Why Randomize?

Adam Osman
J-PAL
Course Overview

1. What is Evaluation?
2. Outcomes, Impact, and Indicators
3. Why Randomize?
4. How to Randomize
5. Threats and Analysis
6. Sampling and Sample Size
7. Project from Start to Finish
8. Generalizability
Methodologically, randomized trials are the best approach to estimate the effect of a program.

1. Strongly Disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly Agree
Session Overview

I. Background

II. What is a randomized experiment?

III. Why randomize?

IV. Conclusions
I - BACKGROUND
What is the impact of this program?

Program starts
What is the impact of this program?

1. Positive
2. Negative
3. Zero
4. Not enough info
“Before vs. After” is rarely a good method for assessing impact.
What is the impact of this program?

Primary Outcome

Program starts

Impact

Counterfactual

Time
How to measure impact?

*Impact* is defined as a comparison between:

1. the outcome some time after the program has been introduced

2. the outcome at that same point in time had the program not been introduced (the "counterfactual")
Impact: What is it?

- Time
- Primary Outcome
- Program starts
- Counterfactual
- Impact
Impact: What is it?

Program starts

Primary Outcome vs. Time

Counterfactual vs. Impact
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)

• **Problem**: Counterfactual cannot be observed

• **Solution**: We need to “mimic” or construct the counterfactual
Constructing the counterfactual

• Usually done by selecting a group of individuals that \textit{did not} participate in the program

• This group is usually referred to as the \textit{control group} or \textit{comparison group}

• How this group is selected is a \textit{key decision} in the design of any impact evaluation
Selecting the comparison group

• Idea: Select a group that is exactly like the group of participants in all ways except one: their exposure to the program being evaluated

• Goal: To be able to attribute differences in outcomes between the group of participants and the comparison group to the program (and not to other factors)
1. Randomized Experiments
   • Also known as:
     – Random Assignment Studies
     – Randomized Field Trials
     – Social Experiments
     – Randomized Controlled Trials (RCTs)
     – Randomized Controlled Experiments
Impact evaluation methods

2. Non- or Quasi-Experimental Methods
   a. Pre-Post
   b. Simple Difference
   c. Differences-in-Differences
   d. Multivariate Regression
   e. Statistical Matching
   f. Interrupted Time Series
   g. Instrumental Variables
   h. Regression Discontinuity
II – WHAT IS A RANDOMIZED EXPERIMENT?
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 program rather than to other factors.
Evaluation of “Women as Policymakers”: Treatment vs. Control villages at baseline

<table>
<thead>
<tr>
<th>Variables</th>
<th>Treatment Group</th>
<th>Control Group</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female Literacy Rate</td>
<td>0.35</td>
<td>0.34</td>
<td>0.01 (0.01)</td>
</tr>
<tr>
<td>Number of Public Health Facilities</td>
<td>0.06</td>
<td>0.08</td>
<td>-0.02 (0.02)</td>
</tr>
<tr>
<td>Tap Water</td>
<td>0.05</td>
<td>0.03</td>
<td>0.02 (0.02)</td>
</tr>
<tr>
<td>Number of Primary Schools</td>
<td>0.95</td>
<td>0.91</td>
<td>0.04 (0.08)</td>
</tr>
<tr>
<td>Number of High Schools</td>
<td>0.09</td>
<td>0.10</td>
<td>-0.01 (0.02)</td>
</tr>
</tbody>
</table>

Standard Errors in parentheses. Statistics displayed for West Bengal
**/***: Statistically significant at the 10% / 5% / 1% level
Source: Chattopadhyay and Duflo (2004)
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

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.
III – WHY RANDOMIZE?
Why randomize? – Conceptual Argument

If properly designed and conducted, randomized experiments provide the most credible method to estimate the impact of a program.
Why “most credible”?

