

## CHAPTER 3

# Using Mobile Phones for High-Frequency Data Collection

Amparo Ballivian\*, João Pedro Azevedo\* and  
Will Durbin\*

\*World Bank, [aballivian@worldbank.org](mailto:aballivian@worldbank.org),  
[jazevedo@worldbank.org](mailto:jazevedo@worldbank.org), [sdurbin@worldbank.org](mailto:sdurbin@worldbank.org)

### Abstract

The 'Listening to Latin America and the Caribbean' ('Listening to LAC' or 'L2L') project was motivated by the financial crisis of 2008, when policy makers in the region asked the World Bank how the crisis would affect their efforts to reduce poverty and what policy responses they could design to mitigate those impacts. Unfortunately, little data existed to answer this question, as poverty data is collected infrequently. The L2L project aimed to answer this key question: Can we use cellular phone communication technology to reduce the time and cost of collecting household survey data from a probabilistic sample without compromising data quality? This paper presents the results of two pilots of this mode of data collection in Peru and Honduras that allowed us to test this question empirically. The results suggest that using mobile phones for short and frequent surveys can produce high-quality data more quickly – and more cheaply on a per survey basis – than traditional methods, and can be a valuable complement to less frequent, more comprehensive, more expensive household surveys. But, in order for mobile data to produce timely information for policy decisions, the system for mobile surveys must be in place before the crisis starts. In other words, the L2L model cannot be launched after the onset of a crisis. This is because: (i) in order to ensure statistical representativeness, an appropriate sample must be drawn; (ii) it takes some time to recruit the panel; and (iii) an initial face-to-face interview is needed to collect data on

---

#### How to cite this book chapter:

Ballivian, A, Azevedo, J P, and Durbin, W. 2015. Using Mobile Phones for High-Frequency Data Collection. In: Toninelli, D, Pinter, R & de Pedraza, P (eds.) *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*, Pp. 21–39. London: Ubiquity Press. DOI: <http://dx.doi.org/10.5334/bar.c>. License: CC-BY 4.0.

the socio-economic characteristics of each household, which cannot be done by mobile phones due to the large number of questions. In addition, several implementation issues explained in this report need to be addressed ahead of time. For this reason it is not possible to initiate the program of data collection immediately after the onset of a crisis and obtain relevant data quickly. Therefore, the most desirable use of the L2L model of mobile surveys may be as a complement to on-going national surveys which collect mobile phone numbers of household members.

### Keywords

poverty, household surveys, mobile data, panel surveys, data quality, SMS, CATI, IVR

## Listening to Latin America and the Caribbean (‘Listening to LAC’)

### *Background*

Evidence-based decision making for poverty alleviation has evolved considerably in the past 30 years. Whereas in the 1980s only 25 countries had regular household surveys, today the World Bank’s external microdata catalog has 1,580 household surveys on 183 countries. But the data collection mechanisms used today are virtually the same as those used since probabilistic survey data started being collected: after the sample is drawn, a number of interviewers travel to peoples’ homes, they ask the household head dozens of questions verbally, they record the answers on a paper form and, several weeks or months later, the answers are transferred to a digital support. Only then can data analysis begin. All this is costly, takes a long time and is prone to error. Recent advances in survey data collection are introducing digital technology to replace paper questionnaires, thereby reducing time and errors in data transcription. But the time and cost involved in traveling to respondents’ living locations remain essentially the same.

Reducing the time to collect data, particularly in crisis situations, may make the difference between adopting policy actions based on evidence or on guesses. Regardless of the nature of the crisis – economic, political, social, natural disasters or other – policy makers and public authorities need to address these situations within days, or at most weeks, after the onset of the crisis. When these crises happen in developing countries, donors that provide financial or technical assistance also find themselves bound by these very narrow timeframes. Traditional data collection methods simply do not produce data and corresponding analyses quickly enough to be used as evidence supporting short-term policy

decisions. Cost considerations are also important drivers of survey frequency, so reducing cost can also lead to more frequent data collection.

In parallel, information and communication technologies, and in particular the signal coverage and rate of use of cellular phones, has expanded exponentially in developing countries. The 'Listening to Latin America and the Caribbean' ('Listening to LAC' or 'L2L') pilot attempted to take advantage of these two trends – an increasingly ubiquitous modern technology and a rise in evidence-based policy making – to produce more frequent data for policy decisions following crises situations. The key question that the L2L project aimed to answer is this: ***Can we use cell phone communication technology to reduce the time and cost of collecting probabilistic sample data without compromising data quality?***<sup>6</sup>

Telephone interviewing has three main problems: (1) obtaining representative samples of the national population; (2) obtaining adequate response rates; and (3) data quality compared to face-to-face interviewing, which is the standard method of survey data collection in developing countries. The L2L pilot tested for the prevalence and seriousness of these problems in a systematic way.

