

There's No Place Like Home? How the Interview Method Affects Results with the Progress out of Poverty Index[®]

Mark Schreiner

25 August 2015

Abstract

This study randomizes the interview method for the Progress out of Poverty Index[®], a short survey for estimating consumption-based poverty rates for participants in pro-poor programs. A face-to-face interview in participants' homes is the most accurate, but it is also the most costly. In the test here in a poor, rural area in India, mis-reporting is disconcertingly frequent, yet the distribution of responses to survey questions—and estimated poverty rates—usually does not differ systematically between a given alternative method and the at-home benchmark. Estimated poverty rates, however, do differ across alternative methods, because completing an interview is linked both with the method and with participants' poverty. To the extent that these results generalize, the PPI[®] can be used with alternative interview methods without affecting results as long as the alternative uses an enumerator and has the same (high) completion rates as with face-to-face, at-home interviews.

Acknowledgements

The Ford Foundation funded this work. Thanks go to the self-help-group participants who completed the survey (twice) as well as to the staff and field teams of the SHG organizers, Pradan and the National Rural Livelihoods Mission. Preeti Anand and Devahuti Choudhury of Grameen Foundation India served as Project Manager and Program Manager. I am grateful for the leadership of Frank DeGiovanni, Anirban Ghosh, Roshan Gupta, Chandni Ohri, Julie Peachey, and Matt Walsh. Mindfield Research Services Private Limited collected the data. The PPI is a poverty-measurement tool that Grameen Foundation promotes to help organizations achieve their social objectives more effectively. It is the same as what Microfinance Risk Management, L.L.C. calls the Simple Poverty Scorecard[™].

Author

Mark Schreiner is the Director of Microfinance Risk Management, L.L.C. and a Senior Scholar at the Center for Social Development in Washington University in Saint Louis.

There's No Place Like Home?

How the Interview Method Affects Results with the Progress out of Poverty Index[®]

1. Introduction

The Progress out of Poverty Index[®] is a low-cost survey with 10 simple questions. Hundreds of pro-poor programs and other organizations in about 60 countries use the PPI to estimate the poverty of their participants so as to improve their internal social-performance management and to prove their poverty outreach to external stakeholders. The questions in a PPI for a given country come from that country's national consumption survey, and responses to PPI questions are linked to consumption-based poverty based on data from the national survey.

The national surveys are done by enumerators in respondents' homes, so the PPI is most accurate when done in-person/at-home. But sending enumerators to participants' homes is costly. Thus, pro-poor programs—unless they can combine the PPI interview with a visit to participants' homes that they would do anyway—would prefer a less-costly interview method. The main alternatives are:

- In-person/away-from-home (an enumerator separately interviews multiple participants in one visit to—for example—a meeting of community savings-and-loan groups, a class at an agricultural-extension center, or a health-post waiting room)
- In-person/by-phone (an enumerator interviews participants by telephone)
- Automated/by-phone (interactive-voice response or SMS/text-messaging)¹

¹ SMS is not tested here because Indian regulation would have required respondents to send one text for each question and to bear all related fees.

Does the interview method affect responses or poverty estimates? This study compares results when participants in self-help groups in a poor, rural area in India's state of Madhya Pradesh are interviewed twice, once in-person/at-home (the benchmark) and once with a randomly-assigned alternative method. The order of the two methods—as well as the enumerator in each interview—is also randomized.

It turns out that the interview method does affect estimated poverty rates even though alternative methods—in spite of disconcertingly frequent mis-reporting—do not often skew the distribution of responses away from the benchmark.² For those randomly selected for the in-person/away-from-home alternative and who completed both the benchmark and the alternative survey, the estimated poverty rate by the \$1.25/day 2005 PPP poverty line is about 53 percent in the benchmark and 54 percent in the alternative. For in-person/by-phone, the benchmark estimate is 46 percent, and the alternative is 45 percent. For IVR, the benchmark is 39 percent, and the alternative is 36 percent.

How is it that poverty-rate estimates vary across alternatives but not—for a given alternative—between the benchmark and the alternative?

For households randomized into a given alternative, poverty rates resemble the benchmark rates because—except for IVR—the distribution of households' responses to PPI questions under the alternative are not skewed away from the distribution of

² IVR is an exception; it not only affects estimates of poverty rates but it also frequently leads to different response distributions.

responses under the benchmark. After applying PPI weights and converting to poverty likelihoods, responses in the alternative that were “more-poor” or “less-poor” than in the benchmark balance each other out.

At the same time, completion rates³ vary by method: 84 percent for the benchmark, 91 percent for in-person/away-from-home, 60 percent for in-person/by-phone, and 12 percent for IVR.⁴ Differences in completion rates reflect differences in methods, and they are also linked with participants’ poverty. In particular, in-person/by-phone and IVR require a household to have access to a telephone, and access to a phone is less common among poorer households.⁵

Different interview methods have different completion rates, and completion rates are linked with poverty. Therefore, the population represented by completed interviews differs by method. Completed interviews with in-person/by-phone and IVR represent less-poor populations (with lower estimated poverty rates) than completed interviews with in-person/away-from-home.

Thus, the interview method affects the estimated poverty rate (across methods) even though each given method’s estimate is about the same as for the benchmark.

³ The *completion rate* is the share of sampled households who complete an interview. It corresponds to definition “RR5” in American Association for Public-Opinion Research (2011): completed interviews divided by the sum of completed interviews, partially completed interviews, refusals and break-offs, non-contacts, and other.

⁴ Most non-completed interviews are not refusals but rather cases in which the household is not found or (for by-phone methods) is not reached by phone.

⁵ Responding under IVR also requires following instructions to press keys on a phone’s dial pad. Because poor people tend to have less education, completing an IVR is more difficult for them than is verbally giving answers to an enumerator.

Poorer participants are less likely to complete phone-based interviews, so the population represented by phone-based methods is skewed toward the less-poor, reducing these methods' poverty-rate estimates, even though participants who complete in-person/by-phone interviews give, on net, about the same responses as they do in-person/at-home.

Can the PPI be applied with less-costly interview methods without sacrificing accuracy? Yes, as long as the methods use an enumerator and achieve the same (high) completion rates⁶ as face-to-face, at-home interviews.⁷ Of course, this assumes high-quality fieldwork by well-trained enumerators.

To the extent that these results generalize, this means that in-person/away-from-home is fine, but in-person/by-phone may not be (its estimated \$1.25/day poverty rate is about 8 percentage points lower), and IVR definitely is not (its estimated poverty rate is about 18 percentage points lower, and IVR responses to individual questions often differ systematically from benchmark responses).

The rest of this paper:

- Presents the PPI
- Discusses theory, hypotheses, possible approaches, and how randomization is used to identify interview-method effects
- Tests whether responses to individual PPI questions for a given method systematically differ from the benchmark
- Reports poverty-rate estimates by method
- Discusses the meaning and generalizability of the results

⁶ Again, this uses a strict definition of *completion rate* (households with completed interviews divided by households in the sample). In particular, this definition counts sampled households who are not found or not contacted as non-completers.

⁷ How high is a *high*? Wikipedia's (2015) standard is 80 percent. Fincham (2008) requires 80 percent for academic publishing. For best accuracy, PPI surveys should match the 90-percent-plus completion rates of most national consumption surveys.

2. The PPI

Figure 1 is the PPI for India (Schreiner 2012a). Its 10 questions come from India’s 2009/10 (Round 66) Socio-Economic Survey. The points are derived statistically by relating responses to the 10 questions with households’ consumption-based poverty status in the SES. For ease of use, the points are scaled and rounded so that, for a given question, the “most-poor” response gets zero points and a “less-poor” response gets more points. The sum of a household’s points—the *score*—ranges from 0 (most-likely poor) to 100 (least-likely poor).

Scores do not definitively mark households as poor or non-poor. Rather, scores are linked with a probability that consumption is below a poverty line. For a given score range, the *poverty likelihood* is defined as the share of households who are poor among SES households in that score range. For the example of scores from 20–24, the poverty likelihood is 49.7 percent for the \$1.25/day 2005 PPP poverty line and 25.9 percent for India’s Tendulkar national poverty line (Figure 2).

A pro-poor program can estimate the poverty rate of its population of participants as the average of poverty likelihoods in a sample. If a program applies the PPI twice, then it can estimate changes in poverty rates over time.⁸ Scores can also be used to segment participants for targeted services. The documentation for each country’s PPI reports the expected error in estimated poverty rates, the precision of estimates, and the accuracy of segmentation.

⁸ This estimate of change is not necessarily the same as an estimate of program impact.

