Listening to LAC: Using Mobile Phones for High Frequency Data Collection

Final Report

February 2013

This paper is a product of the Poverty, Equity and Gender Team of the World Bank's Latin America and Caribbean region. It was produced by a team led by Amparo Ballivian and Joao Pedro Azevedo, and including Will Durbin, 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 authors may be contacted at aballivian@worldbank.org and jazevedo@worldbank.org. The team is grateful to five peer reviewers who provided excellent comments: Robert Tortola, John Newman, Gabriel Demombynes, Katheleen Beegle and William Jack. The team is also grateful to the World Bank trust funds which provided financial support -SFLAC, RSR and ESSD- and to the Brightstar Corporation, which donated 1,000 mobile phones.

Table of Contents

List of Figures:

List of Annexes:

List of Tables:

JEL:
Keywords: Data, mobile phones, survey, Listening to LAC
Sector Board: Poverty Board

Executive Summary

Evidence-based decision making for poverty alleviation has evolved considerably in the past 30 years; whereas in the 1980's only some 25 countries worldwide had regular household surveys. Today the World Bank's microdata catalog has 1019 household surveys on 186 countries; but the data collection mechanisms used to collect these data are virtually the same as those used several decades ago, and both the time and costs of collecting survey data are high. So, while long-term development policy decisions that take months or years to consider can be and are increasingly based on evidence, short-term decisions - in particular those following economic crises or natural disasters - continue to be based on an evidence void. Traditional data collection methods simply do not produce data and corresponding analyses quickly enough to be used in crisis situations.

On a parallel track, information and communications technologies, and in particular the coverage and use of cellular phones, has expanded exponentially in developing countries. Listening to LAC attempts to take advantage of these two trends – a ubiquitous modern technology and evidence-based policy decisions - to produce more frequent data for policy decisions following crises situations.

The key question that the Listening to LAC (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?*** Telephone interviewing has three main problems: (1) obtaining representative samples of the national population, (2) adequate response rates, and (3) data quality compared to face-to-face interviewing, which is the standard method of survey data collection in developing country. The L2L pilot has tested for the validity 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)[1] have different attrition rates? (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?

This 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.

The results suggest that using mobile phones for short and frequent surveys can produce good quality data more quickly –and more cheaply on a per survey basis- than traditional methods. But, this does not mean that it is possible to initiate the data collection effort after the onset of a crisis and obtain relevant data quickly. In order for cell phone collected data to produce enough information for policy decisions, probabilistic sampling and a baseline survey are essential. In addition, several implementation issues explained in this report need to be addressed ahead of time. For this reason the system for frequent data collection must be in place before the frequent

[1] 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.

data collection process starts. Hence, this method can be a valuable complement to less frequent household surveys conducted by official statistical agencies.

I. Background

Data collection mechanisms used today are virtually the same as those used since probabilistic survey data started being collected: a number of surveyors travel to peoples' homes, they ask a number 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 travelling to respondents' living locations remain essentially the same. For surveys employing a nationally representative, statistically-valid samples, while some technological advances are being brought to data collection methods to improve data quality, such as the use of tablets, other technologies to improve time and cost remain untapped.[2]

Reducing the time to collect data, particularly in crisis situations, may make the difference between adopting policy actions based on evidence or not. 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

[2] The use of cellular phones for data collection is common in 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, and while they are extremely valuable in some situations, they are often not an effective tool for making policy decisions when inferences from the sample to the population at large are needed.

narrow timeframes. Traditional data collection methods do not produce data and corresponding analyses quickly enough to be used in these decisions. Cost considerations are possibly the most important drivers of survey frequency, so reducing them can also lead to more frequent data collection.

As Thomas and Purdon (1994) stated in the context of the move from traditional surveys methods to telephone interviewing in the United States: "the main attraction of telephone interviewing is that it enables data to be collected from geographically scattered samples more cheaply and quickly than by field interviewing … [and makes] possible to avoid cluster sampling, which incurs unfavorable statistical design effects but has to be used in field survey designs to control interviewer travel costs." The use of landline telephones to collect survey data is now an accepted practice in countries with large phone coverage. It is not very common, however, in the developing world, because landline phone coverage in poorer countries is not sufficiently large. In fact, mobile phone coverage is now larger than fixed phone coverage in most developing countries, and continues to grow.

However, telephone interviewing has three main problems: (1) obtaining representative samples of the national population, (2) adequate response rates, and (3) data quality compared to face-to-face interviewing. The L2L pilot has tested for the validity and seriousness of these problems in a systematic way.

The key question that the Listening to LAC (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?*** The L2L pilot attempted to answer some subsidiary questions, such as: (1) Do different cell phone technologies (SMS, IVR, CATI)[3] have different attrition rates? (2) Do attrition rates differ between

[3] 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.

countries (Peru and Honduras)? (3) Do attrition rates vary according to observable characteristics, such as age, gender, and the education level of the head of household? (4) Does offering an incentive affect attrition rates, do incentives affect attrition rates differently across different groups and is the impact of incentives country-dependent? (5) 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)? (6) What are the costs of the different methods of cell phone communication for eliciting survey responses? This 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.

The literature on technological innovations and response rates in data collection is summarized in the next section of this paper.[4] Section 3 presents the project concept and the design choices. Implementation issues are discussed in Section 4. The analysis of the pilot data -in terms of response and attrition rates, comparisons by country, type of cellular technology, household characteristics, and impact of incentives on response rates- is presented in Section 5. Lessons learned conclude the document.

II. Literature review

In order for a random sample to be statistically representative of the entire population two important sources of bias need to be minimized: selection bias and attrition bias. Selection bias is eliminated when every potential respondent in the population has the same probability of being surveyed. This is achieved by well-known sample selection techniques, which are explained in the next section and in annex 2. Attrition bias can

---

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.

4. In addition, annex 1 presents a summary review of reports of similar experiments with technological innovations in data collection.

have the same analytical problems as sample selection bias, unless attrition is systematically the same across households with different characteristics. But non-random attrition does not necessarily lead to biased estimates if attrition is related only to exogenous variables (Hausman and Wise, 1979).[5]

The Listening to LAC project had two large sources of potential bias: its panel survey nature and the use of cellular phones to collect data. In panel surveys attrition generally increases from wave to wave, although panel loss generally declines over time. Attrition is also difficult to measure because some panel members drop indefinitely, while others are intermittent respondents. To understand better the attrition rates of the L2L project, we conducted a quick survey of the literature on panel attrition rates for studies conducted in developing countries or using non-standard technologies. Our review reveals that overall panel attrition rates vary widely depending on the country, the length of questionnaire, the sensitivity of the questions, the mode of data collection, the frequency between panel waves, the existence of incentives, and effort (number of attempts to re-contact).[6] These findings should be borne in mind when assessing the results of the L2L pilot response rates presented in section V.

Hill (2004) reports panel attrition rates for 13 studies conducted in developing countries for periods between 2 and 12 years, with sample sizes ranging from 1262 to 9774, and included evaluating interventions and exploring poverty, fertility, and child health and development. The percentage attrition at the end of the panel study ranges from 3 percent (Vietnam) to 52 percent (Peru). These studies used standard face-to-face methods, which are known to have lower attrition rates.

[5] Although there is an extensive literature on the results of panel surveys, the reporting on response rates is remarkably deficient. Johnson and Owens (2003) reviewed 95 audited journal articles and found that only 21 percent provided full documentation of sample dispositions. For the 48 reports on surveys conducted by telephone this percentage dropped to 12.5.

[6] Sykes et. al. (2010) show that even ethnicity of the enumerator affects response rates significantly in the United States.

Armstrong (1977) reviews 16 published studies on mail surveys and found that the part of the sample responding to the first wave ranged from 10 to 75 percent, with a median of 42 percent. In a literature review of response rates for email surveys conducted by Shonlau et. al. (2003) response rates are reported to be as low as 6 percent in some studies but could reach 68 percent in others. Lynn and Kaminska (2011) find only few and small differences in survey response rates between mobile phone interviews and fixed phone interviews. The few differences that they do offer suggest that data quality may be higher with mobile phone interviews, perhaps because survey respondents have greater control over whether other people are within earshot and whether others can listen in from another line.

Dillman (2009) reports on measurement differences and response rates in mixed-modes of data collection, which are believed to improve response rates. De Leeuw (2005) reported that the use of a second or even a third mode of data collection may improve response rates and may also improve coverage, and Lesser et. al. (1999) indicates that response rates to general public surveys can be increased from 15 to 20 percentage points by the inclusion of an incentive.

Nord and Hopwood (2007) assess the comparability of data collected by landline telephones and in person for the US food security set of questions, which is the same as one of the modules used in the L2L questionnaire, and find that the data are comparable: “Response patterns to the indicators that comprise the food security scale were, for practical purposes, invariant between modes” (p. 1478). Interestingly, they also hypothesize that this result may be partly attributed to the name recognition of the interviewing agency and that mode effects may be different depending on the respect that the agency conducting the survey commands on respondents. The L2L pilot found some evidence of this (see section V).

A Pew Research Center study (2012) examines the pattern of response rates for telephone surveys over time for surveys conducted by various institutions and finds that response rates have declined considerably for all types of survey topics, economic, social, political, media and others, both in the United States and abroad. For household surveys done by the Pew Research Center itself, the response rate (percentage of households sampled that yielded an interview) declined from 36 percent in 1997 to 9 percent in 2012.

Finally, it should be noted that a lower response rate is not necessarily an indicator of lower survey quality. Visser et. al. (1996) show that surveys with lower response rates produced more accurate measurements than surveys with higher response rates. Similar findings are reported in Keeter (2006) and Curtin et. al. (2000). Understanding the reasons for non-participation and the characteristics of drop-outs is important. For the L2L project this is done in Section V.C.

III. Project concept and design

The pilots were designed in a way that allowed testing the response rates and the quality of data, while also providing some information on the cost of collecting data using mobile phones. Moreover, we also wanted to make sure that the method worked for the types of questions that interested us regarding poverty and vulnerability, since we foresaw the L2L model being used to provide a valid way to collect high-frequency data immediately following shocks in developing countries, so that authorities can make decisions based on this evidence.[7]

For these reasons, during the pilot we decided to interview households instead of individuals. Another reason to study households and not individuals is that, unlike a

[7] It is possible that mobile phone surveys produce good quality data for some types of survey questions, such as those typically asked in marketing research, but not for household survey questions aimed at eliciting poverty and vulnerability, which are typically more sensitive.

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.[8] 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 in Peru and the Gallup World Poll sampling framework in Honduras.

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 surveys designed to track welfare over time, we also used this initial survey to recruit the panel.

We were particularly interested in studying welfare impacts of crisis in two segments of the population: (i) the vulnerable population, loosely defined as those households that 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 to minimize 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 received no financial incentive (the control group).

[8] 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.

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.

III.A Technological choices

The first decision was 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 (see figure 1). Further, 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 internet can be very challenging. In contrast, mobile phone coverage is already very high in Latin America and the Caribbean (see Table 1) and it keeps increasing (see figures 1, 2 and 3). So the first decision was to collect high-frequency data using cellular phones. [9]

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

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

* = 2011 data

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

[9] 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.

Figure 1: Internet and mobile use in Latin America and the Caribbean

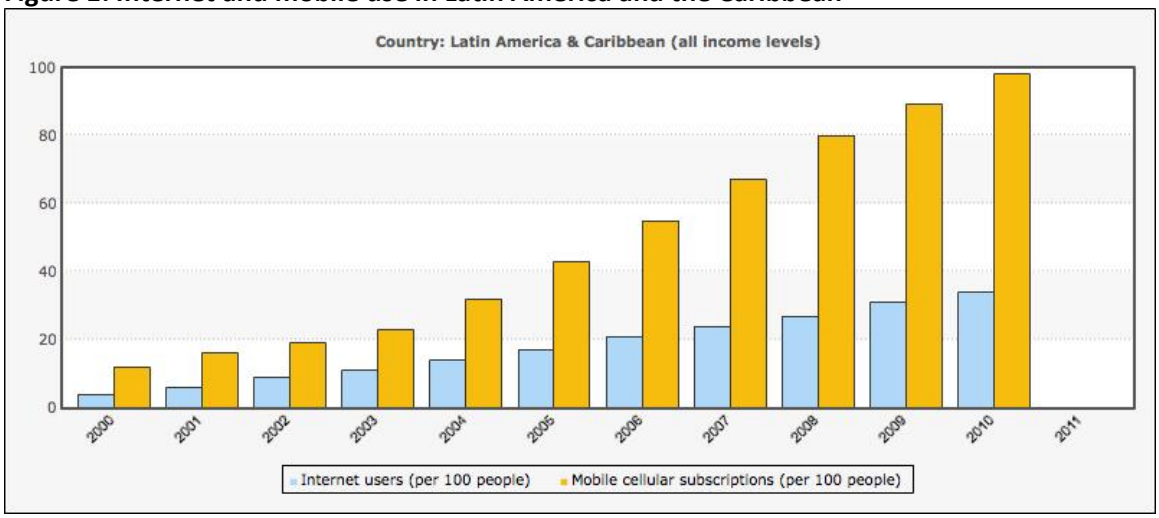


Figure 2: Internet and mobile use in Honduras

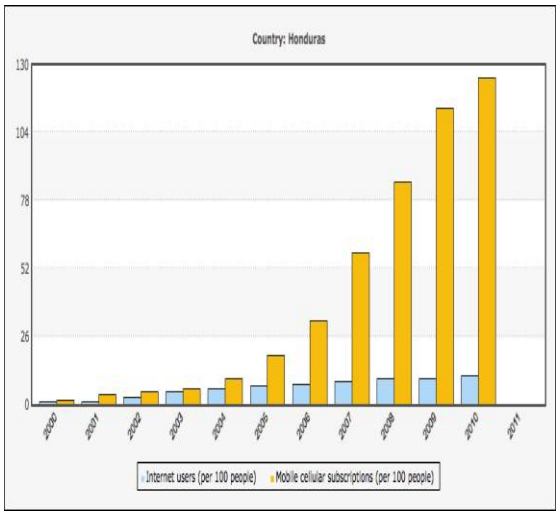
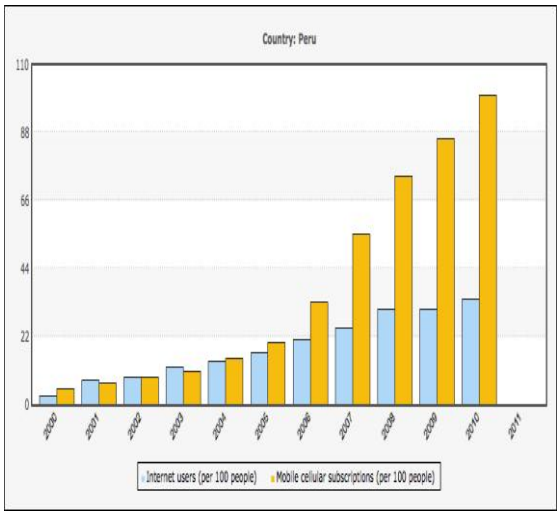


Figure 3: Internet and mobile use in Peru



Source: World Bank Databank (www.data.worlbank.org)

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 particularly difficult to reach (though not necessarily because the interviewers were using mobile phones)[10]. 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.

