

Measurement & Data Collection

Specifying good outcomes and the indicators we will use to measure them requires a deep understanding of the program being developed, the objectives of those implementing the program, and potential pathways through which the program or policy can impact lives, both positively and negatively.

– Rachel Glennerster and Kudzai Takavarasha, *Running Randomized Evaluations: A Practical Guide*.

Our primary outcomes of interest are usually defined within our original research question (e.g. what is the impact of class size on learning, where learning is the outcome.) But many questions remain:

1. Theory of Change: What is the pathway through which we expect changes attributable to our program to come about?
2. Measurement: What are our outcomes and indicators?
3. Data Collection: How do we obtain data on our outcomes?
4. Data Quality: How do we ensure that data quality standards are met in the process of collecting data?

Theory of Change

A theory of change describes a strategy or blueprint for achieving a given long-term goal. It identifies the preconditions, pathways, and interventions necessary for an initiative's success.

- See www.theoryofchange.org from the non-profit organization Center for Theory of Change for more

information.

- For a case study on coming up with a theory of change and a logic model, refer to the [J-PAL case study](#).

Measurement

An “outcome” is a change or impact caused by the program we are evaluating, or it could be characteristic of the respondents we want to measure. An “indicator” is an observable signal used to measure outcomes.

- The World Bank’s Living Standards Measurement Study has [extensive documentation on designing questionnaires](#) to measure various outcomes and indicators, as well as other topics in survey design and implementation.

In determining our outcomes and indicators and how they fit into the evaluation’s theory of change, questions we need to consider include:

1. What are the intermediate outcomes that can measure each component of our theory of change (from starting assumptions to process)?
2. What baseline outcomes will measure the various ways in which our sample is heterogeneous, to help us understand the context, to help us measure heterogeneous treatment effects, as well as to help us absorb variance to help with the precision?
3. For an example of an evaluation that estimates heterogeneous treatment effects, see [this paper](#) by Monica Angelucci, Dean Karlan, and Jonathan Zinman looking at microcredit in Mexico.
4. What other environmental outcomes (e.g. school quality, density) can help us assess the context of our evaluation sample, which may help us determine the extent to which results are generalizable to other contexts?
5. How do we ensure that our outcomes are being measured in a

way that ensures both validity – an unbiased, accurate measure of the outcome – and reliability – a consistent, precise measure of the outcome?

6. How can we systematically record difficult-to-measure outcomes, as in a [J-PAL study](#) that tried to measure bias against female policymakers in India?
7. Which outcomes must be measured quantitatively, and which qualitatively?
8. How do we consider proxy measures? An example is [PPI](#), a leading poverty measurement tool used to measure household poverty, improve targeting and social performance, and track changes in welfare

Measurement in specific topic areas

Gender: J-PAL's research resource, [A Practical Guide to Measuring Women's and Girls' Empowerment in Impact Evaluations](#), gathers insights from diverse disciplines and the experiences of J-PAL affiliated researchers around the world and offers practical tips for how to measure women's and girls' empowerment in impact evaluations.

Health: [Measuring Health Outcomes in Field Surveys](#) is an online course that explores the fundamentals of field-based health research through documentary-style learning at ongoing research projects in India and Kenya.

Data Collection

Having determined our indicators based on our theory of change, the next question becomes: How do we obtain data on those indicators from individuals in our sample? There are broadly two sorts of data: *primary* data and *secondary* data.

Primary data are those that researchers collect themselves,

usually as part of the experiment. (Primary data should not be confused with primary outcomes!)

Secondary data come from other sources—perhaps administrative data of a government partner, survey data from another research organization, study, or statistics bureau.

- How do we obtain and use nonpublic administrative data for a randomized evaluation? This [resource](#) provides practical guidance on how to obtain and use nonpublic administrative data for a randomized evaluation. While many of the concepts in this guide are applicable across countries and contexts, certain sections are targeted specifically to research conducted in and pertaining to the United States.
- What are data access procedures for key administrative data sources in the United States? Specific requirements and processes for obtaining data vary by data source. To assist researchers in screening potential data sources, J-PAL North America has [cataloged a number of key US data sets](#). This catalog documents procedures on how to access data based on information provided by the originating agencies.
- [This list](#) compiled by The Guardian links to a number of existing datasets that might be of interest to international development researchers and practitioners.

If our evaluation involves the collection of primary data, the next set of questions to answer includes:

1. Should we build the data collection team in-house, or outsource the task to a survey firm?
2. Should we use paper-based surveys, or electronic data collection (computer-assisted interviewing—CAI)? Determining what type of data collection process to use will affect a host of factors, from the types of questions that can be asked, to the

timeline for collecting and analyzing data.

- [This post](#) on the Development Impact blog covers some of the tradeoffs involved with engaging in paper versus electronic data collection.

3. If engaging in electronic data collection, what software and platform should we use?

Some options used by IPA and J-PAL researchers in the past include:

- [SurveyCTO](#), computer-assisted interviewing on Android phones. [The SurveyCTO blog](#) is also an excellent resource for various data collection and measurement topics.
- [SurveyBe](#), tablet-based CAI software from EDI.
- [Pendragon](#), compatible with Android, iOS, and desktop operating systems.
- [Open Data Kit](#), free, open-source Android software similar to SurveyCTO.
- [Blaise](#), tablet/computer-based software that requires a little more coding ex ante.
- [Fielddata](#), cell-phone based CAI software.

Data Quality

How do we monitor the data collection process, both during the process of data collection and after the data has come in, to ensure data quality? Some resources that IPA and J-PAL have developed include (these will be published at a future date as we are in the process of finalizing them):

- Data Quality Check manual
- Backcheck guide
- [bcstats](#), a Stata program written by an IPA staff member for conducting back checks on survey data
- High frequency checks guide
- High frequency checks Stata code

A related set of questions involves data entry protocols when paper based data collection is used rather than electronic data collection:

1. Should we manage data entry in-house or outsource?
2. How do we ensure data entry quality?
 - [cfout](#), a Stata program written by an IPA staff member for reconciling multiple rounds of data entry.

Please note that the practical research resources referenced here were curated for specific research and training needs and are made available for informational purposes only. Please [email us](#) for more information.

Published on *The Abdul Latif Jameel Poverty Action Lab*
(<https://www.povertyactionlab.org>)

<https://www.povertyactionlab.org/es/node/20818>