There are several other tools that also estimate household-level poverty:

- National consumption surveys (Deaton and Zaidi, 2002)
- USAID’s Poverty Assessment Tool (PAT, Schreiner, 2014a)
- DHS wealth index (Rutstein and Johnson, 2004)
- Multidimensional Poverty Index (MPI, Alkire and Foster, 2011)
- Proxy-means tests (PMT, Grosh and Baker, 1995)
- Poverty maps (Elbers, Lanjouw, and Lanjouw, 2003)

The PPI differs in that it is:

- Low-cost (unlike consumption surveys)
- Simple and transparent (unlike the PAT, wealth index, PMT, and poverty map)
- Designed for program managers (unlike the wealth index, PMT, and poverty map)
- Focused on consumption-based poverty (unlike the wealth index and MPI)
- Uses not only statistics but also judgment to select questions (unlike the PAT, wealth index, PMT, and poverty map)
- Lacks captive users (unlike the PAT, PMT, and poverty map)

Without captive users, the PPI must win voluntary adoption if it is to fulfill its goals of strengthening accountability and making social-performance management more transparent and intentional. This imperative led to tailoring the PPI to the demands of non-specialist managers in local, pro-poor programs.⁹ In particular, the survey is short (and thus low-cost), the questions are simple and common-sense, points are zeros or positive whole numbers, and all the math is simple and transparent. This paves a clear path from questions, to responses, to points, to scores, to poverty likelihoods, and finally to estimated poverty rates. If managers can see how the PPI works, then they

⁹ The school of hard knocks taught the PPI developer (the author) that complex tools are rarely used, and the scoring literature (reviewed in Schreiner, 2012b, and Caire and Schreiner, 2012) teaches that simple tools can be about as accurate as complex ones.

are more likely to adopt it voluntarily, apply it properly, and act confidently on its results.

In the end, cost is king. National consumption surveys provide the most-accurate measures of consumption poverty, but local, pro-poor programs cannot afford to ask their participants for the quantities and prices (market or hypothetical) of hundreds of possible consumption items.

To enable frequent, widespread poverty measurement by non-specialists in local programs, the PPI trades accuracy for cost (and documents its accuracy). It seeks to improve the information available for social-performance management in a context of limited resources, time, and decision-making ability/energy. Estimates that are “good enough for government work” help more than (slightly) more-accurate alternatives which are not made or used due to their cost or complexity (Schreiner, 2014b).

Of course, costs are never low enough. The PPI is relatively low-cost, but it is still absolutely costly. The main cost is the time and toll for enumerators to travel to participant’s residences. Naturally, PPI users would like to reduce this cost, for example, by interviewing away-from-home in a central location (so that participants go to the enumerator, rather than vice versa) or by interviewing by phone.

But the PPI is constructed from national-survey data that was collected in-person/at-home. Therefore, accuracy is highest when the PPI is done in-person/at-home. How much do alternative interview methods affect accuracy?

3. Study design

3.1 Theory

The interview method affects PPI results by affecting how much respondents:

- Believe that “looking poor” will increase their chances of qualifying for aid (leading to *under-reporting* with responses that are “more-poor” than the benchmark)
- Feel embarrassed to reveal their true poverty (leading to *over-reporting* with responses that are “less-poor” than the benchmark)
- Fear being caught in a lie (leading to less lying and greater accuracy)
- Suffer guilt when lying, even if not caught (leading to less lying and more accuracy)
- Feel motivated to answer carefully (leading to greater accuracy)
- Interpret a question as it was interpreted in the national consumption survey (leading to less inadvertent inaccuracy)¹⁰

3.2 Hypotheses

How might an interview method affect these factors? An in-person/at-home interview should be best at catching and preventing lies, as respondents fear that enumerators can verify responses—even though they rarely do—or happen to notice contradictory evidence. Enumerators can also help align the interpretation of PPI

¹⁰ “The experience of being interviewed by another person differs from completing a survey on-line or on-paper. For example, an interviewer can help respondents stay focused and may be able to provide clarification or encouragement at difficult junctures” (Keeler *et al.*, 2015, p. 4). For example, the number of household members is a key PPI question, and national surveys use complex definitions of *household* to capture the complex reality of household membership. Enumerator-based methods elicit more accurate responses because they help respondents apply a definition of *household* that consistently mimics that of the national survey. For employment indicators (such as “How many household members work?”), respondents likewise need enumerators to define *work*. Definitions are needed even for seemingly straightforward questions such as “Do you have a TV?” because a TV may be broken, borrowed, or bought on credit. While automated methods (such as IVR) can provide written or recorded guidelines, respondents ignore them more readily than instructions from an enumerator.

questions with the national survey. Carefulness, honesty, and completion rates should be high as respondents feel obliged to be courteous to a solicitous person on their doorstep.

For in-person/away-from-home, the embarrassment of being poor is greater (if there are on-lookers), increasing over-reporting.¹¹ But enumerators can no longer verify responses, increasing under-reporting meant to qualify for aid. As with the benchmark, a face-to-face enumerator makes the respondent more careful, more honest, and more likely to interpret questions appropriately.

Inaccuracies should be more common with in-person/by-phone. While there is still an enumerator, he/she is not present face-to-face, decreasing the risk of being caught in a lie and dampening guilt when lying. The respondent feels less social pressure to be careful, and the enumerator has fewer non-verbal cues for detecting whether a respondent is mis-interpreting a question.

IVR should be the least accurate.¹² Lies cannot be caught, and it is easier to lie to a machine than to a person. Without an enumerator to be nice to, respondents are less careful, and any mis-interpretations of questions cannot be corrected.

These forces affect not only responses to PPI questions but also completion rates (which depend mostly on whether a household is reached to start an interview). A

¹¹ On-lookers might also expose lies. The study here tried to mute the motive to under-report to qualify for aid by training enumerators to tell households in their introduction that “This survey will not benefit you directly but will help organizations like [the self-help-group organizers] to design better development programs in [this] district.”

¹² SMS shares many of IVR’s weaknesses, although SMS is not tested here.

respondent can hang up on a robo-call more readily than on a human caller, supposing that the household has a phone (and the surveyor has the number) in the first place.¹³

Likewise, respondents can more readily excuse themselves from an interview request from an enumerator on the phone than from one who meets them face-to-face.

For a sampled household, failure to complete the PPI survey means that some—usually all—responses are missing (true values exist, but they are not recorded).

Missing responses are inaccurate; they differ from the responses that would be obtained in the in-person/at-home benchmark, for which completion rates are high.

3.3 Approaches to comparing interview methods

How to measure the effects of these forces? To compare results in the benchmark versus an alternative, households may be interviewed in:

- Two samples once
- One sample twice

Interviewing two samples once reveals the net change in the distribution of responses. Interviewing one sample twice reveals not only the net change but also household-level under-reporting and over-reporting, even though the two types of mis-reporting counterbalance each other in the sample to some extent.

¹³ SMS further requires respondents to read and write. Accuracy falls with SMS and IVR (as well as in-person/by-phone) if the respondent must pay for the call, if line quality is poor, if the respondent takes the call in the presence of third parties, or if the respondent is distracted or multitasking (Lynn and Kaminska, 2011).

Interviewing one sample twice has some disadvantages:

- The greater burden makes households less likely to complete the second interview (or more likely to respond to it carelessly)
- Respondents may correct responses in the second interview because, after the first interview made them aware of their knowledge gaps, they sought to fill the gaps
- Respondents may purposely repeat known inaccuracies to be consistent or to avoid seeming to have lied or to have made a mistake in the first interview
- The respondents may differ across a household's two interviews
- Enumerators—like respondents—may try to keep responses consistent (to hide their mistakes in the first interview), or they may run through the second interview quickly, assuming that responses will be the same

Both approaches—two samples once, and one sample twice—have been used to test for interview-method effects.

For example, Keeter *et al.* (2015) compare the distribution of responses in one sample who were interviewed in-person/by-phone versus in a second sample who took a low-cost, automated survey via web. They find (p. 2) that differences “are fairly common, but typically not large, with a mean difference of 5.5 percentage points and a median of 5.”¹⁴ These are net changes; over-reporting and under-reporting partially balance, so changes are more frequent than the net changes reveal. Most of the largest net changes seem to stem from “social-desirability bias”; people are more likely to report responses that other people might view with disdain or disagreement (such as a low quality of family life or social life) via web than in-person/by-phone.¹⁵

¹⁴ Keeter *et al.* do not say how common is *common* nor how large is *large*.

¹⁵ “The social interaction inherent in a telephone or in-person interview may exert subtle pressures on respondents . . . to present themselves in a more positive light to an interviewer, overstating socially desirable behaviors and attitudes and understating opinions and behaviors they fear would elicit disapproval” (Keeter *et al.*, p. 4).

For one sample interviewed twice, Martinelli and Parker (2009) study responses to a PPI-like survey used in Mexico to qualify households for cash transfers (leading to incentives to under-report). Households first fill out a long¹⁶ paper survey on their own in a government office away-from-home. Those scoring below a cut-off are later visited in-person/at-home by a social worker who applies the same survey to cross-check the on-paper responses. For households who complete both interviews, Martinelli and Parker find that the away-from-home, on-paper survey—vis-à-vis the at-home, face-to-face benchmark—leads to:

- Wide variation across questions in the rates of mis-reporting
- Net under-reporting for 13 questions on durable assets.¹⁷ The incidence of under-reporting goes up with the level of potential cash aid
- Over-reporting for some “status goods” such as type of toilet, source of water, type of cooking fuel, and type of floor. This is consistent with embarrassment, and it is more common for households with more income or more education

To check not only net differences but also the full extent of both under-reporting and over-reporting, this study interviews one sample twice.

¹⁶ Skoufias, Davis, and Behrman (1999, Appendix A).

¹⁷ Martinelli and Parker do not say whether the 13 questions are the only ones tested.

