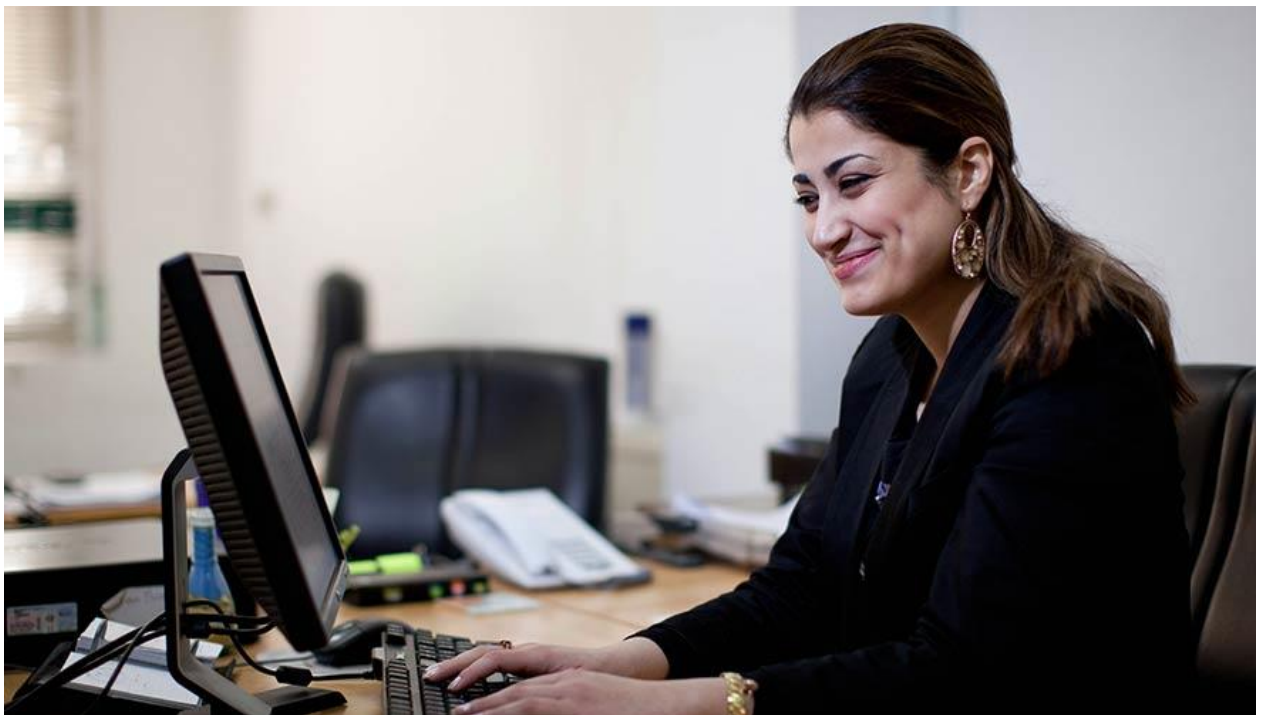# CASE STUDY 4:
WAGE SUBSIDY PROGRAM IN JORDAN

Understanding Threats to Experimental Integrity

This case study is based on Matthew Groh, Nandini Krishnan, David McKenzie, and Tara Vishwanath: "Soft Skills or Hard Cash? The Impact of Training and Wage Subsidy Programs on Female Youth Employment in Jordan", *The World Bank Policy Research Working Paper* 6141, July 2012.

J-PAL thanks the authors for allowing us to use their paper as a teaching tool.

