



## **Case 1: Get Out the Vote Phone Calls to Encourage Voting: Do they “work”?**



**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. We thank the authors for allowing us to use their paper and for sharing their data with us.**

## **Introduction**

*Hello, may I speak with Joe Iowa please? Hi. This is Marc Shotland calling from Vote 2002, a non-partisan effort working to encourage citizens to vote. We just wanted to remind you that elections are being held this Tuesday. The success of our democracy depends on whether we exercise our right to vote or not, so we hope you'll come out and vote this Tuesday. Can I count on you to vote next Tuesday?*

In the week preceding the 2002 U.S. congressional elections, volunteers for the Vote 2002 Campaign, launched a voter mobilization drive. During this week, they called 60,000 potential voters and gave the preceding message.

Did the Vote 2002 Campaign increase the voter turnout rate? How would we find out? This case study tackles these questions by examining different methods people may use to evaluate the effect of a program or intervention. While the context of this case study is electoral politics in the U.S., the issues that arise here are also present when evaluating the effect of social programs in the context of developing countries.

## **I - Background**

### **Historical Voter Turnout in the US**

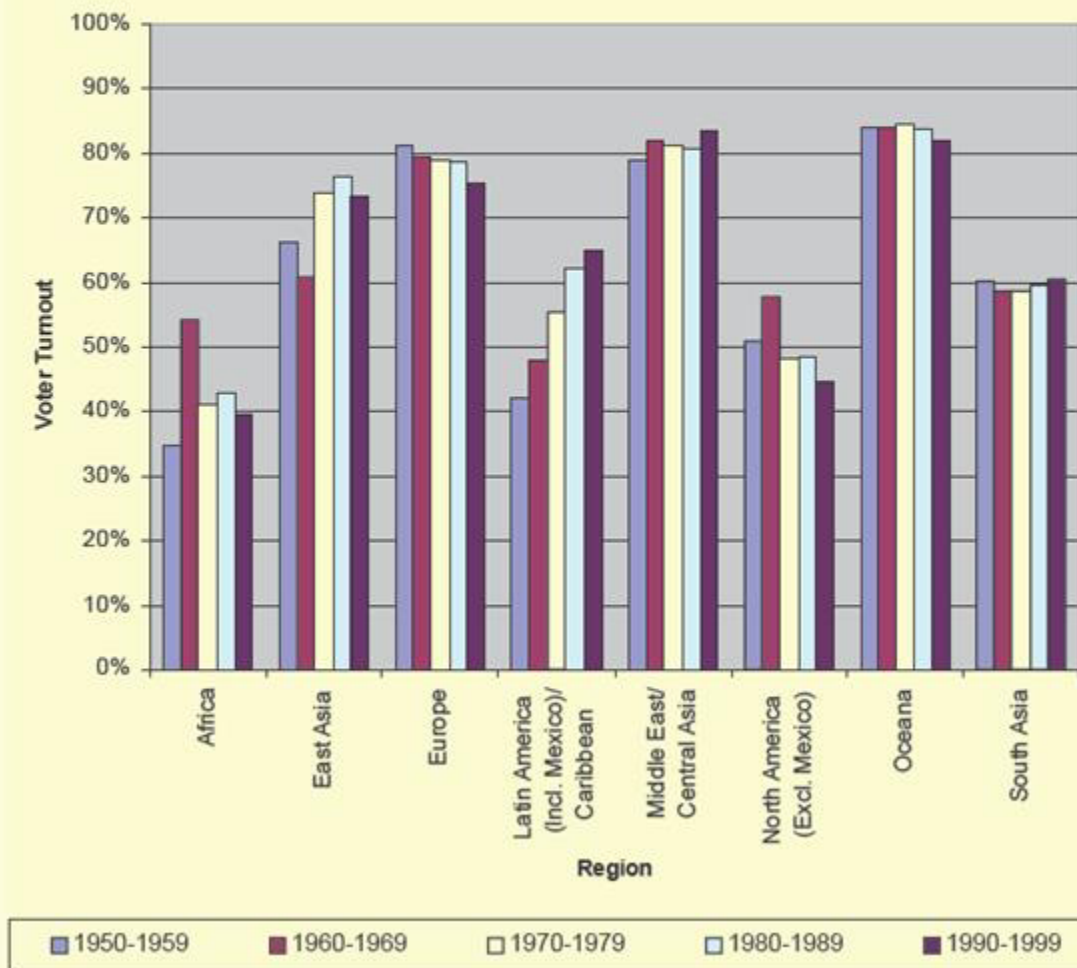
The United States has seen wide variations in the percent of the population that participates in elections (known as voter turnout). Much of this variation is the result of laws that changed who was eligible to vote. In 1868, after the Civil War, citizenship was extended to “all persons born or naturalized in the United States.”<sup>1</sup> The right to vote, regardless of “race, color or previous condition of servitude,” soon followed with the passage of the 15th Amendment.<sup>2</sup> These new constitutional amendments resulted in nearly one million recently-freed slaves registering to vote.<sup>3</sup> Women won the right to vote in 1920 with the 19th Amendment to the Constitution.<sup>4</sup> And in 1965, obstacles erected to prevent certain populations from voting were deemed illegal by the Voting Rights Act.<sup>5</sup> Since then, voter turnout has gradually declined, suggesting a growing apathy towards the democratic process. <sup>6</sup>

