Measurement

Nava Ashraf
Harvard Business School
Course Overview

1. What is evaluation?
2. Measuring impacts (outcomes, indicators)
3. Why randomize?
4. How to randomize
5. Sampling and sample size
6. Threats and Analysis
7. Cost Effectiveness Analysis and Scaling Up
8. RCT: Start to Finish
Course Overview

1. What is evaluation?
2. Measuring impacts (outcomes, indicators)
3. Why randomize?
4. How to randomize
5. Sampling and sample size
6. Threats and Analysis
7. Cost Effectiveness Analysis and Scaling Up
8. RCT: Start to Finish
Lecture Overview

• What to Measure
  – Case study review (Theory of Change)

• How to measure it (well)
  – Validity, Reliability
  – How to measure the immeasurable
  – Sources of data
  – Data collection
  – Other considerations
Case study

- Women as Policymakers
Public good investments

- Public goods reflect Women’s preferences
- Women have different preferences
- Investments reflect women’s preferences
  - Pradhan’s preferences matter
  - Imperfect democracy
  - Some democracy
  - Reservations for Women
  - More female pradhans
  - Women are empowered
## Log frame

<table>
<thead>
<tr>
<th>Objectives Hierarchy</th>
<th>Indicators</th>
<th>Sources of Verification</th>
<th>Assumptions / Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Impact</strong> (Goal/ Overall objective)</td>
<td>Public good investment represents women’s preferences</td>
<td>Government spending</td>
<td>Pradhan preferences matter: imperfect/some democracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Administrative data: Budgets, Balance Sheets</td>
<td></td>
</tr>
<tr>
<td><strong>Outcome</strong> (Project Objective)</td>
<td>Women voice political views</td>
<td>Number of times a woman spoke</td>
<td>Women develop independent views</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transcript from village meeting</td>
<td></td>
</tr>
<tr>
<td><strong>Outputs</strong></td>
<td>More female Pradhans</td>
<td>Whether or not a Panchayat had a female Pradhan</td>
<td>The law is implemented, there is no backlash</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Administrative records</td>
<td></td>
</tr>
<tr>
<td><strong>Inputs</strong> (Activities)</td>
<td>Reservations for women</td>
<td>Law is passed</td>
<td>The government realizes the need for women representation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The constitution</td>
<td></td>
</tr>
</tbody>
</table>

## Data used

<table>
<thead>
<tr>
<th>Sources of Measurement</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household (HH) Survey</td>
<td>• Declared HH preferences&lt;br&gt;• HH perceptions of quality of public goods and services</td>
</tr>
<tr>
<td>Village Leader Interview</td>
<td>• Political experience&lt;br&gt;• Investments undertaken</td>
</tr>
<tr>
<td>Village PRA</td>
<td>• Village infrastructure + investments&lt;br&gt;• Perception of public good quality&lt;br&gt;• Participation of men and women&lt;br&gt;• Issues</td>
</tr>
<tr>
<td>Administrative Data</td>
<td>• Budgets&lt;br&gt;• Balance sheets</td>
</tr>
<tr>
<td>Transcript from village meeting</td>
<td>• Who speaks and when (gender)&lt;br&gt;• Issues raised</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Issue</th>
<th>Investment</th>
<th>West Bengal</th>
<th>Rajasthan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Issue</td>
<td>Reserved Investment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W</td>
<td>M</td>
</tr>
<tr>
<td>Drinking Water</td>
<td># facilities</td>
<td>31%</td>
<td>17%</td>
</tr>
<tr>
<td>Road Improvement</td>
<td>Road Condition (0-1)</td>
<td>31%</td>
<td>25%</td>
</tr>
<tr>
<td>Irrigation</td>
<td># facilities</td>
<td>4%</td>
<td>20%</td>
</tr>
<tr>
<td>Education</td>
<td>Informal education center</td>
<td>6%</td>
<td>12%</td>
</tr>
</tbody>
</table>
Women empowerment?

