ABDUL LATIF JAMEEL
Poverty Action Lab

TRANSLATING RESEARCH INTO ACTION

Why Randomize?
And Common Critiques

Shawn Powers
J-PAL

Course Overview

1. What is Evaluation?
2. Outcomes, Impact, and Indicators
3. **Why Randomize and Common Critiques**
4. How to Randomize
5. Sampling and Sample Size
6. Threats and Analysis
7. Project from Start to Finish
8. Cost-Effectiveness Analysis and Scaling Up

Review: 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)

I – 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.

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.

II – 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"?

In all other impact evaluation methods, we need to assume that the two groups do not differ systematically at the outset or that any differences between them have been statistically accounted for.

BUT we have no way to test this assumption.

Randomization assures that there is no systematic difference between groups.

Why randomize? – Empirical Argument

We can evaluate the same program with different methodologies and get different answers

e.g. the case of flipcharts in Kenya

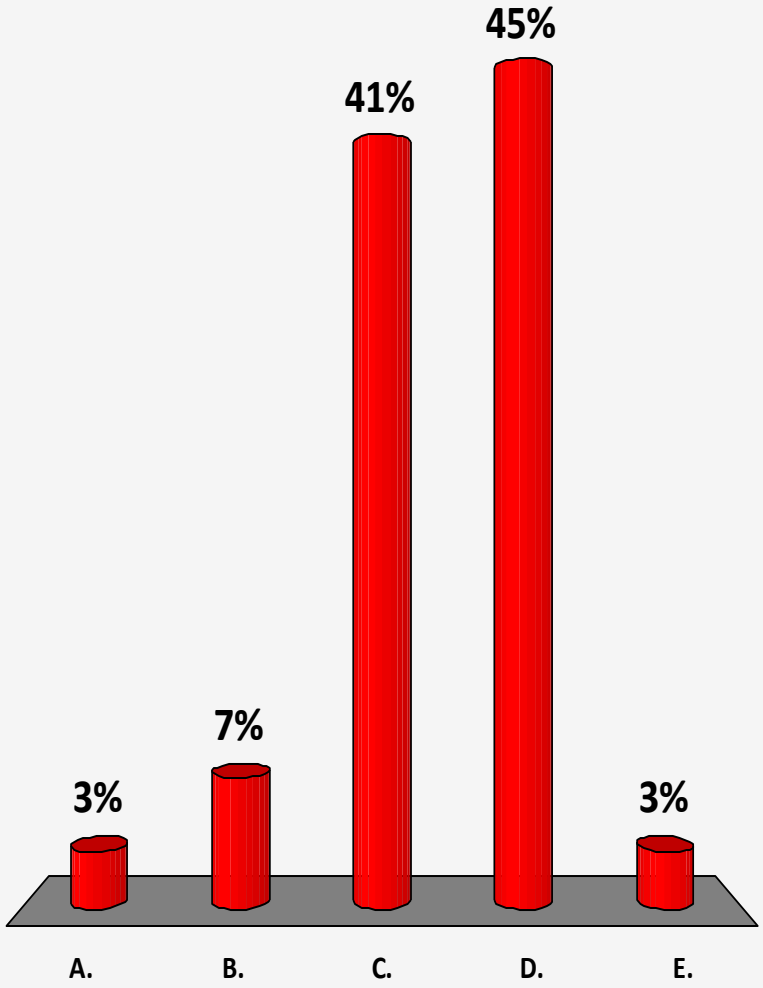
Example - Pratham's "Learn to Read" Program

Which of these methods do you think is closest to the truth?

Method	Impact Estimate
(1) Pre-post	0.60*
(2) Simple Difference	-0.68*
(3) Difference-in-Difference	0.24*
(4) Regression	0.04

*: Statistically significant at the 5% level

A. Pre-Post
B. Simple Difference
C. Difference-in-Differences
D. Regression
E. Don't know



Example - Pratham's "Learn to Read" Program

Method	Impact
(1) Pre-Post	0.60*
(2) Simple Difference	-0.68*
(3) Difference-in-Differences	0.24*
(4) Regression	0.04
(5) Randomized Experiment	

Example - Pratham's "Learn to Read" Program

Method	Impact
(1) Pre-Post	0.60*
(2) Simple Difference	-0.68*
(3) Difference-in-Differences	0.24*
(4) Regression	0.04
(5) Randomized Experiment	0.88*

