Measurement: Outcomes, Impacts, and Indicators

Vincent Pons
Professor, Harvard Business School
Course Overview

1. What is Evaluation?
2. Measurement: Outcomes, Impacts, and Indicators
3. Why Randomize?
4. How to Randomize?
5. Sampling and Sample Size
6. Threats and Analysis
7. Start to Finish
8. Generalizability
Measurement:

Outcomes, Impacts, and Indicators
Example: Tuberculosis in India

- **Low compliance of TB patients with their treatment**

  - **Setting:**
    - Urban slums in India; high rates of TB, patients often stop taking pills before the end of their 6-months treatment
    - Partnership with Operation Asha, a health NGO in Northern Indian states

  - **Program:**
    - Biometric devices installed in treatment centers to keep track of health workers’ presence and patients’ compliance with the treatment

  - **Measurement:**
    - Data on health workers’ presence, patients’ visits to the center, and their health
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
## Theory of Change

<table>
<thead>
<tr>
<th>Needs assessment</th>
<th>Inputs</th>
<th>Outputs</th>
<th>Intermediary Outcomes</th>
<th>Outcome</th>
</tr>
</thead>
</table>
| • TB patients fail to comply with the treatment and fall ill again  
  • Patients’ health and rate of TB treatment success | • Biometric devices installed in TB centers  
  • Fraction of devices which actually work | • Health workers are more likely to attend centers and follow-up with patients who forgot to come  
  • Observation days, spent with health workers | • Patients are more likely to take pills and finish treatment  
  • Surveys and observation days | • Healthier patients, rate of relapse  
  • Height and weight, health symptoms |
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
### Purpose of Measurement (I)

<table>
<thead>
<tr>
<th>Needs assessment</th>
<th>• What are outcomes at baseline?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs</td>
<td>• Are high-quality inputs provided?</td>
</tr>
<tr>
<td>Outputs</td>
<td>• Do intended recipients get the full intervention?</td>
</tr>
<tr>
<td>Intermediary Outcomes</td>
<td>• Do recipients change behavior?</td>
</tr>
<tr>
<td>Outcome</td>
<td>• What are long-term outcomes?</td>
</tr>
</tbody>
</table>
Purpose of Measurement (II)

• Covariates, context for generalizability

• Treatment compliance (individual and group level)
  – Predictors of compliance

• Heterogeneous treatment effects
  – Predictors of heterogeneity

• Cost effectiveness

• Qualitative information (The “Why?” and “How?”)
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
Where can we get data?

- From existing sources (Secondary data)
  - Publicly available
  - Administrative data
  - Other secondary data
- Collected by researchers (Primary data)
  - Surveys
  - Non-survey methods
Types and Sources of Data

Information about a person/household

- Cognition, anthropometrics
- Demographics
- Behavior, beliefs
- Patience, risk aversion, psychometrics
- Knowledge
- Income, expenditure

NOT about a person/household

- Farming inputs and outputs
- Quality of medical care
- Business income taxes

Information reported by a person

- Bank transactions
- Phone data
- Sales records
- School/university records, criminal record

Automatically generated

- Prices
- Weather, air quality
- Stock markets
- VAT records
Primary Data Collection

- Surveys
- Exams, tests, etc.
- Games
- Vignettes
- Direct observation
- Diaries/logs
- Focus groups
- Interviews
Primary Data: Modes

- **Interviewer administered**
  - Paper-based
  - Computer-assisted/digital
  - Telephone-based

- **Self-administered**
  - Paper
  - Computer/digital
  - Task platforms like Amazon Mechanical Turk
Primary Data: Who, Where, When?

• Survey respondent
  – Who has the most information for question of interest?
  – Who is most likely available?
  – Who is least likely to be biased?
• Example: head of household vs. person at home, mother vs. healthcare provider

• Survey setting
  – Is the respondent distracted?
  – Will peers influence responses?
• Example: interview at clinic vs. at home, interview at school vs. at home

• Survey timing
  – Could the survey affect behavior?
• Example: conduct health test just before or after health worker visit
Suppose you want to learn how often child patients came to treatment centers to take pills. Who would you ask?

