Measuring impact

William Parienté
UC Louvain
J-PAL Europe

povertyactionlab.org
Course overview

1. What is evaluation?
2. Measuring impact
3. Why randomize?
4. How to randomize
5. Sampling and Sample Size
6. Analysis and inference
7. RCTs: Start to Finish
Introduction

• Measurement is central in an impact evaluation
• Objective of this lecture is to discuss the process to obtain credible indicators that we want to use for the evaluation
  – How to choose these variables/indicators?
  – How to collect them without errors
• Impact evaluation methods (and randomization) in the next course
Lecture overview

- An example of intervention
- Keys to successful measurement
  - Relevance of the outcomes
  - Choice of indicators
  - Sources of data
  - Reliability of data collection
  - Timing
An example of intervention

- Improvement of access to water
- In many rural areas, springs are the main source of water
  - But water may be contaminated
- High prevalence of waterborne diseases
- Intervention: protecting the source
  - Concrete base and pipe
Keys to Successful Measurement

• Relevance of the outcomes
• Choice of indicators
• Sources of data
• Reliability of data collection
• Timing
Relevance

• How do we ensure we measure the relevant outcomes?
Relevance: Drawing the chain of causality

- We want to answer more than:
  - how effective is the intervention?
- We also want to answer:
  - why it is effective?
- We want to draw the link

  inputs ➔ intermediary outcomes ➔ final outcomes

Defining and measuring intermediate outcomes will enrich our understanding of the program, reinforce our conclusions, and make it easier to draw general lessons.
Relevance: The Critical Role of Theory

1. Map out a theory of change
2. Use the theory to generate hypotheses that you can test in your project
3. What (theoretical) final outcomes are needed to demonstrate the validity of hypothesis
4. What (theoretical) intermediary outcomes needed to distinguish various hypotheses?
Theory of change

Access to clean water at source → Choose to collect only this water → Know that this source is clean → Understand the benefits of clean water → Have access to clean water at home → No recontamination → Hygiene practices → Contaminated water is the primary source of illness → Drink clean water → Choose to drink only clean water → Less diarrhea
## Logical framework

<table>
<thead>
<tr>
<th></th>
<th>Objectives Hierarchy</th>
<th>Assumptions / Threats</th>
<th>Indicators</th>
<th>Sources of Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Final outcome/Impact</strong></td>
<td>Lower rates of diarrhea</td>
<td>Waterborne disease is primary cause of diarrhea</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Int. Outcome (Project Objective)</strong></td>
<td>Households drink cleaner water</td>
<td>Shift away from dirty sources. No recontamination</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outputs (intermediary outcomes)</strong></td>
<td>Source water is cleaner; Families collect cleaner water</td>
<td>continued maintenance, knowledge of maintenance practices</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inputs (Activities)</strong></td>
<td>Source protection is built</td>
<td>Sufficient materials, funding, manpower</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Keys to Successful Measurement

• Relevance of the outcomes
• **Choice of indicators**
• Sources of data
• Reliability of data collection
• Timing
Relevance to Reality: choice of indicators

- Our theory and hypothesis helps us define the set of outcomes
- Need to find indicators that map well the outcomes
- Sometimes these are “basic” indicators which are relatively straightforward (e.g. height and weight)

- Other indicators may be more complicated
  - Income/savings
Relevance to Reality

• In some cases it is difficult to translate theoretical measures into real variables
  – Ability
  – Risk aversion, impatience, selfishness
  – Happiness,
  – Etc.

• But many creative ways to collect indicators
Hints on outcomes and indicators

• Choose those with a reasonable chance of being “moved” within the evaluation timeline

• Chose those that are not too difficult to collect and measure

• Chose those that occur with enough frequency to detect an impact given your sample size
## Logical framework

<table>
<thead>
<tr>
<th>Final outcome/Impact</th>
<th>Objectives Hierarchy</th>
<th>Assumptions / Threats</th>
<th>Indicators</th>
<th>Sources of Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower rates of diarrhea</td>
<td>Waterborne disease is primary cause of diarrhea</td>
<td>Rates of diarrhea</td>
<td>Household survey</td>
</tr>
</tbody>
</table>

### Int. Outcome (Project Objective)

<table>
<thead>
<tr>
<th></th>
<th>Households drink cleaner water</th>
<th>Shift away from dirty sources. No recontamination</th>
<th>(Δ in) drinking water source; E. coli</th>
<th>Household survey, water quality test at home storage</th>
</tr>
</thead>
</table>