These are two related considerations: the characteristics of the mobile phone and cellular coverage. The characteristics of the different modes of communicating between enumerators and respondents, and their advantages and disadvantages for the purpose of collecting survey data, are summarized in table 2.

[10] This difficulty was encountered for 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).

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

	Audio/ Text	Self-Reporting	Pros	Cons
SMS (Short Message ..)	Text	Yes	Low cost	Maximum 160 characters Requires literacy Does not allow visual aids
IVR (interactive Voice Recognition)	Audio	Yes	No need for operators	Often viewed as annoying Medium cost Does not allow visual aids
CATI (Computer -Assisted Telephone Interview)	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. Does not allow visual aides
USSD	Text	Yes	No length limitations.	Requires close collaboration and approval by telecom companies. Not commonly marketed in LAC Does not allow visual aids
Mobile internet	Text	Yes	- No length limitations. - Lower cost than voice communications. - Allows use of visual aids	Limited GPRS coverage in LAC. Requires smart phones.

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 in Honduras used the three remaining communication technologies (SMS, IVR and CATI). But the survey designs (sample segmentation and contact frequency) were deliberately different. Annex 1 provides a

graphic representation of the two designs. In Peru, households were randomly assigned to a communication mode (SMS, IVR, CATI), which stayed constant for all rounds (waves) of the survey. In Honduras, all the survivor group of households (the households that responded to the first questionnaire) was exposed to all three communication modes, as explained in Box 1. The questionnaires were worded exactly the same way, regardless of the mode, which meant short questions, since SMS is limited to 160 characters, unlike IVR and CATI. The Honduras design was a test-retest design, which is closely related to the difference-in-difference methodology of experimental evaluation, and which allows conducting the reliability tests as explained in Box 1 in Annex 1.

III.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.[11] 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 -$0, $1 or $5- which were distributed after completion of each mobile survey. Unfortunately, mobile payments are not very developed in Latin America,[12] so instead of money transfers the pilot transferred the equivalent in free air-time minutes to each respondent's mobile phone account. When the respondent used a pre-paid phone, which do not allow to credit air time, the free air time was credit to other monthly phone in the household. On average, respondents received this incentive within 72 hours of completing each monthly survey.

In order to ensure that the economic incentives were being noticed by panelists who responded the surveys, upon survey completion the each panelist received a text message notification thanking them for their participation, indicating the amount of

[11] A generous donation from Brightstar Corporation made this possible.
[12] See, for instance: http://mobilereadiness.mastercard.com/the-index

airtime credited, and encouraging them to keep participating in the upcoming surveys. This notification was sent via text message regardless of the methodology group (SMS, IVR or CATI) the panelist was assigned to.

As part of the survey administration process Gallup implemented a number of mechanisms to maximize the response rate and panelist retention. Based on previous experiences Gallup expected the majority of the responses to occur within 24 hours after the surveys were sent (in the case of SMS surveys) and in a coincidental manner (at the time of the first call attempt) in the case of IVR and CATI surveys. However, it was anticipated that many panelists would need to be re-contacted. Therefore, in order to stimulate a response from panelists who failed to respond within this time frame, Gallup implemented the following actions:

- The surveys were left open for responses for up to 2 weeks after the original transmission of the survey (from original call in the case of IVR and CATI).
- First reminder was sent within 72 hours of first attempt (SMS and IVR).
- Second reminder was sent within 144 hours of first attempt (SMS and IVR).
- Call backs were made within 72 and 144 hours of first attempt (CATI); or
- Up to 2 call backs were made per appointment with respondent (CATI).

If the panelist failed to respond after the two week period and re-contact attempts, Gallup closed the survey for that month and sent (call for) a new survey the following month. If the panelist failed to respond to this subsequent survey attempt within 24 hours (SMS)/ failed to answer the phone (IVR and CATI), Gallup implemented the following actions:

- Send an immediate “urgent” reminder indicating that they missed last month’s survey but that we want to keep them as panelists and to hear from them this

time around. Gallup repeated that if they responded they could earn their free airtime (in case panelists belonged to the two groups that received incentives).

- Send/call for the scheduled 72 hour reminder; and
- Send/call for the scheduled 144 hour reminder

Non-responsive panelists who failed to respond to four consecutive monthly surveys were placed on a "Drop out List" for further panelist recovery actions.

In Peru, Gallup also attempted to re-contact and recover all panelists who dropped from the panel. However, panelists who responded to any contact attempts with a hard refusal (i.e. strong statements indicating they wish to terminate their participation), were excluded from the recovery list and no further re-contact attempts were made. Panelists who were re-contacted and refused to continue taking part in the panel exercise, but did not respond with a "hard refusal" were interviewed in order to learn about the reasons for their decision to terminate their participation (see Section V.C). Recovery attempts for drop out panelists from the SMS and IVR groups were performed via phone using human interviewers. Drop out panelists from the CATI group were contacted for recovery attempts by means of SMS (first time) and IVR (second time). No more than two recovery contacts were made per participant.

III.C Sample design and power analysis

The sample size was 1,500 households in each country. 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 (INEI), the L2L sample was based on the sampling frame for the household survey (ENAHO) conducted by the national statistical agency (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 vulnerable. For the purposes of this project, "close to poverty line" was defined as 40 percent of consumption distribution that symmetrically band 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, those households whose monthly per capita consumption falls 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 not the most vulnerable in times of crisis. Further, 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 households below but close to the poverty line (who may escape poverty in case of positive shocks). Therefore, in Honduras, the sample was drawn proportionately across the country by population size. The Honduras sample is only representative at the national level. Power analyses were also done for each country. These and additional technical details for the sample selection in each country are presented in Annex 3.

III.D Questionnaire design[13]

This section describes the process to select the questionnaires for: (a) the initial face-to-face survey in Peru, (b) the initial face-to-face survey in Honduras, (c) the monthly mobile surveys, and (d) the final face-to-face surveys in Peru and Honduras.

(a) Initial face-to-face (F2F) surveys

In Peru, the starting point was the ENAHO 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, using the latest household survey deployed by the Honduran Statistics Institute, except that only best predictors of income were chosen, because Honduras did not have a recent consumption aggregate.

(b) Monthly 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. In addition, two questions sought to ensure the validity of the responses by testing if the respondent was a member of the household. To accomplish this, the first two questions in each monthly questionnaire asked the respondent for their gender and year of birth, and the answers were compared to the household roster obtained during the face-to-face interview.

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 are likely to change in times of crisis. To simplify the

[13] Please see www.worldbank.org/lacpoverty/l2l for copies of the questionnaires and related materials.

questionnaire and avoid 'recency' effects[14] 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 (see section…)

IV. Pilot implementation

IV.A Field work

Through a competitive bidding process, Gallup won the contract to implement L2L on the ground. Working with affiliates in Peru and Honduras, Gallup successfully completed the initial face-to-face interviews, monthly or bi-monthly surveys using mobile phones, and a final face-to-face interview of a sub-group of the initial sample. Gallup completed the face-to-face interviews as well as the IVR and CATI interviews in-house, and subcontracted the SMS interviews to an SMS aggregator.[15]

By and large, the project was completed as envisioned, with a few exceptions. First, a number of restive communities in Peru did not allow Gallup's interviewers to enter the area. Where possible, these were replaced following INEI's standard methodology.[16] Second, despite extensive planning by the World Bank and Gallup teams, no SMS messages went out during the week of April 23 due to an error by the SMS aggregator, which simply forgot to program the system to send out messages that week. However, in this case, the messages went out the next week and the response rates did not suffer as a result. Other problems may have occurred during this project, only some of which

[14] 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).

[15] Training materials for the Gallup interviewers are available in the website provided in footnote 14.

[16] These replacements were chosen using the methodology employed by the Peruvian Institute of Statistics (INEI, its acronym in Spanish) for its ENAHO survey. When confronted with a problem in a particular location, INEI moves to the next "Centro Poblado" in the same "Conglomerado."

are definitively knowable. For instance, survey respondents may have lost their mobile phone, or it may have broken; other respondents may have been out of mobile phone range or their phones may have lost battery power and were not re-charged until after the question period was over. There may also have been problems with the system of the mobile phone operator, causing lost calls or messages. A final face-to-face interview asked respondents about these issues and the results are reported in section V.C.

IV.B Costs

The implementation of both face-to-face and mobile phone surveys also provided actual cost data for implementing each method. These are summarized in table 3.

Table 3: Costs for a sample of 1500 surveys (in US dollars)

Methodology	Cost Per Interview	Cost Per Year*
Face to Face	40	720,000
CATI**	25	450,000-720,000
IVR	17	306,000
SMS	8	144,000

* Assuming a monthly survey for 12 months.
** The CATI cost in Honduras is more expensive than other countries (about US$40 per interview), because the telecommunication cost is very high. In the LAC region, the cost distribution is more similar to Peru, which is the amount shown in this table.

IV.C Ethical issues

As with any survey involving human subjects, this project raised a number of ethical issues, such as data disclosure (privacy) and obligation for action. For instance, if an interviewer learns that a family is suffering from hunger: Should the government authorities be alerted? The L2L project team raised these issues with its advisory board (within the World Bank). The consensus was that the ethical issues raised by the L2L project were no different than those confronted by countless other survey projects. The only real difference is that mobile phone surveys gather data nearly instantaneously,

which could allow for a quicker redress of serious problems but any face-to-face interview also must confront this problem. To deal with anonymity issues, the World Bank team followed the guidance of the survey industry, which includes not disclosing any personal identifying information. Accordingly, Gallup delivered already anonymized data to the Bank.

V. Pilot Results [17]

The main interest of this section is not analyzing the answers to the survey questions per-se, but 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. There is no intention in this report to make inferences about the Peruvian or Honduran population in terms of welfare, education, health or other questions asked, based on the answers to the L2L survey questions.[18]

V.A Panel attrition rates

(a) Characteristics of attrition rates in the Peru pilot

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.

Table 4: Overall Attrition in Peru

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

[17] This section is largely based on Gallup's reports, available at http://www.worldbank.org/lacpoverty/l2l, which present the definitions of the terminology used in the following results reporting.

[18] Although there is nothing wrong with doing these inferences, given the probabilistic nature of the sample, and we hope other researchers will do so.

The post-recruitment response rate was higher among urban, relatively affluent households and more educated panelists (see Table 5). While 73 percent of lower income panelists (households 20 percent below or above the national poverty line) who completed the face-to-face interview and accepted to join the panel did not participate in the first wave of the survey, the non-response rate among more affluent households was 58 percent. Furthermore, the non-response rate among residents of urban areas was 57 percent from the initial face-to-face interview to wave 1, and 78 percent among residents of rural areas.

However, with each wave the gap between poor and more affluent households narrowed. When comparing those respondents who still participated in wave 6 with those who gave their approval to be contacted in the initial Face to Face interview, we see that the level of attrition among those with low household income was 76 percent, close to that of more affluent households with 73 percent. The differences in the level of attrition between city dwellers and residents of rural areas also decreased with each wave. However, in wave 6 the level of attrition of residents of rural areas was still 10 percentage points higher than among city dwellers (80 percent vs. 70 percent, respectively).

Table 5 also shows that in Peru, respondents with lower education were less likely to participate in any of the 6 waves than were those with higher levels of education. Female respondents were slightly more likely to participate in wave 1 than were male panelists (the level of attrition was 64 percent vs. 69 percent, respectively). However, this difference disappeared with consecutive survey rounds. Heads of household were less likely to participate in all waves of the survey than others (spouse, children, grandchildren etc.). This finding can probably be explained by the fact that head of households are usually older -and perhaps busier and/or less familiar with new technologies- than younger respondents. The mean age for heads of household was 45 years, and for others 33 years. The study shows that respondents older than 45 years were clearly less likely than younger respondents to participate in all waves of the study.

Final attrition rates remained higher among older respondents, those from rural areas, and those with lower levels of education.

Table 5: Attrition by demographics in Peru

	Poorer	More affluent	Urban	Rural
Wave 1	**73%**	**58%**	**57%**	**78%**
Wave 2	72%	61%	60%	76%
Wave 3	73%	65%	62%	77%
Wave 4	71%	68%	64%	76%
Wave 5	73%	71%	68%	77%
Wave 6	**76%**	**73%**	**70%**	**80%**

	Gender		Relationship with head of household HH	
	Male	**Female**	**HH**	**Other**
Wave 1	**69%**	**64%**	**72%**	**61%**
Wave 2	69%	66%	70%	65%
Wave 3	69%	69%	72%	67%
Wave 4	72%	68%	72%	66%
Wave 5	74%	70%	76%	68%
Wave 6	**75%**	**74%**	**78%**	**71%**

	Age			Education		
	15-30	**30-45**	**Older than 45**	**Lower**	**Average**	**Higher**
Wave 1	**59%**	**67%**	**72%**	**80%**	**64%**	**56%**
Wave 2	65%	66%	71%	75%	68%	57%
Wave 3	66%	68%	74%	79%	68%	61%
Wave 4	68%	67%	74%	77%	68%	63%
Wave 5	71%	70%	75%	78%	72%	65%
Wave 6	**73%**	**72%**	**78%**	**81%**	**75%**	**67%**

(b) Characteristics of attrition rates in the Honduras pilot

As in Peru, the post-recruitment response rate in Honduras was higher among women, city dwellers, panelists who do not consider themselves poor,[19] panelists who are not the head of the household, 15-30 year-olds, and those with average and higher education (see Table D on the following page). For example, while 47 percent of panelists with lower levels of education that completed the face-to-face interview and

[19] Please note that in the case of Honduras, we look at self-perceived poverty – in contrast to Peru.

accepted to join the panel did not participate in the first wave of the survey, the non-response rate among those with higher levels of education was 26 percent.
However, among most groups these gaps narrowed and some differences even reversed when looking at final attrition rates (from initial face-to-face until the end of the panel study). Only in the case of those who consider themselves poor versus those who do not did the same gap in response rates persist.

