

**Authors**

Thoai Ngo  
Population Council

Jessie Pinchoff  
Population Council

## Everybody Talks About Big Data - What About Good Data?



At IPA, we care deeply about improving the quality of data collected in international development research. Bad data is at best a waste of resources, but may also misinform policies or programs. To begin to address this problem more systematically, [IPA](#) and Yale co-hosted a Field Research Measurement workshop that took place on a recent Friday at Yale University. A group of 32 researchers from universities such as Harvard and the London School of Tropical Medicine and Hygiene, research organizations such as the [Abdul Latif](#)

[Jameel Poverty Action Lab \(J-PAL\)](#), [World Bank](#), [Center for Effective Global Action \(CEGA\)](#), [University of Michigan's Survey Research Center](#), the [Bill & Melinda Gates Foundation](#), and survey software designer [SurveyCTO](#) discussed a variety of topics related to field measurement and data quality in a round table format. The goal was to convene a group of experts to discuss the challenges that come up in field research but are often not the main research priority. However, the question was raised: should they be?

We discussed five major topics including electronic data collection, survey design, enumerator effects, how to measure difficult concepts, and behavioral responses to research. Anyone who has done research knows very well these topics can often impact data quality, the external validity of study findings, and our understanding of results. While there has been growing investment in research in international settings, there has not been a strong effort to ensure that the data collected are of high quality. Motivating researchers, donors, and implementers to focus on these measurement-related questions will result in more data-driven technology and strategies to ensure accuracy and efficiency in data collection that can be used for better policy or programmatic decision-making.

"Everyone wants high-quality data without taking the time to understand how to get it."

Participants shared past work, current projects, and remaining questions that may be the focus of future work. Berk Ozler, a workshop participant, summarized the discussion in detail on his [World Bank Blog](#).

Questions that came up included:

- What are some of the opportunities and challenges to using crowd sourced non-survey data, such as cell phone records?
- What is the best way to design a survey - should the order of questions or modules be fixed? What are some of the pros and cons, and how can we use data, metadata, and paradata to make some of these decisions?
- How can we measure enumerator effects and should these effects be included in regression models?
- How can you validate new measurement tools if you don't know what the 'truth' really is?
- What is our responsibility to study participants to share findings from our work?

Across topics, two major themes stood out to us at IPA throughout the day: 1) How can we better use technology to improve the quality of our data collection; and 2) What are the incentives for studying measurement to improve data quality and how can we share this information?

**Technology for data collection:** There has been widespread and increasing use of technology to collect data, but no clear consensus that what we get is always 'better'. We can harness crowd-sourced data using cell phones or mobile platforms, but do we lose representativeness? Is it always better or worse to collect our survey data on tablets and cell phones? On paper, enumerators could move around the survey and change the order of

modules and questions depending on who was home to take the survey. Now, surveys are more linear, and we can put ‘speed bumps’ or checks into place (For example, if someone says they are 25, and next says their birth year is 1940, the tablet can calculate this and show the enumerator an error to fix). Is this loss of flexibility better or worse? The answer is often that it depends. Our colleagues at [SurveyCTO](#) are extremely interested in these questions and collaborating with researchers to figure out how to improve the design and use of their software to maximize data quality, and recently blogged about it [here](#). At IPA, we use SurveyCTO as a platform to collect data for 95% of our 250+ ongoing impact evaluations and this has improved the timeliness and the quality controls of our data markedly. We have even started embedding some measurement related studies into our fieldwork using SurveyCTO. For example, in our [Maximum Diva Woman’s Condom](#) study in Zambia we are looking at how participants respond to sensitive questions; you can read more about [here](#).

**Incentives:** While all of the participants felt strongly that measurement and data quality were important, almost none had made this a specific focus of their work or projects (with the exception perhaps of colleagues from the [Survey Research Center](#) at University Michigan). Without strong incentives from donors or other partners, however, it is difficult to prioritize this work. Everyone wants high quality data without taking the time to understand how to get it. Embedding a measurement experiment within a larger project might be one strategy, but there are concerns that the embedded measurement experiment could undermine larger survey results. This is compounded by the fact that many journals will not publish articles related purely to measurement or papers on studies that have negative findings. So, if you are a researcher about to embark on a large-scale data collection, carefully consider what you are measuring and how, and how you can optimize data quality, and if there are opportunities to share your data. The [IPA Research Transparency Initiative](#) was developed to advocate for improving data quality, reporting, and sharing of data for re-use. We suspect that most people don’t share their data because they did not put a strong data quality plan in place before starting data collection.

So, what’s next? IPA won’t be able to improve the quality of field research alone. Therefore, we are calling on our research partners, implementing partners, and donors to invest effort and resources to begin to focus on quality and measurement. If more funding is made available to ask some of these important questions, IPA (along with its collaborating researchers and implementing partners) can design and implement studies to answer them. IPA is in a unique position to progress this agenda as we collect and share data across many countries and projects, with a strong commitment to research transparency, and making methods and data such as these publicly available. Any measurement or data quality findings and innovations can be disseminated across our research network of over 400 researchers to create a feedback loop in which we conduct research on measurement and apply the findings to future research projects. Although incentives are not currently in place to encourage this type of work, these are necessary steps to improve data quality and ultimately strengthen our research findings that ultimately drive global policy.

Contact us at [researchsupport@poverty-action.org](mailto:researchsupport@poverty-action.org) if you are interested in collaborating with the Research Department at IPA to conduct measurement experiments.

*Jessie Pinchoff is a Research Manager & Thoai Ngo is Senior Director, both with IPA's Research and Knowledge Management team.*

May 25, 2016