- Voters respect women’s voices (gender bias)
- Parents invest in daughters
- Women’s empowerment
- More female pradhans
- Women can be re-elected
- Career opportunities for women (in politics)
- Reservations for women
Perceptions and Attitudes

• “How effective is your leader?” (ineffective, somewhat effective, effective, very…)  
  – Survey response: explicit distaste for female leaders (Feeling Thermometer)

• Listen to a Vignette (Male v. Female)  
  – Bias large and significant for male listeners

• Revealed preference – voting behavior

• Implicit Association tests  
  – Increased likelihood of associating women with leadership activities
Implicit Association Test
Taking a real IAT

If the keys do not work, click the mouse inside the white box and try again.
If the red appears, press the other key to make the red go away.
HOW TO MEASURE IT (WELL)

• The basics
The Basics

• Data that should be easy?
  – E.g. Age, # of rooms in house, # in hh

• What is the survey question identifying?
  – E.g. Are hh members people who are related to the household head? People who eat in the household? People who sleep in the household? Bobcats?

• Pre-test questions in local languages
When the obvious is not so obvious...

• Let’s think about the people who eat from the same pot in the household where you usually stay. There are how many adults, adolescents, and children? Adults are age 18 and older, adolescents are ages 13 to 17, and children are ages 12 and younger.

  – So in total there are how many people in the household where you usually stay? DON’T ADD TOTAL FOR RESPONDENT.
Validity, Reliability

• How to measure it (well)
The main challenge in measurement

- Accuracy
- Precision
The main challenge in measurement

• Validity

• Reliability
Validity

- In theory:
  - How well does the indicator map to the outcome? (e.g. intelligence $\rightarrow$ IQ tests)

- In practice:
  - Are your survey questions unbiased?
  - Potential biases:
    - Social desirability bias
    - Demand bias (response bias)
    - Framing effect
    - Recall bias
    - Anchoring bias
Reliability

• In theory:
  – The measure is consistent, precise, but not necessarily valid

• In practice:
  – Length, fatigue
  – “How much did you spend on broccoli yesterday?” (as a measure of annual broccoli spending)
  – Ambiguous wording (definitions, relationships, recall period)
  – Answer choice (open/closed, Likert, ranked)
General noise

- Surveyor training/quality
- Data entry
- Poor translation
- How do you generalize from certain questions?
Measuring the immeasurable

• How to measure it (well)
What is hard to measure?

(1) Things people do not know very well

(2) Things people do not want to talk about

(3) Abstract concepts

(4) Things that are not (always) directly observable

(5) Things that are best directly observed
Why the Hard to Measure?

• Missing key characteristics that:
  – interact with policies to change their impact
  – help us tailor policies and programs to better reach stated objectives

• If we can't measure it, we can't evaluate its importance
1. Things people do not know very well

**What:** Anything to estimate, particularly across time. Prone to recall error and poor estimation

- **Examples:** distance to health center, profit, consumption, income, plot size

**Strategies:**

- Consistency checks – How much did you spend in the last week on x? How much did you spend in the last 4 weeks on x?
- Multiple measurements of same indicator – How many minutes does it take to walk to the health center? How many kilometers away is the health center?
2. Things people don’t want to talk about

What: Anything socially “risky” or something painful

Examples: sexual activity, alcohol and drug use, domestic violence, conduct during wartime, mental health

Strategies:

• Don’t start with the hard stuff!
• Consider asking question in third person
• Always ensure comfort and privacy of respondent
Choosing Fruit vs. Chocolate
Read and van Leeuwen (1998)

If you were deciding today, would you choose fruit or chocolate for next week?
Patient Choices for the Future

Choosing Today  Eating Next Week  Time

Today, 74% of subjects choose fruit for next week.
Choosing & Eating Simultaneously

If you were deciding today, would you choose fruit or chocolate for today?
Today, 70% of subjects choose chocolate for today.
“...but you must bind me hard and fast, so that I cannot stir from the spot where you will stand me... and if I beg you to release me, you must tighten and add to my bonds.”

- The Odyssey
I, __________________, commit to save for ____________________.

I have opened a SEED savings account with a

Goal Date / Goal Amount of ________.

I will try everything in my power to accomplish my SEED Savings Goal by

saving ____________ Pesos a day / a week.

If I achieve this goal, I will be able to enjoy my savings to ____________

by ________________.

__________________________________  _________________________
Name                                      Date
Results

Average bank account savings after 6 months, 46%; after 12 months, 80%

Increase in average savings for those who took up: after 6 months, 192%; after 12 months, 337%

Similar product subsequently used by the Bank to help clients keep commitment to stop smoking
3. Abstract concepts

**What:** Potentially the most challenging and interesting type of difficult-to-measure indicators

- **Examples:** empowerment, bargaining power, social cohesion, risk aversion

- **Strategies:**
  - Three key steps when measuring “abstract concepts”
    - Define what you mean by your abstract concept
    - Choose the outcome that you want to serve as the measurement of your concept
    - Design a good question to measure that outcome

- Often choice between choosing a self-reported measure and a behavioral measure – both can add value!