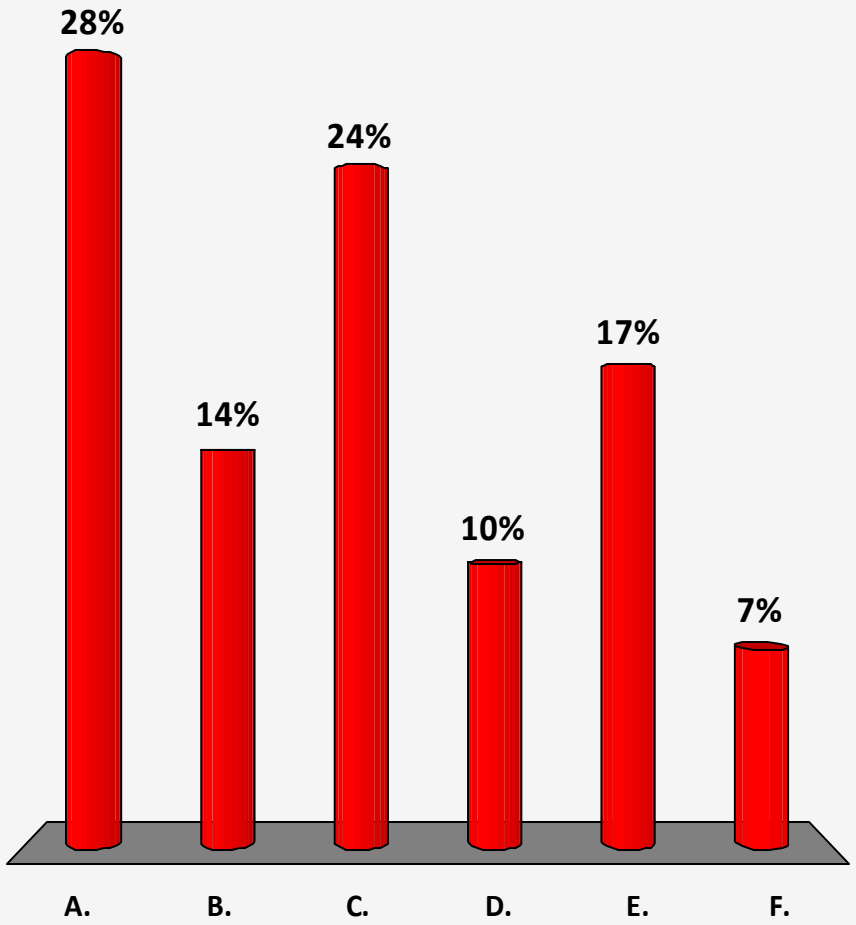
Summing Up - Why Randomize?

- 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

III – COMMON CRITIQUES

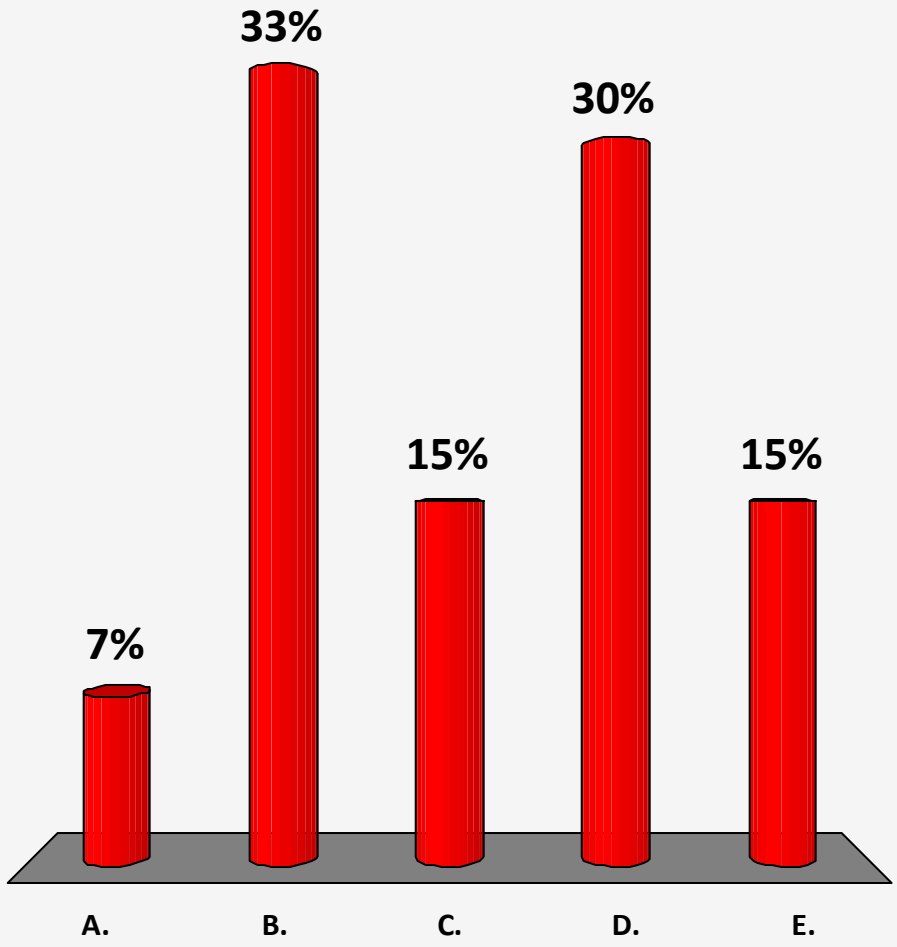
What is the most convincing argument you have heard **against** REs?

