

## GUIDE 2: GET OUT THE VOTE

Why Randomize?

---



This case study is based on “Comparing Experimental and Matching Methods Using a Large-Scale Field Experiment on Voter Mobilization,” by Kevin Arceneaux, Alan S. Gerber, and Donald P. Green, *Political Analysis* 14: 1-36.

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

---

## DESCRIPTION

---

In late 2002, a non-partisan civic group, Vote 2002 Campaign, ran a get-out-the-vote initiative to encourage voting in that year's U.S. congressional elections. In the 7 days preceding the election, Vote 2002 placed 60,000 phone calls to potential voters, encouraging them to "come out and vote" on election day.

## LEARNING OBJECTIVE

---

- Learn to identify evaluation methods without being told the specific method.
- Further develop the intuition of bias.
- Explore the problem of causal inference, and the various ways of estimating the impact of a program using comparison group designs.
- Introduce the concept of selection bias and how comparison group designs are only as good as their ability to get rid of selection bias
- Show how random assignment gets rid of selection bias

## SUBJECTS COVERED

---

Causality, counterfactual, impact, comparison groups, selection bias, omitted variables, randomization, equivalence and comparability

## GENERAL GUIDANCE

---

This case is fictional. These press releases and other things didn't actually happen. But this is typical of a "methodological debate" you may see over programs. One problem people may have is understanding some of the graphs or tables. Particularly the graphs in Method 1 and the regression table in Method 4. Be prepared to explain those.

## INTRODUCTION

---

The main goal of impact evaluation is to estimate what would have happened in the absence of the program. A number of methods have been used to estimate the counterfactual and to understand why experimental designs are the gold standard. We must understand the shortcomings of other approaches, which is the purpose of this case.

The main point of this case is to try to see evaluation from the perspective of 90% of IPA and JPAL's partners, and to be able to explain – IN NON-TECHNICAL LANGUAGE – why non-randomized evaluation methods may fail at estimating the counterfactual.

## ESTIMATING THE IMPACT OF THE GET OUT THE VOTE PROJECT

---

### METHOD 1

#### Discussion Topic 1 (15 minutes)

1. What type of evaluation does this news release imply?

**Answer**

Pre-post evaluation

2. What represents the counterfactual?

**Answer**

Potential voters reached BEFORE the program.

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

**Answer**

We do not know what potential voters who were reached would have done if they HAD NOT been reached. The counterfactual implies that these potential voters would have not changed their behavior throughout the course of the year. It is quite plausible, however, that potential voters would have been more likely to vote even without being reached, due to other factors. In that same election year, for example, the political atmosphere might have been more charged, which in turn could have lead more people to get out and vote.

On the contrary, it might have been that the people who were reached were all unemployed and disillusioned with the government and political process anyway that they didn't care to vote. Thus, we would see a negative impact in that case and end up attributing it to our program.

### METHOD 2

#### Discussion Topic 2 (15 minutes)

1. What type of evaluation is this opinion piece employing?

**Answer**

Post comparison of program participants and program non-participants (i.e. simple comparison of means)

2. What represents the counterfactual?

**Answer**

Potential voters who were not reached, whose voting outcomes were measured after the implementation of the program.

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

**Answer**

The potential voters who were reached may be systematically different from the potential voters who were not reached. Potential voters who were reached (had time to answer the phone call), might all be older people who spend a lot of their time at home, and thus are more likely to vote anyway. As a result, we might attribute the impact to our program, where really the higher voting rate was just due to the fact that the people who were reached would have voted anyway because they have more time and so we'll get an upward bias.

On the other hand, it might be completely possible that those people who are more likely to be at home are all unemployed bums that don't feel any civic responsibility whereas the ones that were not reached are more educated employed (and thus not likely to be at home) civic human beings who will make a special effort to go and vote. If this was the case, we would end up getting an estimate that is biased downwards.

### METHOD 3

#### Discussion Topic 3 (15 minutes)

1. What type of evaluation is this letter using?

**Answer**

Difference-in-Differences

2. What represents the counterfactual?

**Answer**

Potential voters that were not reached, whose voting outcomes were measured both before and after the implementation of the program (in order to obtain the improvement in their voting outcomes over the course of the two election periods).

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

**Answer**

The mix of the potential voters in the reached vs. not reached groups might have changed over time. Let's assume that all the people who were reached were college graduates who were looking for employment. Last year since they were in college they didn't have time to vote but this time since they are unemployed they would have gone and voted anyway. So, we would have seen a positive improvement in their voting rates regardless of the fact that they were reached. However, in the not reached group, everyone was employed both last time and this time and their voting rates didn't change by much. We'll end up finding a positive impact and attribute it to the program, whereas the difference was only due to the change in composition of one of the groups.

## METHOD 4

### Discussion Topic 4 (15 minutes)

1. What type of evaluation is this report utilizing?

**Answer**

Multivariate regression

2. What represents the counterfactual?

**Answer**

Potential voters who were not reached, controlling for (or holding constant) their age, gender, household size, district, past voting behavior etc.

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

**Answer**

Despite controlling for many confounding variables, it is likely that some (potentially unmeasured or

immeasurable) variables that are correlated with whether someone was reached or not have not been included (i.e. omitted variables bias). Examples include time spent at home, civic responsibility or other unobserved factors. The potential voters who were reached were not randomly assigned to the program, so we are unable to account for these other factors. Due to this "selection effect", our results could be biased.

## METHOD 5

### Discussion Topic 5 (15 minutes)

1. What type of evaluation is this report utilizing?

**Answer**

Matching

2. What represents the counterfactual?

**Answer**

Potential voters who were not reached, but had the same characteristics (age, gender, location, past voting behavior etc.) as the ones that were reached.

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

**Answer**

Despite matching on many confounding variables, it is likely that some (potentially unmeasured or immeasurable) variables that are correlated with whether someone was reached or not, have not been used in the matching process. Examples include time spent at home, civic responsibility or other unobserved factors. So, the potential voters who were reached were not, on average, identical to the ones that were not reached. So, unless we randomize, we cannot be sure that the two groups were equivalent, on average.

## METHOD 6

### Using randomized experiments

As you can see, not all methods give the same result. Hence, the choice of the appropriate method is crucial. The purpose of this case study was not to evaluate one particular voter mobilization campaign, but to evaluate evaluation methods in this particular context.

In the analysis of the Vote 2002 Campaign, we found that people who happened to pick up the phone were more likely to vote in the upcoming (and previous) elections. Even though we statistically accounted for some observable characteristics, including demographics and past voting behavior, there were still some inherent, unobservable differences between the two groups, independent of the get-out-the-vote campaign. Therefore, when our non-randomized methods demonstrated a positive, significant impact, this result was due to “selection bias” (in this case, selection of those who pick up the phone) rather than a successful get-out-the-vote campaign.

**TABLE 1**  
Comparing All Six Methods

Method	Estimate d impact
Pre-Post	17.9 pp*
Simple Difference	10.8 pp*
Difference-in-Differences	1.9 pp*
Multivariate Regression with Panel Data	4.6 pp*
Matching (All Covariates)	2.8 pp*
Randomized Evaluation <sup>‡</sup>	0.4 pp

**NOTES:** pp means “percentage points” and \* indicates statistically significant at the 5% level  
<sup>‡</sup> Randomized evaluation estimate is adjusted to reflect that only 25,000 of 60,000 in the treatment were treated (i.e. the Treatment on Treated effect)