Why randomize?

Abdul Latif Jameel
Poverty Action Lab
I. Background
II. What is a randomized evaluation?
III. Advantages and limitations of experiments
IV. How wrong can you go: “Vote 2002” campaign
V. Conclusions
How to measure impact?

- What would have happened in the absence of the program?

- Since counterfactual is not observable, the key goal of all impact evaluation methods is to construct or “mimic” the counterfactual.
Constructing the counterfactual

- Counterfactual is often constructed by selecting a group not affected by the program

- Randomized:
  - Use random assignment of the program to create a control group which mimics the counterfactual.

- Non-randomized:
  - Argue that a certain excluded group mimics the counterfactual.
Types of impact evaluation methods

1. Randomized Evaluations
Also known as:
- Random Assignment Studies
- Randomized Field Trials
- Social Experiments
- Randomized Controlled Trials (RCTs)
- Randomized Controlled Experiments
Types of impact evaluation methods (Cont.)

2. Non-Experimental or Quasi-Experimental Methods

Examples:
- Pre – Post
- Differences-in-Differences
- Statistical Matching
- Instrumental Variables
- Regression Discontinuity
- Interrupted Time Series
Validity

• A tool to assess credibility of a study

  – **Internal validity**: relates to ability to draw causal inference, i.e. can we attribute our impact estimates to the program and not to something else

  – **External validity**: relates to ability to generalize to other settings of interest, i.e. can we generalize our impact estimates from this program to other populations, time periods, countries, etc?
Outline

I. Background
II. What is a randomized evaluation?
III. Advantages and limitations of experiments
IV. How wrong can you go: “Vote 2002” campaign
V. Conclusions
The basics

Start with simple case:

- Take a sample of program applicants
- *Randomly* assign them to either:
  - **Treatment Group** – is offered treatment
  - **Control Group** - not allowed to receive treatment
    (during the evaluation period)
Key advantage of experiments

Because members of the groups (treatment and control) do not differ systematically at the outset of the experiment,

any difference that subsequently arises between them can be attributed to the treatment rather than to other factors.
Some variations on the basics

- Assigning to multiple treatment groups
- Assigning of units other than individuals or households
  - Health Centers
  - Schools
  - Local Governments
  - Villages
Key steps in conducting an experiment

1. **Design** the study carefully
2. **Randomly** assign people to treatment or control
3. Collect **baseline** data
4. **Verify** that assignment looks random
5. **Monitor** process so that integrity of experiment is not compromised
Key steps in conducting an experiment (cont.)

6. Collect follow-up data for both the treatment and control groups in identical ways.

7. Estimate program impacts by comparing mean outcomes of treatment group vs. mean outcomes of control group.

8. Assess whether program impacts are statistically significant and practically significant.
“Random”

• What does the term *random* mean?

• Is *random assignment* the same as *random sampling*?
Basic setup of a randomized evaluation

Based on Orr (1999)
Random assignment vs. random sampling

• Random assignment:
  – Relates to internal validity.

• Random sampling:
  – Relates to external validity.
Outline

I. Background
II. What is a randomized evaluation?
III. Advantages and limitations of experiments
IV. How wrong can you go: “Vote 2002” campaign
V. Conclusions
Random assignment

- Implies that the distribution of both observable and unobservable characteristics in the treatment and control groups are statistically identical.
- In other words there are no systematic differences between the two groups.
Key advantage of experiments

Because members of the groups (treatment and control) do not differ systematically at the outset of the experiment,

any difference that subsequently arises between them can be attributed to the treatment rather than to other factors.
Other advantages of experiments

- Relative to results from non-experimental studies, results from experiments are:
  - Less subject to methodological debates
  - Easier to convey
  - More likely to be convincing to program funders and/or policymakers
Limitations of experiments

• Despite great methodological advantage of experiments, they are also potentially subject to threats to their validity. For example,
  – Internal Validity
    (e.g. Hawthorne Effects, survey non-response, no-shows, crossovers, duration bias, etc.)
  – External Validity
    (e.g. are the results generalizable to other populations?)
• It is important to realize that some of these threats also affect the validity of non-experimental studies
Other limitations of experiments

- Measure the impact of the offer to participate in program
- Costly
- Ethical issues
- Partial equilibrium
Outline

I. Background

II. What is a randomized evaluation?

III. Advantages and limitations of experiments

IV. How wrong can you go: “Vote 2002” campaign

V. Conclusions

- Source: “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
## Summary table

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimated Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – Simple Difference</td>
<td>10.8 pp *</td>
</tr>
<tr>
<td>2 – Multiple regression</td>
<td>6.1 pp *</td>
</tr>
<tr>
<td>3 – Multiple regression with panel data</td>
<td>4.5 pp *</td>
</tr>
<tr>
<td>4 – Matching</td>
<td>2.8 pp *</td>
</tr>
<tr>
<td>5 – Randomized Experiment</td>
<td>0.4 pp</td>
</tr>
</tbody>
</table>

Outline

I. Background
II. What is a randomized evaluation?
III. Advantages and limitations of experiments
IV. How wrong can you go: “Vote 2002” campaign
V. Conclusions
Conclusions

- If properly designed and conducted, social experiments provide the most credible assessment of the impact of a program.

