Outcomes, indicators, and measuring impact

Rema Hanna

Harvard University

povertyactionlab.org

1

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

- <u>Needs Assessment:</u>
  - Identifying problems/constraints that might help us choose among possible interventions/experiment

- <u>Background Information</u>:
  - Describing the environment within which the intervention/experiment takes place

- <u>Process Evaluation:</u>
  - Measuring the inputs into the intervention
  - Assessing the implementation of the intervention

- <u>**Impact Evaluation:**</u>
  - **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 <u>before</u> 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
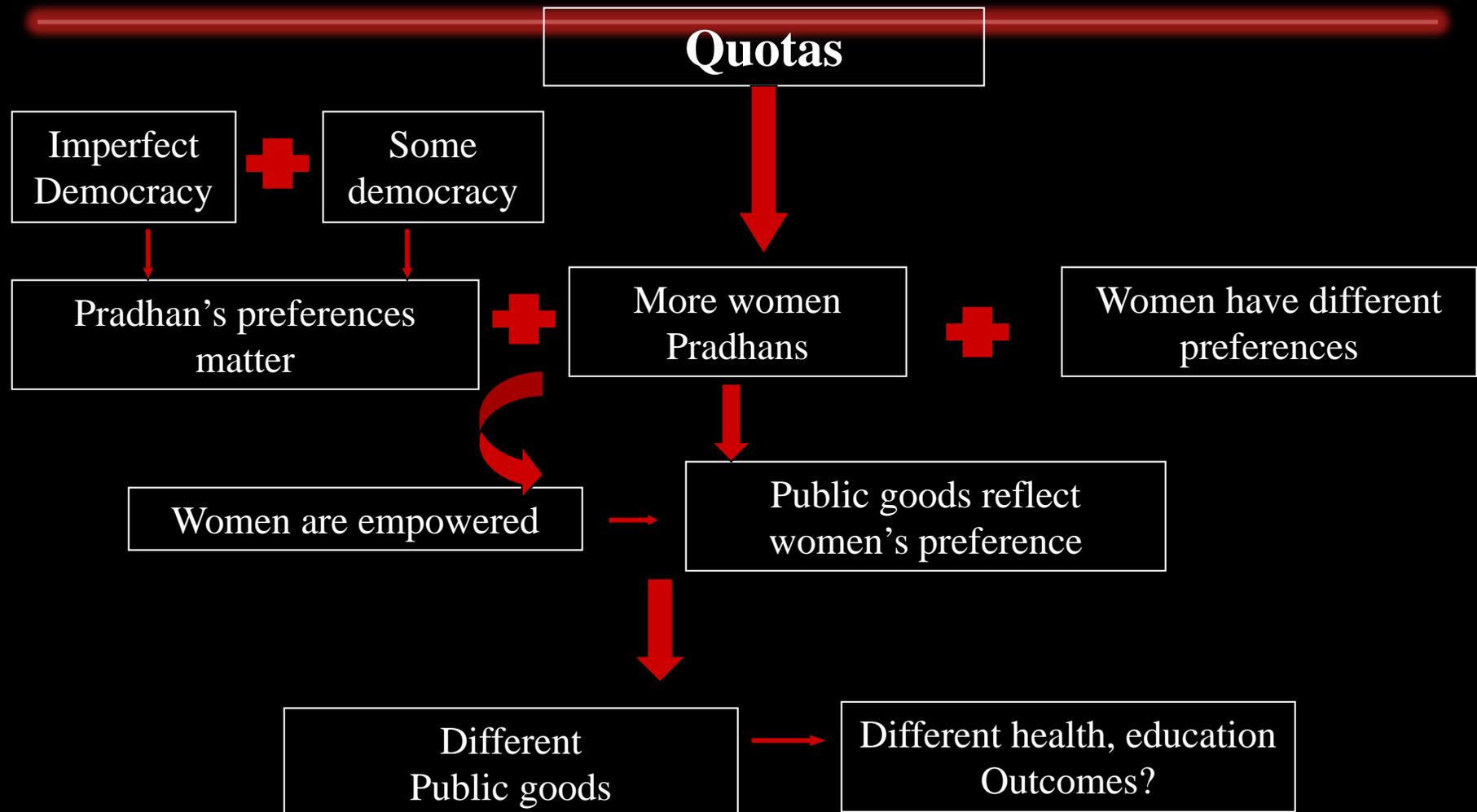
Modeling the effects of quotas