The L2L pilot showed that it is possible to conduct nationally representative surveys using cell phones provided that an adequate sampling frame is used. To examine data quality issues, the L2L pilot attempted to answer some subsidiary questions, such as: (1) Do different cell phone technologies (SMS, IVR, CATI)<sup>7</sup> have different attrition rates (L2L used a panel of respondents; attrition refers to the drop-out rate over panel waves)? (2) What is the quality of the data collected, in terms of external validity (comparison with traditional methods), internal validity (internal consistency of answers) and reliability (consistency of answers over time/methods)? (3) Do attrition rates differ between countries (Peru and Honduras)? (4) Do attrition rates vary according to observable characteristics, such as age, gender and the education level of the head of household? (5) Does offering an incentive affect attrition rates? Do incentives affect attrition rates differently across different groups and is the impact of incentives country-dependent? (6) What are the costs of the different methods of cell phone communication for eliciting survey responses?

---

<sup>6</sup> The use of cellular phones for data collection commonly involves using crowd-sourcing, but this method is not viable when analysis needs a statistically valid, representative sample that allows researchers to make statistical inferences about the population. Crowd-sourced surveys suffer from selection bias. For this reason, while they are extremely valuable in some situations, they are often not an effective tool for making policy decisions concerning the population at large.

<sup>7</sup> SMS is the well-known acronym for Short Message Service, which allows communications between two mobile phones using short messages (maximum 160 characters). IVR is a lesser known acronym for Interactive Voice Recognition, an audio message sent over telephone lines by a computer application. CATI is the acronym for Computer-Assisted Telephone Interview, in which a person interviews another by voice communication using a telephone. The last two can be used with landline or mobile telephones.

This summary paper presents the results of two pilots of this mode of data collection in two developing countries, Peru and Honduras, and the analysis of the characteristics of the resulting response rates and data quality attributes.

### *Project concept and design*

The pilots were designed to test the response rates and the quality of data, while also providing some information on the cost of collecting data using mobile phones. Moreover, while mobile phone surveys may produce high-quality data for some types of survey questions, such as those typically asked in marketing research, it was important to test whether the method would work as well with survey questions aimed at eliciting information on poverty and vulnerability, which are typically more sensitive.

Because traditionally poverty rates are calculated at the household level, we decided to interview households instead of individuals. Another reason to study households and not individuals is that, unlike in a face-to-face interview, in a mobile phone survey it may be very difficult (indeed, impossible in some situations) to know precisely who is answering the questionnaire.<sup>8</sup>

Because we did not know the distribution of phone ownership, coverage or actual use per socio-economic characteristics, to minimize bias we did not sample from telephone records. Instead, we used two different nationally representative sampling frameworks: the official one provided by the national statistical agency in Peru and the Gallup World Poll sampling framework in Honduras.<sup>9</sup>

We started with an in-person visit to households, following traditional sampling techniques. During this initial face-to-face interview, we gathered baseline information on household characteristics and recruited participants. Since we adopted a panel design in order to test data quality issues in tracking welfare over time, we also used this initial survey to recruit the panel. Interviews were only conducted with households who gave expressed consent to do so. During the face to face interview, households also were asked about their willingness to participate in the follow up surveys via cell phone. Those who accepted signed a written consent form.

We were particularly interested in studying the welfare impacts of a potential crisis in two segments of the population: (i) the vulnerable population, loosely

---

<sup>8</sup> For each mobile phone survey, we attempted to ensure that the respondent was a member of the household by asking two validation questions (year of birth and gender) to match the answers with the household roster obtained at the initial face-to-face interview. We have not reviewed this data yet, but the initial results are not very encouraging, in the sense that there appears to be a significant amount of discrepancy between the household roster and the data provided in the mobile survey for year of birth and gender of the respondent.

<sup>9</sup> Through a competitive bidding process, Gallup won the contract to implement L2L on the ground.

defined as those households that may fall into poverty following a negative shock (e.g., a financial or food-price crisis); and (ii) the upwardly-mobile, loosely defined as that segment of the population that may escape poverty following a positive shock (e.g., a boom in commodity prices). This affected sampling choices.

We also wanted to explore the impact of incentives on the minimization of panel attrition. For this purpose, we randomly assigned households to three groups: one third of households received US\$1 in free airtime for each questionnaire they answered, one third received US\$5 in free airtime and one third (the control group) received no financial incentive.

In summary, the design of the projects mixed some elements of traditional surveys, such as probabilistic sampling and an initial face-to-face interview to recruit the panel, with modern technology to collect frequent data.