## International Voter Turnout

~Voter Turnout in other Democracies~

The decay in voter interest is a concern for many countries. In the more developed countries, voter turnout has been declining in the past half century. Figure 1, below, shows the trend in voter turnout for a number of regions since 1950.<sup>7</sup>

Figure 1



### Voter Mobilization Drives

In the 1998 and 2000 U.S. congressional elections, voter turnout was particularly low. Among the Voting Age Population (VAP) in the U.S., only 46.6% voted in the 2000 congressional (and corresponding presidential) elections and a record-low 34.7% voted in 1998, the prior mid-term election.<sup>7</sup> With another mid-term election approaching in 2002, individuals in Iowa and Michigan, concerned with the downward trend in participation, organized to launch the Vote 2002 campaign—a telephone drive reminding voters of the

election date and encouraging them to vote. With enough volunteers, and by contacting enough potential voters, they expected to dramatically improve voter turnout in their respective states.

There is some debate, however, over whether telephone campaigns are actually effective. In the past 40 years of voter mobilization, the traditional, person-to-person campaign tactics such as canvassing have been replaced by mass marketing techniques like direct mail and telephone calls.<sup>8</sup> This period has seen a decline in the number of political party volunteers as well as membership in nonpartisan organizations such as the Lions, Rotary and Kiwanis Clubs. Alan Gerber and Donald Green, in a study conducted in 2000, contend that “a decline of personal mobilization has arguably contributed to the erosion of voter turnout in the United states since the 1960s.”<sup>9</sup> This contention seems reasonably convincing. The correlation between the rise of telephone campaigning and the decline in voter turnout, however, does not mean that the telephone campaign *caused* the decline in voter turnout. It is possible that had canvassing been replaced by *nothing* as opposed to telephone calls and direct mail, even fewer people would be voting than currently are. So does mass marketing at least slow the decline? The next section explores whether this was indeed the case by asking the more specific question of whether the Vote 2002 Campaign had an effect on voter turnout.

## II – Did the Vote 2002 Campaign work?

Vote 2002 members obtained the telephone numbers of 60,000 individuals. They called all 60,000, but were able to speak with only about 25,000. For each individual, they recorded whether or not the call was completed successfully. They also collected data on household size, age, gender, whether the voter was newly registered, which state and district the voter was from, and the competitiveness of that district. And from official voting records, they were able to determine whether these individual households actually voted in the 2002 election.

### Discussion Topic I: Analyzing 2002 Data

Vote 2002 has agreed to share its data with you on the 60,000 households in their campaign. You are asked to use these data to estimate the impact of the Vote 2002 campaign on voter participation. You are asked to consider the two methods described below. You may want to refer to the appendix for some background on estimating the impact of a program.

#### Method 1 – Difference in proportion of voters between reached and not-reached

Assume the 25,000 households who received the call constitute the participant group and the other 35,000 households (who were called but not reached) represent the comparison group. If you want to see what the impact of receiving a call has on voter turnout, you could check whether those who received the call were more likely to vote than those who did not receive the call. Compare the proportion of people who voted in the treatment group and that of the comparison group.

*Discussion question 1* – What might the problem be in using this method to evaluate the effect of the program? Would it give you an accurate measure of the true impact?

#### Method 2 – Use multiple regression to control for differences between reached and not-reached

If you were concerned that people reached might have different inherent characteristics from those who were not reached, you could control for these differences by using a multivariable regression as follows: The participant and comparison groups are defined in the same way as in method 1. To estimate the impact of the program, you run a regression where the “dependent variable” is a zero/one variable indicating whether the person voted or not (i.e. 0 = did not vote, 1 = voted). The “key explanatory variable” is a zero/one variable indicating whether the person received the call or not (i.e. 0 = did not receive the call, 1 = received a call). Potential differences in characteristics can be controlled for using other “explanatory variables” such as age, gender, newly registered voter, etc. The coefficient on the key explanatory variable (i.e. received the call) represents the estimated impact of the program.

