CASE STUDY 2: VOCATIONAL TRAINING FOR DISADVANTAGED YOUTH

Why Randomize?

This case study is based on "Training Disadvantaged Youth in Latin America: Evidence from a Randomized Trial" by Orazio Attanasio, Adriana Kugler and Costas Meghir, American Economic Journal: Applied Economics 3 (July 2011)

J-PAL thanks the author for allowing us to use their paper as a teaching t

KEY VOCABULARY

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

Comparison Group: in an experimental design, a randomly assigned group from the same population that does not receive the intervention that is the subject of evaluation. Participants in the comparison group are used as a standard for comparison against the treated subjects in order to validate the results of the intervention.

Program Impact: estimated by measuring the difference in outcomes between comparison and treatment groups. The true impact of the program is the difference in outcomes between the treatment group and its counterfactual.

Baseline: data describing the characteristics of participants measured across both treatment and comparison groups prior to implementation of intervention.

Endline: data describing the characteristics of participants measured across both treatment and comparison groups after implementation of intervention.

Selection Bias: statistical bias between comparison and treatment groups in which individuals in one group are systematically different from those in the other. These can occur when the treatment and comparison groups are chosen in a non-random fashion so that they differ from each other by one or more factors that may affect the outcome of the study.

Omitted Variable Bias: statistical bias that occurs when certain variables/characteristics (often unobservable), which affect the measured outcome, are omitted from a regression analysis. Because they are not included as controls in the regression, one incorrectly attributes the measured impact solely to the program.

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.

Participating youth were given US$2.20 per day to defray transportation and lunch costs; women with children under seven years of age were given US$3.00.

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.

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.

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 questions: 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). Immediate

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

	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

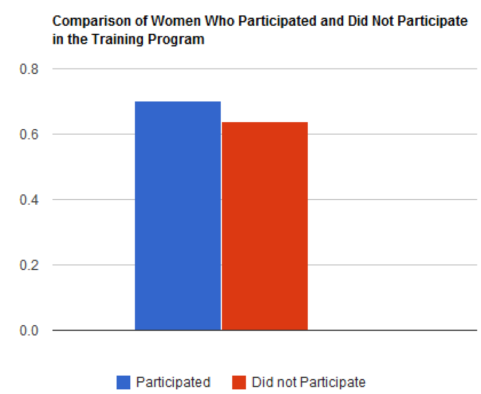
1. What type of evaluation does this opinion piece imply?
2. What represents the counterfactual?
3. What are the problems with this type of evaluation?

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 seems to be working 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.

FIGURE 1



Discussion Topic 2

1. What type of evaluation is this opinion piece using?
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 that 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 the 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 is this analysis using?
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.

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.

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)

Discussion Topic 4

1. What type of evaluation is this report utilizing?
2. What represents the counterfactual?
3. What are the problems with this type of evaluation?

Note: Data used in this case are real. The "articles" presented have been artificially produced for the purpose of the case.

Methodology	Description	Who is in the comparison group?	Required Assumptions	Required Data
Pre-Post	Measure how program participants improved (or changed) over time.	Program participants themselves—before participating in the program.	The program was the only factor influencing any changes in the measured outcome over time.	Before and after data for program participants.
Simple Difference	Measure difference between program participants and non-participants after the program is completed.	Individuals who didn't participate in the program (for any reason), but for whom data were collected after the program.	Non-participants are identical to participants except for program participation, and were equally likely to enter program before it started.	After data for program participants and non-participants.
Differences in Differences	Measure improvement (change) over time of program participants *relative to* the improvement (change) of non-participants.	Individuals who didn't participate in the program (for any reason), but for whom data were collected both before and after the program.	If the program didn't exist, the two groups would have had identical trajectories over this period.	Before and after data for both participants and non-participants.
Multivariate Regression	Individuals who received treatment are compared with those who did not, and other factors that might explain differences in the outcomes are "controlled" for.	Individuals who didn't participate in the program (for any reason), but for whom data were collected both before and after the program. In this case data is not comprised of just indicators of outcomes, but other "explanatory" variables as well.	The factors that were *excluded* (because they are unobservable and/or have been not been measured) do not bias results because they are either uncorrelated with the outcome <u>or</u> do not differ between participants and non-participants.	Outcomes as well as "control variables" for both participants and non-participants.
Statistical Matching	Individuals in control group are compared to similar individuals in experimental group.	<u>Exact matching</u>: For each participant, at least one non-participant who is identical *on selected characteristics*. <u>Propensity score matching</u>: non-participants who have a mix of characteristics which predict that they would be as likely to participate as participants.	The factors that were *excluded* (because they are unobservable and/or have been not been measured) do not bias results because they are either uncorrelated with the outcome <u>or</u> do not differ between participants and non-participants.	Outcomes as well as "variables for matching" for both participants and non-participants.
Regression Discontinuity Design	Individuals are ranked based on specific, measureable criteria. There is some cutoff that determines whether an individual is eligible to participate. Participants are then compared to non-participants and the eligibility criterion is controlled for.	Individuals who are close to the cutoff, but fall on the "wrong" side of that cutoff, and therefore do not get the program.	After controlling for the criteria (and other measures of choice), the remaining differences between individuals directly below and directly above the cut-off score are not statistically significant and will not bias the results. A necessary but sufficient requirement for this to hold is that the cut-off criteria are strictly adhered to.	Outcomes as well as measures on criteria (and any other controls).
Instrumental Variables	Participation can be predicted by an incidental (almost random) factor, or "instrumental" variable, that is uncorrelated with the outcome, other than the fact that it predicts participation (and participation affects the outcome).	Individuals who, because of this close to random factor, are predicted not to participate and (possibly as a result) did not participate.	If it weren't for the instrumental variable's ability to predict participation, this "instrument" would otherwise have no effect on or be uncorrelated with the outcome.	Outcomes, the "instrument," and other control variables.
Randomized Evaluation	Experimental method for measuring a causal relationship between two variables.	Participants are randomly assigned to the control groups.	Randomization "worked." That is, the two groups are statistically identical (on observed and unobserved factors).	Outcome data for control and experimental groups. Control variables can help absorb variance and improve "power".