3.4 Randomization

The population of this study is the 8,434 members of self-help groups in the data bases in late 2014 of two SHG organizers (Pradan and the National Rural Livelihoods Mission) in a poor, rural area of India (Madhya Pradesh state, Jabalpur division, Dindori district, Amarpur and Samnapur tehsils).¹⁸ To get a clean estimate of effects of the interview method, three aspects of the study were randomized:

- Alternative interview method, at the SHG level,¹⁹ with:
 - 1,940 participants for in-person/away-from-home
 - 3,497 for in-person/by-phone
 - 2,997 for IVR²⁰
- Method order, with the benchmark preceding the alternative for half the sample^{21,22}
- Enumerator, with interviews randomly divided among 12 enumerators²³

¹⁸ Anand (2015) describes the details of fieldwork. The initial survey firm faked an aspect of randomization and, in the pilot, did not respect the randomized method order and enumerator. Randomization was re-done, and a new firm did the fieldwork.

¹⁹ Clustering methods by SHG decreases the cost of fieldwork. Standard errors in the analysis here do not account for clustering, assuming that changes in responses across interview methods are not correlated among households in a given SHG.

²⁰ For each alternative, the goal was for 1,000 households to complete both a benchmark and alternative interview. More households were assigned to the two by-phone methods in anticipation of lower completion rates. When SMS was dropped because India regulations require respondents to send—and pay for—one SMS per response, the SMS households were re-assigned to IVR.

²¹ A household's responses in its first interview are not known to the enumerator in its second interview unless, by chance, the same enumerator was assigned to both.

²² With IVR, the benchmark is always second. This cut costs, as the benchmark was skipped for the 88 percent of IVR households who did not complete an IVR interview.

²³ Each in-person/by-phone household is randomly assigned an enumerator who called each sampled household once a day for three days at different times of day. For the benchmark visit for households assigned to the in-person/by-phone method, two enumerators each make two attempts to complete the interview in the residence. For each SHG assigned to in-person/away-from-home, three enumerators are randomly assigned to three visits. (Naturally, the same enumerator could be drawn for two or three visits.) Three visits are planned because some households are not found in a given

On average in repeated samples, randomization breaks any links between an interview method and other things that might affect responses or completion. This allows confidently attributing differences in results to differences in interview methods.

For example, randomizing the alternative interview method breaks any link between phone ownership and assignment to by-phone methods. If by-phone methods are applied only with households known to have phones, then the study's completion rate could not be extrapolated to households not known to have a phone.

Randomizing the method order breaks any links between a method and effects on responses due to a household's completing the same survey twice. If, for example, the alternative interview is always after the benchmark, then the natural net increase in durable goods over time will look like over-reporting caused by the alternative method when in fact it is an artifact of the study's need to interview each household twice.²⁴

Randomizing the enumerator breaks any links between a method and aspects of a given enumerator (such as carefulness). Specific enumerators are not concentrated in

visit and because the benchmark and the alternative are not both done on the same day. The benchmark is done—first or second as randomized—in one visit, and the away-from-home interview is done in different visit. For IVR SHGs, three enumerators are likewise randomly assigned to three visits. Benchmark interviews for 166 IVR cases were done by six supervisors at the end of the fieldwork because the 12 regular enumerators had been mistakenly sent home. It is not known how this might affect the IVR results.

²⁴ Some households dispose of durables between interviews, but on average, the expected net change is to add durables. Here, the average days between interviews is 4.8, so this bias would be small, even without randomization.

any specific interview method, so any effects due to the characteristics of the enumerators do not get confused with effects due to the method itself.

The population is the households of SHG participants. The respondent may be any adult household member. The benchmark respondent is randomly selected on-the-spot (Kish, 1949) and usually turns out to be the SHG participant. For by-phone alternatives, the respondent is whoever answers the phone (usually not the SHG participant). For away-from-home, the respondent is the first adult household member who the enumerator finds away-from-home (usually the SHG participant). Thus, the respondent is not fully randomized. In particular, the respondent is more likely to vary across a household's pair of interviews for the by-phone alternatives than for the away-from-home alternative. This probably increases the frequency of differences for by-phone methods, although it is unknown whether they balance out across households or whether they tend toward under-reporting or over-reporting.

Given randomization (with the caveats above), the impact of a given alternative interview method is simply the difference in results vis-à-vis the benchmark.

4. Effects of alternative interview methods

The interview method affects three PPI results: completion rates, responses to questions, and estimated poverty rates.

4.1 Completion rates

A method's *completion rate* is the number of households who respond to all 10 questions in India's PPI, divided by the number of sampled households for which an interview is attempted.

For in-person methods (at-home or away-from-home), an *interview attempt* is defined as being sought by an enumerator for an interview.

For by-phone methods (in-person or automated/IVR), an *interview attempt* is being sought by an enumerator in-person to collect a phone number for a later by-phone interview.²⁵ Attempts thus count participants randomized into a by-phone method who:

- Are sought but not found
- Are found but do not provide a number²⁶
- Provide a number but are not reached when called²⁷

²⁵ Enumerators go to SHGs to collect phone numbers from participants randomly assigned to a by-phone method. When participants provide numbers, the enumerator:

- Advises in-person/by-phone participants that they will be called by a person
- Advises IVR “pull” participants that they will receive an automated call
- Provides IVR “push” participants with an identifier and a number for them to call
- Shows all IVR participants how to respond using a phone key pad

²⁶ Households can provide the number of a phone owned by a non-household member.

²⁷ For push IVR (computer calls the respondent), there are up to nine calls, three per day (at different times of day) on three different days. For pull IVR (respondent calls the computer), sampled households who provided phone numbers are sent an introductory text message and two reminder text messages, in addition to the in-person visit that collected the phone number and explained the pull IVR method.

- Are reached when called but do not complete the interview, or
- Complete the interview

Completion rates are higher in-person than by-phone:²⁸

- In-person/away-from-home: 91 percent
- In-person/at-home (benchmark): 84 percent
- In-person/by-phone: 60 percent
- Automated/by-phone (IVR): 12 percent

This study does not directly reveal why the interview methods have these effects on completion, but they are consistent with common-sense explanations:

- In-person, respondents are courteous and cooperate with enumerators. Furthermore, all respondents have a home or community where they can (eventually) be found
- By-phone, households feel a weaker social obligation to cooperate.^{29,30} Furthermore, households without access to a phone cannot be reached by phone.

Accurate data is what would be obtained—like the data used to make the PPI—in-person/at-home. With methods with lower completion rates than the benchmark, some households do not complete the alternative even though they would complete the benchmark. The (missing) data for these households is inaccurate.³¹

²⁸ These completion rates reflect all interview completions and all sampled households, including those who did not complete both the benchmark and an alternative.

²⁹ With IVR, respondents can hang-up before completion without offending anyone.

³⁰ For those assigned to IVR, the IVR call precedes the benchmark interview, so they may be more wary of a robo-call than others who are interviewed first in-person, even though IVR households do receive an in-person visit to collect the phone number and to explain the method. Furthermore, all IVR households receive an introductory text message before the first IVR attempt.

³¹ It is not known why in-person completion rates here are higher away-from-home than at-home. Perhaps people are usually out during the day when enumerators come calling. PPI accuracy is defined relative to that of a national consumption survey. Most national surveys report completion rates in excess of 90 percent, so—all else constant—the away-from-home alternative here is more accurate than the benchmark.

This source of inaccuracy matters only if non-completion is related both with the interview method and with poverty.³² And it is; by-phone methods—because they are by-phone³³—have lower completion rates than in-person methods, and this leads to lower estimates of poverty rates (see below).

4.2 Responses to questions

How do PPI responses differ in the benchmark versus alternatives? This section looks at responses to the 10 scored questions in India’s PPI (Figure 1) as well as eight non-scored questions (Figure 3) that are often in other countries’ PPIs.

4.2.1 Two samples of households interviewed once

For households who completed both the benchmark and a given alternative method, Figure 4 shows the net distribution of responses for the 18 questions under

³² That is, higher completion rates are better because—all else constant—they decrease the risk that the sample is not representative of the population in ways linked with poverty. Re-weighting cannot make households who complete both the benchmark and a given alternative to be representative of the population because the data here has only two indicators for all households in the population (tehsil of residence, and the identity of the SHG organizer). Even with a rich set of indicators, re-weighting rarely achieves its goal fully, and most pro-poor programs lack the expertise to re-weight.

³³ India’s PPI asks about phones, so having a phone both decreases poverty-rate estimates and increases completion of by-phone interviews. Nevertheless, the estimates here are not contaminated by some sort of mutual causation for by-phone households who complete both the benchmark and alternative interviews. These households are less-poor by several other PPI questions, and, in any case, the link between phones and poverty is derived with data from in-person/at-home interviews in India’s nationally representative consumption survey where having a phone is not linked with completion.

both methods.³⁴ This presentation is “net” because it ignores that each household completed an interview under each method, as if the data were from two samples of households interviewed once.

For households assigned to in-person/away-from-home, distributions for all questions are similar in the benchmark versus alternative. This does not necessarily mean that almost all households gave the same responses under both methods; under-reporting by some households may cancel out over-reporting by others. If so, then the interview method may still affect estimated poverty rates.³⁵ This is checked below.

For in-person/by-phone, the response distributions again are mostly similar, with over-reporting of almirahs and bicycles and under-reporting of farm implements.

For IVR, distributions frequently vary between the benchmark and alternative.³⁶ About 10 questions are over-reported (for example, cooking fuel, sewing machines, and vehicles), and three seem under-reported (number of children, energy for lighting, and pressure cookers). Furthermore, the strangeness of the changes in responses for school

³⁴ n is 1,043 for in-person/away-from-home, 1,050 for in-person/by-phone, and 151 for IVR. A few of the eight additional non-scored questions have missing values for IVR. Fieldwork ran from 20 February to 4 April 2015.