Table 1 presents the estimated impact of the Vote 2002 Campaign using each of these methods. Table 2 compares the average characteristics of the participant and comparison groups used in both of these methods.

*Discussion question 2* – For method 2, discuss whether it is reasonable to expect that the estimated impact represents the true causal effect of Vote 2002 on voter participation.

*Discussion question 3* – Why do you think the estimated impact using method 2 was lower than the estimated impact using method 1?

*Discussion question 4* – Can you address the weaknesses of method 1 by taking a random sample from the participant group and a random sample from the comparison group?

*Discussion question 5* – Using the data described above, can you think of more convincing methods to estimate the impact of the Vote 2002 campaign?

<b>Table 1: Percent who voted, Reached and Not-Reached</b>			
<b>Voter turnout of voters who were:</b>			
	<b>Reached</b>	<b>Not Reached</b>	<b>Estimated Impact</b>
<b>Method 1:</b>			
<b>Simple Difference</b>	64.5%	53.6%	10.9 pp*
<b>Method 2:</b>			
<b>Multiple Regression<sup>a</sup></b>			6.1 pp*

pp=percentage points

\*: statistically significant at the 5% level

a: controls include household size, age, newly registered, county, state senate district, state house district, from a competitive district

<b>Table 2: Characteristics of Reached and Not-Reached Groups</b>			
	<b>Reached</b>	<b>Not Reached</b>	<b>Difference</b>
<b>Household Size</b>	1.56	1.50	0.06
<b>Average age</b>	55.8	51.0	4.8
<b>Percent female</b>	56.2%	53.8%	2.4 pp*
<b>Percent newly registered</b>	7.3%	9.6%	-2.3 pp*
<b>Percent from a competitive district</b>	50.3%	49.8%	0.5 pp
<b>Percent from Iowa</b>	54.7%	46.7%	8.0 pp*
<b>Sample Size</b>	25,043	34,929	

pp=percentage points

\*: statistically significant at the 5% level

### **Discussion Topic II: Using Panel Data**

If you are still concerned about differences in characteristics between reached and non-reached individuals, you could use panel data – i.e. track the same person over time.

It turns out that staff members of the Vote 2002 also had data on whether the person voted in the previous elections (1998 and 2000). Past voting behavior is thought to be a strong predictor of future voting behavior. Table 3 indicates past voting behavior for the group of people who were reached by the Vote 2002 Campaign and the group of people who were called but not reached.

*Discussion question 1* - How can these data on past voting behavior be used to improve your analysis?

*Discussion question 2* – Given the information on Table 3, would you expect the method you proposed in the previous question to result in a higher or lower estimate of the impact of the Vote 2002 Campaign on voter turnout?

<b>Table 3: Percent who voted in past elections of Reached and Not-Reached Groups</b>			
<b>Voter turnout in 1998, 2000 and 2002</b>			
	<b>Reached</b>	<b>Not Reached</b>	<b>Difference</b>
<b>Voted in 2002</b>	64.5%	53.6%	10.9 pp*
<b>Voted in 2000</b>	71.7%	63.3%	8.4 pp*
<b>Voted in 1998</b>	46.6%	37.6%	9.0 pp*
<b>Difference 2002 - 1998</b>	17.9%	16.0%	1.9 pp*

pp=percentage points

\*: statistically significant at the 5% level

### **Discussion Topic III: Randomized Experiment**

It turns out that the 60,000 individuals were *randomly* selected from the larger population of about 2 million potential voters. This is akin to the randomization that occurs in a clinical trial, where the treatment/drug is randomly assigned to be received by one group of patients but not the other. We can exploit this randomization to estimate the impact of the Vote 2002 campaign. The idea is that the 60,000 individuals Vote 2002 called (now called the treatment group) should be identical to the 2,000,000 individuals (now called the control group) in everything (observable and unobservable) except for the fact that the first group was called by the Vote 2002 Campaign. Table 4 compares the treatment and control groups along some observable characteristics. Table 5 presents the estimated impact of the Vote 2002 Campaign by comparing the voter turnout of the treatment group with the voter turnout of the control group.

*Discussion question 1* – Notice that the two groups look very similar in Table 4. Is this what you would expect?

*Discussion question 2* – Notice that the impact estimates in Table 5 are not statistically significant. This result is different than those obtained with the previous methods. What may be the rationale behind this difference in results?

Technical note: For reasons we will discuss later in the course, the impact estimate from taking this simple difference needs to be adjusted to take into account the fact that 35,000 individuals in the treatment group were not reached. Table 6 shows both the simple comparison of treatment and control groups—where treatment is all those Vote 2002 called. It also shows the implied impact of the program after an adjustment is made for the fact that not all of this group were reached. (You will learn how to do this adjustment in the course).