A. The health worker
B. The health worker’s supervisor
C. The child
D. The mother
E. No one -- use the health worker’s records
Administrative data

Information collected, used, and stored primarily for administrative (i.e., operational) purposes, rather than research purposes.

- Medical records
- Grade books
- Arrest records
- Bank account data
- Personnel records
- Log books
Administrative Data: Sources

• Health
  – Vital statistics office
  – Health facilities (e.g., hospitals, clinics)

• Finance
  – Banks, credit unions
  – Credit rating agencies

• Education
  – Schools and universities
  – Department of education

• J-PAL’s Catalog of US data sets
• J-PAL’s collaboration with governments, e.g. in Tamil Nadu (India)
• Administrative records of program being evaluated
Why are administrative data useful?

The outcomes and metrics required for a study may already be tracked by a government or organization:

- Available retrospectively
- Enable long-term follow-up
- May include near census of relevant population
- Reduce logistical burden and burden on subjects
- Often less expensive than surveys
- May reduce bias and error
What source of data do you primarily expect to use in your work?

A. Primary survey data
B. Other primary (non-survey) data
C. Administrative data
D. Public data
E. Other secondary data
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
Concept of Measurement

Construct
(Intelligence/cognitive development)

Indicator
(IQ Test result)

Data
(Recorded test scores)

https://commons.wikimedia.org/wiki/File:Red_Silhouette_-_Brain.svg
Concept of Measurement

Data
(Arrest reports)

Construct
(Property crime)

Indicator
(Number of arrests for property crimes)

Handcuffs by Anbileru Adaleru from the Noun Project
Photo: A. Sautmann, NYC 2008
Concept of measurement

- What type of information represents the construct best?
- How do we collect this information?
- How precisely will the recorded information represent the true information?
School attendance records reported by Texas public schools in 2017 are

A. A construct
B. An indicator
C. Data
D. Don’t know
Income inequality is

A. A construct
B. An indicator
C. Data
D. Don’t know
Hospital visits per month from electronic medical records are

A. A construct
B. An indicator
C. Data
D. Don’t know
The Goals of Measurement

- Accuracy/Unbiasedness
- Validity
Validity

Construct

Indicators

In theory: how well does the indicator map to the construct we are trying to measure?
The Goals of Measurement

- Accuracy/Unbiasedness
- Validity

- Precision
- Reliability
Reliability

Is the indicator measured in a way that is consistent and precise?

Would we get the same data if we measured several times?
The Goals of Measurement

- Reliability
- Validity
Which is worst?

A. Low validity, low reliability
B. Low validity, high reliability
C. High validity, low reliability
D. All equally bad
E. Don’t know/can’t say
Outcome: weekly consumption
Indicator: expenditure in last week
Which is likely to be the main problem?
A. Validity
B. Reliability
C. Both
D. Neither
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
Proxy indicators

- Constructs are hard to measure

- Proxy indicators
  - Correlated with construct
  - Correlation is dynamic

- A good proxy for nutrition:
  - Weight for height

- Not a good proxy:
  - Height alone
Indices: Examples

- State of the economy
- Consumer Price Index (CPI)
- Stock market indices
- Human Development Index (HDI)
- Corruption index
- Women’s empowerment

→ Combine several indicators into a weighted summary measure
  → Advantage: can capture different aspects of a construct
  → Disadvantage: opacity
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
The Response Process

Construct → Indicators → Data Collection → Data

Error / Measurement bias
Four-Step Response Process

1. Comprehension of the question
2. Retrieval of Information
3. Judgement and Estimation
4. Reporting an Answer
Step 1: Comprehension

1.1 Total monthly income, before taxes

____________________
____________________
____________________
____________________
Step 2: Retrieval

Social Security benefits, Unemployment or Workers’ Compensation, Pensions...
Step 3: Estimation/Judgement

Social = $200 per month
Workers’ Compensation = 0
Pension = $220 per month
What else??
Step 4: Response

1. Comprehension of the question
2. Retrieval of Information
3. Judgement and Estimation
4. Reporting an Answer

1.1 Total monthly income, before taxes
____________________
____________________
____________________
____________________

$400
Which stage in the response process might produce measurement error?