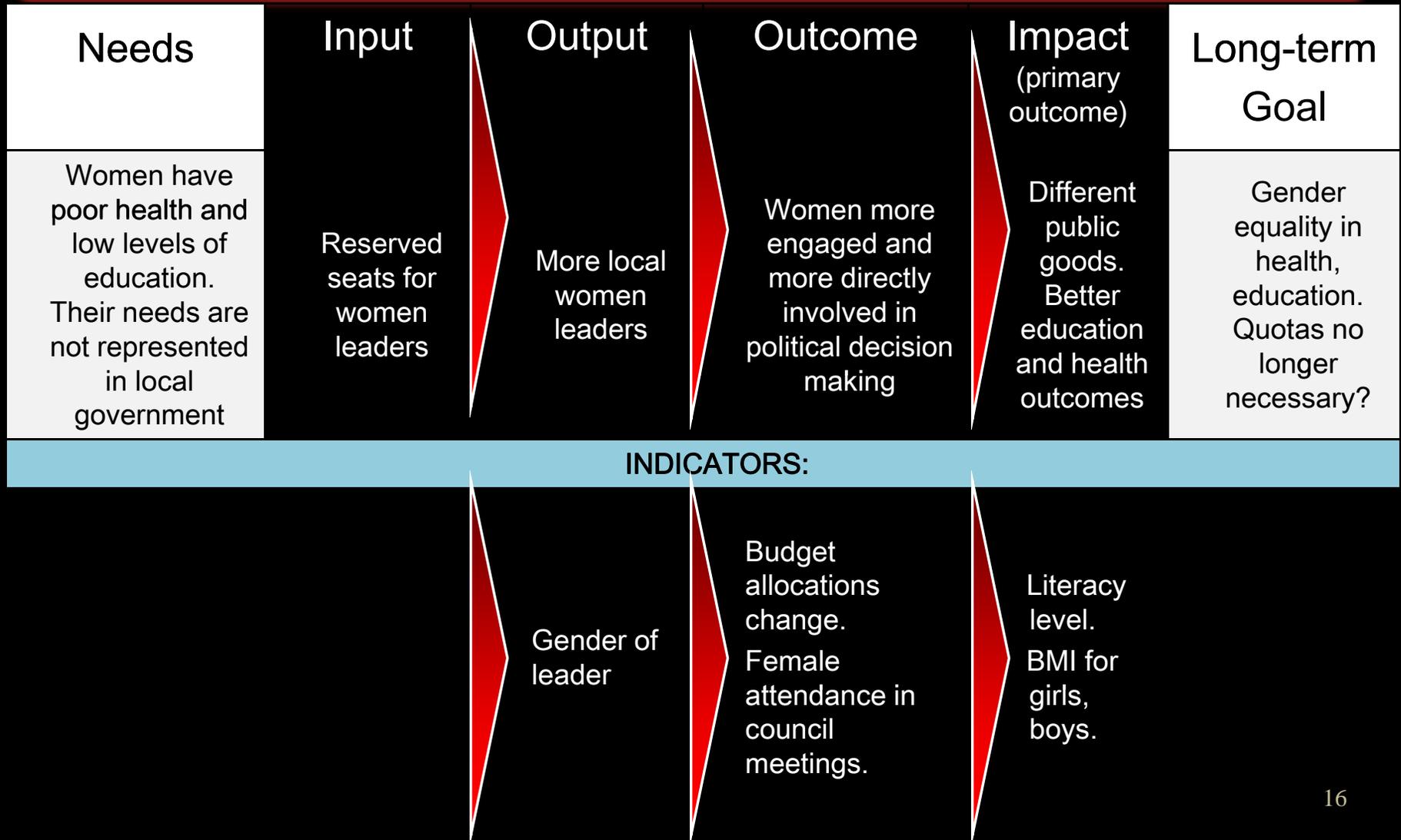- What are the possible chains of outcomes in the case of the quotas?