• Despite advent of modern contraception, more than 50% of births are reported as unwanted.

• Contraceptive methods often stocked out, rationed by wait times.

• Official policy that husband consent is not required to obtain contraceptives; not in rural areas. Many husbands unaware of birth control use (21%).
The baby gave you nine months' notice. Where're the napkins, shawls and everything?

There's nothing at all, sister. In fact we've problems in feeding and dressing the other ten children at home.

Take off your shirt so that we can wrap in the baby.

Okay sister.

You men're to blame because you always discourage your wives to use any of the many family planning methods!

Never again, darling. Family planning makes sense!
Free, Instant Access to Depo Provera or Jadelle

If you bring this voucher with your NRC card to Nurse Grace Daka at Chipata Clinic between xx and xx hours Monday through Saturday, we guarantee that you will receive:

- very quick, first-class personalized family planning services from Grace Daka, your own dedicated family planning nurse only for the lucky women in this study
- one implant of Jadelle or one years' worth of Depo Provera if the nurse deems it medically appropriate
- a wait time of no more than 30 minutes; we will give you a free gift if you wait longer than 30 minutes
- a free, surprise gift for you, the woman of the house, if you are one of the first 50 women to see Grace Daka with this voucher

Name ..............................................................
NRC# ..............................................................
Date ..............................................................

see inside for details
Control

Individual

Couples

Free, Instant Access to Depo Provera or Jadelle

If you bring this voucher with your NICHD card to Nurse Grace Dace or Qopen Clinic between six and six hours Monday through Saturday, we guarantee that you will receive:

1. Same-day, friendly, personalized family planning services from Grace Davis, your own dedicated family planning nurse only for the lucky woman in this story.
2. A free implant of Jadelle or one year's worth of Depo Provera if the nurse deems it medically appropriate.
3. A wait time of no more than 30 minutes. We will give you a free gift if your wait is longer than 30 minutes.
4. A free surprise gift for you, the woman of this house, if you are one of the first 50 women to see Grace Davis with this voucher.

Hand____________________________________
Name:___________________________________
Date:___________________________________
Results

Women who received the voucher alone:

- 23% more likely to visit a family planning nurse
- 38% more likely to take up a concealable form of contraception
- 57% reduction in unwanted births

...than when they were given the voucher with their husband.

Only when women have greater autonomy to adopt contraception does it lead to a decrease in unwanted births.

...but is that the whole story?
How "socially connected" do you feel to the other people in this room?
Things that aren’t directly observable

**What:** You may want to measure outcomes that you can’t ask directly about or directly observe

- **Examples:** corruption, fraud, discrimination

**Strategies:**

- Sometimes you just have to be clever...
- Don’t worry – there have already been lots of clever people before you – so do literature reviews!
Things that are best directly observed

What: Behavioral preferences, anything that is more believable when done than said

Strategies:
- Develop detailed protocols
- Ensure data collection of behavioral measures done under the same circumstances for all individuals
Sources of data
Where can we get data?

• Administrative Data
  – State government census data
  – (Anonymized) voting data
  – Cell phone usage (e.g. anonymized mPesa transfers)

• Other Secondary Data
  – World Bank/UN/IFPRI

• Primary Data
  – Your own survey
Primary Data Collection

• Self-reported Surveys
• Exams, tests, etc
• Games
• Vignettes
• Direct Observation
• Diaries/Logs
Modules

- Income, consumption, expenditure
- Perceptions, expectations, aspirations
- Bargaining power
- Patience, risk
- Behavior (time use)
- Anthropometric
- Cognitive, Learning
- Yields
Why collect your own data?

- The standard RCT design is
  - Baseline
  - During the intervention
  - Endline
  - Scale-up, intervention

- Pros vs. cons of collecting your own data
  - Scale, cost
  - Focus of questions
Data Collection Considerations

• Quality Control
• Surveyor training
• Surveyor (gender) composition
• Human subjects
• Data Security
• Electronic v paper
• Costs
OTHER CONSIDERATIONS

• How to measure it (well)
Don’t forget

• Ethics
• Might affect compliance
• Respondent (and interviewer) fatigue