Introduction to randomized impact evaluations

Seema Jayachandran
Northwestern University
What is Evaluation?

Evaluation

Program Evaluation

Impact Evaluation
Program Evaluation

Evaluation

Program Evaluation

Impact Evaluation
Monitoring and Evaluation

- Evaluation
  - Program Evaluation
  - Impact Evaluation
- Monitoring
Program Evaluation

- Evaluation
- Program Evaluation
- Impact Evaluation
Components of Program Evaluation

• Needs Assessment
• Program Theory Assessment
• Process Evaluation
• Impact Evaluation
• Cost Effectiveness

• What is the problem?
• How, in theory, does the program fix the problem?
• Does the program work as planned?
• Were its goals achieved? The magnitude?
• Given magnitude and cost, how does it compare to alternatives?
Who is this evaluation for?

- Academics
- Donors
  - Their Constituents
- Politicians / policymakers
- Technocrats
- Implementers
- Beneficiaries
- Proponents, Skeptics
How can impact evaluation help us?

• Surprisingly little hard evidence on what works
• Can do more with given budget with better evidence
• If people knew money was going to programs that worked, could help increase pot for anti-poverty programs
• Instead of asking “do aid/development programs work?” should be asking:
  – Which work best, why and when?
  – How can we scale up what works?
Programs and their Evaluations: where do we start?

**Intervention**
- Start with a problem
- Verify that the problem actually exists
- Generate a theory of why the problem exists
- Design the program
- Think about whether the solution is cost effective

**Program Evaluation**
- Start with a question
- Verify the question hasn’t been answered
- State a hypothesis
- Design the evaluation
- Determine whether the value of the answer is worth the cost of the evaluation
Measuring how well it worked

IMPACT EVALUATION
Did we achieve our goals?

- Primary outcome (impact)
- Example: Does intervention to protect springs reduce child diarrhea?

- Also distributional questions:
- Example: What was the impact for households with good v. bad sanitation practices?
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”)
What is the impact of this program?
Impact: What is it?

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

Primary Outcome vs. Time

Program starts

Impact

Counterfactual
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 did not participate in the program

- This group is usually referred to as the control group or comparison group

- How this group is selected is a 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)
Income per person, per month

1457
1000
947
500
0

Treat Compare

Non random assignment
Impact evaluation methods

1. Randomized Experiments
   • Also known as:
     – Random Assignment Studies
     – Randomized Field Trials
     – Social Experiments
     – Randomized Controlled Trials (RCTs)
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
Constructing the Counterfactual

• 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.
Random assignment

Income per person, per month

<table>
<thead>
<tr>
<th></th>
<th>Treat</th>
<th>Compare</th>
</tr>
</thead>
<tbody>
<tr>
<td>1457</td>
<td></td>
<td>1442</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How impact differs from process?

- When we answer a process question, we need to describe what happened.

- When we answer an impact question, we need to compare what happened to what would have happened without the program.
Randomized Evaluation

The “gold standard” for Impact Evaluation

Randomized Evaluation
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)
Some variations on the basics

• Assigning to multiple treatment groups
  – Example: Information campaign on treating water + Provision of public water taps

• Assigning of units other than individuals or households
  ▪ Schools
  ▪ Local Governments
  ▪ Villages or neighborhoods
Random Sampling and Random Assignment

Randomly **sample** from area of interest
Random Sampling and Random Assignment

Randomly sample from area of interest

Randomly assign to treatment and control

Randomly sample from both treatment and control (Example: households within neighborhood)
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.
“Piped water to households in Morocco”: Treatment vs. Control villages at baseline

<table>
<thead>
<tr>
<th>Variables</th>
<th>Control Group</th>
<th>Treatment Group</th>
<th>Difference</th>
<th>Difference in Std Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household size</td>
<td>4.55</td>
<td>4.73</td>
<td>-0.18</td>
<td>7%</td>
</tr>
<tr>
<td>Income index</td>
<td>4.21</td>
<td>4.42</td>
<td>-0.21</td>
<td>8%</td>
</tr>
<tr>
<td>Index of knowledge of diarrhea causes</td>
<td>1.25</td>
<td>1.23</td>
<td>0.02</td>
<td>2%</td>
</tr>
<tr>
<td>Main water source = public tap</td>
<td>0.44</td>
<td>0.43</td>
<td>0.01</td>
<td>2%</td>
</tr>
<tr>
<td>Distance to public tap (meters)</td>
<td>139</td>
<td>134</td>
<td>5</td>
<td>3%</td>
</tr>
</tbody>
</table>