### *A. Technological choices*

The first set of decisions we confronted involved the technology to use to communicate with respondents frequently: internet or cellular phones? Text-based or audio-based? Collecting surveys through free internet programs is very common today. But internet use is still low in developing countries – on average only 32 percent of the population use internet regularly; in Honduras and Peru, the percentage is 18 and 39, respectively, in 2013. Furthermore, internet users tend to be more educated, more urban and wealthier than the population at large. And reaching a pre-defined person or household through the internet can be very challenging. In contrast, mobile phone coverage is already very high in Latin America and the Caribbean (see **Table 1**), so the first decision was to collect high-frequency data using cellular phones.<sup>10</sup>

In order to determine the viability of using cellular phones to collect survey data, pre-tests were carried out in Peru and Nicaragua in 2010. In each country, the World Bank team worked with ad-hoc (not probabilistic) samples of individuals in different settings (e.g., urban, semi-urban, rural) and among different demographic groups (young, old, men, women) to test the facility with which individuals were able to answer survey questions using cellular phones. These pre-tests were implemented using Episurveyor, a software application to collect survey data using internet on mobile phones. The trials suggested that the majority of individuals had little difficulty using cell phones. However, the pre-test showed that the response rates would decline substantially beyond 10 questions. The pre-test also showed that, while most people own a cellular phone in urban areas, some of the poorest households in remote areas did not own a phone. Lastly, the pre-test made it clear that familiarity with cell phone features was more common among the young, and that poor rural women were

---

<sup>10</sup> However, the profile of internet usage in the developing world today is in many ways similar to that of the early adopters of the mobile phone, so internet-based surveys may become an option in the near future.

	Honduras	Peru	LAC average
Mobile cellular subscriptions (per 100 people)	103*	101*	109*
Population covered by a mobile-cellular network (%)	86	97	98
Households with a mobile telephone (%)	81	73	84
Population using mobile internet (%)	2.9	5.8	4.4

**Table 1:** Mobile phone coverage in Honduras, Peru, and LAC average, 2010.

\* 2011 data.

Source: World Bank, *Information and Communications for Development 2012: Maximizing Mobile*; [www.worldbank.org/ict/IC4D2012](http://www.worldbank.org/ict/IC4D2012).

particularly difficult to reach (though not necessarily because the interviewers were using mobile phones).<sup>11</sup> These factors pointed to the use of communication technologies that can work using the simplest possible mobile phone and the cellular technology networks that have the largest coverage.

When choosing a mode of data collection, we considered a variety of factors. One was coverage of the target population. Another practical consideration was cost. The characteristics of the different modes of communicating between enumerators and respondents and some of the advantages and disadvantages of these modes for the purpose of collecting survey data are summarized in **Table 2**.

While internet surveys and mobile survey apps offer many advantages, they can be used only on smartphones, which are concentrated among the wealthy in urban areas. In addition, indicators of overall mobile phone coverage rates can be misleading because, while the *overall* geographic coverage of cellular communications is increasing, the coverage of communication networks used by smartphones (internet on cellular networks) is still very limited in developing countries. So, mobile phone survey programs based on mobile internet technology would probably be biased against the poor and vulnerable, precisely the subjects of policy attention in times of crises. In addition, we learned during the project design phase that USSD is not usually marketed in Latin America, since the regulations for its use have not been approved.

Consequently, the surveys in both Peru and Honduras used the three remaining communication technologies – SMS, IVR and CATI – but the survey designs (sample segmentation and contact frequency) were deliberately different. In Peru, households were randomly assigned to a communication mode (SMS, IVR, CATI), which stayed constant for all rounds, or waves, of the survey. In Honduras, all the survivor group of households (the households that

<sup>11</sup> This difficulty was encountered by our pre-test interviewer (white, American, male) but we simply intend to report it and not draw conclusions. For more information on the effects on responses of the gender, tribe and religion matches of the enumerator and the respondent see Baird et al. (2008).

	Audio/ Text	Self- Administered (Yes/No)	Pros	Cons
<b>SMS (Short Message Service)</b>	Text	Yes	Low cost	Maximum 160 characters Requires literacy Does not allow visual aids Automated
<b>IVR (Interactive Voice Recognition)</b>	Audio	Yes	No need for interviewers	Often viewed as annoying Medium cost Does not allow visual aids
<b>CATI (Computer-Assisted Telephone Interview)</b>	Audio	No	Respondent can ask to clarify questions	Higher cost than SMS or IVR, mainly because: (i) voice is more expensive than text communications; and (ii) operators' salaries need to be paid Does not allow visual aids
<b>USSD (Unstructured Supplementary Services Data)</b>	Text	Yes	No length limitations	Requires close collaboration and approval by telecom companies Not commonly marketed in LAC region Does not allow visual aids
<b>Mobile internet</b>	Text	Yes	No length limitations Lower cost than voice communications Allows use of visual aids	Limited mobile internet coverage in LAC region Requires smartphones

**Table 2:** Pros and cons of mobile technologies for survey data collection.

responded to the first questionnaire) was exposed to all three communication modes. Both designs allow for validity tests, while only the Honduran design allowed for reliability tests.<sup>12</sup> The Honduran design was a test-retest design of the communication mode, which is closely related to the difference-in-difference methodology of experimental evaluation. Importantly, the questionnaires were worded exactly the same way, regardless of the mode, which meant short questions, since SMS is limited to 160 characters.

