CASE STUDY 2: VOCATIONAL TRAINING FOR DISADVANTAGED YOUTH

Why Randomize?

Photo: Shutterstock.com

This case study is based on:

> *Attanasio, Orazio, Adriana Kugler, and Costas Meghir. 2011. "Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial." American Economic Journal: Applied Economics, 3 (3): 188-220.DOI: 10.1257/app.3.3.188*

J-PAL thanks the authors for allowing us to use their paper and for sharing their data.

LEARNING OBJECTIVES

To understand the different methods commonly used to estimate the impact of a given intervention, and to understand their strengths, weaknesses, and underlying assumptions.

SUBJECTS COVERED

Causality, counterfactual, impact, comparison groups, selection bias, omitted variables, randomization, and balance.

KEY VOCABULARY	
Comparison Group:	A group that is as similar as possible to the treatment group in order to be able to learn about the counterfactual. In an experimental design, the comparison group (also called the control group) is a group from the same population as the treatment group that, by random assignment, is not intended to receive the intervention.
Counterfactual:	What would have happened to the participants of an intervention had they not received the intervention. The counterfactual can never be observed; it can only be inferred from a comparison group different from the treatment group.
Estimate:	In statistics, a "best guess" about an unknown value in a population (such as the effect of a program on an outcome) according to a rule (known as the "estimator") and the values observed in a sample drawn from that population.
Impact:	The impact of the intervention is the effect of the treatment on the whole population. The impact is estimated by measuring the differences in outcomes between the treatment group and its counterfactual, i.e., by measuring the difference in outcomes between treatment and comparison groups.
Omitted Variable Bias:	Bias that occurs when relevant (and often unobservable) variables/characteristics are left out of the regression analysis (also known as *confounders)*. When these variables predict both the outcome and participation in an intervention, their omission can lead us to incorrectly over or under estimate the impact of the program. For example, omitting socioeconomic status, which is correlated with test scores, could lead to overestimating the impact of a tutoring intervention on a group of wealthy students.
Treatment Group:	The group that receives the intervention.
Selection Bias:	Selection bias is bias that occurs when the individuals who receive the program are systematically different from those who do not. For example, 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 participate are likely different from those who don't (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, consider a situation where an intervention is making a phone call to a landline:

- Callers may be unable to reach certain participants (for example, participants in rural areas may have poor cell phone service and may be more likely to have landlines than those in urban areas).

- Some participants may be less likely to pick up the phone depending on the time of day they are called (for example, calling a home phone during standard business hours).

INTRODUCTION

All around the world, many young people struggle to find stable employment in both developed and developing countries. It is estimated that by the end of 2010, around 75.1 million young people worldwide were unemployed (ILO). Youth unemployment is commonly blamed on a lack of skills, especially in developing countries where education systems fail to equip young people with the skills they need to get a stable job.

In 2001, the Colombian government started a vocational training program for disadvantaged youth in its seven largest cities to tackle the problem of youth unemployment. The training program included three months of in-classroom training and three months of on-the-job training for people between the ages of 18 and 25[1]. The classroom training was provided by private institutions selected through a competitive bidding process, while the on-the-job training was provided by legally registered companies operating in various sectors, including manufacturing, retail and trade, and services.

WHAT IS THE IMPACT OF VOCATIONAL TRAINING?

What is required in order for us to measure whether the vocational training worked – whether it had any impact on the probability of employment of participating youth?

In general, to ask if a program works is to ask if the program achieves its goal of changing certain outcomes for its participants, and ensure that those changes are not caused by some other factors or events happening at the same time. To show that the program causes the observed changes, we need to simultaneously show that if the program had not been implemented, the observed changes would not have occurred (or would have been different). But how do we know what would have happened? If the program happened, it happened. Measuring what would have happened requires entering an imaginary world in which the program was never given to these participants. The outcomes of the same participants in this imaginary world are referred to as the counterfactual. Since we cannot observe the true counterfactual, the best we can do is to estimate it by mimicking it.

1 While both men and women participated in the program, the sample of men in the evaluation was not balanced at the baseline, so we present data only for women.

The key challenge of program impact evaluation is constructing or mimicking the counterfactual. We typically do this by selecting a group of people that resemble the participants as much as possible but who did not participate in the program. This group is called the comparison group. Because we want to be able to say that it was the program and not some other factor that caused the changes in outcomes, it is important that the only difference between the comparison group and the participants is that the comparison group did not participate in the program. We then estimate "impact" as the difference observed at the end of the program between the outcomes of the comparison group and the outcomes of the program participants.