At the end of the Honduran panel study the non-response rates of panelists from rural areas, men and those with lower levels of education were very similar when compared with their socio-demographic counterparts. For example, 51 percent of those with lower levels of education who had originally agreed to join the panel failed to participate in wave 2, compared to 48 percent of those with higher levels of education.

Interestingly, despite lower attrition among 15-30 year-olds for the first survey wave (37 percent vs. 44 percent of others), the youngest panelists were less likely than others to participate in the second wave (with attrition reaching 55 percent among 15-30 year-olds vs. 47 percent of others).

A similar pattern was seen when comparing the non-response rates of heads of households versus those of other household members. While, the non-response rate for panelists who are the head the of household was 43 percent for the first wave, compared to 38 percent of other household members – a mirror image was seen for non-response rates for wave 2: 47 percent of heads of households vs. 54 percent of others.

Table 6: Attrition by demographics in Honduras

	Consider himself/herself poor	Does not consider himself/herself poor	Urban	Rural
From F2F to Time One	42%	38%	35%	45%
F2F to End of Panel	51%	46%	49%	50%

	Gender		Relationship with head of household	
	Male	**Female**	**Head of household**	**Other**
From F2F to Time One	44%	39%	43%	38%
F2F to End of Panel	51%	49%	47%	54%

	Age			Education		
	15-30 year-olds	**30-45 year-olds**	**Older than 45 years**	**Lower education**	**Average education**	**Higher education**
From F2F to Time One	37%	43%	44%	47%	30%	26%
F2F to End of Panel	55%	47%	47%	51%	48%	48%

(c) Comparison between Peru and Honduras

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.

Table 7: Overall Comparison of Initial/Final Attrition in Peru and Honduras

	Peru	Honduras
From F2F to Time One/Week 1 for Honduras	67%	41%
From F2F to End of Panel Study	75%	50%

There are several plausible explanations as to why the Honduran panel performed better than the Peruvian one in terms of panelist response and retention: (i) transportation of paper surveys to the central processing center was more expeditious, allowing for a shorter time gap between panelist recruitment and administration of the first follow-up surveys, it is likely that some Peruvian panelist forgot about the survey and disregarded the invitations to participate in the successive survey waves, (ii) Gallup's field operations partner in Honduras (CID-Gallup) is a very well-known and trusted firm in that country, less so Peru's fieldwork partner - it is possible that panelists in Honduras felt more confident responding to the follow-up surveys than their Peruvian counterparts, (iii) panelists in Honduras were surveyed more frequently and for a shorter period of time (four and a half months vs. six months in Peru) than their Peruvian counterparts and were also surveyed through different means (IVR, SMS and CATI), so it is possible that they became more aware and expectant of the follow-up surveys (this could partially explain the lower final attrition of Honduran panelists, but it cannot account for the higher response rate to the first follow-up surveys we saw in Honduras), and (iv) fieldwork in Peru took place in December, while in Honduras it happened in February and March, and December is not an ideal month to conduct surveys in LAC.

Panelists from urban areas in Honduras were more likely to participate in Time One than those from rural areas (the respective non-response rates were 35 percent and 45 percent). However, this difference was not observed in Time Two (see Table 8). In contrast, in Peru this gap still existed at the end of the study (70 percent vs. 80 percent, respectively).

Table 8: **Attrition by Level of Urbanization in Honduras**

	Level of Urbanization	
	Urban	Rural
Initial F2F to Wave 1 (weeks 1-6)	35%	45%
Initial F2F to Wave 2 (weeks 12-16)	49%	50%

V.B Attrition by data collection mode

(a) Attrition by data collection mode in Peru

The Peruvian L2L study clearly produced a lower non-response rate for CATI when compared to IVR and SMS, as Table 9 shows. Over the course of the 6 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 (81percent). 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.

Table 9: Attrition by mode in Peru

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

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[20]. 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").

(b) Attrition by data collection mode in Honduras

As Tables 10 and 11 below reveal, all three survey modalities showed lower levels of initial and final attrition in Honduras when compared with the Peruvian figures. In both countries, CATI surveys generated the lowest attrition in "Time One" and "Time Two", followed by SMS and IVR. However, in the case of Honduras "Time Two" attrition levels increased quite significantly for CATI while remaining stable for SMS and IVR. Still, attrition for CATI was much lower than for the other two modes throughout the whole study in both countries, especially in Honduras.

Table 10: Initial Attrition/Non-Response by data collection mode (Peru vs. Honduras)

Initial F2F to Time 1/ (Week 1 for Honduras)	**Peru**	**Honduras**
IVR	80%	60%
SMS	70%	55%
CATI	49%	12%
Overall	67%	41%

Table 11: Final Attrition/Non-Response by Mode (Peru vs. Honduras)

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

[20] 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.

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. Like in Peru, CATI surveys generated the lowest attrition, but the difference was more pronounced in the case in Honduras.

Table 12: Initial Attrition/Non-Response by Mode (Peru vs. Honduras)

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

Table 13: Final 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%
Overall	**75%**	**50%**

V.C Impact of Incentives

(a) Attrition by incentive level in Peru

Economic incentives in the form of mobile phone credit for every completely answered survey did not seem to have a big effect on the post-recruitment response rate in Peru (see Table 14).

Table 14: Attrition by incentive level in Peru

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

However, as the panel exercise progressed, incentives had some effect on minimizing attrition. With each wave, the level of attrition increased for all 3 groups, with the biggest increase being registered among panelists who received no economic incentive at all. In wave 6, the level of attrition for this group was 80 percent compared to 73 percent among those who received 1 dollar worth of mobile phone credit and 71 percent of those who received 5 dollars' worth of mobile phone credit per completed survey. 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).

Other findings include:

- Receiving a phone was a better incentive than the free air time incentive.
- Giving a phone to a household did not change the CATI nor IVR response rates, but increased SMS response rates considerably. These could be explained if most people receiving a phone were assigned to SMS group, but this was not the case.
- Most people who received phones were in the 0 incentive group.

(b) Attrition by incentive level in Honduras

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. Among those who did not receive any incentive to participate, 45 percent failed to participate in the first

wave of the survey, in comparison with 41 percent of those who received one dollar worth of mobile phone credit and 38 percent of those who received five dollars.

When fielding the second wave the gap in non-response between those who did not receive any incentive and those who received one dollar worth of phone credit narrowed (54 percent vs. 52 percent). Yet, the difference between these two groups and those who received five dollar increased, with the non-response rate of the latter being 43 percent for the second wave. Therefore, in contrast to Peru the size of the incentive mattered in the Honduran study, with higher incentives being more effective in minimizing attrition.

It is worth noting that Honduran panelists had to work harder than their Peruvian counterparts. The former had to answer up to three surveys per month, while the latter only answered one survey per month.

Table 15: Attrition by Incentive Level in Honduras

	No incentive	1 USD	5 USD
From F2F to Time One	45%	41%	38%
F2F to End of Panel	54%	52%	43%

V.D Understanding attrition

Panelists in Peru who did not participate in any of the six monthly waves were asked (by phone) for the reasons why they had not done so, despite their agreement to participate in the initial F2F interview. This was an open-ended question and the most common responses were technological (lack of phone signal or not receiving the survey or damaged phone accounted for 28 percent of the responses), not understanding the questions or not knowing how to answer them (20 percent), lack of time (13 percent),

and loss/theft of phone (12 percent). Also, some respondents thought that they were charged when answering the survey (7 percent). Surprisingly, only 2 percent reported problems charging the phone, although some of the above answers regarding technology problems may include this.

These answers point to the importance of trying to re-contact households several times and improving the simplicity of the questions, as well as the credibility of the promise that participation entails no cost to respondents.

Among those who had participated in all monthly surveys in Peru (n=169) 95 percent said that they would participate in a project like this in the future. Of those who responded some monthly surveys (n=260) 73 percent indicated their willingness to participate in the future. The corresponding proportion among panelists who had not taken part in any of the monthly surveys (n=271) was 51 percent.

Table 16: **Reasons for Not Responding Among Panelists Not Participating in Any of the Six Waves in Peru**

(Could you tell me the reasons why you could NOT answer ANY of the questions we sent you?	Frequency	Valid Percent
Did not have phone signal	62	15%
Did not have time/was busy	54	13%
Did not receive the surveys	52	13%
Does not have this cell phone anymore (loss/theft)	51	12%
Did not know how to answer	44	11%
Did not understand the questions	38	9%
Cell phone was damaged	34	8%
Was charged when answering the survey	29	7%
Other reason	17	4%
Did not want to answer	15	4%
Had problems with electricity/cell phone could not be charged	8	2%
Answered all surveys received/does not believe to have skipped some	4	1%
Questions all seemed the same	3	1%
Don't know	1	0%
Total	412	100%

Besides the post-exercise telephone surveys, Gallup conducted face to face closing surveys among 700 panelists. As part of this survey, panelists were asked what would motivate them to keep on participating in a project like this in the future. The single most frequently mentioned motivation to participate in the future was an economic incentive either in monetary form, cell phone credit or a gift (20 percent). However, when adding up all the answers related in some way to the "project's mission", the motivation that this project will improve people's lives in the long-term seems to be significantly more important for panelists than one-time monetary incentives. These obviously include 14 percent of mentions of people who hoped that the project improved people's or their own living standards and the 13 percent who would be motivated if they could report about the socio-economic reality of people's lives.

However, it seems that many other frequently named motivations are related as well to this hope for better future lives. For instance, some panelists would be encouraged to participate if their opinion was heard (8 percent), others wished to be asked questions that would be more targeted towards the panelists' lives (6 percent) and some said they would be motivated if the study's result were published (5 percent). Moreover, some panelists stated that they would like to take part in future surveys if they were a continuation of this World Bank study (5 percent). Others would be motivated if questions were simpler and instructions clearer (5 percent), cell phone credit gifts were higher (4 percent) and more frequent updates about the project were given (2 percent).

Table 17: Top 10 Motivations to Take Part in the Future Panel Surveys (Peru)

What would motivate you the most to keep on participating in a project like this in the future? I would be motivated if I…	Frequency	Valid percent
Received monetary incentive/ credit/ gift	105	20%
Project improves people's or their own living standards	75	14%
Could report about socio-economic reality of people's lives	69	13%
My opinion would be heard	42	8%
The questions asked more about my life	34	6%
Results were published	28	5%
Project would be this World Bank study	27	5%
Survey questions were simplified, instructions clarified	24	5%
Received higher cell phone credit	21	4%
Received frequent updates about project	13	2%
Other motivations	87	17%
Total	525	100%

V.E Data quality

This section addresses the following questions related to two dimensions of 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, 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.[21]

(a) L2L Comparison with standard benchmark

To answer the first question the team applied a criterion validity test. Criterion validity refers to the 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

[21] This question was analyzed only for the Honduras pilot.

Face survey was adopted as the criterion measurement for the analyses presented in this report. In the case of Peru, the National Household Survey (ENAHO) was used as the criterion variable. The World Bank was granted access to the most recent micro data from this survey. The questions asked by SMS surveys, which were identical to those asked by the Face to Face survey, were used as the test measurement. Since the SMS sample was affected by a high level of attrition – fifty-five percent of participants who originally agreed to join the panel did not respond to the first SMS survey sent to them – for the purpose of this analysis only households that responded to both surveys (45% of the sample) were included. This analytic decision was made 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).

In order to test the criterion validity of the SMS measurements in Honduras, 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. the possession of TV, and sanitary infrastructure), factual information on access to the internet inside or outside the household, and perceptual information (i.e. whether the respondents considered themselves poor).

Table 18 shows that responses to all questions by SMS differ from those collected via Face to Face by at least 7.4 percent points, a margin that is statistically significant at 95% confidence level. Interestingly, the responses given via SMS significantly underestimate facts regarding household infrastructure, while over-estimating Internet access and self-perceptions on poverty. Gallup (2012, pages 35 and 36) presents a number of hypothesis that may explain these results.

Table 18: Comparative results SMS vs. F2F in Honduras (percent responding "yes")

	F2F (only households that answered question in SMS)	SMS	Difference (F2F-SMS)
Do you currently have a TV at home?	87.9	72.6	15.3
Is the property or house equipped with plumbing for water?	98.7	86.5	12.2
Does your house have any type of sanitary/bathroom facilities?	96.5	88	8.5
Do you have access to internet from somewhere outside your home, such as work, school, internet café or room, or library?	19.5	35.1	-15.6
In the last 30 days, have you had access to internet thorough any available computer, or not?	17.4	28.9	-11.5
Do you consider yourself as poor?	65.3	72.7	-7.4
When you were 15 years old, do you think you and your parents were poor?	69.2	77.6	-8.4

Similar criterion validity tests were performed for IVR and CATI. The results are presented in tables 19 and 20.

