Outcomes, indicators, and measuring impact

Rema Hanna
Harvard University

povertyactionlab.org
Course Overview

1. Why evaluate? What is evaluation?
2. Outcomes, indicators and measuring impact
3. Impact evaluation – why randomize
4. How to randomize
5. Sampling and Sample Size
6. Analysis and inference
7. RCTs: Start to Finish
Goals of measurement

• **Needs Assessment:**
  – Identifying problems/constraints that might help us choose among possible interventions/experiment

• **Background Information:**
  – Describing the environment within which the intervention/experiment takes place

• **Process Evaluation:**
  – Measuring the inputs into the intervention
  – Assessing the implementation of the intervention

• **Impact Evaluation:**
  – Measuring the outcomes/impact of the intervention
Lecture Overview

• Outcomes and indicators
• Logical Model
• Data Collection
Lecture Overview

• Outcomes and indicators
  – Intended goals
    Unintended consequences
  – Possible outcomes and indicators
• Logical Model
• Data Collection
The setting:
Quotas in the Village Councils
The setting: Quotas in the Village Councils

• What are the main goals of the Village Council?

• What are the main characteristics of the quota policy?
The controversy about quotas

• Why were quotas deemed to be desirable in this context?

• Why did some people doubt that quotas would be effective?
The possible effects

• Let’s start by drawing a list of everything we think quotas for women may affect.
Multiple (primary/final) outcomes

- Suppose that you collect data on 20 different outcomes
  - You find one is significantly positive
  - You find one is significantly negative
  - You find that for 18, the outcomes are very similar, and not significantly different

- What can you conclude?

Define key hypotheses before the beginning of the experiment.
Defining key hypotheses

• What might be examples of a few key hypotheses to test?

• Which variables, or combinations of variables, might you use to test these key hypotheses?
Lecture Overview

• Outcomes and indicators
• Logical Model
  – Key hypotheses and chain of causality
  – Hints on data collection
• Data Collection
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 → primary outcome

Defining and measuring intermediate outcomes will enrich our understanding of the program, reinforce our conclusions, and make it easier to draw general lessons.
Modeling the effects of quotas

- What are the possible chains of outcomes in the case of the quotas?
- What are the critical steps needed to obtain the final results?
- What variables should we try to obtain at every step of the way to discriminate between various models?
A theory of change...

- Imperfect Democracy
  - Pradhan’s preferences matter
  - Women are empowered
    - Public goods reflect women’s preference
      - Different Public goods
        - Different health, education Outcomes?

- Some democracy
  - More women Pradhans
    - Public goods reflect women’s preference

- Women have different preferences
## Model with indicators: quotas for local women leaders

<table>
<thead>
<tr>
<th>Needs</th>
<th>Input</th>
<th>Output</th>
<th>Outcome</th>
<th>Impact (primary outcome)</th>
<th>Long-term Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women have poor health and low levels of education. Their needs are not represented in local government</td>
<td>Reserved seats for women leaders</td>
<td>More local women leaders</td>
<td>Women more engaged and more directly involved in political decision making</td>
<td>Different public goods. Better education and health outcomes</td>
<td>Gender equality in health, education. Quotas no longer necessary?</td>
</tr>
</tbody>
</table>

### INDICATORS:
- Gender of leader
- Budget allocations change. Female attendance in council meetings.
- Literacy level. BMI for girls, boys.
Lecture Overview

- Outcomes and indicators
- Logical Model
- Data Collection
  - In practice
  - Data Entry
  - Ethics
## Data collected in panchayat study

<table>
<thead>
<tr>
<th>Tool</th>
<th>Target Respondent</th>
<th>Target Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP Interview</td>
<td>Village Leader</td>
<td>- Pradhan’s background (socioeconomic status, education)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Political ambitions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Political experience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Investments undertaken</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Public records. such as GP balance sheets</td>
</tr>
<tr>
<td>Transcript of Gram Sabha</td>
<td>GP</td>
<td>- Who speaks and when (gender)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- For how long do they speak?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- What issues do they raise?</td>
</tr>
</tbody>
</table>
# Data collected in panchayat study

<table>
<thead>
<tr>
<th>Tool</th>
<th>Target Respondent</th>
<th>Target Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Village Participatory Resource Appraisal</td>
<td>10 to 20 villagers per village</td>
<td>○ Village infrastructure (schools, roads, wells, SC and ST areas, cultivated land, irrigation, energy projects)</td>
</tr>
<tr>
<td>(village mapping exercise and focus groups)</td>
<td></td>
<td>Perception of quality of different public goods</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participation of men and women in activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>What issues villagers have raised with GP</td>
</tr>
<tr>
<td>Household interviews</td>
<td>Head of household (the male in some HH; the female in other HH)</td>
<td>○ HH demographic and socioeconomic data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HH outcomes (child health, measurement of height and weight, etc.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HH perceptions of quality of public goods and services</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Declared HH preferences</td>
</tr>
</tbody>
</table>
## Data collected in panchayat study

<table>
<thead>
<tr>
<th>Tool</th>
<th>Target Respondent</th>
<th>Target Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing administrative data</td>
<td>Public data archives (national, GP, and Village)</td>
<td>- A snapshot of village characteristics—population, public goods, demographics, etc.—at the time of the 1991 and 2001 census</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Expenditures on public goods and services in GP (from GP balance sheets)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Issues addressed at GP public assemblies (from Gram Sabha minutes)</td>
</tr>
</tbody>
</table>
Why collect your own data?