Source: Devoto et al. (2011)
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.
Example
The Problem

• 13% of world population lacks access to “improved water sources”

• Nearly 2 million children die each year from diarrhea

• 20% all child deaths (under 5 years old) are from diarrhea
The Goal

- MDG: “reduce by half the proportion of people without access to sustainable drinking water”
The Solution(s)
Spring cleaning in Kenya

• **Key question:** What is the impact of the clean springs program on the rate of water borne illness?

• **Methodological Question:** How should we estimate the impact of the program?
Spring Cleaning Sample

Total Population (562 springs) → Target Population (200) → Evaluation Sample (200) → Random Assignment

Year 1 (50) → Year 2 (50) → Years 3,4 (100)
Data collection

- Data from both treatment and control areas
- Measured water quality (\textit{E. coli}) at both springs and houses
- Representative sample of households using a spring (7-8 households per spring)
- Household survey: child anthropometrics, mother-reported diarrhea, hygiene practices, etc.
Impact of clean springs intervention

• 66% reduction in source water E. coli concentration
• 24% reduction in household E. coli concentration
• 25% reduction in incidence of diarrhea for children under 3 years old
• No significant change for 5-12 year olds
Is contaminated source water really the main problem?

• Quantity of water: Could be a stronger determinant of health than quality of water (Curtis et al, 2000)

• Hygiene: Water quality helps little without hygiene (Esrey, 1996)
  – 42% in the Kenya live without a toilet at home

• Low demand: People are more willing to pay for convenient water than clean water
Alternative Solution(s)?
<table>
<thead>
<tr>
<th>Intervention</th>
<th>Impact on Diarrhea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring protection (Kenya)</td>
<td>25% reduction in diarrhea incidence for ages 0-3</td>
</tr>
</tbody>
</table>
## Making Policy from Evidence

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Impact on Diarrhea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring protection (Kenya)</td>
<td>25% reduction in diarrhea incidence for ages 0-3</td>
</tr>
<tr>
<td>Source chlorine dispensers (Kenya)</td>
<td>20-40% reduction in diarrhea</td>
</tr>
<tr>
<td>Home chlorine distribution (Kenya)</td>
<td>20-40% reduction in diarrhea</td>
</tr>
<tr>
<td>Hand-washing (Pakistan)</td>
<td>53% drop in diarrhea incidence for children under 15 years old</td>
</tr>
<tr>
<td>Piped water in (Urban Morocco)</td>
<td>0.27 fewer days of diarrhea per child per week</td>
</tr>
</tbody>
</table>
III – Why randomize?
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.
IV – Conclusions
Conclusions - Why Randomize?

• There are many ways to estimate a program’s impact

• This workshop 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
When to do a randomized evaluation?

- When there is an important question you want/need to know the answer to
- Timing--not too early and not too late
- Program is representative not gold plated
  - Or tests a “proof of concept”
- Time, expertise, and money to do it right
- Develop an evaluation plan to prioritize
When NOT to do an RE

• When the program is premature and still requires considerable “tinkering” to work well

• When the project is on too small a scale to randomize into two “representative groups”

• If a positive impact has been proven using rigorous methodology and resources are sufficient to cover everyone

• After the program has already begun and you are not expanding elsewhere
Developing an evaluation strategy

- Start with a question
- Verify the question hasn’t been answered
- State a hypothesis
- Design the evaluation
- Determine whether the value of the answer is worth the cost of the evaluation

- With key questions answered from impact evaluations, process evaluation can give your overall impact
- A few high quality impact studies are worth more than many poor quality ones

If you ask the right question, you’re more likely to care
Questions?