Table 19: Comparative results IVR vs. F2F in Honduras (percent responding "yes")

	F2F (only households that answered question in IVR)	IVR	Difference (F2F – IVR)
Do you currently have a TV in your home?	86.4	75.6	10.8
Is the property or house equipped with plumbing for water?	97.1	84.4	12.7
Does your house have any type of sanitary/bathroom services?	97.1	88.1	9
Do you have access to internet from somewhere outside your home, such as work, school, internet café or room, or library?	19.7	34.3	-14.6
In the last 30 days, have you had access to internet thorough any available computer, or not?	20.5	29.3	-8.8
Do you consider yourself as poor?	68.3	75	-6.7
When you were 15 years old, do you think you and your parents were poor?	69.9	77.4	-7.5

Table 20: Comparative results CATI vs. F2F in Honduras (percent responding "yes")

	F2F (only those who answered question in CATI)	CATI	Difference (F2F – CATI)
Do you currently have a TV in your home?	83.2	84.7	-0.9
Is the property or house equipped with plumbing for water?	97.7	97.7	0
Does your house have any type of sanitary/bathroom facilities?	96.4	96.8	-0.4
Do you have access to internet from somewhere outside your home, such as work, school, internet café or room, or library?	14.7	16.3	-1.6
In the last 30 days, have you had access to internet thorough any available computer, or not?	12.7	12.9	-0.2
Do you consider yourself as poor?	72	73.9	-1.9
When you were 15 years old, do you think you and your parents were poor?	72.4	74.5	-2.1

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, while the items related to "Internet access" and "self-perceptions on poverty" received higher scores. Like in the case of SMS, the observed differences between IVR and Face to Face are statistically significant. The answers collected via CATI (in Table 19), 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 between SMS, CATI and IVR, or between any of these and Face to Face,[22] cannot be attributed to demographic differences between them.

(b) Test-Retest Reliability Results

In order to test the reliability of SMS measurements, we conducted two identical SMS measurements of the same questions analyzed in the criterion validity analysis

[22] While the analyses presented in tables 18, 19 and 20 are theoretically based on the same panelists, some panelists failed to respond to some questions in one or more modes. This explains the slight differences in the Face to Face responses across tables.

discussed above. For comparative purposes, Gallup performed repeated administrations of these questions by means of Face to Face, IVR and CATI on the same group of panelists.[23] In all cases, the repeated measurements were performed within a minimum of 10 weeks from the first administration. Table 21 shows the results of the test-retest analysis performed by computing a Cronbach Alpha reliability coefficient for each survey method in Honduras.[24]

Table 21: Test-Retest Reliability for SMS in Honduras

	n	Percent "Yes" Time 1	Percent "Yes" Time 2	Pearson Correlation	Cronbach Alpha
Do you currently have a TV at home?	158	72%	73%	0.74	0.87
Is the household equipped with plumbing for water?	156	89%	87%	0.65	0.79
Does household have sanitary/bathroom facilities?	152	89%	88%	0.58	0.74
Do you have access to internet from somewhere outside your home?	153	33%	32%	0.61	0.76
In the last 30 days, have you accessed the internet, or not?	153	24%	29%	0.54	0.70
Do you consider yourself poor?	153	76%	76%	0.40	0.57
When you were 15 years old , do you think you and your parents were poor?	151	81%	82%	0.58	0.73
Total Reliability					**0.74**

Overall, the SMS measurements seem to have been quite consistent, as shown by the "yes" scores collected at "time 1" and "time 2" for each question. The Cronbach Alpha

[23] The actual number of panelists for each mode varies due to differences in attrition rates across modes.

[24] Cronbach's alpha is 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. The formula for the standardized Cronbach's alpha is: $\alpha = \frac{N \cdot \bar{c}}{\bar{v} + (N-1) \cdot \bar{c}}$, where N is equal to the number of items, c-bar is the average inter-item covariance among the items and v-bar equals the average variance (SPSS FAQ). Only the Honduras survey design allows a reliability test.

scores also suggest a very good level of overall reliability (0.74).[25] 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 on poverty. Tables 22 to 24 show the test-retest reliability analysis for IVR and CATI, as well as the comparative Cronbach Alpha coefficients for all three methodologies.

Table 22: Test-Retest Reliability for IVR in Honduras

	n	Percent "Yes" Time 1	Percent "Yes" Time 2	Pearson Correlation	Cronbach Alpha
Do you currently have a TV at home?	146	75%	74%	0.88	0.93
Is the household equipped with plumbing for water?	137	88%	87%	0.77	0.87
Does household have sanitary/bathroom facilities?	141	87%	87%	0.78	0.88
Do you have access to internet from somewhere outside your home?	139	35%	32%	0.71	0.83
In the last 30 days, have you accessed the internet, or not?	136	29%	29%	0.65	0.79
Do you consider yourself poor?	135	79%	77%	0.72	0.84
When you were 15 years old, do you think you and your parents were poor?	134	79%	83%	0.84	0.91
Total Reliability					**0.86**

[25] The Cronbach Alpha reliability coefficient obtained in an identical test-retest analysis performed with the Face to Face method was quite close (0.77). Face to Face was held as the benchmark methodology in this study.

Table 23: Test-Retest Reliability for CATI in Honduras

	n	Percent "Yes" Time 1	Percent "Yes" Time 2	Pearson Correlation	Cronbach Alpha
Do you currently have a TV at home?	411	87%	73%	0.50	0.65
Is the household equipped with plumbing for water?	**411**	**99%**	**91%**	**0.38**	**0.55**
Does household have sanitary/bathroom facilities?	411	96%	92%	0.49	0.65
Do you have access to internet from somewhere outside your home?	411	16%	28%	0.69	0.81
In the last 30 days, have you accessed the internet, or not?	411	12%	19%	0.79	0.86
Do you consider yourself poor?	409	73%	82%	0.51	0.68
When you were 15 years old, do you think you and your parents were poor?	409	74%	83%	0.46	0.63
Total Reliability					**0.69**

Table 24: Test-Retest Reliability for IVR, SMS and CATI in Honduras (Cronbach Alpha Coefficients)

	IVR	SMS	CATI	All Methods Combined
Do you currently have a TV at home?	0.93	0.87	0.65	0.93
Is the household equipped with plumbing for water?	0.87	0.79	0.55	0.89
Does household have sanitary/bathroom facilities?	0.88	0.74	0.65	0.91
Do you have access to internet from somewhere outside your home?	0.83	0.76	0.81	0.92
In the last 30 days, have you accessed the internet, or not?	0.79	0.70	0.86	0.89
Do you consider yourself poor?	0.84	0.57	0.68	0.91
When you were 15 years old, do you think you and your parents were poor?	0.91	0.73	0.63	0.92
Total Reliability	**0.86**	**0.74**	**0.69**	**0.91**

As can be seen in the tables above, 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.

It is also interesting that both IVR and CATI, outperformed SMS in those items that inquire about personal internet access, which could be explained by the pattern observed in the criterion validity analysis, where SMS surveys were most often responded 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 phones keypad 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 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 understand in a more satisfactory manner.

Importantly, for all methodologies the "yes" responses were quite consistent (as shown by Tables 21 to 23 above), 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.

V.E Multivariate analysis

In order to understand if there are systematic differences in household characteristics between responding and non-responding households, we conducted a multivariate analysis where the probability of completing the survey (1 or 0) is regressed against a number independent variables, such as the survey mode (IVR, CATI or SMS), urban or rural location, education, gender incentive level, whether the household has children and/or elders, an asset index representing a proxy for household wealth,[26] and others.

The results of the Honduras multivariate analysis are presented in annex 4. They show that age of the head of household, household composition and wealth are all statistically significant variables affecting the response rate. Households were most likely to complete mobile surveys if the primary respondent in the initial face to face interview or if the head of the household was between 31 and 40 years old. These households were statistically significantly more likely to respond than households where the primary respondent in the initial face to face interview or if the head of the household was 30, though there were no significant differences in response rates where the age was over 40. Household composition also matters, though not necessarily in an expected way: households with children *and* elderly are more likely to complete surveys (in six models, statistically significant at the 1 percent level in one, the 5 percent level in two, and the ten percent level in one.) Just having kids in the household, or just having elders, has no impact on response rate. Wealthier households are more likely to complete surveys, though response rates do not increase monotonically with wealth (the eight decile was most likely to respond in all six models, the tenth (richest) decile ranked third in four of six models).

[26] This index is not the same as the household income variable in the previous section. In the previous section household income corresponds to the average income in the household sample cluster as per the ENAHO, whereas in this section we construct a wealth index from the answers to our baseline survey questions on access to electricity, water, transport and similar items, using principal components' analysis.

Households in which men responded to the initial face-to-face interview were less likely to complete the mobile phone questionnaires (this finding was significant at the ten percent level in two of six models). Households in which the primary respondent in the initial face-to-face interview knew how to text using a mobile phone were more likely to respond to the mobile phone surveys – importantly, not only for SMS surveys but for CATI and IVR as well, which do *not* involve texting (in six models, this finding is significant at the ten percent level in three and at five percent in one).

The principal findings regarding mobile mode of communication in Honduras indicate that households only exposed to SMS respond at a lower rate than those exposed to SMS, CATI and IVR (however, this finding was statistically significant at the ten percent level in one of two models). The order in which households were exposed to SMS, CATI and IVR did not impact response rates.

Incentives increase completion rates: US$5 of free airtime increases participation among participants using all three modes -SMS, IVR and CATI- (significant at the 1 percent level), while a US$1 incentive only improves the response rates of SMS users (significant at the 5 percent level).

Households in small communities were significantly less likely to respond to surveys whether or not controlling for location (urban/rural). However, controlling for the size of the community, urban residents were less likely to complete questionnaires. Not controlling for community size, there was no statistically significant difference between urban and rural households' response rates.

The results for Peru are also presented in annex 4. The principal finding regarding the mode of communication show that having a CATI operator is positively correlated with a higher take-up rate (acceptance to participate in the panel), **but only after controlling for other variables, in particular, wealth.** Knowing how to use SMS is highly correlated with higher take up rates. This indicates that it may not be desirable to assign survey

mode randomly. However, knowing how to use SMS affects the likelihood of responding equally, regardless of the survey mode assigned to the household.

There are no systematic differences in take-up rates between urban and rural households or between different asset classes. Knowing how to read and write (head of household) makes more likely that the household will accept to participate, but not so when one controls for geographical location. Having electricity increases the response rate over time. Controlling for everything else, the odds of staying in the panel are 3 times higher for households that have access to electricity.

VI. Conclusions and Lessons Learned

The main lesson is that it is possible to collect household survey data in a short period of time using cellular phones, but the system to do so must be in place before frequent data collection starts. 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 the socio-economic characteristics of each household, which cannot be done by mobile phone due to the large number of questions. For these reasons, the most desirable use of mobile surveys may be as a complement to on-going national surveys and, if the mobile phone number is collected as part of these, the time to set up the panel and conduct rapid mobile surveys can be considerably reduced.

A mobile phone survey does not have to be subject to sample selection bias. The project has shown that it is possible to select samples on the basis of either a national official sampling frame or other representative sample frame, rather than on the basis of phone records or random digit dialing, and use mobile phones subsequently for follow up short surveys.

Regarding the mode of communication with respondents, higher attrition rate and lower survey completion rate was found among panelists who were exposed to self-administered modes (IVR and SMS). Of those Peruvian panelists surveyed about their lack of response to the panel, 26 percent said they prefer to be interviewed by a person. Also, panelists responding the surveys via IVR or SMS showed a higher propensity to leave questions unanswered than did respondents answering via CATI. Interviewers are important for getting higher response rates and ensuring that respondents give consideration to all the survey questions.

The results of the L2L pilot program indicate that the SMS surveys performed quite satisfactorily in terms of generating reliable measurements, as part of a test-retest study. This conclusion is supported by the fact that the Cronbach 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. SMS failed the criterion validity test at 95 percent confidence interval. There are reasonable indications that this result arises from respondent switching behavior (different respondents answering the survey every month). If this is confirmed, the results would not reflect on the comparison between SMS and the benchmark (face to face), but would combine these differences with differences in the respondent. This suggests that allowing different household members to respond the frequent survey may not be a good option and additional efforts to identify the respondent are warranted.

Although CATI is the method that has the lowest non-response rate, it is also the most expensive of the three methods used in the L2L project. SMS has lower cost of the 3 methods explored, but resulted in twice the attrition rates as CATI in this project. A combination of both may is suggested to be explored in future research, in the same spirit as 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). Although, IVR responses proved the most reliable data in the test-retest the small differences in terms of reliability with the other modes suggests that IVR is not a very suitable mode to communicate with respondents (at least for these types of surveys).

Economic incentives did not seem to have led to a higher response rate in first re-contact surveys, but they did help containing attrition thereafter. Incentives do not need to be very large to make a difference, and their effect on attrition varies from country to country and from mode of data collection (SMS vs. IVR vs. CATI). The project showed that most attrition occurs between recruitment and first re-contact. Once panelists respond to the first survey, attrition is marginal and frequent contact and economic incentives seem to be important for minimizing it.

Higher attrition rates were found among older, less educated, less affluent panelists and households living in rural areas. 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.[27]

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

[27] For details on weighting schemes, please refer to "Baseline Face-to-Face Surveys in Honduras and Peru. Methodological Report," by Gallup.

electricity to re-charge phones, but it overcomes the security problems in approaching respondents in person 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 that this leads to more evidence-based policy decisions when governments and donors are confronted with sudden shocks.

LIST OF ANNEXES

ANNEX 1: Survey design – graphic representation

Table A1-1: Peru sequential design

Group	Oct 2011-Dec 2011	Date of F2F+1 week	Date of F2F+5 weeks	Date of F2F+10 weeks	Date of F2F+15 weeks	Date of F2F+20 weeks	Date of F2F+25 weeks	Date of F2F+35 weeks
1	F2F1	SMS1	SMS2	SMS3	SMS4	SMS5	SMS6	F2F2
2	F2F1	IVR1	IVR2	IVR3	IVR4	IVR5	IVR6	F2F2
3	F2F1	CATI1	CATI2	CATI3	CATI4	CATI5	CATI6	F2F2

F2F = Face-to-Face interview
The numbers refer to the number of a particular questionnaire, not a time period. If a questionnaire is repeated, it retains the same number, even though it happens at a later time.

