.