The impact estimate is only as accurate as the comparison group is successful at mimicking the counterfactual. If the comparison group poorly represents the counterfactual, the impact is (in most circumstances) poorly estimated. Therefore, the method used to select the comparison group is a key decision in the design of any impact evaluation.
That brings us back to our question: What impact does a vocational training program have on the probability of employment of disadvantaged youth in Colombia?

In this case, the intention of the program is to equip participating youth with skills valued by employers and the outcome measure is probability of employment. Asking if the training program "worked" is to ask if it increased the probability that participating youth would be employed following the program. The impact is the difference between the probability of employment of those who participated in the program to what that probability of those same participants would have been had they not participated in the training program.

What comparison groups can we use? The following experts illustrate different methods of evaluating impact. (Refer to the table on the last page of the case for a list of different evaluation methods).

ESTIMATING THE IMPACT OF VOCATIONAL TRAINING

METHOD 1:

Newspaper Article: Huge Gains for Women in Training Program

Statistics released today by a government agency indicate that the government-sponsored vocational training program, which has been running since 2001 in the seven largest cities of Colombia, increased the probability of employment of participating women by 49.66 percent, a huge and important gain for young disadvantaged women. Before participating in the program, women were only 46.92 percent likely to be employed, and when these women were surveyed several months after completing the training program, they were 70.22 percent likely to have a job. These numbers provide evidence in support of vocational training programs, which governments all over the world have adopted to resolve the pressing problem of youth unemployment. Governments should take note of these results and start training programs or scale up existing ones

Table 1: Employment differences

	Mean	Standard Error
Baseline employment	46.92%	.017
Endline employment	70.22%	.015
Difference	23.30***	49.66% increase

Note: Statistically significant at the 95 percent level. Sample size: 910 women.

DISCUSSION TOPIC 1

1. What type of evaluation does this opinion piece imply?

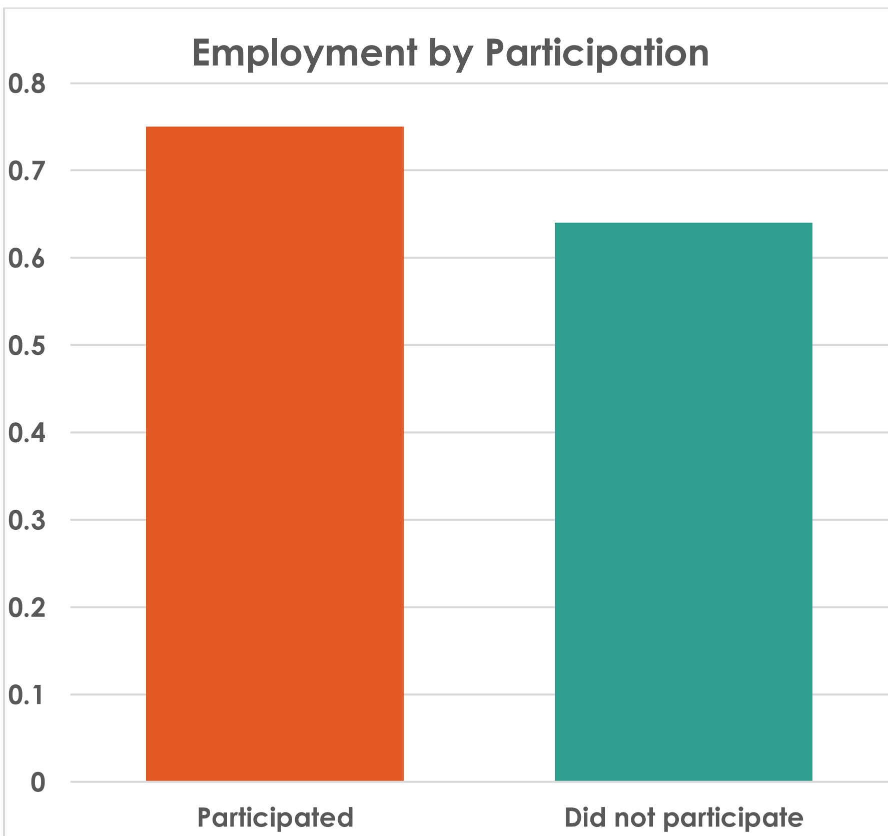
2. What represents the counterfactual?