Table A1-2: Honduras Test-Retest design

	Time 1											Time 2					
Group	Feb.13	Feb.20	Feb.27	Mar.5	Mar.12	Mar.19	Mar.26	Apr.2	Apr.9	Apr.16	Apr.23	Apr.30	May 7	May 14	May 21	May 28	June 4
1	F2F1	SMS1	IVR1	CATI1			SMS2	SMS3-A	SMS3-B	SMS4		SMS1	IVR1	CATI1		SMS2	F2F1
2		F2F1	CATI1	SMS1	IVR1		SMS2	SMS3-A	SMS3-B	SMS4			CATI1	SMS1	IVR1	SMS2	F2F1
3			F2F1	IVR1	CATI1	SMS1	SMS2	SMS3-A	SMS3-B	SMS4			IVR1	SMS1	CATI1	SMS2	F2F1
Extra 1	F2F1	SMS1					SMS2	SMS3-A	SMS3-B	SMS4			SMS1			SMS2	
Extra 2		F2F1	SMS1				SMS2	SMS3-A	SMS3-B	SMS4				SMS1		SMS2	
Extra 3			F2F1	SMS1			SMS2	SMS3-A	SMS3-B	SMS4					SMS1	SMS2	

* A household was invited to take part in a survey using each methodology at least twice during the study. The questionnaires for time 1 and time 2 were identical within and across methodologies.
* After the first face-to-face administration, each group was exposed to the remaining 3 methodologies according to a randomization scheme (3 rotations, one methodology per week).
* All households were interviewed face-to-face upon panel recruitment (and some at the very end of the study). Therefore, face-to-face could not be part of the random rotation scheme.
* Any additional household that remained in the panel was only interviewed via SMS (Groups Extra 1, Extra 2 and Extra 3 above).
*The data collection process was carefully controlled to ensure that all the groups within the sample were representative of the population.

Box 1: The Honduras Test-Retest design

One part of Honduran panel (n=600 households) was exposed to the three survey methods studied. They were also surveyed on a more frequent basis and for a relatively shorter period of time (four and a half months vs. six months in Peru). The surveys were administered in three time blocks, as follows:

Time One: This first administration took place between weeks 1 and 6 of the panel study. As part of "Time One," each panelist was surveyed three times. Each time was with the same exact questionnaire, but using a different survey method (SMS, CATI or IVR).
Time Two: This administration took place between weeks 12 and 16 of the panel study. As part of the "Time Two" block, each panelist was surveyed three times. Each time was with the same exact questionnaire used in Time One, and with the same three methodologies presented in the same order as in Time One.
In-Between: During the period between Time One and Time Two (weeks 7 through 11), panelists received four SMS surveys with questionnaires different than the ones used in "Time One" and "Time Two". However, for the purpose of this attrition analysis, only "Time One" and "Time Two" administrations are being considered.

The other part of the Honduran panel (n=900) was only interviewed by SMS after the initial face-to-face interview.

ANNEX 2: Literature Review on Using Mobile Phones to Conduct Research

Harold Alderman, et al., 2001. "Attrition in Longitudinal Household Survey Data," Demographic Research, Max Planck Institute for Demographic Research, Vol. 5(4), pp. 79-124.

Focusing on surveys that report the highest per-year attrition rates between survey rounds, this paper considers the extent and implications of attrition for three longitudinal household surveys from Bolivia (a childcare program), Kenya (informal networks and AIDS prevention), and South Africa (with a broad array of socioeconomic info). Their main finding was that, in contrast to often-expressed concerns about attrition, for many estimates the coefficients on standard variables in equations are unaffected by attrition; attrition apparently is not a general problem for obtaining consistent estimates of the coefficients of interest → multivariate estimates of behavioral relations may not be biased due to attrition, so the collection of longitudinal data for developing countries is not a bad idea

After conducting three sets of tests of attrition as they relate to observed variables in the data, the report estimates indicate that: (a) the means for a number of critical outcome and family background variables differ significantly between those who are lost to follow-up and those who are re-interviewed; (b) a number of family background variables are significant predictors of attrition; but (c) nevertheless, the coefficient estimates for standard family background variables in regressions and probit equations for a majority of the outcome variables considered in all three data sets are not affected significantly by attrition. Other conclusions include: (i) Neither family background variables nor outcome variables measured in the first of two surveys reliably predict attrition in multivariate probits; (ii) Attrition does not generally significantly affect the estimates of the association between family background variables and outcome variables, and, (iii) Attrition apparently is not a general problem for obtaining consistent estimates of the coefficients of interest for most of the child development outcomes in

the Bolivian data, for the fertility/social network outcomes in the Kenyan data, and for some of the anthropometric indicators in the South African data.

Sarah Baird, et al., 2008. "Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)," CIDER Working Paper, University of California-Berkeley.
The paper reports the results of a longitudinal survey of 7,500 Kenyan children who had received de-worming drugs in the western Kenyan district of Busia during seven-year period (1998-2005) of very high migration. It describes and analyzes data quality issues, such as difficulties in tracking respondents and high survey attrition, enumerator recording errors (data entry errors) and enumerator bias, respondent reporting error, and respondent recollection mistakes. A comparison between surveys in developing countries that track migrants and those that did not suggests that following movers is critical to reducing sample attrition. Furthermore, the substantial monetary cost associated with tracking KLPS-1 movers was worthwhile since migrants have substantially different characteristics than non-migrants, and these appear to interact with de-worming treatment in non-trivial ways. For this study, tracking paid off.

Focusing on issues with enumerators, the extent of enumerator error was low, with an average of less than one recording error per survey. Errors decrease over time as enumerator experience with the survey instrument increases; they increase over the course of multiple interviews within a single day, presumably due to fatigue. Also, more educated enumerators tend to make somewhat fewer data errors, while females on the author's team committed slightly more errors. Some evidence indicated that the enumerator-respondent match in terms of gender, ethnicity, and religion correlated with responses regarding trust of others and religious activities, suggesting some field officer bias on sensitive questions.

Overall, enumerator bias and survey completion mistakes were minimized in the KLPS-1 due to intensive enumerator training and continuous quality checks throughout data collection. Other data quality issues concerned the presence of an individual in the room

in addition to the enumerator, as well as respondent and retrospective questions and the problem with respondent recall error, with the elapsed time of recalling events opening up the possibility of errors and thus bias in statistical analysis. It does appear that having someone else present at the time of enumeration potentially affects survey responses, reinforcing the importance of consistently holding private interviews, and suggesting it may be useful to control for the presence of other individuals at the interview. Concerning retrospective questions, the authors show that although survey responses relating to the respondent's own characteristics are highly reliable, reports on the outcomes of other family members (in this specific study, parents) are far less reliable, with less educated respondents having the least consistent reporting. These findings are worthy of serious consideration by development economists who utilize LSMS-style household surveys, which are often conducted among poorly educated respondents and where one household member typically reports information (on consumption expenditures, among other variables) for many other household members.

Peter Lynn and Olena Kaminska, 2011. "The Impact of Mobile Phones on Survey Measurement Error," Working Paper 2011-07, Institute for Social and Economic Research (University of Essex).

The authors present a theory of the ways in which mobile phone interviews and fixed phone interviews may differ and how these differences may affect survey measurement. They identify four key features that may vary between mobile and fixed phone interviews. These are: line quality; the extent of multi-tasking amongst survey respondents; the extent to which survey respondents are distracted from the task of answering questions, and the extent to which other people are present and able to overhear what the survey respondent is saying. Their findings, based on a sampling of people residing in the Greater Budapest region and consisting of 724 people, with 342 interviews achieved on fixed phones and 202 on mobile phones, are taken from a randomized experiment, carried out in 2005 as part of a research program by European Social Survey and Gallup-Europe.

Overall, the authors find only few and small differences in survey measures between the mobile phone interviews and the fixed phone interviews. The few differences that they do offer suggest that data quality may be higher with mobile phone interviews, perhaps because survey respondents have greater control over whether other people are within earshot and whether others can listen in from another line. When other people can hear the responses being given—which may be more likely when responding on a fixed line—respondents may have a tendency to censor their responses to avoid socially undesirable answers. The authors' findings suggest that measurement error differences between interviews carried out on a mobile phone and those carried out on a fixed phone are small. Of the differences that they find, most are in the direction of indicating less social desirability bias with mobile phone interviews.

Brian Dillon (2010) – "Using Mobile Phones to Conduct Research in Developing Countries," Cornell University Working Paper: 1-20.

The paper describes and analyzes the ongoing experience of a research project [Research on Expectations in Agricultural Production (REAP)] that uses mobile phones to collect detailed demographic, agricultural and economic data from rural households in Tanzania on a high frequency basis. The sample was composed of a total of 195 farmers drawn from 15 villages in northwest Tanzania; villages were chosen at random from among the primary cotton-growing districts; within each village, 20 sample farmers were interviewed over-the-phone from September 2009 to July 2010, once every 3 weeks on a pre-arranged day. Everyone in the sample was given a cell phone. Prior ownership of a mobile phone did not exclude participants from receiving an additional phone from the project. Response rate for the study has been very high (an average of 192 of the 195 respondents has been interviewed in the first 6 waves.

Problems with phone surveys include selection bias (b/c in developing countries mobile phone registries are non-existent, very incomplete, and/or likely to be unavailable to researchers); non-response; patchy mobile network coverage could introduce bias into the selection of research villages, with the availability of power another consideration:

in the study, the researchers made arrangements with a "charging station", where survey participants can charge their phones.

On balance, the experience of the REAP study suggests that phone-based enumeration of relatively complex economic surveys in low-income countries is not only feasible, but also, under the right circumstances, superior to traditional face-to-face enumeration. Relative to a traditional survey, the cost savings of a phone survey are substantial, as long as the questions of interest call for repeated enumeration of the same households. In addition, the centralized nature of phone-based data collection allows for rapid detection and correction of errors, interactive participation by the primary researchers in real time, and streamlined data entry. The limits to the feasibility of phone-based research appear to be those imposed by the usefulness of high frequency data collection, the sensitivity of the questionnaire material, and the coverage area of the available mobile phone networks.

Kevin Croke, et al., 2012. "Collecting high frequency panel data using mobile phones. Does timely data lead to accountability?"

As African mobile phone ownership rates have risen, interest in using mobile telephony as a data collection platform has increased. This paper proposes the creation of nationally representative (panel) surveys that ask mobile phone questions every two weeks and whose (anonymized) data is publicly released within four weeks after their collection. It draws on two pilot projects in Tanzania and South Sudan that use mobile phone interviews for data collection. The survey in Tanzania has been running longer (33 rounds/waves to date), while the survey in South Sudan is one that operates under the more difficult conditions. The experience was largely a success. High frequency panel data have been collected on a wide range of topics in a manner that is cost effective, flexible (questions can be changed over time), and rapid. After responding to the mobile phone interviews, households tend not to drop out: even after 33 rounds of interviews in the Tanzania survey, respondent fatigue proved not to be an issue. In the Tanzania survey, attrition and non-response have been an

issue, but in ways related to the way the survey was originally set up and that are fixable. Both surveys are still running.

In South Sudan, 1,000 respondents in 10 urban areas were selected by mobile survey selected respondents and handed out mobile phones (half of them with integrated solar chargers) and called respondents on a monthly basis using a call center operating from Nairobi using interviewers capable of speaking South Sudan's main languages. Respondents who successfully completed an interview were rewarded with an amount varying from $2 to $4. In Tanzania, the 550 households in Dar es Salaam were not given phones; only recently, after round 33, were some phones distributed to respondents who had never before participated. Households were assigned one of four technologies: IVR, USSD (an approach allowing direct transmission of questions from a phone company server to the respondent's phone; this technology also works on low-end phones), WAP (Web-based mobile phone surveys, suited for high-end phones with Internet capability) and voice (respondent is called by a person at a call center). After a relatively short period of time (7 rounds), voice became the technology of choice (the others were facing too many problems) and all who were reachable and accessible to mobile phones were put through the call center.

The authors found economic status to be a significant predictor of survey participation: households without a phone, those using non-premium phone providers and those in the second poorest income quintiles were significantly less likely to participate relative to households of median wealth. At the same time data has been collected on a wide range of topics in a manner that is cost effective, flexible (i.e. questions can be changed over time) and rapid. And once households are included in the survey, they tend to stick with it: respondent fatigue is not an issue. Attrition and non-response were an issue in the Tanzania survey, but in ways that are related to the how the survey was originally set up and fixable. An important conclusion is that, while selection can never be controlled based on unobservables, re-weighting based on observables should be a standard procedure after every survey round in a mobile phone survey. While years of

education, gender and even the amount given as reward do not explain non-response, the evidence from the Tanzania and Sudan surveys suggests that mobile surveys can collect quality data in a very timely way that is of use to a wide range of data users. The Tanzania survey pointed towards the importance of putting in place mechanisms that avoid attrition and non-response right from the implementation of the baseline. Much attrition in the Tanzania survey can be explained by choices made in the survey's organization (such as not distributing mobile phones). In both, it was found that once households are included in the mobile survey, they are likely to remain in the survey: respondent fatigue was not found to be an issue. The research also suggests that because of their high frequency, quality control of mobile surveys is dynamic in that issues identified in one round can be corrected in the next. Finally, when interviews are lengthy, face-to-face interviews are probably more cost effective.

Don Dillman, et al., 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.

The study examines the potential for improving response rates by changing from one mode of data collection to another. It evaluates the sequential use of modes rather than offering respondents a choice of modes by which to respond. It also discusses the consequences for measurement and non-response errors with changing from one mode to another. Data collection from 8999 households was done in two phases. Phase 1 data collection was conducted by telephone interview, mail, interactive voice response, or the Internet; Phase 2 focused on non-respondents to Phase 1, and was done by a different mode, either telephone or mail. The results of the study suggest that switching to a second mode is an effective means of improving response. It found that for the satisfaction-dissatisfaction questions, respondents to the aural modes (telephone and IVR) were significantly more likely than were respondents to the visual modes (mail and Web) to give extreme positive responses, a difference that cannot be accounted for by a tendency towards recency effects with telephone. In general, switching to a second

mode of data collection was not an effective means of reducing non-response error based on demographics.

The study found little doubt that a mixed-mode strategy of following a complete data collection strategy by one mode, with a short pause, followed by an attempt to collect data by another mode, can increase response rates substantially. If one's intent is to measure change, however, switching modes could have major negative consequences for measures of satisfaction. Using an alternative mode that depends upon a different channel of communication, i.e., aural vs. visual, to increase response may introduce measurement differences issues that cannot be ignored. The study also showed that non-responding households differed from those that responded. In each case, non-responding households had sample frame characteristics of significantly lower education than did the responding households. Non-responding households tended to be less educated, were more likely to be female, and were younger than those responding to the survey. Income differences—the only household variable—was an exception with only one significant difference (for the IVR treatment) between respondents and non-respondents.