### *B. Incentives*

In order to minimize non-response, three types of incentives were given. First, households that did not own a mobile phone were provided one for free.<sup>13</sup> Approximately 127 phones were donated in Honduras and 200 in Peru. Second, all communications between the interviewers and the households were free to the respondents. Finally, households were randomly assigned to one of three incentive levels – US\$0, US\$1 or US\$5 – which were distributed after completion of each mobile survey. Unfortunately, mobile payments are not very developed in Latin America,<sup>14</sup> so instead of money transfers the pilot transferred the equivalent in free airtime minutes to each respondent's mobile phone account.

### *C. Sample design*

The sample size was 1,500 households in each country, though sampling was done in different ways in Peru and in Honduras. In Peru, where the World Bank has a very close working relationship with the National Statistics Institute (Instituto Nacional de Estadística e Informática, INEI), the L2L sample was based on the sampling frame for the national household survey (Encuesta Nacional de Hogares, ENAHO) conducted by INEI every three months. In Honduras, the sampling was done deliberately without using the National Statistics Institute's sampling frame, in order to test the feasibility of replication of the L2L model in countries where a strong relationship with the statistics office is absent. Instead, the sampling frame used was the Gallup World Poll sampling frame, which is regularly conducted in 160 countries.

In Peru, the sample selection was guided by the following criteria: (i) the sample should be representative nationally, and in urban and rural areas; and (ii) households close to poverty line should be oversampled because policy decisions in time of crises need to be especially mindful of the poor and vulner-

---

<sup>12</sup> We tested for reliability using Cronbach's alpha, a measure of internal consistency, that is, how closely related a set of items are as a group. A 'high' value of alpha is often used as evidence that the items measure an underlying construct. Please see [www.worldbank.org/lacpoverty/l2l](http://www.worldbank.org/lacpoverty/l2l) for further details.

<sup>13</sup> A generous donation from Brightstar Corporation made this possible.

<sup>14</sup> See, for instance: <http://mobilereadiness.mastercard.com/the-index>

able. For the purposes of this project, 'close to poverty line' was defined as the 40 percent of consumption distribution that symmetrically bands the national poverty line: 20 percent above and 20 percent below. In 27 percent of Peruvian households monthly per capita consumption was below the moderate poverty line in 2010 (ENAHO). Consequently, households whose monthly per capita consumption fell between 7 and 47 percent of the national distribution were oversampled.

Honduras did not have an income oversample because the poverty rate is 60 percent, so oversampling 20 percent above the poverty rate would include a large portion of the middle class, which is likely not the most vulnerable in times of crisis. Furthermore, in countries with high poverty rates the poverty line would likely be very close to the average income, so the income distribution would already include a large percentage both of the vulnerable (just above the poverty line) and of households below but close to the poverty line (who may escape poverty in case of positive shocks).

#### *D. Questionnaire design<sup>15</sup>*

For the initial face-to-face surveys the starting point was the official national household survey questionnaire. Step-wise regressions were done to select the set of questions that best predicted consumption. For the purposes of robustness, the regressions were also done with questions that best predicted income, which yielded the same results. A similar procedure was done in Honduras, except that only best predictors of income were chosen, because Honduras did not have a recent consumption aggregate. For the monthly cell phone surveys the pre-test results and other mobile surveys done elsewhere revealed that attrition and non-response increase significantly with the number of questions, and especially after 10 questions. So a maximum of 10 questions had to be chosen for the monthly questionnaire.

Most questions were time-variant and each questionnaire was repeated to observe if answers changed over time. All questions related to variables that strongly affect household welfare and that are likely to change in times of crisis. To simplify the questionnaire and avoid 'recency' effects<sup>16</sup> in the CATI and IVR modes, only questions admitting yes/no answers were chosen. In addition, one set of questions was the food security module developed by the U.S. Department of Agriculture specifically to test the internal validity of the responses using Rasch analysis.

---

<sup>15</sup> Please see [www.worldbank.org/lacpoverty/l2l](http://www.worldbank.org/lacpoverty/l2l) for copies of the questionnaires and related materials.

<sup>16</sup> Recency is the tendency for respondents to answer the last option in a list of possible answers due to low memory retention. Recency is more common in audio modes of survey deployment. See Krosnick and Alwin (1987).

### *E. Costs*

The implementation of both face-to-face and mobile phone surveys also provided actual cost data for implementing each method. The implementing agency, Gallup, provided this information on the basis of actual costs incurred, as summarized in **Table 3**.

### **Pilot results**

There was no intention of the pilot to make inferences about the Peruvian or Honduran populations in terms of welfare, education, health or other questions asked, based on the answers to the L2L survey questions.<sup>17</sup> Rather, we analyzed the results of the different data collection modes along two lines: (i) attrition rates and the behavior of these rates in relation to household characteristics, survey mode and incentive level; and (ii) data quality.

### *Peru*

Two thirds of recruited households in Peru failed to answer the first round of follow-up surveys. As **Table 4** shows, attrition slightly increased with each wave of the survey (between 1 and 3 percentage points per wave), reaching 75 percent in wave 6.

