An Introduction to Randomization*

TAF – CEGA Impact Evaluation Workshop Day 1

* http://ocw.mit.edu (J-PAL Impact Evaluation Workshop materials)

Outline

- Background
- What is randomized evaluation?
- Advantages and limitations of experiments
- Conclusions

An example: The “Vote 2002” campaign

- Arceneaux, Gerber and Green (2006)
- Intervention: get-out-the-vote phone calls to increase voter turnout in Iowa and Michigan, 2002 midterm elections
- Treatment group = 60,000 individuals (35,000 actually reached by phone)
- Control group = >2,000,000 individuals
- Main outcome: turnout (did the individual vote?)

Effect sizes using experimental v. non-experimental methods

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

How to measure impact?

- What would have happened in the absence of the intervention program?
  - Since the counterfactual is not observable, key goal of all impact evaluation methods is to construct of “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.

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

- Background
- **What is randomized evaluation?**
- Advantages and limitations of experiments
- Conclusions

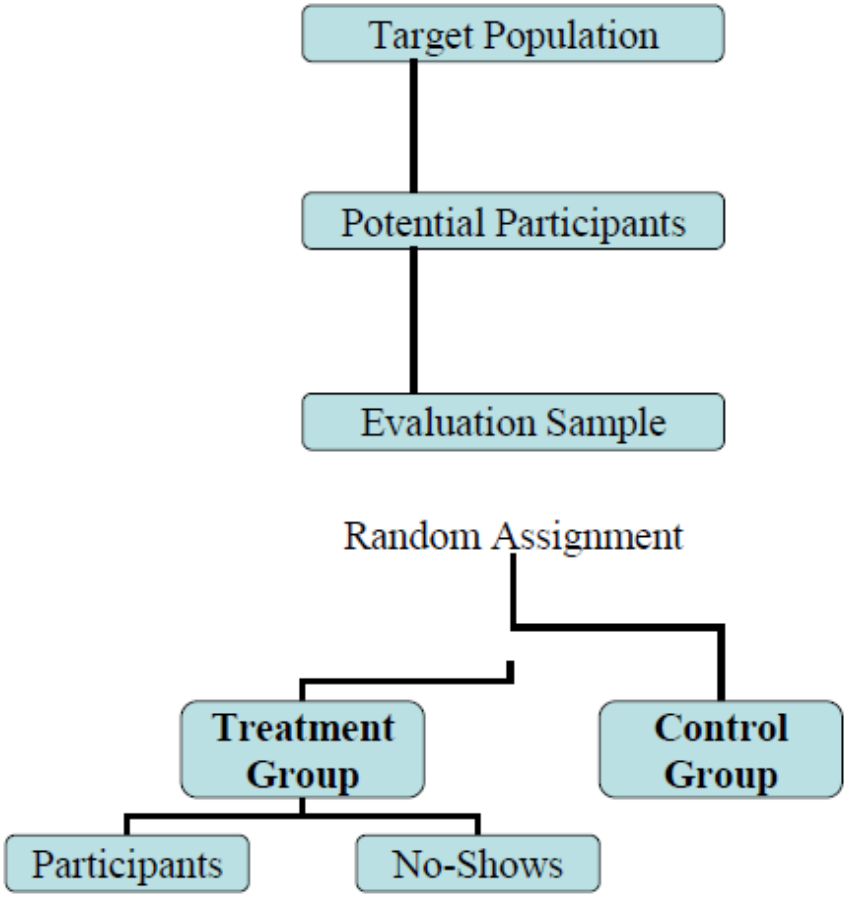
Key steps in conducting an experiment

- Design the study carefully
- Randomly assign people to treatment or control
- Collect baseline data
- Verify that assignment looks random
- Monitor process so that integrity of experiment is not compromised

Key steps (continued)

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

Basic setup of a randomized evaluation



Random assignment v. random sampling

- Random assignment
  - Relates to internal validity
- Random sampling
  - Relates to external validity

Outline

- Background
- What is randomized evaluation?
- **Advantages and limitations of experiments**
- Conclusions

Key advantage of experiments

- Successful random assignment implies characteristics of treatment and control groups are statistically identical
  - In other words, there are no systematic differences between the two groups
  - Thus, any difference that subsequently arises between them can be attributed to the treatment rather than to other factors

Other advantages

- 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

- Costly
- Ethical issues
- Despite methodological advantage, they are also potentially subject to threats to validity:
  - Internal Validity
    - e.g. Hawthorne Effects, survey non-response, no-shows, crossovers, duration bias, etc.
  - External Validity
    - e.g. do results apply to other populations?
- Some of these threats also affect the validity of non-experimental studies

Outline

- Background
- What is randomized evaluation?
- Advantages and limitations of experiments
- **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 (continued)

- 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

Thanks!

Study design (back)

- Key question:
  - Does the experimental design answer the policy question you seek to answer?
- Random sampling
  - Select your “universe” carefully: who does your sample represent?
- Stratified random sampling
  - Guarantee that important sub-populations are represented to the degree that allows for statistically significant comparisons across groups

Random assignment to treatment and control: basics

- Start with simple case:
  - Take a sample of program applicants
  - Randomly assign them to either:
    - Treatment Group (offered treatment)
    - Control Group (not allowed to receive treatment, at least during the initial evaluation period)

Random assignment to treatment and control: variations

- Stratified randomization
  - Useful, when there are small sample size issues, to ensure balancing across treatment and control groups
- Assigning of individuals to multiple treatment arms
- Assigning of units other than individuals or households:
  - Health centers
  - Schools
  - Local governments
  - Villages

Random assignment: continued

- Phased roll-out of interventions
  - Exploit budgetary limitations to roll out interventions in phases, randomizing who receives intervention first
    - Addresses some ethical concerns related to randomization
  - Example: *Progresa*

Random assignment: continued

- Promotion of existing programs
  - Example: Voter turnout intervention (get-out-the-vote)
    - Target individuals who may be influenced by promotion
    - Randomize the promotion
  - What policy question do we answer through randomized promotion?

Verify balancing across treatment/control (back)

- We can test the validity of the randomization using the baseline data
  - For a start, can compare means of observed variables across treatment and control--check for no significant differences
  - Can also compare other moments, or statistically compare the distribution