3. What are the problems with this type of evaluation?

METHOD 2:

Letter to the Editor: Let's Not Jump to Conclusions

Newspapers tend to exaggerate many claims and this is exactly what the article "Huge Gains for Women in Training Program" did last week when reporting about the impact of the government's vocational training program. As an economist interested in labor markets, I have been following this training program since the government first announced it. Obviously, I hoped that the program would work and I am really happy to see positive results coming out from it. But the claims that the program had such a massive impact are very misleading. After all, many things could have happened to these women between the start and end of the training program. The Colombian economy has been experiencing healthy growth rates since 2002 and cities across the country have become safer. These confounding aspects could affect the results of the program's evaluation, so we should get rid of these and focus instead on how women who participated in the training compare to women who did not participate in the training. I've gone ahead and done this calculation. You will see that this shows that the program increased the probability of employment of trained women by 10 percent, a far cry from the almost 50 percent increase claimed by the article, but still an increase nonetheless.



DISCUSSION TOPIC 2

1. What type of evaluation does this opinion piece imply?

2. What represents the counterfactual?

3. What are the problems with this type of evaluation?

METHOD 3:

Donor Report: Comparing apples to apples

The government's vocational training program has received a lot of press coverage recently. Some have claimed that the program has an enormous impact, while others argue that the impact is significantly more moderate. This report seeks to provide a more accurate measure of the impact of the program using a more appropriate method. Previous analyses have used the wrong metrics to calculate the training program's impact – possibly overestimating by how much the probability of employment is actually increased by the program. For instance, if you compare the probability of employment of those women who participated in the training program and those who did not, you might be introducing selection bias into the estimate. These two groups of women might be very different for many reasons beyond just participating or not in the training program.

What you need to do to get a more accurate estimate is to compare changes in the probability of employment of the two groups. This way, we can see how fast the probability of employment changes for each group. When we repeat the analysis using this more appropriate outcome measure, we see that women participating in the program experienced an increase in their probability of employment of 5.85 percent, showing that participating in a vocational training program does increase probability of employment, but not by the magnitudes claimed by other analyses.

DISCUSSION TOPIC 3

1. What type of evaluation does this opinion piece imply?

2. What represents the counterfactual?

3. What are the problems with this type of evaluation?

METHOD 4:

The numbers don't lie, unless your statisticians are asleep

Over the last few weeks, the public has received conflicting information about the impact of the Colombian government's vocational training program. Those who support the program assert that vocational training successfully equips young women with valuable skills, resulting in a substantially higher chance of being employed. Others, however, believe that this impact is grossly inflated and that actual gains are more modest, and perhaps driven by external factors and not the vocational training itself.

Unfortunately, both camps are using flawed instruments of analysis and the question of whether vocational training increases the chance of getting a job among women remains unanswered. This report uses sophisticated statistical methods to measure the true impact of the vocational training program. We are concerned with other factors that might influence the results. As a result, we carried out a survey to collect information about age, marriage status, education levels, and the city where participants lived. All these variables can potentially affect the employability of the person, so our analysis controls for them, allowing us to separate out the true effect of the vocational training.

Table 2: Probability of Employment

	(1)	(2)
Training	0.065 ** (0.022)	0.057* (0.022)
Age		0.004 (0.005)
Marriage		-0.066* (0.026)
Education Level		0.007 (0.006)
City		-0.036*** (0.005)
Constant		0.63 ** (0.14)

Looking at Table 2, we notice that the results change and our impact estimate drops when we control for additional variables. The results from column (1) suggest that the training program increased the probability of employment by 6.5 percent – this is significant at the 10 percent level. If we look at column (2), which includes controls for confounding variables, the impact is diminished to 5.7 percent, significant at the 10 percent level as well. More importantly, however, marriage and city are both significant as well (though in the opposite direction).

By controlling for variables that can affect chances of employment, we discover that the actual impact of the training program is modest. While this increase indicates that vocational training is no panacea for youth unemployment, it is still an increase that can make a difference in the lives of many.

DISCUSSION TOPIC 4

1. What type of evaluation does this opinion piece imply?

2. What represents the counterfactual?

3. What are the problems with this type of evaluation?

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:

> J-PAL. "Case Study: Vocational Training for Disadvantaged Youth" Abdul Latif Jameel Poverty Action Lab. 2020. Cambridge, MA.

The original paper can be cited as follows:

> Attanasio, Orazio, Adriana Kugler, and Costas Meghir. 2011. "Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics*, 3 (3): 188-220.DOI: 10.1257/app.3.3.188