³⁵ This is because over-reporting (say) chairs decreases a poor household’s poverty likelihood more than under-reporting chairs increases a less-poor household’s poverty likelihood. Thus, the net effect on an estimated poverty rate may not be zero even if the number of under-reports is the same as the number of over-reports.

³⁶ This is an eye-ball judgment. The statistical tests that could be used with the data in Figure 4 are less powerful than those used below that account for the fact that each household completed both a benchmark and an alternative interview.

attendance and the number of workers suggest that—without an enumerator—IVR respondents were confused by these (relatively) complex questions.

Comparing across methods (looking at responses to either the benchmark or the alternative for a given alternative), in-person/away-from home is the poorest, in-person/by-phone is less poor, and IVR is the least poor. The causes and effects of this pattern are discussed below.

4.2.2 One sample of households interviewed twice

Each household in this study is interviewed twice, in-person/at-home and with an alternative method. This reveals not only net changes in the distribution of responses across households (Figure 4) but also how each household’s responses differ from the benchmark and whether they tend towards under- or over-reporting.

In-person/away-from-home

Figure 5 reports the direction of changes in responses for households assigned to in-person/away-from-home, vis-à-vis the benchmark.³⁷

All questions³⁸ have both some under-reporting (the alternative is “more poor” than the benchmark) and some over-reporting (the alternative is “less poor” than the benchmark). Total mis-reporting (under-reporting plus over-reporting) is disconcertingly frequent, averaging about 18 percent. About half of households report a different main

³⁷ Figures 5, 8, and 9 look at direction of change, not distance. For example, they count a change from (say) 1 worker to 4 workers the same as a change from 2 workers to 3 workers. For the effects of changes in responses on poverty likelihoods (and thus on estimated poverty rates), both direction and distance matter.

³⁸ Except two questions for which under-reporting was not possible.

source of energy for lighting; about one-third change labour type, number of workers, or agricultural implements; and about one-fifth change the male head's education, household size, school attendance, or telephone. In general, there is more mis-reporting about aspects of household members and less mis-reporting about assets or aspects of the residence.

Figure 6 looks at the distribution (not just the direction) of changes for the main source of energy for lighting. Most switches are:

- From home-made lantern to electricity (5 percent of the sample)
- From purchased lantern to none/firewood/torch, or vice versa (4 percent)
- From home-made lantern to none/firewood/torch, or vice versa (14 percent)
- From electricity to none/firewood/torch, or vice versa (17 percent)

The two switches that do not mostly balance out are:

- From electricity (benchmark) to none/firewood/torch (alternative)
- From home-made lantern (benchmark) to electricity (alternative)

Overall, there is no consistent pattern. If some lanterns use electricity, it might lead to confusion.³⁹ In addition, respondents (or enumerators) may not understand (or explain) the meaning of *main* in “What is the main source . . . ?” This is more likely if a household uses multiple sources, depending on purpose, price, or availability.

A key question in all PPIs is the number of household members. Here, the response changes for one-fifth of households: 11.6 percent over-report (seeming less-poor

³⁹ Indeed, the question is flawed, as a lantern is not a source of energy at all. The Hindi term used in the survey refers only to non-electric lanterns.

by reporting fewer members in the alternative) and 10.1 under-report (Figure 7). Net mis-reporting (over-reporting minus under-reporting) is -1 percent (rounded, Figure 5).

What is the risk that this seeming tendency to over-report household size reflects the population's reality rather than a non-representative sample due to unusual luck-of-the-draw? Figure 5 reports this risk as the p value (0.32) from McNemar's (1947) test of correlated percentages. In conventional—and often mis-used (Cowger, 1984)—academic parlance, the fact that the p value exceeds 0.10 means that over-reporting of household size is not “statistically significant” at the 90-percent level.

Thirteen of the 18 questions have non-zero net mis-reports (Figure 5). Of the 13, the largest (in absolute value) is 3 percent. Twelve of the 13 are under-reports, and four are statistically significant ($p < 0.10$). In sum, interviewing in-person/away-from-home leads to net under-reporting. Whether this under-reporting is “low” or “high” depends on its effect on poverty likelihoods and thus on estimated poverty rates (discussed below).

If a household gives the most-poor (or least-poor) response in the benchmark, then it is not possible to under-report (or over-report) in the alternative. The last three columns of Figure 5 show mis-reporting rates conditional on mis-reporting's being possible. For in-person/away-from-home, conditional under-reporting is always at least 10 percent. For three questions, it is more than one-in-three, and for six questions, it is more than one-in-two. Over-reporting is less common, but it still exceeds one-in-ten for 12 questions and one-in-five in four questions.

In the end, under- and over-reports mostly cancel out. Still, mis-reporting is unexpectedly frequent.⁴⁰ If PPI questions are simple and straightforward, then why—even with alternative interview methods—would responses change so frequently?

This study has no direct explanation. The data are consistent with a small share of respondents trying to look “more-poor” (when they are away-from-home and think that they can get away with it) and with most mis-reports being due to random measurement error.⁴¹ Even if this is the case, however, frequent mis-reports—even if they cancel out—are a concern, and one that could be addressed. After all, random measurement error is “random” not because nothing causes it but rather because its causes are not known. If some causes are identified, then random error can be reduced.

The data are also consistent with high rates of deliberate under-reporting and a good deal of deliberate over-reporting. This would be worrisome because in a less-poor population than the one here (in which under-reporting is more frequently possible than it is here), net mis-reporting would be more strongly skewed toward under-reporting and could thus make estimated poverty rates too high.

⁴⁰ The results are in the same ballpark as Onwujekwe, Hanson, and Fox-Rushby (2006). Asking PPI-like questions, they find mis-reporting by at least 27 percent for households interviewed at-home by two different enumerators and by at least 30 percent for households interviewed twice at-home on different days by the same enumerator.

⁴¹ Random measurement error can be measured if some households are randomized into the same method—be it benchmark or alternative—in both of their interviews. But this was not done here.

In-person/by-phone

Figure 8 reports the direction of changes in responses for in-person/by-phone. While the benchmark respondent is usually a female SHG member, the by-phone respondent is usually a male household member. All else constant, changing respondents should increase the rate of mis-reporting, regardless of method.

Nevertheless, average total mis-reporting is again about 18 percent. The main source of energy for lighting is again mis-reported the most, consistent with the idea that it did not make sense or that respondents (or enumerators) did not understand it (or explain it) well. Total mis-reporting exceeds 30 percent for three questions, and it exceeds 20 percent for six other questions.

Bicycles, motorcycles, and cars are over-reported, on net, by 7 percentage points ($p < 0.01$). It may be that male respondents (in particular) are embarrassed not to own these “male assets” and thus claim to have them, knowing that the enumerator cannot check. At the same time, carts, ploughs, and sprayers are under-reported (5 percentage points, $p < 0.01$).

The household labour type (being “self-employed/wage/salary”) is also over-reported (4 percentage points, $p < 0.02$). This is a complex question, so the more-limited ability of the enumerator to explain it by-phone may lead to the over-reporting. It may also be partly a consequence of India’s rural right-to-work/social-security program that guarantees 100 days of minimum-wage employment per household per year for adults willing to do unskilled manual labor on public-works projects. More than

half of sampled households are in this program, and they may worry about affecting their qualifying status (and so under-report non-agricultural employment) more in the at-home interview than in the by-phone interview.

As with the away-from-home alternative, the number of workers is under-reported by-phone (3 percentage points, $p < 0.07$). School attendance is over-reported (2 percentage points, $p < 0.10$). This question is complex, and households may be embarrassed if children do not go to school (and lying is easier by-phone than at-home).

Overall, “net net” mis-reporting across all questions for in-person/by-phone is +3 across the 18 questions. In contrast, it is -22 for in-person/away-from-home.

Conditional under-reporting is again unexpectedly high, exceeding 8 percent for all but one question. For four questions, it exceeds one-in-two.

In broad terms, mis-reporting in-person/by-phone is like mis-reporting in-person/away-from-home in that it is more common for aspects of household members and less common for assets or aspects of the residence. This is consistent with the idea that—for households who complete both interviews—most mis-reports are due to the question or factors related to the household itself rather than the interview method.

Automated/by-phone (IVR)

For IVR, mis-reporting is rampant (Figure 9). Average total mis-reporting is 38 percent (versus 18 percent in the other two alternatives). Total mis-reporting exceeds 80 percent for two questions, 50 percent for three others, and 20 percent for eight others.

Net mis-reporting is statistically significant ($p < 0.10$) for 13 of 18 questions, and it is severe for school attendance (57 percent) and the number of workers (68 percent).⁴² Seven questions have statistically significant over-reporting,⁴³ yet under-reporting for other questions is so great that “net net” mis-reporting is –99.

In sum, IVR responses are all over the place. The IVR sample—because it is IVR—both under-reports on net and is skewed to be less-poor than the population. Both these factors make estimated poverty rates too low. IVR is not recommended as a way to apply the PPI.