<b>Table 4: Characteristics of Treatment and Control Groups</b>			
	<b>Reached</b>	<b>Not Reached</b>	<b>Difference</b>
<b>Voted in 2000</b>	56.7%	56.4%	0.3 pp
<b>Voted in 1998</b>	22.7%	23.1%	-0.4 pp
<b>Household Size</b>	1.50	1.50	0.00
<b>Average age</b>	52.0	52.2	-0.2
<b>Percent female</b>	54.6%	55.2%	-0.6 pp
<b>Percent newly registered</b>	11.6%	11.6%	0.0 pp

pp=percentage points

\*: statistically significant at the 5% level

<b>Table 5: Randomized Analysis</b>			
<b>Voter turnout of voters who are in the:</b>			
	<b>Treatment</b>	<b>Control</b>	<b>Estimated Impact</b>
<b>Method 4a: Randomized</b>			
<b>Simple Difference</b>	58.2%	58.0%	0.2 pp
<b>Method 4b: Randomized</b>			
<b>Multiple Regression</b>			0.2 pp

pp=percentage points

\*: statistically significant at the 5% level

## Conclusion

Table 6 presents impact estimates of the Vote 2002 Campaign using the different methods discussed in this case study.

<b>Table 6 - Summary Table of Impact Estimates</b>	
<b>Estimated Impact of the Vote 2002 Campaign</b>	
<b>Method</b>	
<b>Simple Difference</b>	10.8 pp*
<b>Multiple Regression</b>	6.1 pp*
<b>Difference in Difference with Panel Data</b>	1.9 pp*
<b>Randomized Experiment</b>	0.2 pp

pp=percentage points

\*: statistically significant at the 5% level

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.

### **Discussion Topic V: Application to Development**

Selection bias is a problem that arises in many program evaluations. Think about some of the non-randomized development programs you have, or have seen, evaluated. Discuss how the participant group was selected, and how “selection” may have affected the ability to estimate the true impact of the program.

## Appendix – Estimating the impact (or causal effect) of the program

- Estimating a causal effect (a.k.a impact) of a program or intervention involves a comparison between the outcome had the intervention been introduced and the outcome had the intervention not been introduced. The latter is usually referred to as the *counterfactual*
- The *counterfactual* represents the state of the world that program participants would have experienced in the absence of the program (i.e. had they not participated in the program)
  - The counterfactual does not represent the state of the world in which participants receive absolutely no services, but rather the state of the world in which participants receive whatever services they would have received had they not participated in the program being evaluated
- The counterfactual can never be directly observed
  - Hence, the main goal of an impact evaluation can be viewed as an effort to *construct* or *mimic* the counterfactual
  - This is usually done by selecting a group of individuals that did not participate in the program
  - This group is usually referred to as the *control group* (in case of a randomized experiment) or *comparison group* (in case we are using non-experimental methods to estimate the impact)
  - How this group is selected is a key decision in the design of any impact evaluation
  - The idea is to select a group that is exactly like the group of participants in all ways except one: their exposure to the program being evaluated
  - The goal in the end is to be able to attribute differences in outcomes between the group of participants and the control/comparison group to the program (and not to other factors)

## References

- 1 14th Amendment of the US Constitution
- 2 15th Amendment of the US Constitution
- 3 US Department of Justice Civil Rights Division, Voting Section, Introduction To Federal Voting Rights Laws, Before the Voting Rights Act, Reconstruction and the Civil War Amendments  
[http://www.usdoj.gov/crt/voting/intro/intro\\_a.htm](http://www.usdoj.gov/crt/voting/intro/intro_a.htm)
- 4 19th Amendment to the US Constitution
- 5 US Department of Justice Civil Rights Division, Voting Section, Introduction To Federal Voting Rights Laws, The Voting Rights Act of 1965 [http://www.usdoj.gov/crt/voting/intro/intro\\_b.htm](http://www.usdoj.gov/crt/voting/intro/intro_b.htm)
- 6 Gerber, Alan and Donald Green, 2000. "The Effects of Canvassing, Telephone calls, and Direct mail on Voter Turnout: A Field Experiment" *American Political Science Review* 94 (3): 653-663
- 7 International Institute for Democracy and Electoral Assistance:  
[http://www.idea.int/vt/country\\_view.cfm?CountryCode=US](http://www.idea.int/vt/country_view.cfm?CountryCode=US)
- 8 Arceneaux, Kevin, Ian Gerber, and Donald Green 2004. "Comparing Experimental and Matching Methods using a Large-Scale Field Experiment on Voter Mobilization" *Preliminary Draft*
- 9 Gerber, Alan and Donald Green, 2000: 653