	Method	Description	What assumptions are required, and how demanding are the assumptions?	Required data
Randomization	Randomized Evaluation/ Randomized Control Trial	Measure the differences in outcomes between randomly assigned program participants and non-participants after the program took effect.	*The outcome variable is only affected by program participation itself, not by assignment to participate in the program or by participation in the randomized evaluation itself*. Examples for such confounding effects could be information effects, spillovers, or experimenter effects. As with other methods, the sample size needs to be large enough so that the two groups are statistically comparable; the difference being that the sample size is chosen as part of the research design.	Outcome data for randomly assigned participants and non-participants (the treatment and control groups).
Basic Non-Experimental Comparison Methods	Pre-Post	Measure the differences in outcomes for program participants before the program and after the program took effect.	*There are no other factors (including outside events, a drive to change by the participants themselves, altered economic conditions, etc.) that changed the measured outcome for participants over time besides the program*. In stable, static environments and over short time horizons, the assumption might hold, but it is not possible to verify that. Generally, a diff-in-diff or RDD design is preferred (see below).	Data on outcomes of interest for program participants before program start and after the program took effect.
	Simple Difference	Measure the differences in outcomes between program participants after the program took effect and another group who did not participate in the program.	*There are no differences in the outcomes of participants and non-participants except for program participation*, and both groups were equally likely to enter the program before it started. This is a demanding assumption. Non-participants may not fulfill the eligibility criteria, live in a different location, or simply see less value in the program (self-selection). Any such factors may be associated with differences in outcomes independent of program participation. Generally, a diff-in-diff or RDD design is preferred (see below).	Outcome data for program participants as well as another group of non-participants after the program took effect.
	Differences in Differences	Measure the differences in outcomes for program participants before and after the program *relative* to non-participants.	*Any other factors that may have affected the measured outcome over time are the same for participants and non-participants, so they would have had the same time trajectory absent the program*. Over short time horizons and with reasonably similar groups, this assumption may be plausible. A "placebo test" can also compare the time trends in the two groups before the program took place. However, as with "simple difference," many factors that are associated with program participation may also be associated with outcome changes over time. For example, a person who expects a large improvement in the near future may not join the program (self-selection).	Data on outcomes of interest for program participants as well as another group of non-participants before program start and after the program took effect.

More advanced statistical non-experimental methods

Method	Description	What assumptions are required, and how demanding are the assumptions?	Required data
Multivariate Regression/OLS	The "simple difference" approach can be—and in practice almost always is—carried out using multivariate regression. Doing so allows accounting for other observable factors that might also affect the outcome, often called "control variables" or "covariates." The regression filters out the effects of these covariates and measures differences in outcomes between participants and non-participants while holding the effect of the covariates constant.	Besides the effects of the control variables, *there are no other differences between participants and non-participants that affect the measured outcome*. This means that any unobservable or unmeasured factors that do affect the outcome must be the same for participants and non-participants. In addition, the control variables cannot in any way themselves be affected by the program. While the addition of covariates can alleviate some concerns with taking simple differences, limited available data in practice and unobservable factors mean that the method has similar issues as simple difference (e.g., self-selection).	Outcome data for program participants as well as another group of non-participants, as well as "control variables" for both groups.
Statistical Matching	Exact matching: participants are matched to non-participants who are identical based on "matching variables" to measure differences in outcomes. Propensity score matching uses the control variables to predict a person's likelihood to participate and uses this predicted likelihood as the matching variable.	Similar to multivariable regression: *there are no differences between participants and non-participants with the same matching variables that affect the measured outcome*. Unobservable differences are the main concern in exact matching. In propensity score matching, two individuals with the same score may be very different even along observable dimensions. Thus, the assumptions that need to hold in order to draw valid conclusions are quite demanding.	Outcome data for program participants as well as another group of non-participants, as well as "matching variables" for both groups.
Regression Discontinuity Design (RDD)	In an RDD design, eligibility to participate is determined by a cutoff value in some order or ranking, such as income level. Participants on one side of the cutoff are compared to non-participants on the other side, and the eligibility criterion is included as a control variable (see above).	*Any difference between individuals below and above the cutoff (participants and non-participants) vanishes closer and closer to the cutoff point.* A carefully considered regression discontinuity design can be effective. The design uses the "random" element that is introduced when two individuals who are similar to each other according to their ordering end up on different sides of the cutoff point. The design accounts for the continual differences between them using control variables. The assumption that these individuals are similar to each other can be tested with observables in the data. However, the design limits the comparability of participants further away from the cutoff.	Outcome data for program participants and non-participants, as well as the "ordering variable" (also called "forcing variable").
Instrumental Variables	The design uses an "instrumental variable" that is a predictor for program participation. The method then compares individuals according to their predicted participation, rather than actual participation.	*The instrumental variable has no direct effect on the outcome variable. Its only effect is through an individual's participation in the program*. A valid instrumental variable design requires an instrument that has no relationship with the outcome variable. The challenge is that most factors that affect participation in a program for otherwise similar individuals are also in some way directly related to the outcome variable. With more than one instrument, the assumption can be tested.	Outcome data for program participants and non-participants, as well as an "instrumental variable".