Other findings include: (a) The telephone interview brought in respondents with somewhat less education and smaller households with higher incomes; (b) Mixed-modes of data collection are believed to improve response rates. De Leeuw (2005) reported that the use of a second or even a third mode of data collection may improve response rates and may also improve coverage. However, mixed mode surveys could potentially produce different answers to the same questions, such as more positive responses to scale questions on telephone than on Web surveys (Dillman and Christian, 2005; Christian et al., 2008); (c) Past research has shown that response rates to general public surveys can be increased from 15 to 20 percentage points by the inclusion of an incentive (Lesser et al., 1999). In addition, research on the use of incentives with telephones has suggested that incentives contingent upon completion of the interview may be less effective than unconditional incentives sent ahead of time (Singer, 2002);

(d) Telephone response rates have declined significantly in recent years and coverage problems are increasing as noted by Singer (2006). Mail coverage remains a concern for general public surveys but response rates seem not to have suffered the large decline experienced for the telephone, and (e) Experiments have shown that respondents to surveys are more likely to offer socially desirable answers and to demonstrate acquiescence in the presence of an interviewer than in the self-administered situation (de Leeuw, 1992, 2005; Schuman and Presser, 1981).

ANNEX 3: Sample design

A. Peru sample

Since the fourth trimester of ENAHO 2010 is used as a sampling frame, a brief description of ENAHO sampling design is in order.

ENAHO sample design

The ENAHO sample is selected in three stages in urban and in more densely populated rural areas and in two stages in less densely populated areas. In the first stage, selection of the Primary Sampling Units (PSUs) occurs. All PSUs are grouped in 8 strata, defined by the size (number of residents). Strata 1 through 5 correspond to urban areas, strata 6 through 8 correspond to rural areas. The selection occurs within department-specific strata, and the probability of selection of PSUs is proportional to the number of households in them.

The second stage depends on which of the two large categories a PSU belongs to. The first category is comprised of urban PSUs and more populous rural PSUs. The second category consists of less populous rural PSUs. For simplicity, this document refers to the first category as “urban” and the second as “rural”.

In urban category, Secondary Sampling Units – conglomerados - are selected from each PSU with the probability proportional to their size, and with implicit stratification. Implicit stratification is based on a number of socio-economic variables. In rural category PSUs consist of one conglomerado (SSU) only. For these PSUs, the second stage is omitted.

At the third stage, households are randomly selected from SSUs. In rural areas 8 households are randomly drawn from each SSU; in urban areas 6 households per SSU are selected.

ENAHO survey includes a panel component. The SSUs (conglomerados) of the panel component are selected randomly (from the entire sample of ENAHO SSUs). In order to

be able to compare L2L data to the future rounds of ENAHO at the SSU level, L2L sample will be drawn from panel conglomerados only.

L2L sampling procedure

There are multiple ways to oversample households 20 percent above and 20 percent below. One can oversample households from SSUs with higher proportion of households of interest (i.e. households which belong to the band from 7th to 47th percentile of income distribution). Alternatively, it is possible to oversample PSUs with higher proportions of households of interest. The number of households drawn from each SSU will be the same as in ENAHO (6 households per urban SSU and 8 per rural SSU). In order to reduce design effects (i.e. maintain greater number of clusters), the second option was chosen.

The L2L sample frame comprises all the panel conglomerados from the fourth trimester of ENAHO 2010, or 281 conglomerados. The sampling procedure involved the following steps:

(1) Additional stratification (based on the proportion of households from 7th to 47th income percentile). All PSUs are divided into two strata: PSUs where the proportion of households from 7th to 47th percentiles of the income distribution exceed 40 percent – stratum I, and PSUs where this proportion is less than 40 percent – stratum II. These two strata are treated as separate samples. 60 percent of L2L sample (or 900 households) will be drawn from stratum I, and 40 percent (or 600 households) from stratum II. Due to the size of the strata and the number of households to be drawn from each, further steps of the sampling procedure will differ depending on the stratum.

(2a) For stratum I: The group of panel conglomerados visited during the fourth trimester of ENAHO 2010, where the households from 7th to 47th percentile of national income distribution exceed 40 percent is small: there are only 888 households in this group (Table 1). Therefore, all the conglomerados from stratum I will be included into L2L sample.

Table A3-1: L2L strata (based on 7th-47th percentiles of income distribution)

	L2L stratum I		L2L stratum II	
ENAHO stratum	Number of conglomerados	Number of households	Number of conglomerados	Number of households
			48	288
2	9	54	46	276
3	3	18	19	114
4	11	66	15	90
5	17	102	16	96
6	6	48	5	40
7	55	440	8	64
8	20	160	3	24
Total number of households:		888		992

(2b) For stratum II: As Table 1 shows, there are 992 households in all the conglomerados in the second stratum. Only 600 of them will be included into L2L sample. The selection procedure consists of two stages: PSU selection, and SSU selection.

(3) PSU selection. 60 percent of PSUs from stratum II will be selected: 60 = (600/992)*100. 60 percent of PSUs will be drawn with probability proportional to size from each of 8 ENAHO strata. Table 2 shows the number of PSUs selected from each stratum.

Table A3-2: PSU selection from L2L stratum II

ENAHO stratum	PSUs selected	Total number of PSUs
1	3	4
2	8	12
3	6	9
4	6	9
5	8	12
6	4	5
7	5	8
8	2	3

Note: some of the PSUs are selected twice (in the strata where there is high variation in the size of PSUs)

(4) SSU selection is different in rural and in urban areas.

(4a) SSU selection – rural areas. In the rural PSUs (strata 6 – 8), none of the selected PSUs included more than 1 conglomerado. Consequently, there is no need for (or possibility of) SSU selection in rural areas. Following INEI practice, 8 households will be randomly drawn from each of rural PSU/SSU. Given that there are 11 selected rural PSUs, and 8 households will be drawn from each, L2L sample will include 88 rural households.

(4b) SSU selection – urban areas. The number of SSUs (conglomerados) per PSU in urban strata varies highly: from 1 to 32. 512 households[5] will be selected; all urban PSUs treated as a single stratum, according to probability proportional to the size of SSU. Prior to selection, all SSUs will be sorted according to socio-economic index, as in ENAHO sampling, to ensure implicit stratification along socio-economic dimension.

(5) Household selection: 6 households will be randomly selected from each urban SSU, and 8 households from each rural SSU. It is important to note, that in the PSU where the difference in the size of SSUs is high, some of the SSUs were selected more than once. In this case, the number of households drawn from such SSUs was multiplied by the number of times the SSU was selected. Similarly, in the strata where the difference in the size of PSU is high, some PSUs were selected more than once. In this case, a PSU was included twice in the sample from which conglomerados were drawn. As a result of these operations the number of households to be drawn from different conglomerados ranges from 6 to 24.

Calculation of the expansion factors

The expansion factors for the households in the L2L sample are calculated as

FL2Lspch = 1/Pfinspch, where Pfinspch denotes the probability of selection into L2L sample.

Pfinspch is calculated as

Pfinspch = P01spc*Pcongspc*PL2Lspch, where

P01spc is the probability of selection of a conglomerado into the 4th trimester of the panel component of ENAHO 2010. As selection into panel and non-panel components, as well as interview trimesters is random, this probability is equal to 0.082 (the number

of panel conglomerados in the fourth trimester of ENAHO 2010/all conglomerados in ENAHO 2010).

Pcongspc is the probability that a conglomerado c was selected into ENAHO 2010. It is equal to p1*p2 for urban areas and rural areas of strata 7 and 8 (if muesusmdptoregmuesmaest is not equal to missing in the sampling sheet), and equal to p1 for stratum 6, and strata 7 and 8 if muesusmdptoregmuesmaest is equal to missing).

PL2Lspch = P3L2Lspch for households in L2L stratum 1;

PL2Lspch = P1L2Lspc * P3L2Lspch for rural households in L2L stratum 2;

PL2Lspch = P1L2Lspc * P2L2Lspc * P3L2Lspch for urban households in L2L stratum 2;

P3L2Lspch – probability of selection of a household h from conglomerado c;

P2L2Lspc – probability of selection of a conglomerado c (from all PSU selected in (3));

P2L2Lspc – probability of selection of a PSU p from stratum s.

Sample validity checks

In order to check the quality of the sample, ENAHO 2010 survey is used. However, due to the differences in the sampling procedure, the number of households to be selected for L2L sample exceeds the number of households interviewed in ENAHO 2010 by approximately 20 percent. While L2L sample consists of 1522 households, only 1212 households were interviewed in L2L conglomerados (or 79 percent of the L2L sample). This discrepancy probably accounts for the fact that weighted number of all ENAHO observations in L2L conglomerados (using L2L weights) is only 22,535,360 – only 76 percent of the population of Peru.28

[28] For a list of selected cluster in Peru please visit http://www.worldbank.org/lacpoverty/l2l.

B. Honduras sample

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 are not the most vulnerable in times of crisis. 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 the close to non-poor (who can escape poverty in case of positive shocks). Therefore, in Honduras, the sample was drawn proportionately across the country by population size. The steps were:

1) Post-stratification weighting
   a. Using data supplied by Gallup's Honduras vendor, I created targets for the distribution of the age of the head of household and the highest level of education of the head of household.
   b. We used a SAS post-stratification raking macro in order to adjust the household level data so that it aligned with the target on head of household age and education.
   c. Note: there was no selection probability correction base weight as the sample was drawn proportionately across the country by population size, unlike Peru which had an oversample by income.
2) Trimming and renormalizing
   a. I trimmed the weights so as to balance bias (how close the demographics of the weighted data align to the targets) and efficiency (the size of the design effect from weighting)
   b. I then renormalized the data so that the sum of the weights equals the number of observations

Table A3-3: Honduras - Face to Face weighting results

Age group	Unweighted Frequency	Unweighted Percent	Weighted Percent	Target percent	Target - Weighted Percent
15-24	101	6.69	9.327	9.3273	-0.000314
25-34	324	21.46	23.4414	23.4421	-0.000673
35-44	411	27.22	24.1671	24.1672	-0.0001
45-54	330	21.85	18.8084	18.8082	0.0002221
55+	344	22.78	24.2561	24.2552	0.0008647
Education group	**Unweighted Frequency**	**Unweighted Percent**	**Weighted Percent**	**Target percent**	**Target - Weighted Percent**
Less than Primary	310	20.53	27.6525	27.6525	0.00
Primary	818	54.17	51.9965	51.9965	0.00
Secondary +	382	25.3	20.351	20.351	0.00

C. Power analysis

Power analysis is the ability to find a statistically significant difference between sample groups when a real difference exists in the population. The power of a study is determined by three factors: the sample size, the alpha level, and the effect size.29 As the sample size increases, the power of the test increases because differences between the population groups can be detected by smaller effect sizes, or differences between the sample groups. Power was calculated for the three incentive groups, for the three modes of communications, and for the interaction between the two. The power calculations compared incentive groups within each data collection mode group in Peru and Honduras. As the sample sizes are nearly identical for each treatment group, US$1 or US$5, the results of the power analysis are the same for either treatment group used. The results show that, given two groups of 233 households each (half of the sample size in each group, after attrition) the minimum effect size for a statistical test, with power of approximately .8 and an alpha of .05, is 13.5 percent. The minimum-effect size given two groups of 133 households (if additional households drop out) is 18 percent.

The power calculations below (based on Lenth, 2006-9) compare incentive groups within each data collection mode group in Peru and Honduras. Within each data collection mode group, $0, $1 and $5 incentives are offered. For the sake of this exercise comparisons are made between the $0 incentive group and one other treatment groups. As the sample sizes are nearly identical for each treatment group, $1 or $5, the results of the power analysis are the same for either treatment group used. Again the significance level $\alpha = 0.05$ is used and the power was as close to .8 as possible given the software limitations.30 The results, presented in table…, show that, given two groups of 233 (half of the sample size in each group, after attrition) the minimum effect size for a statistical test, with power of approximately .8 and an alpha of .05, is 13.5 percent.

[29] See http://psych.wisc.edu/henriques/power.html.

[30] Java Applets for Power and Sample Size [Computer software]. Retrieved *August* 25, 2011 from http://www.stat.uiowa.edu/~rlenth/Power.

Under the same conditions, the minimum effect size given two groups of 133 (if additional households drop out) is 18 percent, between two groups of 200 is 14.5 percent and between two groups of 300 is 12 percent.

Table A3-4 Power test for incentives only

Peru SMS Group		
Treatment	Sample Size	Proportions
$0	233	0.465
$1 or $5	233	0.6
	Difference	0.135
	Significance Level (α)	0.05
	Power (1 - β)	0.8094

Peru IVR and CATI Groups		
Treatment	Sample Size	Proportions
$0	133	0.465
$1 or $5	133	0.645
	Difference	0.18
	Significance Level (α)	0.05
	Power (1 - β)	0.8119

Honduras SMS/IVR/CATI Group		
Treatment	Sample Size	Proportions
$0	200	0.465
$1 or $5	200	0.61
	Difference	0.145
	Significance Level (α)	0.05
	Power (1 - β)	0.8043

Honduras SMS Group		
Treatment	Sample Size	Proportions
$0	300	0.465
$1 or $5	300	0.585
	Difference	0.12
	Significance Level (α)	0.05
	Power (1 - β)	0.818

Comparisons between modes of communication, disregarding incentive groups, were made for each of the countries. In Peru this would be between the IVR, CATI and SMS groups. The IVR and CATI groups both have an initial n size of 400, while the SMS group has an initial n size of 700. In Honduras this would be between the groups interviewed with the three modes versus the group contacted only by SMS. The former has an initial

n size of 600 while the latter has an n size of 900. The significance level α= 0.05 is used and the power was as close to .8 as possible given the software limitations. The results, shown in Table … ., indicate that, given two groups of 400 households, the minimum effect size for a statistical test, with power of approximately .8 and an alpha of .05, is 10.2 percent. Under the same conditions, the minimum effect size given one group of 400 and another of 700 is 9 percent, and between one group of 600 and another of 900 is 7.5 percent.