Regarding the mode of communication with respondents in Peru, higher attrition rate and lower survey completion rate was found among panelists who were exposed to self-administered modes (IVR and SMS), as **Table 5** shows. Over the course of the six waves the level of attrition for SMS increased to 79 percent (initial face-to-face compared with wave 6) and to 61 percent for CATI, with attrition for IVR remaining stable (81 percent). It should be noted that the L2L project deliberately sent out more invitations to take part via SMS ( $n = 677$ ), compared to IVR ( $n = 383$ ) and CATI ( $n = 384$ ). Since the level of attrition for SMS is relatively high compared to the CATI group, the higher  $n$ -size of the SMS group drives up the overall attrition of the panel.

Moreover, IVR and SMS have the disadvantage of a large proportion of respondents only answering some of the questions in any given survey, meaning that respondents completely skipped some questions.<sup>18</sup> IVR and SMS are both self-administered methods, while CATI relies on an interviewer whose job it is to ensure all questions are read, understood and answered by the respondents (recording even legitimate 'Don't Know' responses or 'Refusals').

---

<sup>17</sup> Although there is nothing wrong with doing these inferences, given the probabilistic nature of the sample, and we hope other researchers will do so.

<sup>18</sup> Giving a 'don't know' answer or refusing to answer a question is not considered as a skip. If a respondent skips a question no data were obtained at all.

Methodology	Cost Per Interview	Cost Per Year (12 monthly surveys)
Face-to-face	40	720,000
CATI	25	450,000
IVR	17	306,000
SMS	8	144,000

**Table 3:** Costs for a sample of 1,500 surveys (in US dollars).

Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
67%	68%	69%	70%	72%	75%

**Table 4:** Overall attrition rates in successive waves, Peru.

	IVR	SMS	CATI
<b>Wave 1</b>	<b>80%</b>	<b>70%</b>	<b>49%</b>
Wave 2	75%	75%	47%
Wave 3	78%	76%	49%
Wave 4	78%	75%	52%
Wave 5	84%	76%	53%
<b>Wave 6</b>	<b>81%</b>	<b>79%</b>	<b>61%</b>

**Table 5:** Attrition by Methodology in Peru.

Among Peruvian panelists surveyed about why they did not respond (after the completion of the six waves in the panel survey), 26 percent said they would prefer to be interviewed by a person. Also, panelists responding to the surveys via IVR or SMS showed a higher propensity to leave questions unanswered than did respondents answering via CATI. In short, interviewers were important for getting higher response rates and ensuring that respondents gave consideration to all the survey questions.

Economic incentives in the form of mobile phone credit for every completed survey did not seem to have a big effect on the post-recruitment response rate in Peru (see **Table 6**). However, as the panel exercise progressed, incentives seemed to have had some effect on minimizing attrition. It should be noted that a considerably higher incentive (5 dollars) did not prove much more successful in reducing attrition than a smaller amount (1 dollar).

	No incentive	1 USD	5 USD
<b>Wave 1</b>	<b>68%</b>	<b>66%</b>	<b>66%</b>
Wave 2	70%	67%	66%
Wave 3	73%	68%	68%
Wave 4	72%	70%	67%
Wave 5	76%	71%	69%
<b>Wave 6</b>	<b>80%</b>	<b>73%</b>	<b>71%</b>

**Table 6:** Attrition by Incentive Level in Peru.

### *Honduras*

The initial attrition rate – that is, the proportion of respondents who agreed to participate in the panel after the initial face-to-face survey but did not answer the first round of surveys – was considerably lower in Honduras than in Peru in the three survey modes. The final attrition rate was also lower in Honduras, across all survey modes. As in Peru, CATI surveys generated the lowest attrition, but the difference was more pronounced in the case of Honduras (see **tables 7 and 8**).

While in Peru 67 percent of recruited households failed to answer the first round of follow-up surveys, in Honduras this percentage was only 41. However, the gap between the initial and final attrition – that is, the additional number of panelists that dropped out of the panel between the first follow-up survey and the end of the study – was similar in both countries: in Peru the final attrition was eight percentage points higher than the initial attrition rate and in Honduras it was nine percentage points higher than the initial attrition rate.

In contrast to Peru, economic incentives in the form of mobile phone credit did have a considerable effect on post-recruitment response rate in Honduras. Also in contrast to Peru, the size of the incentive mattered in the Honduran study, with higher incentives being more effective in minimizing attrition (see **Table 9**). It is worth noting that Honduran panelists had to work harder

Initial F2F to Wave 1 (Peru)/ Week 1 (Honduras)	Peru	Honduras
IVR	80%	60%
SMS	70%	55%
CATI	49%	12%
<b>Overall</b>	<b>67%</b>	<b>41%</b>

**Table 7:** Initial Attrition/Non-Response by Methodology (Peru vs. Honduras).

Initial F2F to End of Panel Study	Peru	Honduras
IVR	81%	62%
SMS	79%	60%
CATI	61%	28%
<b>Overall</b>	<b>75%</b>	<b>50%</b>

**Table 8:** Final Attrition/Non-Response by Methodology (Peru vs. Honduras).

	No incentive	1 USD	5 USD
<b>From F2F to Wave 1</b>	45%	41%	38%
<b>F2F to end of panel</b>	54%	52%	43%

**Table 9:** Attrition by Incentive Level in Honduras.

than their Peruvian counterparts: the former had to answer up to three surveys per month, while the latter only answered one survey per month.