Do different responses by method translate into different estimated poverty rates? It depends—whether or not over-reporting balances under-reporting—because:

- Direction matters; a one-category move toward “more-poor” has a different effect on the score than does a one-category move toward “less-poor”
- Distance matters; a one-category move affects a score less than a two-category move
- The effect on a score due to a move in a given direction of a given distance varies by question and by a household’s response to that question in the benchmark
- The effect on a household’s poverty likelihood due to a change in the score depends on the benchmark score and the direction and size of the score change. For the example of India’s PPI, a one-to-three-point increase in the score can change the poverty likelihood from 0 to 15 percentage points

The effects of an interview method on a household’s poverty likelihood—and thus the effects on estimated poverty rates—depend complexly on all of the responses.

⁴² This probably results from the lack of an enumerator to explain complex questions.

⁴³ This is consistent with Martinelli and Parker’s (2009) finding that less-poor households are more likely than more-poor households to over-report.

4.3 Estimated poverty rates

Figure 10 shows the effects of the interview method on estimated poverty rates with India’s national poverty line (Tendulkar MMRP) and with the international “extreme” poverty line of \$1.25/day 2005 PPP. The estimates use only the first 10 questions, as they are the ones scored in India’s PPI.

Among households who complete both a benchmark and an alternative interview, the alternative method has little effect on estimated poverty rates. For the \$1.25/day line,⁴⁴ the in-person/away-from home estimate (54.0 percent) differs from the in-person/at-home estimate (53.0 percent) by +0.9 percentage points. This difference is so small that a result this large or larger would show up in 68 percent ($p < 0.68$) of samples of this size even if the interview method had no effect at all.

Likewise, the in-person/by-phone estimate (44.5 percent) differs from the benchmark (45.6 percent) by -1.1 percentage points, but this difference is not statistically different from zero at the 90-percent level ($p < 0.63$).

The automated/by-phone (IVR) estimate (36.0 percent) differs from the benchmark (39.4 percent) by -3.5 percentage points. This still is not strong evidence of an interview effect; even in the absence of an interview effect, a difference this large would occur in more than half of samples ($p < 0.55$).

In sum, the differences in responses discussed above between an alternative and the benchmark do not translate into large differences in estimated poverty rates.

⁴⁴ Results for India’s national poverty line are similar.

Nevertheless, completion rates vary by method in ways linked with poverty: completion rates are higher for face-to-face methods than for by-phone methods. As discussed earlier, this means that interview methods do affect estimated poverty rates via their effect on the population represented. In particular, households with access to phones tend to be less-poor. Among those with access to phones, those few with the sophistication to respond to IVR are even less-poor.

This leads to large differences in estimated poverty rates across alternatives (Figure 11). For the poverty lines here, by-phone estimates are 8 to 18 percentage points lower than in-person/away-from-home estimates. Between the two by-phone methods, IVR estimates are 7 to 9 percentage points lower than in-person/by-phone. All differences are statistically significant at the 90-percent level ($p < 0.01$).

The interview method affects poverty rates, not via responses to questions but rather via completion rates.

5. Discussion

PPIs are appropriately applied at-home/face-to-face because this is how data is collected in the national consumption surveys upon which PPIs are based. For many organizations, however, it is costly to send an enumerator to a participant's home. Do lower-cost interview methods affect PPI results?

As a test, this study interviewed SHG participants in a rural area of India twice, once in-person/at-home and once with an alternative method. The alternative was randomized, as was the enumerator and the order of the benchmark/alternative.

There are three main results:

- Mis-reporting is distressingly frequent, with about 18 percent of responses for in-person methods (40 percent for IVR) differing in the alternative versus benchmark
- For households in a given alternative who completed both the benchmark and the alternative interview, mis-reporting—weighted by PPI points and converted into poverty likelihoods—does not cause differences in estimated poverty rates
- Completion rates vary a lot across alternative methods and are linked with poverty. By-phone methods have lower completion rates, leading to lower poverty-rate estimates because less-poor households are more likely to have access to phones

This suggests that—all else constant—in-person/away-from-home can substitute for in-person/at-home. By-phone methods under-estimate poverty rates, and the bias is stronger for automated/by-phone (IVR, and probably SMS) than for in-person/by-phone. If almost all participants have phones and completion rates are as good as they would be with in-person/at-home, then in-person/by-phone would probably be acceptable, but IVR or SMS are still unlikely to give results close to the benchmark.

Some questions remain:

- Why is mis-reporting so frequent?
- Can the PPI be made more robust to alternative interview methods?
- How generalizable are these results to other populations and contexts?

5.1 Frequent mis-reporting

Mis-reporting is startling: 20 to 40 percent of responses. How can this be, if PPI questions are simple?⁴⁵ And what can be done about it?

Even if PPI questions are simple, some are relatively complex. In particular, mis-reporting is higher for questions related to household members (household size, number of workers, and school attendance). This is probably because they involve concepts (*household* and *work*) that are best explained by a face-to-face enumerator who can step in when a respondent seems puzzled or when the enumerator sees (or hears from) someone else in the residence who does not seem to have been properly considered.

Some mis-reporting here is due to having different respondents in the benchmark versus alternative. Other sources of bias aside, different respondents may not know the same things, understand questions the same, nor make the same effort. These errors are random in that they would balance out if the respondent is selected at random.

Still, random error probably is not enough to explain, for example, how one-third of households are inconsistent about whether they own a cart, plough, or agricultural sprayer. And not all error is random, so assuming it is may be just so much phlogiston.

⁴⁵ And what does it say about the accuracy of more complex questions?

It also seems unlikely that all the mis-reporting is explained by many households purposely under-reporting to look “more-poor” in the hopes of qualifying for aid while, at the same time, many other households purposely over-report to avoid the embarrassment of revealing their true poverty.

For a given interview method, mis-reporting also depends on the quality of fieldwork and on the quality of training of enumerators. Exactly what enumerators do in the field is rarely known. The enumerators in the test here received a recommended course of training, but even a couple of week-long field visits by project staff cannot reveal whether the enumerators always strictly follow their training.

As a possible standard, mis-reporting rates are less than 10 percent in a handful of small ($n = 30$ or so) proprietary tests by PPI users of a desired alternative method versus in-person/at-home. These tests, however, do not randomize method order nor enumerators. Instead, some enumerators interview with an alternative method, and then the enumerators’ trainer interviews the same households at-home. The trainer (and the organization that wants to use an alternative method) have incentives to find low mis-reporting, but they may also be more careful than the trainers and enumerators here. High-quality training and careful fieldwork can only reduce mis-reporting.

In sum, the unexpectedly high mis-reporting rates here are due to some mix of complex questions, random errors, changes in respondents, intentional mis-reporting, and enumerators’ training and carefulness.

5.2 More robust PPIs

PPI are more robust to mis-reporting when they avoid more-complex questions (involving counts or activities of household members) in favor of simpler questions (involving asset ownership or aspects of the residence). The simpler the questions, the smaller the effect of an alternative interview method.

Nevertheless, most PPIs include (and will continue to include) some complex questions. In particular, household size is a very powerful predictor, school attendance is a development outcome that matters *per se*, and market work is a basic driver of development. Having a variety of types of questions also makes the PPI more accurate with different sub-national groups.

PPI questions probably cannot be simplified enough to make IVR (or SMS) work well, as conditional mis-reporting with IVR is high for all types of questions.

5.3 Generalizability/external validity

Randomization provides internal validity; for the study population, the changes in results can confidently be attributed to the interview method. But internal validity need not imply external validity, as the results may or may not generalize well to other populations or contexts (Ravallion, 2009; Rodrik, 2008).

Of course, one data point is better than none, especially when the drivers of the results—for example, by-phone methods or complex questions—can be expected to apply widely. So while the results may not always generalize, they do provide a rough-

but-sensible default expectation in the absence of knowledge of case-specific differences in populations or contexts that suggest otherwise. For example, the default assumption should be that SMS will not perform much better than IVR.

A number of factors specific to this test affect its generalizability:

- The India PPI (and its questions and points) differ from other PPIs
- The test is in a very poor⁴⁶ area of rural India
- Less than 40 percent of households in the population here own a phone
- Unlike in many PPI applications, the enumerators here are not field agents whom participants in local pro-poor programs already know well
- Enumerators here receive training of above-average quality, and they have strong incentives to complete a household's second interview if it has already completed one
- Enumerators here had to ask participants for phone numbers, present the idea of responding via IVR, and give a short demonstration on how to respond

In-person/by-phone results should improve if almost all participants own phones and if numbers are known. For example, some microlenders require applicants to provide a phone number because calling is a low-cost way to dun late payers. A PPI user might also give phones—and calling credits—to a representative sample, although completion rates may still fall short of in-person/at-home (Croke *et al.*, 2012).⁴⁷

⁴⁶ Here, conditional under-reporting exceeds conditional over-reporting. In a less-poor population, the scope for increased under-reporting would exceed the scope for increased over-reporting, perhaps leading to downwardly biased estimates of poverty rates.

⁴⁷ worldbank.org/en/topic/poverty/brief/high-frequency-data-collection, retrieved 27 June 2015.

5.4 Can in-person/by-phone be used anyway?

In the test here, in-person/away-from-home has results close to the benchmark and so works well as long as completion rates—under the strict definition used here—resemble those in the benchmark. Automated methods (IVR or SMS) are not recommended (regardless of completion rates). The adequacy of in-person/by-phone, however, is less clear-cut (unless completion rates resemble those in the benchmark).