KEY VOCABULARY

Treatment assignment, Treatment status	An individual's treatment assignment is the group they were randomly assigned to: the treatment or comparison group. An individual's treatment status is what actually happened to them: were they treated or not?
Attrition	Attrition is an occurrence when individuals or groups leave the study. This can happen for many reasons: they move away from the study area, they no longer wish to participate, they are absent on the surveyors' attempt to survey them, and many more. What is key to note is that if a unit *attrits*, they do not appear in your data—regardless of their treatment status and their outcome. *Random* attrition is a concern because it reduces your sample size, which all else equal, makes it harder to detect differences between treatment and comparison groups. *Non-random* attrition, or when certain groups are more likely to attrit than others, is a larger concern, because it introduces selection bias (described below) in your study sample.
Balance	Randomization creates two groups that on average look very similar. This can be tested by collecting some baseline demographic information—such as age, gender, years of education, income, etc.— and comparing the average value of these characteristics in the treatment group to the average value of them in the comparison groups. Even when randomization is done correctly, some of these average values will be different; however, this reflects differences that occur by chance. We say the comparison and treatment groups balance if they have similar average values for baseline characteristics.
Selection bias	Selection bias is bias that occurs when the individuals who receive the program are systematically different from those who do not. Consider an elective after school tutoring program. Is it effective at raising children's exam scores? If we compare those who take up the tutoring program to those who don't, we will get a biased estimate of the effect of the tutoring program, because those who chose to take it up are likely different from those who don't. The two groups likely do not balance (for example, those who took it up may be more motivated, or they may be weaker students). Randomization removes selection bias because it breaks the link between characteristics of the individual and their treatment status.
	Selection bias can occur in other ways in a randomized evaluation. For example:
	- Participants can choose to take up a treatment or refuse it
	- Participants can choose to leave the study
	- Surveyors can choose to only survey the closest houses
	If any of these scenarios happen disproportionately in either the treatment or control groups, the result is selection bias.
Attrition bias	Attrition bias is a type of selection bias that occurs when people choose to leave the study. This can bias the estimate of the treatment impact in two ways:
	1. It may be the case that people with certain characteristics (say, those with the highest levels of education) in both the treatment and comparison groups leave. This means your study population looks less like the general population. The treatment effect you estimate might not represent the true effect for the general population.
	2. The reasons people choose to leave may be correlated with the treatment. Suppose some of the treatment group finds your job training classes to be too difficult and leave the study. This could mean that workers who have higher levels of ability or motivation are more likely to receive the training, which would create bias in your results.
Compliance	Any study sample can be split into three distinct groups:
	1. **Compliers:** This group of people will follow their assignment status. If they are assigned to the treatment group, they will take up the treatment; if they are assigned to the control group they will not take up the program.
	2. **Always-takers:** This group of people will always take up the program, regardless of assignment status.
	3. **Never-takers:** This group of people will never take up the program, regardless of assignment status.
	When respondents do not comply with their treatment assignment, the study has *partial compliance*. In the treatment group, the people who do not comply are never-takers, while in the comparison group, those who do not comply are always-takers. We collectively refer to those who do not comply as *non-compliers*, and the action of not complying with treatment status as *non-compliance*.

Intention-to-treat (ITT):	The ITT is a method for estimating the effect of the *program* where you compare the average outcomes of those *assigned to* the treatment group to the average outcomes of those *assigned to* the comparison group, regardless of whether individuals within those groups have actually received the treatment (also known as treatment status). The ITT measures the impact of delivering a program in the real world, where some people don't take up the program when they are supposed to, and others do take up the program when they are not supposed to.
Local Average Treatment Effect (LATE)	The LATE is a method for estimating the effect of the program *on those who complied with their treatment status*. The LATE divides the ITT by the difference in the proportion of treatment group who took up the program and the proportion of the comparison group who took up the program. Recall that the ITT compares the average outcome of the treatment group to that of the comparison group. This means that under partial compliance, the average changes we measure in the treatment group will be diluted by changes in outcomes among those who did not take it up. Intuitively, you should think of the LATE as a way of adjusting the ITT to reflect that not all of those assigned to treatment were treated while some who were assigned to the comparison group were treated.
Spillovers	Spillovers occur when one individual's action of taking up a treatment impacts another individual, regardless of that individual's assignment status. An illustrative example of spillovers are vaccines: If you are randomly assigned to be offered a vaccine—*and you choose to take it up*—you reduce the risk of others around you contracting the disease. It does not matter if the people around you are vaccinated or not—or even if they are in the study—the fact that you took up the treatment has impacted them.

LEARNING OBJECTIVE

To explore how common threats to experimental integrity can influence the effect of a program.

SUBJECTS COVERED

Balance, attrition, selection bias, compliance, spillovers, intention-to-treat effect (ITT), local average treatment effect (LATE).