### Outputs (intermediary outcomes)

<table>
<thead>
<tr>
<th></th>
<th>Source water is cleaner; Families collect cleaner water</th>
<th>continued maintenance, knowledge of maintenance practices</th>
<th>E. coli</th>
<th>Water quality test at source</th>
</tr>
</thead>
</table>

### Inputs (Activities)

<table>
<thead>
<tr>
<th></th>
<th>Source protection is built</th>
<th>Sufficient materials, funding, manpower</th>
<th>Protection is present, functional</th>
<th>Source visits/surveys</th>
</tr>
</thead>
</table>
Keys to Successful Measurement

• Relevance of the outcomes
• Choice of indicators
• Sources of data
• Reliability of data collection
• Timing
Sources of Measures

• Where do you obtain such data?
• Two main sources
  – Administrative data
  – Survey data
Sources of Measures

• Administrative data:
  – Advantages?
    • Often cheap, fast, easy and “clean” (although you would be surprised how often there are errors)
  – Disadvantages?
    • Data may not be relevant
      – Variables of interest may not be in the data
      – The data may not correspond to the relevant time period
    • May be difficult to get permission
    • May not come in the format you want
      – May be aggregated,
Sources of Measures

• Survey data (collected by you)
  – Advantages:
    • Control- you can ensure it is relevant and comprehensive and collected at the right time
  – Disadvantages:
    • Heavy time investment required
      – Management, designing, piloting and refining
      – Expensive
Keys to Successful Measurement

- Relevance of the outcomes
- Choice of indicators
- Sources of data
- Reliability of data collection
- Timing
Reliability and Accuracy

• Once you have selected the outcomes and the indicators to map the outcome
• need to make sure that the data collected is reliable and accurate.
  – Unreliable data → false or misleading conclusions
• The process of collecting “good” data requires a lot of efforts
Reliability: administrative data

- Administrative data:
  - General perception is that this source should be reliable but it is often full of errors
    - Important to understand HOW the data is collected and the CONTEXT
    - “automated” data collection generally high quality
    - Important to think or ask whether the data could be manipulated for someone’s gain
      - E.g. school enrollment records when schools get per capita student grant
Reliability: survey data

- The survey process:
  - Design of questionnaire → Survey printed on paper/electronic → filled in by enumerator interviewing the respondent → data entry → electronic dataset
  - Where can this go wrong?
Reliability: survey data

• The survey process:

• Design of questionnaire $\rightarrow$ Survey printed on paper/electronic $\rightarrow$ filled in by enumerator interviewing the respondent $\rightarrow$ data entry $\rightarrow$ electronic dataset

• Where can this go wrong? Everywhere
Reliability : survey data

- Start with a pilot
- Paper vs electronic survey
- Surveyors and supervision
- Following up the respondents
- Problems with respondents
- Paper or electronic version?
- Neutrality
Piloting

- Good surveys are developed by trial and error
- Often it’s good to start with a very basic set of questions asked in an open-ended way
  - More of a qualitative or focus group style
- Over time, the lessons from this can be refined into a survey
- Good to find out that respondents don’t understand certain issues before it’s too late
- Good piloting will almost always raise questions that weren’t thought of before
- Sometimes the research design can be changed slightly to capture these issues
Piloting

• As the survey becomes more formalized, working through the details becomes important
  – Phrasing of questions
  – Skip codes (“If no, skip to....”
  – Translation issues
Reliability: Paper, Phone, PDA, Tablet or Netbook

- With prices of electronics falling more surveys are collected via PDAs etc
- Regardless of data collection method it is necessary to implement relevant protocol and procedures to ensure data quality.
Reliability: Paper surveys

• Include ID number on all pages in case pages get separated
• The survey should look nice- easier for the enumerator to fill it out
• Minimize complicated skip patterns:
  – i.e. don’t have instructions like “if the answer to the question 5 pages ago was ‘no’, skip this question.”
• Surveys should be clear and should not leave room for interpretation by the enumerator.
• Minimize data entry errors: double entry and reconcile differences
Reliability: Electronic surveys

- Follow paper design tips
- Thoroughly test the electronic version before launch (esp. skip patterns etc)
- Make sure it is easy to fill out (stylus vs. typing)
- Program logical consistency checks
  - A male should not be a mother, daughter etc
  - Test these checks
- Protocol to transfer data from pda/laptop to central database efficiently and routinely.
Reliability: Electronic Vs Paper