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 program rather than to other factors.
Example #2 - Pratham’s Read India program
Example #2 - Pratham’s Read India program

<table>
<thead>
<tr>
<th>Method</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Pre-Post</td>
<td>0.60*</td>
</tr>
<tr>
<td>(2) Simple Difference</td>
<td>-0.90*</td>
</tr>
<tr>
<td>(3) Difference-in-Differences</td>
<td>0.31*</td>
</tr>
<tr>
<td>(4) Regression</td>
<td>0.06</td>
</tr>
<tr>
<td>(5) Randomized Experiment</td>
<td></td>
</tr>
</tbody>
</table>

*: Statistically significant at the 5% level
Example #1 - Pratham’s Read India program

<table>
<thead>
<tr>
<th>Method</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Pre-Post</td>
<td>0.60*</td>
</tr>
<tr>
<td>(2) Simple Difference</td>
<td>-0.90*</td>
</tr>
<tr>
<td>(3) Difference-in-Differences</td>
<td>0.31*</td>
</tr>
<tr>
<td>(4) Regression</td>
<td>0.06</td>
</tr>
<tr>
<td>(5) Randomized Experiment</td>
<td>0.88*</td>
</tr>
</tbody>
</table>

*: Statistically significant at the 5% level
Example #2: A voting campaign in the USA
A voting campaign in the USA

<table>
<thead>
<tr>
<th>Method</th>
<th>Impact (vote %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Pre-post</td>
<td>-7.2 pp</td>
</tr>
<tr>
<td>(2) Simple difference</td>
<td>10.8 pp *</td>
</tr>
<tr>
<td>(3) Difference-in-differences</td>
<td>3.8 pp*</td>
</tr>
<tr>
<td>(4) Multiple regression</td>
<td>6.1 pp *</td>
</tr>
<tr>
<td>(5) Matching</td>
<td>2.8 pp *</td>
</tr>
<tr>
<td>(5) Randomized Experiment</td>
<td></td>
</tr>
</tbody>
</table>
A voting campaign in the USA

<table>
<thead>
<tr>
<th>Method</th>
<th>Impact (vote %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Pre-post</td>
<td>-7.2 pp</td>
</tr>
<tr>
<td>(2) Simple difference</td>
<td>10.8 pp *</td>
</tr>
<tr>
<td>(3) Difference-in-differences</td>
<td>3.8 pp *</td>
</tr>
<tr>
<td>(4) Multiple regression</td>
<td>6.1 pp *</td>
</tr>
<tr>
<td>(5) Matching</td>
<td>2.8 pp *</td>
</tr>
<tr>
<td>(5) Randomized Experiment</td>
<td>0.4 pp</td>
</tr>
</tbody>
</table>
### A voting campaign in the USA

<table>
<thead>
<tr>
<th>Method</th>
<th>Impact (vote %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Pre-post</td>
<td>-7.2 pp</td>
</tr>
<tr>
<td>(2) Simple difference</td>
<td>10.8 pp *</td>
</tr>
<tr>
<td>(3) Difference-in-differences</td>
<td>3.8 pp*</td>
</tr>
<tr>
<td>(4) Multiple regression</td>
<td>6.1 pp *</td>
</tr>
<tr>
<td>(5) Matching</td>
<td>2.8 pp *</td>
</tr>
<tr>
<td>(5) Randomized Experiment</td>
<td>0.4 pp</td>
</tr>
</tbody>
</table>

**Bottom Line:** Which method we use matters!
Conclusions - Why Randomize?

- There are many ways to estimate a program’s impact.
- This course argues in favor of one: randomized experiments.
  - **Conceptual argument:** If properly designed and conducted, randomized experiments provide the most credible method to estimate the impact of a program.
  - **Empirical argument:** Different methods can generate different impact estimates.
What is the most convincing argument you have heard against RCTs? Enter your top 3 choices.

A. Too expensive
B. Takes too long
C. Not ethical
D. Too difficult to design/implement
E. Not externally valid (Not generalizable)
F. Less practical to implement than other methods and not much better
G. Can tell us what the impact is, but not why or how it occurred (i.e. it is a black box)
THANK YOU!