  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…

Quotas

Imperfect Democracy **+** Some democracy

Pradhan's preferences matter **+** More women Pradhans **+** Women have different preferences

Women are empowered → Public goods reflect women's preference

Different Public goods → Different health, education Outcomes?

Model with indicators:
quotas for local women leaders

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

INDICATORS:

		Gender of leader	Budget allocations change. Female attendance in council meetings.	Literacy level. BMI for girls, boys.	

16

Lecture Overview

- Outcomes and indicators

- Logical Model

- Data Collection

  - In practice

  - Data Entry

  - Ethics

Data collected in panchayat study

Tool	Target Respondent	Target Outcomes
GP Interview	Village Leader	o Pradhan's background (socioeconomic status, education) o Political ambitions o Political experience o Investments undertaken o Public records. such as GP balance sheets
Transcript of Gram Sabha	GP	o Who speaks and when (gender) o For how long do they speak? o What issues do they raise?

Data collected in panchayat study

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

Data collected in panchayat study

Tool	Target Respondent	Target Outcomes
Existing administrative data	Public data archives (national, GP, and Village)	o A snapshot of village characteristics—population, public goods, demographics, etc.—at the time of the 1991 and 2001 census o Expenditures on public goods and services in GP (from GP balance sheets) o Issues addressed at GP public assemblies (from Gram Sabha minutes)

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

24

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.

26

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

33

Piloting the survey

Data entry

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)
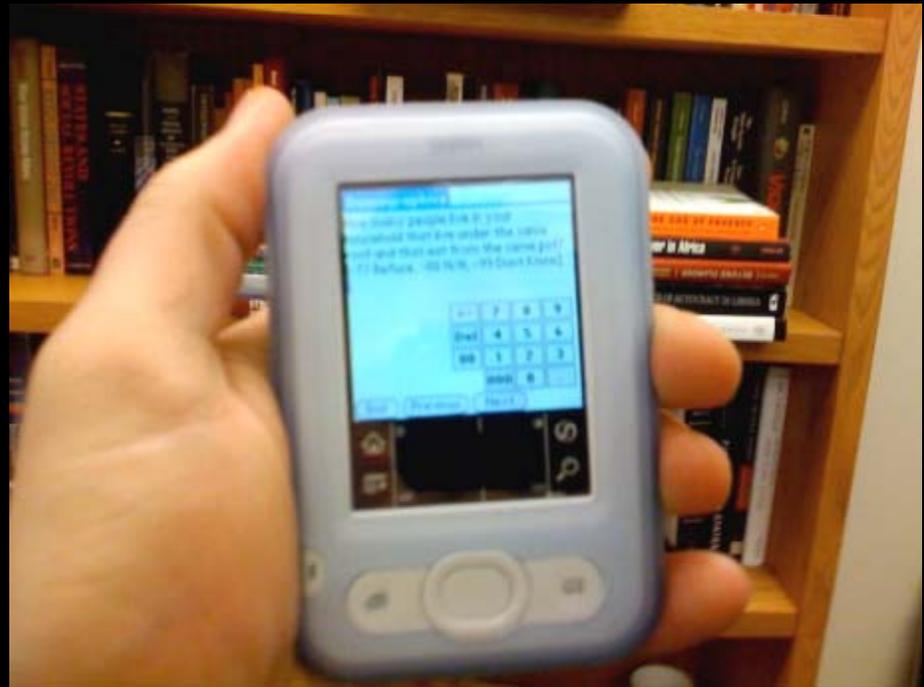
Courtesy of Chris Blattman. Used with permission.

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"
  - $testscore_{normalized} =$
  $(testscore\text{-}average\ testscore) / _{standard\ deviation\ of\ testscore})$

Resource: Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs
Dr. Rachel Glennerster, Prof. Abhijit Banerjee, Prof. Esther Duflo