### Characteristics of non-respondents and the attrite

Higher attrition rates were found among older, less educated, less affluent panelists and among households living in rural areas in both countries. The mobile panel's high attrition among these types of respondents does not necessarily invalidate it as a viable method for nationally representative studies. As long as the attrite population is not systematically different from the respondent population, parameter estimates will not be biased, but their variance will increase as a result of a smaller sample size. This effect can be effectively addressed by increasing the panel size and applying a post-stratification (weighting) scheme.<sup>19</sup>

### Data quality tests

Research aimed to answer two questions related to data quality, validity and reliability:

*Question 1:* Can the SMS method yield valid measurements, i.e. measurements that are comparable, within an acceptable margin of error,

<sup>19</sup> For details on weighting schemes, please refer to 'Baseline face-to-face Surveys in Honduras and Peru. Methodological Report' by Gallup. For additional information on the characteristics of the attrite population please refer to the full report.

to those produced by face-to-face interviews, which is the benchmark standard for surveys?

*Question 2:* Are estimates generated by SMS statistically reliable (that is, stable or consistent across repeated iterations of the same SMS measurement)?<sup>20</sup>

To answer the first question the team applied a criterion validity test, i.e. a comparative analysis between a test and a criterion variable that is supposed to measure the same construct and that is held to be valid. The L2L face-to-face survey was adopted as the criterion measurement for the analyses. Since the SMS sample was affected by a high level of attrition only households that responded to both surveys (45% of the sample) were included in the analysis, in order to ensure that whatever differences might be encountered between the two measures could primarily be attributed to 'mode effects', as opposed to demographic differences between respondents. The difference between the responses given to the test variable and those given to the criterion variable were tested for statistical significance by means of non-parametric analysis of variance (ANOVA).

The results generated by SMS and face-to-face surveys were compared for eight different questions. These questions inquired about factual information on household infrastructure (i.e. possession of TV, type of sanitation facilities, etc.), factual information on access to the internet inside or outside the household, and perceptual information (i.e. whether the respondents considered themselves poor). The responses to all questions by SMS differed from those collected via face-to-face by at least 7.4 percentage points, a margin that is statistically significant at the 95% confidence level. Interestingly, responses given via SMS indicate lower availability of water/sanitation and television ownership, but almost twice as much access to the internet compared to face-to-face responses. In addition, the self-perception of poverty was higher when asked via SMS than face-to-face.

Similar criterion validity tests were performed for IVR and CATI. The responses collected via IVR show a similar pattern as those collected via SMS, with items related to household infrastructure receiving lower 'yes' scores when asked via IVR compared to face-to-face, while the items related to 'internet access' and 'self-perceptions on poverty' received higher scores. As in the case of SMS, the observed differences between IVR and face-to-face are statistically significant. The answers collected via CATI, on the other hand, were almost identical to the ones collected face-to-face, with no item showing a statistically significant difference. Since the same panelists responding to the IVR and CATI surveys also responded to SMS surveys, the differences in responses observed

---

<sup>20</sup> These questions were analyzed only for the Honduras pilot. The numerical results of the validity and reliability tests are available in the full report online, as well as additional multivariate analysis of attrition.

between SMS, CATI and IVR, or between any of these and face-to-face cannot be attributed to demographic differences between them.

Regarding the second data quality question, in order to test the reliability of measurements two identical measurements were performed of the same questions analyzed in the criterion validity analysis discussed above. Gallup performed repeated administrations of these questions by means of face-to-face, IVR and CATI on the same group of panelists in Honduras only. In all cases, the repeated measurements were performed within a minimum of 10 weeks from the first administration.

The results indicate that the SMS surveys performed quite satisfactorily in terms of generating reliable measurements. This conclusion is supported by the fact that Cronbach's Alpha reliability coefficient obtained for SMS (0.74) is very close to the one obtained for face-to-face (0.77) in the same test-retest exercise. Also, as can be expected, the items inquiring about factual information (i.e. on household infrastructure) show a higher reliability than the items measuring perceptions of poverty.

IVR stands out as the method that generated the most reliable responses overall, followed by SMS and CATI which came quite close to each other. Interestingly, IVR responses proved very reliable for all the items tested, outperforming the other two methods in all but one item ('past-30-day access to the Internet), where CATI fared somewhat better. However, while IVR was more reliable, its advantage was relatively small.

It is also interesting that both IVR and CATI outperformed SMS in those items that inquire about personal internet access, a result that could be explained by the pattern observed in the criterion validity analysis, where SMS surveys were most often responded to by younger informants. Therefore, it would appear that the reliability of these questions tends to be affected by an 'informant switching' behavior when asked via SMS.