A. Comprehension
B. Retrieval
C. Judgement/estimation
D. Reporting
E. All of the above
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
Measurement Error: Social Desirability Bias

Example:

Q. Were you arrested in the past month?
• Yes
• No

Respondents might hesitate to admit arrests

Tendency of respondents to answer questions in a manner that emphasize strengths, hide flaws, or avoid stigma

Ask indirectly, ensure privacy
Measurement Error: Framing

Example:
Q1. How many years of education do you have?
Q2. Did you go to public or private school?
Q3. Did your school provide everyone a quality education?
Q4. For the upcoming election, what are the top policy priorities you are looking for from the candidates?

Suddenly, education becomes everyone’s top priority

Phrasing of questions or location in the questionnaire changes respondents focus or way of thinking

Be careful of where questions are placed and if they are phrased to express implicit judgement.
Measurement Error: Telescoping Bias

Example:

Many health surveys in developing countries ask about birth history: Q. Did you give birth to a child who passed away within the last year?

Respondents may report very salient events that happened more than 12 months ago.

People perceive recent events as being more remote than they are (backward telescoping) and distant events as being more recent than they are (forward telescoping).

Visit once at the beginning of the reference period, ask the question. Then ask, “since the last time I visited you, have you...?”
Other things to consider

• Question wording
  – Specific and easy to understand
  – Avoid double negatives
  – Back-translate and pre-test in other languages

• Multiple choice questions: answer options that
  – Cover all possible answers
  – Do not overlap
  – If possible, do not require “check all that apply”

• Survey length and respondent fatigue

• Surveyor training/quality
• Piloting
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
Access to Administrative Data

Regulations that limit access to identified data

- **US: Health Insurance Portability and Accountability Act (HIPAA)**
  - Protects health data and privacy of patients
- **US: Family Educational Rights and Privacy Act (FERPA)**
  - Protects students’ educational records
  - Requires student consent for release of records
- **Other regulations depending on country context – e.g. GDPR**

- **Obligations depend on level of identification and your contract with the data provider**
  - Data security requirements
  - Fines for data leaks
  - Individual authorizations or waivers (similar to informed consent)
Access to Administrative Data: Generating an Electronic File

Records are in an unusable format

- Hand-written records
- PDF file

To address records in an unusable format

Digitize

- Clinic ledgers in Mali
- Tax records in Bangladesh
- Books and academic papers worldwide
Using Administrative Data: Matching

- **Goal:** match administrative with program or survey data

- **Personally Identifiable Information (PII)**
  - Name, Identification numbers, Address
  - Photos or biometric characteristics

- Understand what identifiers the data agency collects
  - Collect those same identifiers from your study sample at baseline

- Use numeric identifiers instead of string variables
  - Identification number and DOB instead of names and addresses

- Participants may not be willing to provide sensitive identifiers
  - E.g. bank account or social security numbers -- emphasize privacy & confidentiality during study enrollment
Using Administrative Data: Separate Identifiers from Outcomes

<table>
<thead>
<tr>
<th>Identified Finder File</th>
<th>Administrative Data File</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study ID</td>
<td>Name</td>
</tr>
<tr>
<td></td>
<td>DOB</td>
</tr>
<tr>
<td></td>
<td>SSN</td>
</tr>
<tr>
<td></td>
<td>Name</td>
</tr>
<tr>
<td></td>
<td>DOB</td>
</tr>
<tr>
<td></td>
<td>SSN</td>
</tr>
<tr>
<td></td>
<td>Outcome 1</td>
</tr>
<tr>
<td></td>
<td>Outcome 2</td>
</tr>
</tbody>
</table>

\[\text{Identified Finder File} + \text{Administrative Data File} = \text{De-identified Analysis File}\]

<table>
<thead>
<tr>
<th>De-identified Analysis File</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study ID</td>
</tr>
<tr>
<td>Treatment Status</td>
</tr>
<tr>
<td>Outcome 1</td>
</tr>
<tr>
<td>Outcome 2</td>
</tr>
</tbody>
</table>
When choosing identifiers for matching study data to administrative data, which of the following identifiers would be preferable to using an individual’s street address?