- Results from social experiments are easy to understand and much less subject to methodological quibbles.

- Credibility + Ease of understanding => More likely to convince policymakers and funders of effectiveness (or lack thereof) of program.
Conclusions (cont.)

• However, these advantages are present only if social experiments are well designed and conducted properly.

• Must assess validity of experiments in the same way we assess validity of any other study.

• Must be aware of limitations of experiments.
The End
Case 1 – “Vote 2002” campaign

- Intervention designed to increase voter turnout in 2002
- Phone calls to ~60,000 individuals
- Only ~35,000 individuals were reached
- Key Question: Did the campaign have a positive effect (i.e. impact) on voter turnout?
  - 5 methods were used to estimate impact
Methods 1-3

• Based on comparing reached vs. not-reached

• Method 1: Simple difference in voter turnout, \((\text{Voter turnout})_{\text{reached}} - (\text{Voter turnout})_{\text{not reached}}\)

• Method 2: Multiple Regression controlling for some differences between the two groups

• Method 3: Method 2, but also controlling for differences in past voting behavior
Impact estimates using methods 1-3

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimated Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1</td>
<td>10.8 pp *</td>
</tr>
<tr>
<td>Method 2</td>
<td>6.1 pp *</td>
</tr>
<tr>
<td>Method 3</td>
<td>4.5 pp *</td>
</tr>
</tbody>
</table>

pp=percentage points; *: statistically significant at the 5% level
Methods 1-3

Is any of these impact estimates likely to be the true impact of the “Vote 2002” campaign?
## Reached vs. not reached

<table>
<thead>
<tr>
<th></th>
<th>Reached</th>
<th>Not Reached</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>56.2%</td>
<td>53.8%</td>
<td>2.4 pp*</td>
</tr>
<tr>
<td>Newly Regist.</td>
<td>7.3%</td>
<td>9.6%</td>
<td>-2.3 pp*</td>
</tr>
<tr>
<td>From Iowa</td>
<td>54.7%</td>
<td>46.7%</td>
<td>8.0 pp*</td>
</tr>
<tr>
<td>Voted in 2000</td>
<td>71.7%</td>
<td>63.3%</td>
<td>8.3 pp*</td>
</tr>
<tr>
<td>Voted in 1998</td>
<td>46.6%</td>
<td>37.6%</td>
<td>9.0 pp*</td>
</tr>
</tbody>
</table>

pp=percentage points
*: statistically significant at the 5% level
Method 4: Matching

• Similar data available on 2,000,000 individuals

• Select as a comparison group a subset of the 2,000,000 individuals that resembles the reached group as much as possible

• Statistical procedure: matching

• To estimate impact, compare voter turnout between reached and comparison group
An illustration of matching

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Precinct</th>
<th>Previous Vote</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>1</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>45</td>
<td>0</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>55</td>
<td>1</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>42</td>
<td>0</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>70</td>
<td>1</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>62</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>Precinct</th>
<th>Previous Vote</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>1</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>0</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>56</td>
<td>1</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>62</td>
<td>0</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

Impact estimates using matching

<table>
<thead>
<tr>
<th>Matching on 4 covariates</th>
<th>Estimated Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.7 pp *</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Matching on 6 covariates</th>
<th>Estimated Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.0 pp *</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Matching on all covariates</th>
<th>Estimated Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.8 pp *</td>
</tr>
</tbody>
</table>

* pp=percentage points; *: statistically significant at the 5% level
Method 4: Matching

- Is this impact estimate likely to be the true impact of the “Vote 2002” campaign?
- Key: These two groups should be equivalent in terms of the observable characteristics that were used to do the matching.

But what about unobservable characteristics?
Method 5: Randomized experiment

- Turns out 60,000 were randomly chosen from population of 2,060,000 individuals.

- Hence, treatment was randomly assigned to two groups:
  - Treatment group (60,000 who got called)
  - Control group (2,000,000 who did not get called)

- To estimate impact, compare voter turnout between treatment and control groups.
  - Do a statistical adjustment to address any other differences.
Method 5: Randomized experiment

• Impact estimate: 0.4 pp, not statistically significant

• Is this impact estimate likely to be the true impact of the “Vote 2002” campaign?

• Key: Treatment and control groups should be equivalent in terms of both their observable and unobservable characteristics

• Hence, any difference in outcomes can be attributed to the Vote 2002 campaign