The CATI responses show an intriguing pattern: both perceptual and factual items behaved somewhat unreliably when compared to the internet-related items for the same method. It should be remembered that CATI was the best-performing method in terms of criterion validity, with almost identical responses to the ones collected via face-to-face.

Another important aspect of this analysis is the fact that the self-administered vs. interviewer-administered dimension does not seem to explain the reliability differences encountered. The top performing method (IVR) is a self-administered method, while SMS and CATI – which fared similarly in the test – are self-administered and interviewer-administered methods, respectively. It should be remembered that the presence of interviewers (or their absence) was a crucial factor in explaining the differences found in the criterion validity analysis. So, since it is no longer the case for the reliability analysis, alternative explanations need to be considered.

A closer look at the survey methods being evaluated suggests that IVR was probably the one that required the shortest time and was least prone to human

error (on the interviewer side). The IVR system would call respondents and play a pre-recorded greeting, followed by instructions and the actual survey questions. Respondents had to press buttons on their mobile phone keypads to answer the questions. The use of a recording guaranteed that the questions were read exactly the same way in each administration, thus controlling for potential errors derived from inconsistent question reading. Besides, it is possible that respondents had to pay close attention to these recordings, as it was obvious that they would not be able to obtain much help or clarification if they missed something.

SMS, on the other hand, relies on the respondent's reading comprehension ability and attention span. Since questions remain in the phone's inbox until the respondent answers them, respondents could conceivably multitask during the survey administration without missing questions. Somewhat similarly, the CATI surveys could have been affected by human factors. Due to logistic considerations, the interviewers who conducted the first surveys were not necessarily the same ones that conducted the second administrations. Thus, although unlikely, there could have been significant variance in speed of reading, intonation, clarity, mastery of the questionnaire, etc.

Alternatively, it could be hypothesized that having a different interviewer re-contact the households to ask the exact same questions could have brought back some anxiety or fear in some respondents. If such was the case, the findings would suggest that, for panel studies such as this one, having no human contact in the administration of repeat surveys is more beneficial for reliability purposes than having inconsistent human contact. This remains, nonetheless, an intriguing set of findings that would require additional research to be understood in a more satisfactory manner.

Importantly, for all methodologies the 'yes' responses were quite consistent, which means most of the variability observed was due to inconsistencies between the 'No' and 'Don't know/Refused' answers. This is an aspect that deserves proper attention as it demonstrates that no methodology performed poorly in terms of consistently accounting for 'presence' of the phenomena inquired.

## Conclusions

The pilots showed that it is possible to do mobile surveys from a nationally representative panel of respondents, but that attrition rates are relatively high and country-specific. Economic incentives can be used to contain attrition, but they work better over repeated waves of the panel.

Data quality tests of validity of the data (for Honduras surveys) showed that answers to SMS and IVR modes were statistically significantly different from answers to the same questions in face-to-face interviews. Responses in the CATI method, on the other hand, were virtually the same as those in face-to-

face interviews. In terms of data reliability tests all methods performed well, with IVR showing the highest reliability.

When choosing between voice and text interviewing, the advantage of text was lower cost; the advantage of CATI was substantially lower attrition rates. A combination of both voice and text is suggested to be explored in future research, in the same spirit as the recent trend of combined landline and mobile phone surveys in the United States.

IVR does not have any advantages over SMS, either in terms of cost or response rates. IVR has the additional inconvenience that survey calls are lost for good when not answered immediately (as opposed to SMS surveys where messages remain in the phone's inbox allowing for a later response). While IVR responses did prove the most reliable data in the test-retest, the small differences in terms of reliability with the other modes suggest that IVR is not a very suitable mode to communicate with respondents (at least for these types of surveys).

Finally, mobile phone surveys have certain practical disadvantages vis-à-vis face-to-face interviews, such as unstable coverage of mobile networks and sometimes lack of electricity to re-charge phones. On the other hand, mobile phone surveys overcome security problems in regions that are prone to conflicts or natural disasters, so they may be a good option in fragile environments.

We are hopeful that disclosing the L2L data to the public will encourage researchers to conduct further analyses. As mobile phone penetration continues to expand in developing countries, we expect this to become an accepted method for collecting survey data more frequently, and hope that this leads to more evidence-based policy decisions when governments and donors are confronted with sudden shocks.

## Acknowledgements

This paper is a summary of a larger report on the project conducted by the Poverty, Equity and Gender Team of the World Bank's Latin America and Caribbean region. The project was led by Amparo Ballivian and João Pedro Azevedo and managed by Will Durbin. Additional team members included Jesus Rios, Johanna Godoy, Christina Borisova and Sabine Mireille Ntsama. This study is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. The full report, alongside with separate technical reports, the dataset, questionnaires, open source software, manuals and other materials produced under this project can be found at [worldbank.org/lacpoverty/l2l](http://worldbank.org/lacpoverty/l2l). The authors encourage further research using these data and materials. The team is grateful to the following individuals who provided excellent comments: Robert Tortola, John Newman, Gabriel Demombynes, Katheleen Beegle and William Jack, as well as to the anonymous peer reviewers selected by the editor of this book. The team

is also grateful to the World Bank trust funds (SFLAC, RSR and ESSD) which provided financial support, and to the Brightstar Corporation, which donated 1,000 mobile phones.