A. An email address
B. A government-issued, unique identification number
C. Date of birth
D. All of the above
E. B and C
Using Administrative Data: Determining Data Accuracy

How can we ensure that the data are accurate?

- Unlike with survey data, the researcher does not have a say in the data collection and processing phase.
Using Administrative Data: Determining Data Accuracy

To address possibly inaccurate data

- **Cross-reference with other sources to ensure accuracy**
- **Identify the data agency’s quality control protocol**
- **Choose indicators that are unlikely to be incorrectly reported**
  - Select variables that are straightforward and less susceptible to human error
  - Request raw variables
- **Communication between program or implementing partner responsible for collecting data and research team**
  - Ask how and why data are collected
Unlike survey data, administrative data are not susceptible to bias.

A. True
B. False
Using Administrative Data: Reporting Bias

• From an individual
  – E.g., under-reporting income to qualify for a social welfare program

• From an administrative organization
  – E.g., schools over-report attendance to meet requirements

- Similar to survey data, but with potentially stronger incentives
+ Much administrative data is not self-reported
Using Administrative Data: Reporting Bias

To address reporting bias

• **Identify the context in which the data were collected**
  – Were there incentives to misreport information?

• **Choose variables that are not susceptible to bias**
  – E.g., hospital visit v. value of insurance claim
Using Administrative Data: Selection Bias

• Administrative records only from individuals or organizations in contact with the administration in question
  – Program recipients
  – Applicants
  – Partner schools and hospitals
  – Etc.

• Ask: what is the reason for the organization to collect these data?

  - Like with survey data, but no control over selection process
  + Some administrative data has near-universal coverage
Using Administrative Data: Differential Coverage

Between treatment and control:

- Differential ability to **link** individuals to administrative records
- Differential probability to **appear** in administrative records
  - E.g., victimization as measured by calls to report crime

### Actual impact:
- **10 unit reduction in victimization**

### Measured impact:
- **10 unit increase in victimization**

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Victims</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>Measured</td>
<td>50</td>
<td>40</td>
</tr>
</tbody>
</table>

- Red: Actual Victims
- Green: Measured Victims
Using Administrative Data: Differential Coverage

To address selection and differential coverage bias

• **Identify the data universe**
  – Which individuals are included in the data and which are excluded, and why?

• **Identify how the intervention may affect the reporting of outcomes**
  – Determine direction in which differential selection might occur and how this might bias effect estimates

• **Collect baseline survey with identifiers for linking**
  – To ensure that you are equally likely to link treatment and control individuals to their records and identify differential coverage
Lecture Overview

• What to measure
  – Theory of change
  – Purpose of measurement

• How to measure
  – Sources of measurement
  – Measurement concepts
  – Indicators and Indices
  – Response process
  – Challenges to measurement
    • Primary data
    • Administrative data
  – Ethics and IRB
Ethics and IRB

- **Two separate questions:**
  - Is the program I am evaluating ethical?
  - Are the research procedures ethical?

- What’s the difference?
- **Research** on **human subjects** is subject to legal oversight (IRB review)
  - Potentially harmful programs implemented by others should be evaluated (even if perhaps not implemented)
  - Data collection and potentially random assignment must be reviewed
Ethics and IRB

- Institutional Review Boards (IRBs) for research on human subjects
- Required for all research that receives support, directly or indirectly, from the United States federal government
- Belmont Principles
  - Respect for persons
    - Informed consent
    - Protection for vulnerable populations
  - Beneficence
    - Maximize benefits and minimize possible harms
    - Do not harm
    - Protect subjects’ privacy
  - Justice
    - Benefits and burdens of research should not go to different groups
IRB Procedures

• **Informed Consent**
  – Subjects must consent to the use of their primary or administrative data
  – They should agree to study procedures with full knowledge of risks and benefits
  – IRBs and/or data providers may require that individuals consent to **each specific data set** that may be used
  – Waiver of informed consent

• Different levels of scrutiny depending on potential risks and populations involved
• All J-PAL studies must be reviewed by an IRB.
Thank you!