Table A3-5: Power tests for modes only

Peru IVR vs. CATI Groups		
Treatment	Sample Size	Proportions
IVR	400	0.5
CATI	400	0.602
	Difference	0.102
	Significance Level (α)	0.05
	Power (1 - β)	0.8089

Peru IVR or CATI vs. SMS Groups		
Treatment	Sample Size	Proportions
IVR or CATI	400	0.5
SMS	700	0.59
	Difference	0.09
	Significance Level (α)	0.05
	Power (1 - β)	0.8071

Honduras SMS, IVR or CATI vs. SMS Groups		
Treatment	Sample Size	Proportions
IVE, CATI or SMS	600	0.5
SMS	900	0.575
	Difference	0.075
	Significance Level (α)	0.05
	Power (1 - β)	0.801

In order to compare both incentive groups and mode groups simultaneously, a two-way ANOVA was performed to test if there was a difference between any of the incentive groups, any of the mode groups and between the groups formed by the interaction of the two, while taking both treatments and their interaction into account. This is done by computing an F statistic, which tells us the variability in the dependent variable between levels of one treatment divided by the variability in the dependent variable not explained by any of the treatments (or their interactions). So this can be thought of as the variance of the dependent variable explained by the treatment in relation to the variance of the dependent variable that cannot be explained at all. The more variance in the dependent variable explained by the differences between the levels of the treatment, the more likely that there are actually differences in the population groups that these levels represent, i.e. that there are statistically significant differences in the mean of the dependent variable between the levels of the treatment group. For the purposes of this study it could be thought of as the variability in the mean years of education explained by the mode used divided by the variability in the mean years of education not explained by the mode, incentive group or the interaction between the two. In other words, is there a statistically significant difference between the mean number of years of education for each mode, taking into account the incentives used and the interaction between mode and incentive?

In order to calculate the power of such an ANOVA an effect size for the ANOVA must be calculated. In this case the effect size is similar to the F statistic but not identical. Instead of the variability in the dependent variable explained by the treatment in relation to the variability in the dependent variable not explained at all, the effect size is the variability in the dependent variable explained by the treatment in relation to the total variability in the dependent variable. Thus an effect size must separately be calculated for each treatment in the ANOVA aside from the default F tests performed.

In the Peru experiment there is the mode treatment with three levels and the incentive treatment with three levels, as well as the interaction between the two. Note that there are actually four modes, but all respondents receive the first mode thus removing it as a treatment level. Power is computed separately for each of these treatments and for their interaction. Given a minimum cell size of 133 the minimum effect sizes that would result in a statistically significant result are displayed below. Effect sizes of this kind less than or equal to 0.1 are considered "small" so the following results should be looked upon favorably.

Table A3-6: Power for interaction of mode and incentive level (cross-sectional look)

	Levels	N Size per level*	Power	Alpha	Minimum Effect Size
Mode	4	399	0.8	0.05	0.09555
Incentive	3	532	0.8	0.05	0.09525
Mode*Incentive	12	133	0.8	0.05	0.1309

**Note* that this assumes all cells to have the n size of the smallest cell for conservative calculation purposes

The above is simply a cross-sectional look at time one (or time two) of the experiment. This is actually a panel of respondents meaning that there are several data collection phases per respondent. With a panel comes attrition and so for a subsequent data collection phase, say T3, we will assume 50 percent attrition. This would reduce the per cell n-size to 66. Given that, the minimum effect sizes that would result in a statistically significant result are displayed below. As expected, the minimum effect size increases as the n size decreases, keeping the alpha and power constant.

Table A3-7: Power for interaction of mode and incentive level (panel look)

	Levels	**N Size per level***	**Power**	**Alpha**	**Minimum Effect Size**
Mode	4	198	0.8	0.05	0.1358
Incentive	3	264	0.8	0.05	0.1353
Mode*Incentive	12	66	0.8	0.05	0.1862

*Note that this assumes all cells to have the n size of the smallest cell for conservative calculation purposes

It is important to note that since this data is collected from a panel, comparisons within the same respondent over time are correlated. A comparison of T2 vs. T3 is not the same as comparing 1500 respondents at T2 to 750 new respondents at T3; they are the same 750 respondents in both time periods plus an additional 750 respondents at T2 who dropped from the panel. The effect of this is to reduce the power of the test as the effective sample size is reduced due to the correlation of the data.

The above is of particular note in Honduras, as the first sub-panel of respondents all received SMS, IVR and CATI. This cannot be analyzed as 600 SMS respondents versus 600 new IVR respondents versus 600 new CATI respondents. Independence of responses is an assumption of ANOVA and thus that test cannot be used. In this case a mixed-model would be used that treats the respondent as a random effect in order to account for the correlation among responses from each respondent. As this is a much more complicated regression based model there is no software to calculate the power of such a test, but it will be less than the power of a test without correlated responses. Despite the lowered power, the Honduran sampling plan allows each respondent to serve as their own control group.

ANNEX 4: Multivariate analysis for Honduras

A brief note about the six models: The first three models treat incomplete questionnaires (those in which some, but not all questions were answered) as though they had no responses at all (non-response). Models four, five and six treat incomplete questionnaires (those in which some, but not all questions were answered) as though they were fully completed questionnaires (complete).

Models one and four do not add controls for wave (time) effects. Models two and five add controls for wave effects among those households engaged in all three types of mobile surveys (CATI, IVR and SMS), while models three and six control for wave effects among those households engaged only in SMS surveys.

The six models were run both on the Head of Household as well as on the respondent during the initial face-to-face interview.

Results are shown below:

	Head of Household					
	model 1	**model 2**	**model 3**	**model 4**	**model 5**	**model 6**
	b/se	b/se	b/se	b/se	b/se	b/se
order: CATI-SMS-IVR	0.991	0.959		0.94	0.898	
	[0.157]	[0.156]		[0.148]	[0.144]	
order: IVR-CATI_SMS	0.986	0.895		1.024	0.929	
	[0.161]	[0.155]		[0.166]	[0.157]	
SMS only	0.717*			0.756*		
	[0.099]			[0.103]		
Incentive = $1	1.339*	1.151	1.727**	1.326*	1.145	1.716**
	[0.159]	[0.187]	[0.321]	[0.160]	[0.186]	[0.328]
Incentive = $5	1.949**	1.738**	2.282**	1.895**	1.647**	2.300**
	[0.232]	[0.285]	[0.409]	[0.227]	[0.270]	[0.426]
Male	0.835+	0.783	0.879	0.814+	0.739*	0.891
	[0.090]	[0.119]	[0.141]	[0.089]	[0.113]	[0.148]
Knowledge of how to send text messages	1.409**	1.634**	1.214	1.449**	1.559**	1.384+
	[0.167]	[0.259]	[0.232]	[0.174]	[0.249]	[0.271]
age under 18	0.429*	0.236**	0.717	0.458+	0.269*	0.75
	[0.175]	[0.132]	[0.357]	[0.184]	[0.141]	[0.367]
age 18-30	0.827	0.707+	0.942	0.798+	0.698+	0.888
	[0.110]	[0.133]	[0.186]	[0.106]	[0.130]	[0.174]
age 41-60	0.819+	0.818	0.792	0.829	0.793	0.853
	[0.097]	[0.132]	[0.145]	[0.097]	[0.127]	[0.157]
age over 60	0.818	0.896	0.702	0.882	0.864	0.837
	[0.179]	[0.278]	[0.222]	[0.197]	[0.278]	[0.269]
household has elders	0.757	0.753	0.691	0.696	0.877	0.58
	[0.269]	[0.339]	[0.377]	[0.257]	[0.416]	[0.337]
household has children	1.013	0.916	0.981	1.003	0.921	0.972
	[0.144]	[0.188]	[0.210]	[0.144]	[0.185]	[0.214]
household has elders and children	1.819	1.218	3.079*	2.076+	1.13	3.764*
	[0.673]	[0.576]	[1.726]	[0.777]	[0.550]	[2.186]
does not receive remittances	1.098	1.342	0.976	1.08	1.199	1.003
	[0.159]	[0.266]	[0.206]	[0.154]	[0.237]	[0.207]
Claro is mobile carrier	0.542**	0.641*	0.415**	0.514**	0.606*	0.403**
	[0.078]	[0.134]	[0.092]	[0.075]	[0.125]	[0.094]
Digicel is mobile carrier	2.115	1.284	5.063	1.513	1.761	1.411
	[1.330]	[0.866]	[5.848]	[0.695]	[1.207]	[1.148]

	Head of Household					
	model 1	**model 2**	**model 3**	**model 4**	**model 5**	**model 6**
	b/se	b/se	b/se	b/se	b/se	b/se
Urban	0.734*	0.691+	0.767	0.703*	0.694+	0.696+
	[0.102]	[0.132]	[0.162]	[0.100]	[0.138]	[0.152]
Asset Index: 2nd decile	0.975	0.907	0.703	1.049	0.938	0.88
	[0.242]	[0.305]	[0.319]	[0.264]	[0.306]	[0.426]
Asset Index: 3rd decile	1.425	1.127	1.881+	1.555+	1.189	2.203+
	[0.359]	[0.401]	[0.697]	[0.404]	[0.421]	[0.895]
Asset Index: 4th decile	1.633*	1.476	1.854	1.898**	1.69	2.141+
	[0.388]	[0.476]	[0.708]	[0.465]	[0.547]	[0.879]
Asset Index: 5th decile	1.446	1.146	1.788	1.711*	1.425	2.021+
	[0.337]	[0.371]	[0.643]	[0.406]	[0.453]	[0.784]
Asset Index: 6th decile	1.428	0.823	2.629**	1.638*	1.002	2.962**
	[0.349]	[0.281]	[0.942]	[0.406]	[0.341]	[1.132]
Asset Index: 7th decile	1.304	0.847	1.955+	1.508+	1.031	2.260*
	[0.299]	[0.276]	[0.677]	[0.355]	[0.337]	[0.842]
Asset Index: 8th decile	2.042**	1.502	3.037**	2.545**	1.826+	3.961**
	[0.476]	[0.495]	[1.049]	[0.604]	[0.591]	[1.485]
Asset Index: 9th decile	1.209	0.83	1.847	1.588+	1.099	2.478*
	[0.306]	[0.297]	[0.691]	[0.408]	[0.386]	[0.991]
Asset Index: 10th decile	1.234	1.083	1.539	1.619+	1.381	2.084+
	[0.315]	[0.397]	[0.576]	[0.419]	[0.494]	[0.843]
Cities 100k-499k	0.986	0.964	1.042	1.038	1.008	1.057
	[0.159]	[0.231]	[0.242]	[0.165]	[0.238]	[0.240]
Cities 50k-99k	1.01	0.938	1.095	0.985	0.955	0.989
	[0.201]	[0.278]	[0.304]	[0.197]	[0.284]	[0.274]
Towns 10k-49k	0.756+	0.678+	0.836	0.756+	0.715	0.767
	[0.124]	[0.159]	[0.205]	[0.124]	[0.166]	[0.191]
Towns under 10k	0.517**	0.463**	0.584+	0.540**	0.503*	0.559*
	[0.099]	[0.128]	[0.170]	[0.105]	[0.140]	[0.164]
less than primary education	0.826	0.89	0.689+	0.790+	0.855	0.662+
	[0.110]	[0.165]	[0.141]	[0.107]	[0.155]	[0.148]
secondary	1.365+	1.478	1.258	1.412*	1.4	1.357
	[0.227]	[0.371]	[0.299]	[0.234]	[0.352]	[0.317]
upper secondary	1.596**	1.385	1.751*	1.634**	1.313	1.921**
	[0.246]	[0.311]	[0.382]	[0.250]	[0.296]	[0.416]
higher (non-university) education	1.148	0.237+	1.541	1.326	0.297	1.856
	[0.617]	[0.178]	[0.876]	[0.689]	[0.246]	[0.984]

	Head of Household					
	model 1	**model 2**	**model 3**	**model 4**	**model 5**	**model 6**
	b/se	b/se	b/se	b/se	b/se	b/se
university or post-graduate	1.414	1.81	1.053	1.463	1.884+	1.114
	[0.358]	[0.661]	[0.392]	[0.359]	[0.653]	[0.420]
Has a job	1.069	1.057	1.082	1.158	1.156	1.125
	[0.122]	[0.173]	[0.186]	[0.135]	[0.191]	[0.200]
technology=ivr	0.142**	0.140**		0.099**	0.100**	
	[0.015]	[0.015]		[0.011]	[0.011]	
technology=sms	0.127**	0.132**		0.092**	0.105**	
	[0.013]	[0.015]		[0.010]	[0.012]	
Wave 2		0.696**			0.793+	
		[0.082]			[0.095]	
Wave 3		0.982			1.124	
		[0.125]			[0.141]	
Wave 4		0.560**			0.639**	
		[0.067]			[0.080]	
Wave 5		0.476**			0.412**	
		[0.057]			[0.055]	
Wave 6		0.829+			0.917	
		[0.094]			[0.110]	
Wave 7		0.688**			0.877	
		[0.082]			[0.104]	
Wave 8		0.455**			0.588**	
		[0.060]			[0.081]	
Wave 9		0.520**			0.701**	
		[0.068]			[0.093]	
Wave 10		0.638**			0.811+	
		[0.076]			[0.100]	
SMS only - wave 2			0.742**			0.859+
			[0.066]			[0.078]
SMS only - wave 3			0.932			0.852+
			[0.075]			[0.078]
SMS only - wave 4			1.014			0.954
			[0.082]			[0.089]
SMS only - wave 5			0.932			1.156
			[0.079]			[0.104]
SMS only - wave 6			0.855+			1.008
			[0.072]			[0.090]
Constant	3.490**	7.450**	0.255**	2.857**	5.633**	0.133**
	[1.298]	[3.926]	[0.134]	[1.072]	[2.912]	[0.074]

	Head of Household					
	model 1	**model 2**	**model 3**	**model 4**	**model 5**	**model 6**
	b/se	b/se	b/se	b/se	b/se	b/se
psuedo r^2	0.11	0.138	0.08	0.132	0.167	0.088
chi2	593.821	601.901	125.823	697.556	702.12	132.408
p	0	0	0	0	0	0
N	8951	4833	4104	8951	4833	4104