## References

- Alderman, H., Behrman, J.R., Kohler, H., Maluccio, J.A., & Watkins, S.C.** (2001). Attrition in Longitudinal Household Survey Data. *Demographic Research*, 5(4), 79–124. DOI: <http://dx.doi.org/10.4054/DemRes.2001.5.4>
- Baird, S., Hamory, J., & Miguel, E.** (2008). *Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)*, (Center for International and Development Economics Research Working Paper). Berkeley: University of California, Berkeley. Retrieved July 29, 2015 from <http://www.escholarship.org/uc/item/3cw7p1hx>
- Croke, K., Dabalen, A., Demombynes, G., Giugale, M., & Hoogeveen, J.** (2012). *Collecting high frequency panel data using mobile phones. Does timely data lead to accountability?* (World Bank Policy Research Working Paper No. 6097). Retrieved July 29, 2015 from [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2087971](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2087971)
- Curtin, R., Presser, S., & Singer, E.** (2005). Changes in Telephone Survey Non-response over the Past Quarter Century. *Oxford Journals/Social Sciences Public Opinion Quarterly*, 69(1), 87–98. DOI: <http://dx.doi.org/10.1093/poq/nfi002>
- de Leeuw, E. D.** (2005). To Mix or Not to Mix Data Collection Modes in Surveys. *The Journal of Official Statistics*, 21(2), 233–255.
- Dillman, D. A., Phelps, G., Tortora, R. D., Swift, K., Kohrell, J., Berck, J., Messer, B. L.** (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research*, 38(1), 1–18. DOI: <http://dx.doi.org/10.1016/j.ssresearch.2008.03.007>
- Dillon, B.** (2010). *Using Mobile Phones to Conduct Research in Developing Countries*. (Cornell University Working Paper). Retrieved July 29, 2015 from [http://www.edi-africa.com/docs/Dillon\\_MobilePhoneSurveys.pdf](http://www.edi-africa.com/docs/Dillon_MobilePhoneSurveys.pdf)
- Gallup.** (2012). *The World Bank Listening to LAC (L2L) Pilot – Final Report*. Retrieved July 29, 2015 from <http://siteresources.worldbank.org/INTLACREGTOPPOVANA/Resources/840442-1355332251652/8976232-1355959000860/Gallup-Final-Report-Oct-2012-L2L.pdf>
- Gallup.** (2012). *The World Bank Listening to LAC (L2L) Pilot Project: Sample Design for Honduras*. Retrieved July 29, 2015 from <http://siteresources.worldbank.org/INTLACREGTOPPOVANA/Resources/840442-1355332251652/8976232-1355959000860/Gallup-Honduras-Sample-Specifications-L2L.pdf>

- Gallup.** (2012). *The World Bank Listening to LAC (L2L) Pilot Project: Sample Design for Peru*. Retrieved July 29, 2015 from <http://siteresources.worldbank.org/INTLACREGTOPPOVANA/Resources/840442-1355332251652/8976232-1355959000860/Gallup-Peru-Sample-Specifications-L2L.pdf>
- Hausman, J.A. & Wise, D.** (1979). Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment. *Econometrica*, 47(2), 455–473.
- Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P.** (2006). Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey, *Oxford Journals/Social Sciences Public Opinion Quarterly* 70(5), 759–779. DOI: <http://dx.doi.org/10.1093/poq/nfl035>
- Lesser, V., Dillman, D., Carlson, J., Lorenz, F., Mason, R., & Willits, F.** (2001, August). *Quantifying the Influence of Incentives on Mail Survey Response Rates and Their Effects on Nonresponse Error*. Paper presented at Proceedings of the Annual Meeting of the American Statistical Association. Retrieved July 29, 2015 from <http://www.amstat.org/sections/srms/proceedings/y2001/Proceed/00474.pdf>
- Lynn P., & Kaminska O.** (2013). The Impact of Mobile Phones on Survey Measurement Error. *Public Opinion Quarterly* 77(2), 586–605. DOI: <http://dx.doi.org/10.1093/poq/nfs046>
- Nord, M., & Hopwood, H.** (2007). Does Interview Mode Matter for Food Security Measurement? Telephone versus In-Person Interviews in the Current Population Survey Food Security Supplement. *Public Health Nutrition* 10(12), 1474–1480. DOI: <http://dx.doi.org/10.1017/S1368980007000857>
- Thomas, R., & Purdon, S.** (1994). Telephone methods for social surveys. *Social Research Update*, Issue 8. Surrey: Department of Sociology, University of Surrey.