INTRODUCTION

Throughout the Middle East, unemployment rates of educated youth have persistently been high, and female labor force participation has been low. Only 23% of female community college graduates in Jordan are employed 16 months after graduating, despite 93% saying they want to work at the time of graduation. This enormous gap between expectations and reality highlights the challenge facing young women who want to work in the Middle East.

The problems faced by young women in the labor market are twofold. First, firms are often reluctant to hire younger workers, regardless of gender, since they lack experience and are of untested quality. Second, employers have qualms about hiring women because they believe that they are less committed to their jobs and might leave if they get married or have children.

The Jordan New Opportunities for Women (Jordan NOW) designed an intervention to get at these labor market frictions: **wage subsidy vouchers** to reduce the cost of employing women.[1] Employers may see females as having a higher probability of leaving early, which lowers any estimated returns from training them and from the experience females accumulate over their tenure with the employer. If the expected benefit is lower, wage subsidies can keep the expected return of investing in female employees positive by partially offsetting the costs of employing them.

A randomized evaluation was designed to test the effectiveness of the intervention. Female students from community colleges who had passed their second year exams were randomized into a treatment group and a comparison group. Graduates assigned to the wage subsidy program were given a non-transferrable job voucher that they could take to a firm while searching for jobs. The voucher paid the employer the minimum monthly wage for a maximum of six months if they hired the worker.

This case study will take us through different threats to experimental integrity. It draws from the evaluation but incorporates hypothetical examples not present in the paper.

---

[1] A second intervention (skills training) was also tested, but for this case study we will only use the wage subsidy program.

Randomization creates groups that on average are balanced at the start of the intervention. However, external influences can make them unbalanced at the end of the program. People within each group may not comply with the treatment to which they were assigned, or we may lose track of some of them before the post-intervention outcomes are measured. These events can potentially reintroduce selection bias, diminishing the validity of the impact estimates, and are threats to the integrity of the experiment.

DISCUSSION TOPIC 1: BALANCE BETWEEN GROUPS

1. Can you check if the groups are balanced at the beginning of the program? How?

The following table is the baseline balance table of the study in Jordan:

Table 1: Baseline Balance Test

(1) Characteristic:	(2) Treatment average	(3) Comparison average	(4) Difference (Treat. − Comp.)
Age	21.1	21.3	-0.2
Married	0.16	0.13	0.03
Mother works	0.06	0.06	0
Father works	0.61	0.53	0.08
Has previously worked	0.18	0.16	0.02
Has taken English training	0.26	0.30	-0.04
Household owns car	0.66	0.64	0.02
Household has Internet	0.18	0.26	-0.08*
Prefers Gvt work to Private sector	0.81	0.81	0
Sample size	300	449	
Notes:	Standard errors in parenthesis. Stars indicate statistical significance: * =0.10, ** =.05. ***=0.01		

2. Are there any characteristics for which the treatment and control groups are different? If so, which ones? Which differences are statistically significant? Are you worried that they indicate the groups are not balanced?

3. Can you check if the groups are balanced at the end of the program? How might this be different from checking in the beginning?

DISCUSSION TOPIC 2: UNDERSTANDING ATTRITION

Attrition is an occurrence when people drop out of the sample over the course of the experiment. Attrition is a concern for several reasons:

First, attrition—whether in the treatment or comparison group—reduces the sample size in the study. Barring any other changes to the study design, a smaller sample size makes it harder to detect the effect of the program.

Second, attrition can cause bias. This bias can arise when certain types of people leave the study (e.g., those who live farthest from the village center, those from the richest households, etc.). If a specific type of person leaves the study in *both* the treatment and comparison group, then the study remains unbiased but the sample looks less like the general population, meaning the results of the study are harder to generalize. Conversely, if a specific type of person disproportionally leaves from either the treatment or comparison group, then the two groups will no longer be biased, leading to selection bias in the estimation of the treatment effect.

Suppose there are 600 female graduates randomized into treatment and comparison groups (300 in each group).[2] Suppose all jobseekers in the treatment group use their wage subsidy vouchers and, because these vouchers are non-transferrable, none of the jobseekers in the comparison group do so. The employment situation[3] for jobseekers in each group are shown for both baseline and endline.