A. External validity
B. Cost
C. Political or ethical considerations
D. Hard to implement in fragile context
E. Hard to do right
F. Cannot be used to answer all the questions we care about



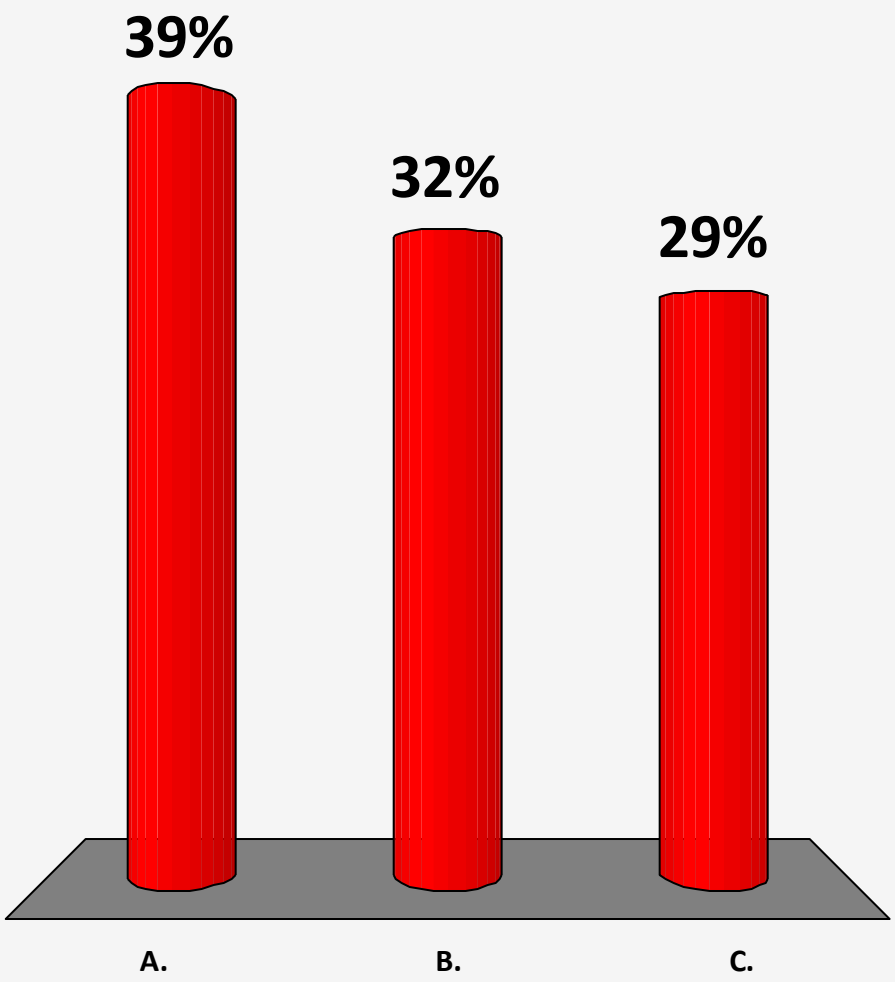
Not externally valid/generalizable

A. Can still be useful if context similar
B. Not unique to Res
C. Can see patterns if you do many Res
D. External validity can be tested
E. Not all evaluations need to be externally valid



Not ethical to randomly assign

A. Limited resources
B. Control group can get program eventually
C. We don't know the intervention works, could even be harmful



Critique #3

A. Response 1
B. Response 2