- To get the data on the variables that you are interested in
- To get adequate coverage of the treated (and control) population
- To get coverage of the treated population at the appropriate time
Why not collect field data?

A surveyor in Udaipur, India, searches for a respondent.
Why not collect field data?

• Time-consuming, risky, drawn-out process
  • High turnover of civil servants ➔ may lose your advocate in an office/ministry ➔ may lose permission to do research
  • Survey company fails to honor contract
  • Poor data quality and need to return to field
  • Natural/human disasters
  • Lots of management
Should you do a baseline?

• Technically you do not need to:
  – Treatment vs comparison endline is unbiased, so consider putting budget into larger sample

• However, baseline allow us to:
  – Check that randomization worked
  – Control for baseline characteristics, especially lagged value of outcome of interest (Other covariates could potentially soak up variation, but usually not much)
  – Baseline allows interactions
When to do the endline

- Outcomes for educational interventions seem to change over time
- Multiple waves of measurement
  - Results at one stage can help in securing funding for later stages
  - Increased precisions
  - Collect data at each stage to help find respondents later
Consider constraints when surveying

• Financial resources – tradeoff between sample size and amount of information obtained from each household

• Human resource capacity of organization implementing the survey – research coordinators, interviewers, data entry staff

• Willingness and ability of respondents to provide desired information
  – For example, some people may not know how old they are.
Respondent willingness and ability

• Willingness to provide desired information:
• Use objective measures if treatment or act of measurement may influence measurement
• Ability to provide information:
• Perceptions vs. reality
Data collection, I

- If sampling a larger target population, you will want a household-level census.
- Data entry format must be clear and should not leave room for interpretation by the enumerator.
Data collection, II

• Contracting out vs. employing enumerators directly?
• Training enumerators in these procedures is essential. Create manuals for all survey instruments.

New surveyors learning to use health equipment in Udaipur, India. Training of 45 surveyors lasted two weeks.
Data collection, III

- Need daily or weekly check of all the forms by a supervisor, and a re-check on a random basis by the research manager
- Re-survey sample of respondents on a random basis
Data collection, IV

- Data collected in multiple rounds
- Names and dates of enumerators on forms and dates of survey
- Forms whose pages can be separated
Data collection, V

- Can do some interesting randomizations during data collection (especially the pilot)
  - Order of questions
  - Nature of question
  - Framing of question
  - Characteristics of enumerator
  - Frequency of data collection
  - Form and value of compensation for respondents (if applicable)

- Check if these systematically affect responses
  - If they do, potential problems with measurement
  - Might not be measuring parameters you’re intending to measure
Data collection, VI

- Field team includes interviewers and supervisors
- A supervisor role
- Ideal team size under one supervisor depends on area of survey, length of questionnaire (8 to 10 typical)
- Payment structure for team and ensuring high data quality

Grosh and Glewwe, 2005
Pilots vary in size & rigor

Pilots & qualitative steps are important

Something always goes wrong

– Respondents might not understand a question

Better to find this out before study begins

Often discover other interesting questions in process
Data entry

- Enter data quickly to catch problems
- When entering data assign a Survey Number to each questionnaire
- Scan documents
- Invest in good data entry database (can use software tool)
- Do double entry of all data and reconcile with the hard copies to detect mistakes.
- Re-enter some entries a third time (supervisor) and seek to achieve less than 0.5% error rate.
- After re-entry clean data
- After re-entry, clean data.
Data entry

Increasingly possible to do things on laptops, PDAs, and cell phones in the field. Goes directly to a data base (e.g. Datadine and Google’s Android platforms)

Courtesy of Chris Blattman. Used with permission.
Budgeting

- Budget adequately
- Things go wrong
  - Exchange rate movements
  - Enumerator downtime
  - Resurveys needed
- New opportunities
Human subjects

• Check what approvals needed
  – Country IRBs
  – University IRBs
• Allow adequate time
• Oral vs. written consent
• Permissions
  – National government
  – Local authorities
  – Relevant ministry
Hints on outcomes and 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
Comparing outcomes across studies

• What do we do if some tests are easier than others?
• We “standardize” or “normalize”
  \[
  \text{testscore}_{\text{normalized}} = \frac{(\text{testscore}-\text{average testscore})}{\text{standard deviation of testscore}}
  \]