Table 2: Employment situation at baseline and endline

Outcome	Baseline		Endline	
	Treatment	Comparison	Treatment	Comparison
Unemployed	300	300	100	200
Employed	0	0	200	100
Sample Size	300	300	300	300

1. Using the table above, calculate the following:

a. At baseline, what is the employment rate for each group?

b. At endline, what is the employment rate for each group?

c. What is the impact of the program on jobseekers employment rate?

[2] We saw in table 1 that there were more individuals in the comparison group, but we changed the figure to make calculations easier.

[3] For the purpose of the case study, employment situation is either "employed" or "unemployed", without distinguishing between full-time and part-time jobs.

Suppose now that in the comparison group, half of the jobseekers who remain unemployed at the end of the year feel disillusioned and refuse to respond to the survey. The employment situation for jobseekers in each group are shown for both baseline and endline:

Table 3: Employment situation at baseline and endline with attrition in the comparison group

	Baseline		Endline	
Outcome	Treatment	Comparison	Treatment	Comparison
Unemployed	300	300	100	100
Employed	0	0	200	100
Sample Size	300	300	300	200

2. Using the table above, calculate the following:

a. What is the impact of the program?

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

c. If it is not accurate, does it overestimate or underestimate the impact? By how much?

d. Does this threat of attrition only present itself in randomized evaluations?

DISCUSSION TOPIC 3: UNDERSTANDING PARTIAL COMPLIANCE

Some people assigned to the treatment group may in the end not actually get treated, either because the program is not implemented properly or because they choose not to enroll. Similarly, some people assigned to the comparison could end up being treated. This is called "partial compliance" or "diffusion" or, less benignly, "contamination." How, then, can we deal with the complications that arise from partial compliance?

Reminder of table 2: Employment situation at baseline and endline

	Baseline		Endline	
Outcome	Treatment	Comparison	Treatment	Comparison
Unemployed	300	300	100	200
Employed	0	0	200	100
Sample Size	300	300	300	300

Suppose you realize that in the treatment group, 60 out of the 100 women that remain unemployed at endline didn't use the voucher because they were intrinsically demotivated and didn't want to get a job.

1. Imagine you compare the employment situation of those *assigned to* the treatment group to the employment situation of those *assigned to* the comparison group, regardless of the treatment status of the individuals within those groups. What is the impact of the treatment?

2. How is called this estimate of the impact of the program? In what ways is it useful and in what ways is it not useful?

You are now interested in learning the effect of treatment on those actually treated ("Local Average Treatment Effect" (LATE) 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 240 women jobseekers who were treated, and dropping the 60 that didn't take up the program. 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. 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). Is this advice sound? Why or why not?

6. Use your estimate of the ITT from question 1 to estimate the LATE, as follows:

$$LATE = \frac{ITT}{\%\ of\ treatment\ group\ who\ took\ up\ treatment - \%\ of\ comparison\ group\ who\ took\ up\ treatment}$$

7. Is the LATE bigger or smaller than the ITT? Does that surprise you?

8. In what ways LATE estimate is useful, and in what ways is it not useful?

DISCUSSION TOPIC 4: UNDERSTANDING SPILLOVERS

Spillovers occur when one individual's action of taking up a treatment impacts another unit, regardless of that unit's assignment status. Spillovers are tricky to measure—they can often occur in people outside the study design, who you don't survey, or can occur in the comparison group, which reduces the measured treatment effect. Spillovers are not inherently good or bad, but they change the way we think of a program's effectiveness.

1. In the case of our voucher program, can you think of positive spillovers? Describe how they could happen.

2. Can you think now of negative spillovers? Describe how they could happen.

3. What are the two strategies that a research team can use regarding spillovers? At what stage of the project should they be conceived and implemented?

REUSE AND CITATIONS

To request permission to reuse this case study or access the accompanying teachers' guide, please email training@povertyactionlab.org. Please do not reuse without permission. To reference this case study, please cite as:

Abdul Latif Jameel Poverty Action Lab (J-PAL). 2019. "Case Study: Wage Subsidy Program in Jordan: Understanding Threats to Experimental Integrity." J-PAL Case Study. Last Modified 19 November, 2019.