Of course, what is *adequate* depends on the purpose. If the question is whether a pro-poor program is on the right track in terms of reaching the poor to the extent believed or claimed,⁴⁸ then in-person/by-phone may be adequate. The bias here of about –8 percentage points will not make a material difference if the PPI is used for this purpose.

If the PPI is to inform decisions that require a point-in-time estimate within 10 percentage points of the true poverty rate, then in-person/by-phone is not adequate.

When measuring change over time, bias washes out, so in-person/by-phone may be adequate even with less-than-benchmark completion rates.

When segmenting participants for targeted services, participants have strong incentives to mis-report, so there is no substitute for in-person/at-home interviews.

⁴⁸ This is the most common use of the PPI.

That said, some PPI users will still want to reduce costs by using in-person/by-phone (with response rates below the benchmark’s) or even IVR or SMS, and it would be delusional to pretend to prohibit that and arrogant to promulgate a one-size-fits-all prescription for what PPI users should do in general.

Users must decide for themselves. But anything does not go; users should still be careful and transparent about how they choose a method based on the evidence here and their specific purpose, population, and context, recognizing their limited time and resources and that, when the benefit of better decision-making is balanced against the cost of improved decision-making inputs, it is rarely optimal to maximize accuracy.

The PPI trades (a little) accuracy for (a lot of) cost. It is both “good enough for government work” and rigorous because it is transparent about its accuracy and about the aspects of its performance for which there is uncertainty.

Make no mistake; alternative interview methods are *off-label*, and the only correct (*on-label*) way to apply the PPI is in-person/at-home. But just as it is sometimes appropriate for a medical doctor—after careful consideration of the available evidence on costs, benefits, and risks in a specific case—to prescribe a drug for a use that lacks the conclusive evidence needed to gain regulatory approval, it may also sometimes be appropriate for a pro-poor program to apply the PPI off-label as long as it has carefully weighed the risks and as long as it reports its methods transparently.

This study does not condone nor condemn off-label methods; it seeks to help users make intentional, transparent judgments about an appropriate method, rather than blindly adopting the least costly (or most accurate) method. The goal is less academic (to maximize accuracy and to prescribe a universal answer) and more practical (to provide a tool to improve real-world decisions).

How can a user “be careful” when using an off-label method? First, do a small pilot test (with one sample interviewed twice and $n = 50$ or so) to check mis-reporting and completion rates. Try to control for method order, and be aware that enumerator effects may dominate. In particular, do not have one enumerator do all the alternative interviews nor all the benchmark interviews. Analyze how responses change between methods (as in Figures 6 and 7). Try to ferret out the causes of mis-reporting. Is it enumerator error (which can be reduced with better/more training)? Do different enumerators make different—and equally valid—judgments in cases without guidelines from the national consumption survey upon which the PPI is based? Do respondents perceive incentives to mis-report? Does the interview method itself affect incentives to mis-report or otherwise cause errors? Then try to address the drivers.

Second, when analyzing and reporting PPI estimates, document the interview method. Speculate on how the method might affect the estimates and the decisions based on them. Do not use an off-label method and act like the estimates are not affected. Be transparent about the aspects of a decision that—like all real-world

decisions—are based on assumptions or judgment rather than data, and discuss the weaknesses of the data.

The essence of both science and good management is being open about how conclusions are reached so that assumptions and judgments can be discussed and perhaps improved. Rigor is not certainty, but rather transparency about uncertainty.

References

- Alkire, Sabina; and James Foster. (2011) “Counting and Multidimensional Poverty Measurement”, *Journal of Public Economics*, Vol. 95, pp. 476–487.
- American Association for Public-Opinion Research. (2011) “Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys”, aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/StandardDefinitions2011_1.pdf, retrieved 25 August 2015.
- Anand, Preeti. (2015) “PPI: Out-of-home Testing”, memo.
- Caire, Dean; and Mark Schreiner. (2012) “Cross-Tab Weighting for Credit Scorecards in Developing Markets”, dean_caire@hotmail.com.
- Cowger, Charles D. (1984) “Statistical Significance Tests: Scientific Ritualism or Scientific Method?”, *Social Service Review*, Vol. 58, No. 3, pp. 358–372.
- Croke, Kevin; Dabalén, Andrew; Demombynes, Gabriel; Giugale, Marcelo; and Johannes Hoogeveen. (2012) “Collecting High-Frequency Panel Data in Africa Using Mobile-Phone Interviews”, World Bank Policy Research Working Paper No. 6097, Washington, D.C., documents.worldbank.org/curated/en/2012/06/16645535/collecting-high-frequency-panel-data-africa-using-mobile-phone-interviews, retrieved 28 June 2015.
- Deaton, Angus; and Salman Zaidi. (2002) “Guidelines for Constructing Consumption Aggregates for Welfare Analysis”, World Bank Living Standards Measurement Survey Working Paper No. 135, Washington, D.C., documents.worldbank.org/curated/en/2002/05/1944189/guidelines-constructing-consumption-aggregates-welfare-analysis, retrieved 28 June 2015.
- Elbers, Chris; Lanjouw, Jean O.; and Peter Lanjouw. (2003) “Micro-Level Estimation of Poverty and Inequality”, *Econometrica*, Vol. 71, No. 1, pp. 355–364.
- Fincham, Jack E. (2008) “Response Rates and Responsiveness for Surveys, Standards, and the *Journal*”, *American Journal of Pharmaceutical Education*, Vol. 72, No. 2, p. 43.
- Grosh, Margaret; and Judy L. Baker. (1995) “Proxy-Means Tests for Targeting Social Programs: Simulations and Speculation”, World Bank Living Standards Measurement Survey Working Paper No. 118, Washington, D.C., go.worldbank.org/W90WN57PD0, retrieved 28 June 2015.

- Keeter, Scott; McGeeney, Kyle; Igielnik, Ruth; Mercer, Andrew; and Nancy Mathiowetz. (2015) "From Telephone to the Web: The Challenge of Mode of Interview Effects in Public-Opinion Polls", Pew Research Center, pewresearch.org/2015/05/13/from-telephone-to-the-web-the-challenge-of-mode-of-interview-effects-in-public-opinion-polls/, retrieved 28 June 2015.
- Kish, Leslie. (1949), "A Procedure for Objective Respondent Selection within the Household", *Journal of the American Statistical Association*, Vol. 44, No. 247, pp. 380–387.
- Lynn, Peter; and Olena Kaminska. (2013) "The Impact of Mobile Phones on Survey Measurement Error", *Public Opinion Quarterly*, Vol. 77, No. 2, pp. 586–605.
- Martinelli, César; and Susan W. Parker. (2009) "Deception and Misreporting in a Social Program", *Journal of the European Economic Association*, Vol. 4, No. 6, pp. 886–908.
- McNemar, Quinn. (1947) "Note on the Sampling Error of the Difference between Correlated Proportions or Percentages", *Psychometrika*, Vol. 17, pp. 153–157.
- Onwujekwe, Obinna; Hanson, Kara; and Julia Fox-Rushby. (2006) "Some Indicators of Socio-Economic Status May Not Be Reliable and Use of Indices with These Data Could Worsen Equity", *Health Economics*, Vol. 15, pp. 639–644.
- Ravallion, Martin. (2009) "Should the Randomistas Rule?", *Economists' Voice*, Vol. 6, No. 2.
- Rodrik, Dani. (2008) "The New Development Economics: We Shall Experiment, But How Shall We Learn?", in Jessica Cohen and William Easterly (eds) *What Works in Development: Thinking Big versus Thinking Small*, Washington, D.C.: Brookings Institution Press.
- Rutstein, Shea Oscar; and Kiersten Johnson. (2004) "The DHS Wealth Index", DHS Comparative Reports No. 6, Calverton, MD: ORC Macro, measuredhs.com/pubs/pdf/CR6/CR6.pdf, retrieved 28 June 2015.
- Schreiner, Mark. (2014a) "How Do the Poverty Scorecard and the PAT Differ?", microfinance.com/English/Papers/Scorecard_versus_PAT.pdf, retrieved 28 June 2015.

- (2014b) “The Process of Poverty-Scoring Analysis”, microfinance.com/English/Papers/Scoring_Poverty_Outline_Analysis.pdf, retrieved 28 June 2015.
- (2012a) “A Simple Poverty Scorecard™ for India”, microfinance.com/English/Papers/Scoring_Poverty_India_2009_EN.pdf, retrieved 28 May 2015.
- (2012b) “An Expert-Based Poverty Scorecard for Rural China”, microfinance.com/English/Papers/Scoring_Poverty_China_EN.pdf, retrieved 28 June 2015.
- Skoufias, Emmanuel; Davis, Benjamin; and Jere R. Behrman. (1999) “An Evaluation of the Selection of Beneficiary Households in the Education, Health, and Nutrition Program (PROGRESA) of Mexico”, ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/125727, retrieved 28 June 2015.
- Wikipedia. (2015) “Response Rates”, en.wikipedia.org/wiki/Response_rate, retrieved 28 June 2015.