	Respondent in Initial Face-to-Face Interview					
	model 1	**model 2**	**model 3**	**model 4**	**model 5**	**model 6**
	b/se	b/se	b/se	b/se	b/se	b/se
order: CATI-SMS-IVR	1.023	0.943		0.955	0.865	
	[0.164]	[0.154]		[0.153]	[0.141]	
order: IVR-CATI_SMS	0.987	0.885		1.021	0.92	
	[0.161]	[0.150]		[0.166]	[0.152]	
SMS only	0.764+			0.799		
	[0.106]			[0.110]		
Incentive = $1	1.323*	1.162	1.745**	1.311*	1.157	1.745**
	[0.157]	[0.189]	[0.321]	[0.158]	[0.189]	[0.330]
Incentive = $5	1.912**	1.732**	2.231**	1.861**	1.646**	2.275**
	[0.227]	[0.285]	[0.398]	[0.223]	[0.273]	[0.414]
Spouse (or partner)	0.967	0.806	1.236	0.962	0.784	1.254
	[0.136]	[0.157]	[0.250]	[0.137]	[0.155]	[0.254]
Children from oldest to youngest	1.15	0.999	1.318	1.192	1.02	1.352
	[0.197]	[0.238]	[0.337]	[0.201]	[0.236]	[0.343]
Step-children from oldest to youngest	0.342	0.382		0.245	0.265+	
	[0.313]	[0.291]		[0.216]	[0.201]	
Parents	0.887	1.935		1.214	2.409*	
	[0.435]	[0.810]		[0.600]	[1.003]	
Siblings	0.556	0.418	0.775	0.483	0.31	0.702
	[0.302]	[0.359]	[0.465]	[0.264]	[0.241]	[0.412]
Sons-in-law and daughters-in-law	1.304	5.303**	0.8	1.493	5.933**	0.869
	[0.723]	[2.035]	[0.496]	[0.828]	[2.340]	[0.595]
Other relatives (grandchildren, grandpa	1.192	2.488*	0.791	1.33	2.453*	0.906
	[0.316]	[0.936]	[0.334]	[0.350]	[0.890]	[0.380]
Other non-relatives (mothers/fathers-in	1.829	4.621	0.498	1.426	3.288	0.186*
	[0.951]	[4.832]	[0.230]	[0.988]	[3.850]	[0.152]
Domestic staff	1	1	1	1	1	1
	[.]	[.]	[.]	[.]	[.]	[.]
External contributor	1	1	1	1	1	1
	[.]	[.]	[.]	[.]	[.]	[.]
Male	0.817+	0.906	0.739+	0.86	0.914	0.799
	[0.095]	[0.149]	[0.129]	[0.102]	[0.154]	[0.144]
Knowledge of how to send text messages	1.265+	1.477*	1.074	1.267+	1.365+	1.198

	Respondent in Initial Face-to-Face Interview					
	model 1	**model 2**	**model 3**	**model 4**	**model 5**	**model 6**
	b/se	b/se	b/se	b/se	b/se	b/se
	[0.155]	[0.243]	[0.211]	[0.159]	[0.228]	[0.242]
age under 18	0.594*	0.420**	0.8	0.636*	0.512*	0.792
	[0.128]	[0.136]	[0.241]	[0.136]	[0.159]	[0.242]
age 18-30	0.814	0.817	0.752	0.782+	0.838	0.684+
	[0.109]	[0.150]	[0.155]	[0.103]	[0.152]	[0.139]
age 41-60	0.826	0.97	0.645*	0.790+	0.895	0.661+
	[0.110]	[0.175]	[0.140]	[0.106]	[0.161]	[0.146]
age over 60	0.66	0.876	0.438*	0.745	0.938	0.532+
	[0.168]	[0.325]	[0.162]	[0.192]	[0.350]	[0.198]
household has elders	0.697	0.641	0.597	0.615	0.682	0.496
	[0.238]	[0.275]	[0.290]	[0.218]	[0.319]	[0.258]
household has children	1.02	1.008	0.94	1.008	1.002	0.931
	[0.146]	[0.218]	[0.203]	[0.146]	[0.211]	[0.208]
household has elders and children	1.953+	1.23	3.980*	2.277*	1.185	5.031**
	[0.742]	[0.574]	[2.168]	[0.880]	[0.592]	[2.853]
does not receive remittances	1.131	1.33	0.987	1.123	1.223	1.008
	[0.166]	[0.277]	[0.214]	[0.162]	[0.253]	[0.217]
Claro is mobile carrier	0.541**	0.662+	0.426**	0.520**	0.625*	0.422**
	[0.079]	[0.142]	[0.095]	[0.077]	[0.134]	[0.097]
Digicel is mobile carrier	2.342	1.034	7.829+	1.573	1.27	1.981
	[1.549]	[0.785]	[9.600]	[0.702]	[0.951]	[1.648]
Urban	0.752*	0.731+	0.774	0.729*	0.737	0.716
	[0.106]	[0.138]	[0.165]	[0.105]	[0.144]	[0.158]
Asset Index: 2nd decile	0.925	0.899	0.692	1.018	0.952	0.885
	[0.223]	[0.300]	[0.307]	[0.246]	[0.305]	[0.404]
Asset Index: 3rd decile	1.285	1.079	1.788	1.426	1.188	2.064+
	[0.314]	[0.387]	[0.637]	[0.359]	[0.424]	[0.795]
Asset Index: 4th decile	1.578+	1.445	2.039+	1.847*	1.674	2.379*
	[0.377]	[0.470]	[0.774]	[0.449]	[0.546]	[0.946]
Asset Index: 5th decile	1.362	1.224	1.572	1.654*	1.574	1.798
	[0.311]	[0.388]	[0.563]	[0.383]	[0.496]	[0.683]
Asset Index: 6th decile	1.311	0.771	2.398*	1.535+	0.983	2.740**
	[0.319]	[0.266]	[0.870]	[0.377]	[0.341]	[1.033]
Asset Index: 7th decile	1.222	0.804	1.739	1.436	1.009	2.044*
	[0.279]	[0.265]	[0.595]	[0.336]	[0.336]	[0.740]
Asset Index: 8th decile	1.903**	1.503	2.773**	2.426**	1.855+	3.708**
	[0.435]	[0.489]	[0.948]	[0.563]	[0.601]	[1.346]

	Respondent in Initial Face-to-Face Interview					
	model 1	**model 2**	**model 3**	**model 4**	**model 5**	**model 6**
	b/se	b/se	b/se	b/se	b/se	b/se
Asset Index: 9th decile	1.103	0.789	1.559	1.476	1.073	2.119*
	[0.270]	[0.275]	[0.559]	[0.367]	[0.374]	[0.800]
Asset Index: 10th decile	1.06	0.913	1.289	1.422	1.204	1.788
	[0.273]	[0.342]	[0.484]	[0.370]	[0.439]	[0.713]
Cities 100k-499k	0.973	1.024	0.949	1.024	1.09	0.97
	[0.157]	[0.238]	[0.223]	[0.163]	[0.253]	[0.223]
Cities 50k-99k	0.995	0.911	1.072	0.971	0.931	0.987
	[0.200]	[0.264]	[0.305]	[0.195]	[0.269]	[0.278]
Towns 10k-49k	0.739+	0.707	0.739	0.741+	0.749	0.696
	[0.124]	[0.163]	[0.190]	[0.123]	[0.172]	[0.181]
Towns under 10k	0.513**	0.470**	0.512*	0.531**	0.507*	0.508*
	[0.100]	[0.129]	[0.153]	[0.104]	[0.140]	[0.155]
less than primary education	0.708*	0.695+	0.652+	0.671*	0.731	0.540*
	[0.109]	[0.151]	[0.155]	[0.107]	[0.157]	[0.142]
secondary	1.233	1.38	1.181	1.222	1.31	1.148
	[0.186]	[0.303]	[0.259]	[0.184]	[0.293]	[0.251]
upper secondary	1.845**	1.673*	2.141**	1.876**	1.663*	2.191**
	[0.264]	[0.351]	[0.440]	[0.267]	[0.350]	[0.447]
higher (non-university) education	0.518	0.259+	0.538	0.606	0.326	0.644
	[0.303]	[0.212]	[0.404]	[0.358]	[0.287]	[0.481]
university or post-graduate	1.715*	2.036*	1.425	1.775*	2.144*	1.507
	[0.409]	[0.655]	[0.514]	[0.413]	[0.670]	[0.532]
Has a job	1.059	1.035	1.079	1.092	1.061	1.119
	[0.115]	[0.160]	[0.174]	[0.122]	[0.168]	[0.186]
technology=ivr	0.140**	0.136**		0.096**	0.097**	
	[0.015]	[0.015]		[0.011]	[0.011]	
technology=sms	0.124**	0.128**		0.089**	0.101**	
	[0.013]	[0.014]		[0.010]	[0.012]	
Wave 2		0.695**			0.795+	
		[0.082]			[0.097]	
Wave 3		0.99			1.15	
		[0.128]			[0.145]	
Wave 4		0.562**			0.645**	
		[0.067]			[0.081]	
Wave 5		0.476**			0.414**	
		[0.057]			[0.055]	
Wave 6		0.835			0.929	
		[0.095]			[0.112]	

	Respondent in Initial Face-to-Face Interview					
	model 1	**model 2**	**model 3**	**model 4**	**model 5**	**model 6**
	b/se	b/se	b/se	b/se	b/se	b/se
Wave 7		0.692**			0.885	
		[0.083]			[0.106]	
Wave 8		0.458**			0.595**	
		[0.061]			[0.083]	
Wave 9		0.514**			0.698**	
		[0.069]			[0.094]	
Wave 10		0.641**			0.821	
		[0.077]			[0.102]	
SMS only - wave 2			0.727**			0.856+
			[0.066]			[0.079]
SMS only - wave 3			0.904			0.842+
			[0.075]			[0.079]
SMS only - wave 4			0.993			0.953
			[0.083]			[0.092]
SMS only - wave 5			0.917			1.168+
			[0.079]			[0.108]
SMS only - wave 6			0.833*			1.008
			[0.072]			[0.092]
Constant	3.816**	6.522**	0.369+	3.199**	4.729**	0.195**
	[1.431]	[3.480]	[0.189]	[1.211]	[2.489]	[0.105]
psuedo r^2	0.117	0.148	0.099	0.139	0.175	0.106
chi2	650.032	821.38	153.291	739.316	754.748	164.613
p	0	0	0	0	0	0
N	8985	4843	4116	8985	4843	4116

Bibliography

- Brian Dillon (2010) – "Using Mobile Phones to Conduct Research in Developing Countries," Cornell University Working Paper: 1-20.

- Don Dillman, et al., 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.

- E. D. de Leeuw, 2005. "To Mix or Not to Mix Data Collection Modes in Surveys," The Journal of Official Statistics, pp. 233-255.

- Gallup Inc., 2010. "The World Bank Listening to LAC (L2L) Pilot Project, Sample Design for Peru," Deliverable for the World Bank, Listening 2 LAC Projects, Poverty, Gender and Equity Sector.

- Gallup Inc., n.a. "The World Bank Listening to LAC (L2L) Pilot Project, Power Analysis Report," Deliverable for the World Bank, Listening 2 LAC Projects, Poverty, Gender and Equity Sector.

- Gallup Inc., 2012. "The World Bank Listening to LAC (L2L) Pilot Project, Report on Attrition of Panel Participants in Peru and Honduras," Deliverable for the World Bank, Listening 2 LAC Projects, Poverty, Gender and Equity Sector.

- Gallup Inc., 2012. "The World Bank Listening to LAC (L2L) Pilot Project, Final Report," Deliverable for the World Bank, Listening 2 LAC Projects, Poverty, Gender and Equity Sector.

- Gallup Inc., n.a. "The World Bank Listening to LAC (L2L) Pilot Project Sample Design for Honduras," Deliverable for the World Bank, Listening 2 LAC Projects, Poverty, Gender and Equity Sector.

- Harold Alderman, et al., 2001. "Attrition in Longitudinal Household Survey Data," Demographic Research, Max Planck Institute for Demographic Research, Vol. 5(4), pp. 79-124.

- Jerry A. Hausman and David A. Wise, 1979. "Attrition Bias in Experimental and Panel data: The Gary Income Maintenance Experiment," Econometrica Journal of the Econometric Society, pp. 455-473.

- J. Scott Amstrong and Terry S. Overton, 1977. "Estimating nonresponse bias in mail surveys," Journal of Marketing Research, Vol. 14, Issue 3, pp. 396-402. Posted on ScholarlyCommons, University of Pennsylvania .

- Kevin Croke, et al., 2012. “Collecting high frequency panel data using mobile phones. Does timely data lead to accountability?”

- L.L. Sykes et al, 2010. “A Systematic Literature Review on Response Rates Across Racial and Ethnic Populations,” Canadian Journal of Public Health, pp. 213-219.

- Mark Nord and Heather Hopwood, 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, The Authors, pp. 1474-1480.

- Matthias Schonlau et al, 2002. “Conducting Research Surveys via Email and the Web,” RAND Corporation, pp. 81-94.

- Penny Visser et al, 1996. “Mail Surveys for Election Forecasting? An Evaluation of the Colombia Dispatch Poll,” Public Opinion Quarterly 60, pp. 181-227.

- Peter Lynn and Olena Kaminska, 2011. “The Impact of Mobile Phones on Survey Measurement Error,” Working Paper 2011-07, Institute for Social and Economic Research (University of Essex).

- Pew Research Center, 2012. “Assessing the Representativeness of Public Opinion Surveys,” Pew Research Center for the People and the Press, Released May 15, 2012.

- Richard Curtin, Stanley Presser and Eleanor Singer, 2002. “The Effects of Response Rate Changes on the Index of Consumer Sentiment,” Public Opinion Quarterly 64(4), pp. 413-428.

- Roger Thomas and Susan Purdon, 1994. “Telephone methods for social surveys,” Social Research Update, University of Surrey, Issue 8.

- Sarah Baird, et al., 2008. “Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1),” CIDER Working Paper, University of California-Berkeley.

- Scott Keeter et al, 2006. “Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey,” Public Opinion Quarterly 70(5), pp. 759-779.

- Virginia M. Lesser et al, 1999. “The influence of Financial incentives on Mail Questionnaire response rates. Paper presented at the Meeting of the Rural Sociological Society, Portland, Oregon.

- Zelee Hill, 2003. “Reducing attrition in panel studies in developing countries,” International Journal of Epidemiology, pp. 493-498.