- Trade off that should be evaluated on case by case
- Electronic surveys:
  - Huge upfront investment (time etc)
    - Makes more sense for surveys with many respondents
  - “Less flexible” (harder to change things on the go)
  - Faster- (no need to do data entry)
  - Less error (if programmed well)
  - Risky for enumerators to carry?
Training and supervision of Enumerators

• Training of enumerators is essential
  – Create manuals for all survey instruments.
  – Training long enough
  – Test enumerators in the field

• Good to do regular “back checking” – resurveying respondents to make sure that they were actually interviewed
  – Re-survey sample of respondents on a random basis

• Importance of close supervision
Follow up information

• Important to collect accurate tracking information
• Make sure that we can follow the respondents over time
  – Directions to home, GPS location, etc.
  – Get correct mobile phone number including that of parents, siblings and friends
Potential bias in answers from respondent

- Respondent might not be willing/able to give the “true” answer
- Why?
- Affects the accuracy of the data
- Survey design should try to minimize this bias
Social desirability bias

• Respondent might feel badly to tell the truth about some socially undesirable subjects
  – How many drinks did you have last week?
  – Do you always use a condom when having sex?
• We call this “social desirability bias”
Manipulation of Data

- Often occurs when there is something valuable at stake
  - CCT eligibility on the basis of a poverty score
  - CCT payouts on the basis of attendance
  - Teacher incentives on the basis of attendance measured by head-teacher
- Can lead to false conclusions
Manipulation of data

- Various strategies can be used to circumvent this problem
  - CCTs and enrollment and teacher attendance: unannounced random spot checks can be used
- Lesson: measures should be designed to be free of manipulation
Recall issues

• Some variables are easier to remember than others e.g. birth of a child
• Often have to collect data that involves recall where things are not so easy:
  – Time use, consumption, health, finances, exam scores
• How to ensure reliability?
Recall Issues

- **Easiest: shorten the recall period**
  - One strategy is to do high frequency short surveys where you ask respondents to recall events from prior day.

- **Diaries: e.g. time-use or financial diaries**
  - Can be difficult to implement - respondents don’t always fill them out (or fill them out when you ask for them)

- **Administrative data: e.g. for test scores**

- **One survey on nutrition** had enumerator live with family and weighed each individual's meals prior to eating and the amount wasted
  - Expensive
Neutrality

• The act of measurement can actually influence behavior of respondents
  – Live in enumerator measuring each person’s food
  – Financial diaries
• If treatment and control group respond differently to measurement then it would be problematic
Keys to Successful Measurement

• Relevance of the outcomes
• Choice of indicators
• Sources of data
• Reliability of data collection
• Timing
Timing

- When should you collect outcomes and variables?
- Related to theory and hypotheses
- Crucial question is “how long do you need for intervention effects to materialize?”
Research design: Baseline

• Although you can sometimes avoid collecting baseline data it is very risky to skip this step.
  • Sample might not be balanced, especially if sample isn’t very big
  • Having a baseline allows the researcher to control for baseline characteristics
    • E.g. initial test scores are very important to have in education
• Baseline allows you to look at effects for subgroups
Research design: Follow-up

- The amount of data that needs to be collected at endline really depends on the type of experiment that is being run
  - Sometimes take-up itself is the outcome of interest
    - for example, take up for microcredit
Research design: Follow-up

• Generally, getting follow-up data is crucial
• Impacts may take time to materialize
  • This may require to track individuals over several years
  • not always easy, respondents may migrate
    • Attrition
• Possible to conduct regular follow up
  – Understand the short term/longer term dynamics
  – Minimize attrition
  – increase precision
**Data collection examples: consumption**

**F3.0**

**INTERVIEWER:** Ask the respondent about their consumption of the following items *in the past 30 days*.

Ask about **goods purchased** (how many times they bought and for how much).

Don’t forget **home production**.
- Ask about quantity consumed in the last month out of own production.
- Ask about value of quantity consumed out of home production (how much it would have cost to buy that much).
- Do not forget to ask about fishing, hunting and foraging (this is part of home production).

Make sure to ask about all items the household is likely to consume.

<table>
<thead>
<tr>
<th>1. Cereals &amp; cereal products (rice, chakki flour, maida flour, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item</strong></td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2. Pulses &amp; pulse products (Bari daal, Moong ki daal, Channa daal, Masoor ki daal)</td>
</tr>
<tr>
<td><strong>Item</strong></td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Conclusion

• Important to choose the right outcome variables when we do an impact evaluation
  – need to understand the possible theory of change
• Find indicators that map these outcomes
• Data collection needs to be done very carefully to get reliable information
  – there might be problems at every stage of the process