Figure 1: PPI for India

Interview ID: _____	<u>Name</u>	<u>Identifier</u>
Interview date: _____	Participant: _____	_____
Country: <u>IND</u>	Field agent: _____	_____
Scorecard: <u>003</u>	Service point: _____	_____
Sampling wgt.: _____	Number of household members: _____	

Indicator	Response	Points	Score
1. How many household members are 17-years-old or younger?	A. Four or more	0	
	B. Three	7	
	C. Two	11	
	D. One	17	
	E. Zero	26	
2. What is the general education level of the male head/spouse?	A. No male head/spouse	0	
	B. Not literate, no formal school, or primary or below	0	
	C. Middle	3	
	D. Secondary or higher secondary	5	
	E. Diploma/certificate course, graduate, or postgraduate and above	7	
3. What is the household type?	A. Labour (agricultural, casual, or other)	0	
	B. Self-employed (agriculture or non-agriculture), regular wage/salary-earning, or others	5	
4. What is the primary source of energy for cooking in the last 30 days?	A. Firewood and chips, dung cake, kerosene, charcoal, coke or coal, gobar gas, or others	0	
	B. LPG or electricity	3	
	C. No cooking arrangement	9	
5. Does the household possess any casseroles, thermos, or thermoware?	A. No	0	
	B. Yes	5	
6. Does the household possess a television and a VCR/VCD/DVD player?	A. No, neither one	0	
	B. Yes, only one	4	
	C. Yes, both	9	
7. Does the household possess a mobile handset and a telephone instrument (landline)?	A. No, neither one	0	
	B. Yes, only a mobile	9	
	C. Yes, a landline, regardless of mobile	15	
8. Does the household possess a sewing machine?	A. No	0	
	B. Yes	1	
9. Does the household possess an almirah/dressing table?	A. No	0	
	B. Yes	5	
10. Does the household possess a bicycle, motorcycle/scooter, or motor car/jeep?	A. No, none	0	
	B. Yes, bicycle only, no motorcycle/scooter, or car	1	
	C. Motorcycle/scooter, but no car (regardless of bicycle)	13	
	D. Motor car/jeep (regardless of others)	18	

Figure 2: Conversion of scores to poverty likelihoods

<u>Score</u>	<u>Tendulkar Natl. poverty line</u>	<u>\$1.25/day 2005 PPP line</u>
0–4	73.7	91.9
5–9	63.5	83.6
10–14	53.5	76.7
15–19	38.0	62.0
20–24	25.9	49.7
25–29	21.9	41.8
30–34	14.6	30.5
35–39	9.4	21.8
40–44	6.5	15.9
45–49	3.6	10.8
50–54	2.0	6.4
55–59	1.1	3.7
60–64	0.5	2.6
65–69	0.2	1.6
70–74	0.2	0.7
75–79	0.0	0.4
80–84	0.0	0.4
85–89	0.0	0.0
90–94	0.0	0.0
95–100	0.0	0.0

Poverty likelihoods are percentages. Consumption is in terms of MMRP expenditure. On average across all of India from 1 July 2009 to 30 June 2010, the Tendulkar national poverty line is Rs24.36 per person per day, giving a national head-count poverty rate of 23.0 percent. The \$1.25/day line is Rs29.29 per person per day, giving an all-India head-count poverty rate of 38.7 percent (Schreiner, 2012a).

Figure 3: Additional typical PPI questions tested

Indicator	Response
11. How many household members are there?	A. Eight or more B. Seven C. Six D. Five E. Four F. Three G. Two H. One
12. Do all household members ages 6-to-17 attend school?	A. No B. Yes C. No one ages 6-to-17
13. In the past week, how many household members did any work for at least one hour?	A. None B. One C. Two D. Three or more
14. What is the main source of energy used by the household for lighting?	A. None/firewood/flaming torch B. Candles C. Flashlight D. Home-made lantern E. Purchased lantern F. Electricity G. Other
15. Does the household possess a chair?	A. No B. Yes
16. Does the household possess any cattle (cow, bull, buffalo, ox, donkey)?	A. No B. Yes
17. Does the household possess a pressure cooker?	A. No B. Yes
18. Does the household possess an animal-drawn cart, an animal-drawn plough, or an agricultural sprayer?	A. No B. Yes

The effects of alternative interview methods is tested for responses to these questions that are not in India's most-recent PPI (Schreiner, 2012a) but that often appear in other countries' PPIs. They are not scored here, and they are not used in the tests of the effects of alternative methods on completion rates nor on estimated poverty rates. As in Figure 1, responses are ordered by decreasing poverty likelihood.

Figure 4: Distribution (%) of responses by method

Indicator	Response	In-person/ Away-from-home		In-person/ by-phone		Automated/ IVR	
		Home	AFH	Home	Call	Home	IVR
1. How many household members are 17-years-old or younger?	A. Four or more	8	9	10	10	6	17
	B. Three	21	22	20	19	13	19
	C. Two	27	26	28	30	42	28
	D. One	18	17	18	19	23	21
	E. Zero	26	25	23	23	17	16
2. What is the general education level of the male head/spouse?	A. No male head/spouse	9	9	8	8	11	17
	B. Not literate, or primary or below	61	61	54	55	40	30
	C. Middle	20	19	21	21	28	22
	D. Secondary or higher secondary	9	9	14	14	17	23
	E. Higher than higher secondary	2	1	2	2	3	9
3. What is the household type?	A. Labour	70	71	68	64	62	58
	B. Self-employed, or wage/salary	30	29	32	36	38	42
4. What is the primary source of energy for cooking in the last 30 days?	A. Firewood/chips/low-quality fuel	100	100	100	100	99	87
	B. LPG or electricity	0	0	0	0	1	4
	C. No cooking arrangement	0	0	0	0	0	9
5. Does the HH own any casseroles, thermos, or thermoware?	A. No	100	100	100	100	100	94
	B. Yes	0	0	0	0	0	6
6. Does the household possess a television and a VCR/VCD/DVD player?	A. No, neither one	92	92	86	87	78	65
	B. Yes, only one	7	7	12	11	17	25
	C. Yes, both	1	1	2	3	4	11
7. Does the household possess a mobile handset or a landline telephone?	A. No, neither one	62	62	28	28	11	13
	B. Yes, only a mobile	38	37	72	72	89	83
	C. Yes, landline	0	0	0	0	0	4
8. Does the household possess a sewing machine?	A. No	97	97	96	96	92	88
	B. Yes	3	3	4	4	7	12
9. Does the HH possess an almirah/dressing table?	A. No	95	95	94	88	93	83
	B. Yes	5	5	6	12	7	17
10. Does the HH possess a bicycle, motorcycle, or motor car/jeep?	A. No, none	82	83	77	70	69	56
	B. Yes, bicycle only	15	14	17	24	23	30
	C. Motorcycle, no car (regardless bike)	3	3	5	5	8	13
	D. Motor car/jeep (regardless others)	0	0	0	0	0	1
11. How many household members are there?	A. Eight or more	3	3	2	1	1	1
	B. Seven	12	12	9	7	5	4
	C. Six	13	14	11	10	9	12
	D. Five	23	23	24	26	34	25
	E. Four	25	26	25	26	24	19
	F. Three	13	12	17	18	17	20
	G. Two	6	6	8	7	8	11
	H. One	5	5	5	5	3	8
12. Do all household members ages 6-to-17 attend school?	A. No	9	12	7	5	7	50
	B. Yes	54	52	57	58	57	40
	C. No one ages 6-to-17	37	36	36	36	36	10
13. In the past week, how many household members did any work for at least one hour?	A. None	2	2	1	3	0	49
	B. One	9	8	4	6	5	26
	C. Two	51	53	47	42	54	11
	D. Three or more	39	36	48	49	42	14
14. What is the main source of energy used by the household for lighting?	A. None/firewood/flaming torch	27	31	24	24	23	34
	B. Candles	0	0	0	0	1	4
	C. Flashlight	0	0	0	0	0	4
	D. Home-made lantern	16	16	14	13	5	6
	E. Purchased lantern	5	4	5	7	1	3
	F. Electricity	51	49	57	55	70	50
	G. Other	0	0	0	0	1	0
15. Does the household possess a chair?	A. No	79	82	71	72	62	62
	B. Yes	21	18	29	28	38	38
16. Does the household possess any cattle?	A. No	97	97	95	96	94	91
	B. Yes	3	3	5	4	7	9
17. Does the household possess a pressure cooker?	A. No	26	26	19	19	21	30
	B. Yes	74	74	81	81	79	72
18. Does the HH possess a cart, plough, or sprayer?	A. No	43	44	33	39	37	38
	B. Yes	57	56	66	61	63	62

Figure 5: Direction of changes in responses, benchmark vs. in-person/away-from home

Question	Mis-reports (%)					Conditional mis-reports (%)		
	Under	Over	Total	Net	<i>p</i> value	Under	Over	Total
1. How many HH members are ≤ 17 -years-old?	12	9	21	-3	0.03 *	13	12	12
2. What is the education of the male head?	11	9	20	-2	0.12	12	9	11
3. What is the HH labour type?	16	15	31	-1	0.54	53	21	31
4. What is the main fuel for cooking?	0	0	0	0	0.32	N/A	0	0
5. Does the HH own any casseroles or thermos?	0	0	0	0	0.09 *	N/A	0	0
6. Does the HH have a TV and a VCR/VCD/DVD?	4	5	9	+1	0.54	54	5	8
7. Does the HH have a telephone?	13	12	25	-1	0.63	34	12	18
8. Does the HH have a sewing machine?	2	2	4	0	0.87	56	2	4
9. Does the HH have an almirah/dressing table?	4	3	7	-1	0.40	67	3	7
10. Does the HH have a bicycle, motorcycle, or car?	10	9	19	-1	0.30	58	9	17
11. How many HH members are there?	12	10	22	-1	0.32	12	11	11
12. Do all HH members ages 6-to-17 attend school?	15	11	26	-3	0.04 *	16	18	17
13. How many HH members work?	17	14	32	-3	0.12	18	24	20
14. What is the main source of energy for lighting?	26	22	48	-3	0.14	35	22	28
15. Does the HH have a chair?	9	6	15	-3	0.04 *	42	8	15
16. Does the HH have any cattle?	3	2	5	0	0.69	81	3	5
17. Does the HH have a pressure cooker?	8	8	16	0	0.88	11	32	16
18. Does the HH have a cart, plough, or sprayer?	17	16	33	-1	0.71	29	37	33

Under-reporting is when the alternative response is "more poor" than the benchmark response.

