Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs
Spring 2009

Why randomize?

Abdul Latif Jameel
Poverty Action Lab

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

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.)

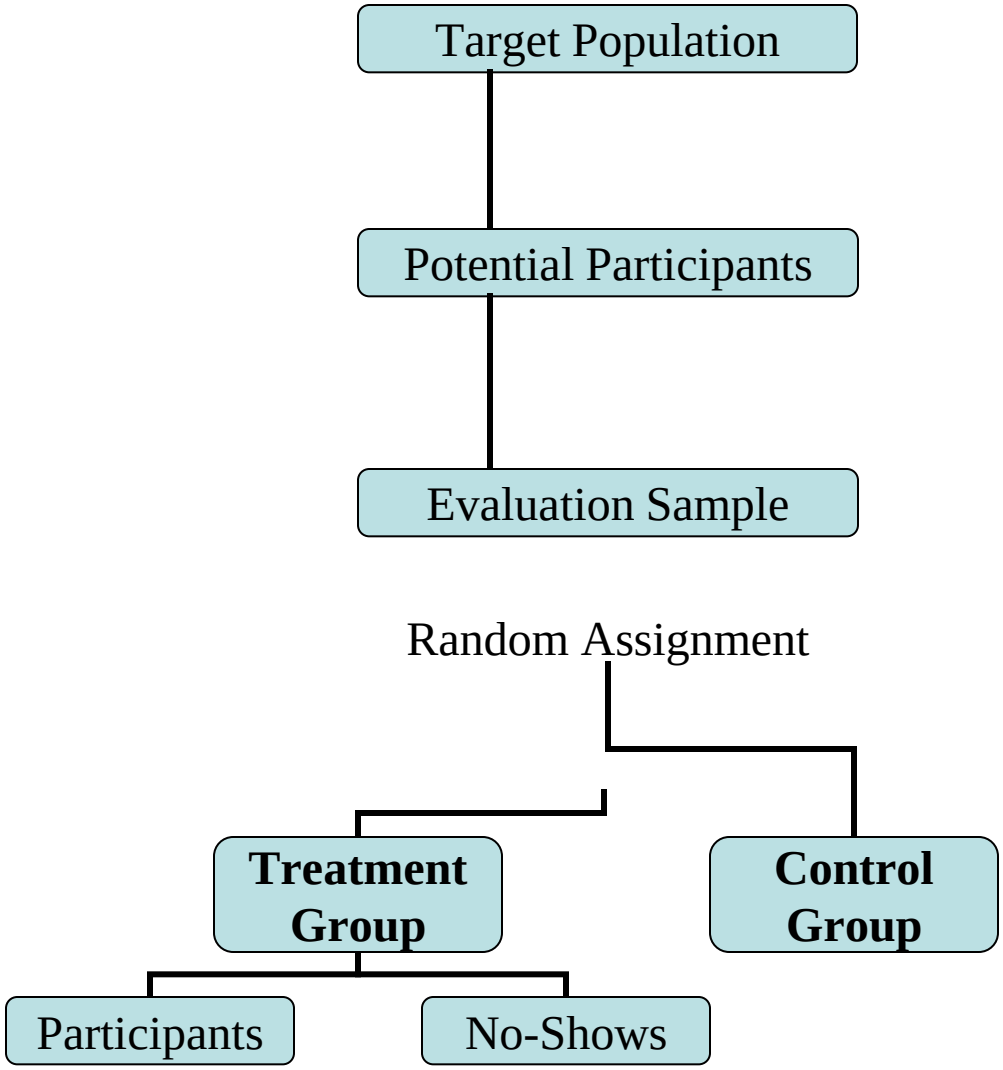
6. Collect follow-up data for both the treatment and control groups in identical ways.
7. Estimate program impacts by comparing mean outcomes of treatment group vs. mean outcomes of control group.
8. Assess whether program impacts are statistically significant and practically significant.

"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

Method	Estimated Impact
1 – Simple Difference	10.8 pp *
2 – Multiple regression	6.1 pp *
3 – Multiple regression with panel data	4.5 pp *
4 – Matching	2.8 pp *
5 – Randomized Experiment	0.4 pp

Source: Arceneaux, Gerber, and Green (2004)

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 same way we assess validity of any other study
- Must be aware of limitations of experiments

ABDUL LATIF JAMEEL
Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

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

	Estimated Impact
Method 1	10.8 pp *
Method 2	6.1 pp *
Method 3	4.5 pp *

pp=percentage points; *: statistically significant at the 5% level
Source: Arceneaux, Gerber, and Green (2004)

Methods 1-3

Is any of these impact estimates likely to be the true impact of the “Vote 2002” campaign?

Reached vs. not reached

	Reached	Not Reached	Difference
Female	56.2%	53.8%	2.4 pp*
Newly Regist.	7.3%	9.6%	-2.3 pp*
From Iowa	54.7%	46.7%	8.0 pp*
Voted in 2000	71.7%	63.3%	8.3 pp*
Voted in 1998	46.6%	37.6%	9.0 pp*

pp=percentage points

*: statistically significant at the 5% level

Source: Arceneaux, Gerber, and Green (2004)

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

Treated Subjects			
Age	Gender	Precinct	Previous Vote
30	1	10	1
45	**0**	**15**	**1**
19	**0**	**12**	**0**
32	1	16	1
55	**1**	**16**	**0**
42	0	15	1
70	1	10	0
24	1	12	0
21	**0**	**14**	**1**
34	1	14	0
62	0	10	0

Untreated Subjects			
Age	Gender	Precinct	Previous Vote
55	**1**	**16**	**0**
45	**0**	**15**	**1**
19	0	12	1
56	1	14	0
28	1	12	0
18	1	12	0
19	**0**	**12**	**0**
21	**0**	**14**	**1**
21	**0**	**14**	**1**
25	0	10	1
62	0	10	1

Source: Arceneaux, Gerber, and Green (2004)

Impact estimates using matching

	Estimated Impact
Matching on 4 covariates	3.7 pp *
Matching on 6 covariates	3.0 pp *
Matching on all covariates	2.8 pp *

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

Source: Arceneaux, Gerber, and Green (2004)

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

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