Over-reporting is when the alternative response is "less poor" than the benchmark.

p values are for the net direction of misreporting from McNemar's test (1947) for differences in correlated percentages.

Mis-reports are changes in a given direction, regardless of whether a change in that direction is possible.

Conditional mis-reports are changes for households for whom a change in a given direction is possible.

n = 1,043 for all questions.

Figure 6: Distribution of changes in response to the main source of energy for lighting, benchmark vs. in-person/away-from-home

		In-person/away-from-home							
Benchmark	Response	None/ firewood/ torch	Candle	Flash- light	Home- made lantern	Purchased lantern	Electricity	Other	Row total
	None/firewood/torch	120	1	—	70	18	76	1	286
	Candles	—	—	—	—	—	—	—	—
	Flashlight	—	—	—	—	—	—	—	—
	Home-made lantern	75	—	—	41	6	47	2	171
	Purchased lantern	26	—	—	9	5	11	—	51
	Electricity	98	—	—	7	7	378	1	531
	Other	2	—	—	1	1	1	—	4
Column total:	321	1	—	167	37	513	4	1,043	

Figure 7: Distribution of changes in response to the number of household members, benchmark vs. in-person/away-from-home

		In-person/away-from-home							Row total	
		Eight or more	Seven	Six	Five	Four	Three	Two		One
Benchmark	Eight or more	19	3	2	—	2	—	—	2	28
	Seven	4	93	13	3	3	6	2	1	125
	Six	—	9	114	8	6	1	1	0	139
	Five	1	3	8	206	9	2	4	4	237
	Four	—	5	4	16	217	8	6	2	258
	Three	1	3	3	6	18	97	10	2	140
	Two	2	3	4	2	4	7	38	6	66
	One	1	2	—	1	7	2	5	32	50
	Column total:	28	121	148	242	266	123	66	49	1,043

Figure 8: Direction of changes in responses, benchmark vs. in-person/by-phone

Question	Mis-reports (%)					Conditional mis-reports (%)		
	Under	Over	Total	Net	<i>p</i> value	Under	Over	Total
1. How many HH members are ≤17-years-old?	12	11	23	0	0.80	13	15	14
2. What is the education of the male head?	11	10	21	-1	0.59	12	10	11
3. What is the HH labour type?	14	18	31	+4	0.02 *	42	26	31
4. What is the main fuel for cooking?	0	0	0	0	0.16	0	0	0
5. Does the HH own any casseroles or thermos?	0	1	1	+1	0.02 *	N/A	1	1
6. Does the HH have a TV and a VCR/VCD/DVD?	8	7	14	-1	0.47	52	7	13
7. Does the HH have a telephone?	9	10	19	+1	0.58	13	10	11
8. Does the HH have a sewing machine?	2	2	5	0	0.89	61	3	5
9. Does the HH have an almirah/dressing table?	0	1	1	+1	0.01 *	50	9	12
10. Does the HH have a bicycle, motorcycle, or car?	8	15	23	+7	0.01 *	36	15	19
11. How many HH members are there?	13	14	27	+1	0.38	13	15	14
12. Do all HH members ages 6-to-17 attend school?	8	10	18	+2	0.10	8	16	11
13. How many HH members work?	16	13	29	-3	0.07 *	16	25	19
14. What is the main source of energy for lighting?	20	19	39	-1	0.66	26	19	22
15. Does the HH have a chair?	12	11	22	-1	0.56	40	15	22
16. Does the HH have any cattle?	3	2	4	-1	0.11	55	2	4
17. Does the HH have a pressure cooker?	7	7	14	0	0.69	9	37	14
18. Does the HH have a cart, plough, or sprayer?	17	12	30	-5	0.01 *	26	36	30

Under-reporting is when the alternative response is "more poor" than the benchmark response.

Over-reporting is when the alternative response is "less poor" than the benchmark.

p values are for the net direction of misreporting from McNemar's test (1947) for differences in correlated percentages.

Mis-reports are changes in a given direction, regardless of whether a change in that direction is possible.

Conditional mis-reports are changes for households for whom a change in a given direction is possible.

n = 1,050 for all questions.

Figure 9: Direction of changes in responses, benchmark vs. IVR

Question	<i>n</i>	Mis-reports (%)				<i>p</i> value	Conditional mis-reports (%)		
		Under	Over	Total	Net		Under	Over	Total
1. How many HH members are ≤17-years-old?	151	40	27	67	-13	0.06 *	42	33	38
2. What is the education of the male head?	151	26	32	58	+5	0.40	30	33	31
3. What is the HH labour type?	151	21	25	46	+3	0.55	55	40	46
4. What is the main fuel for cooking?	151	0	13	13	+13	0.01 *	0	13	13
5. Does the HH own any casseroles or thermos?	151	0	7	7	+7	0.01 *	N/A	7	7
6. Does the HH have a TV and a VCR/VCD/DVD?	151	12	25	37	+13	0.01 *	56	26	32
7. Does the HH have a telephone?	150	11	11	22	+1	0.87	12	11	12
8. Does the HH have a sewing machine?	150	2	7	9	+5	0.06 *	27	7	9
9. Does the HH have an almirah/dressing table?	150	3	13	17	+10	0.01 *	45	14	17
10. Does the HH have a bicycle, motorcycle, or car?	144	13	27	40	+15	0.01 *	42	27	30
11. How many HH members are there?	142	16	31	47	+15	0.02 *	16	32	24
12. Do all HH members ages 6-to-17 attend school?	137	69	12	80	-57	0.01 *	73	19	51
13. How many HH members work?	134	75	7	83	-68	0.01 *	75	13	52
14. What is the main source of energy for lighting?	134	37	19	56	-17	0.01 *	46	20	31
15. Does the HH have a chair?	131	15	17	32	+2	0.76	42	27	32
16. Does the HH have any cattle?	129	4	6	10	+2	0.41	56	7	10
17. Does the HH have a pressure cooker?	129	16	8	24	-9	0.05 *	20	40	24
18. Does the HH have a cart, plough, or sprayer?	128	35	9	45	-26	0.01 *	55	26	45

Under-reporting is when the alternative response is "more poor" than the benchmark response.

Over-reporting is when the alternative response is "less poor" than the benchmark.

p values are for the net direction of misreporting from McNemar's test (1947) for differences in correlated percentages.

Mis-reports are changes in a given direction, regardless of whether a change in that direction is possible.

Conditional mis-reports are changes for households for whom a change in a given direction is possible.

Figure 10: Comparison of estimated poverty rates under an alternative interview method versus under the benchmark, by alternative, for India’s national poverty line and the \$1.25/day 2005 PPP line

Poverty line	Estimated poverty rates (%)		Difference	
	<u>In-person/away-from-home</u>	<u>Benchmark</u>	% points	<i>p</i> value
National	34.0	33.1	+0.9	0.49
\$1.25/day	54.0	53.0	+0.9	0.51
	<u>In-person/by-phone</u>	<u>Benchmark</u>		
National	26.2	26.9	-0.7	0.57
\$1.25/day	44.5	45.6	-1.1	0.42
	<u>Automated/by-phone (IVR)</u>	<u>Benchmark</u>		
National	19.6	21.5	-1.9	0.64
\$1.25/day	36.0	39.4	-3.5	0.48

n is 1,043 for in-person/away-from-home, 1,050 for in-person/by-phone, and 144 for automated/by-phone (IVR).

p values are for a two-sided test that the difference is zero between estimated poverty rates for the alternative versus the benchmark.

See the notes to Figure 2 for details on the two poverty lines.

Figure 11: Comparison of estimated poverty rates among alternative interview methods, for India’s national poverty line and the \$1.25/day 2005 PPP line

Poverty line	Estimated poverty rates (%)		Difference	
	<u>In-person/away-from-home</u>	<u>In-person/by-phone</u>	% points	<i>p</i> value
National	34.0	26.2	+7.8	0.01
\$1.25/day	54.0	44.5	+9.4	0.01
	<u>In-person/away-from-home</u>	<u>Automated/by-phone (IVR)</u>		
National	34.0	19.6	+14.4	0.01
\$1.25/day	54.0	36.0	+18.0	0.01
	<u>In-person/by-phone</u>	<u>Automated/by-phone (IVR)</u>		
National	26.2	19.6	+6.6	0.01
\$1.25/day	44.5	36.0	+8.6	0.01

n is 1,043 for in-person/away-from-home, 1,050 for in-person/by-phone, and 144 for automated/by-phone (IVR).

p values are for a two-sided test that the difference is zero between estimated poverty rates for two alternatives.

See the notes to Figure 2 for details on the two poverty lines.