Do Mobile Phone Surveys Work in Poor Countries?

Ben Leo, Robert Morello, Jonathan Mellon, Tiago Peixoto and Stephen Davenport

Abstract

In this project, we analyzed whether mobile phone-based surveys are a feasible and cost-effective approach for gathering statistically representative information in four low-income countries (Afghanistan, Ethiopia, Mozambique, and Zimbabwe). Specifically, we focused on three primary research questions. First, can the mobile phone survey platform reach a nationally representative sample? Second, to what extent does linguistic fractionalization affect the ability to produce a representative sample? Third, how effectively does monetary compensation impact survey completion patterns?

We find that samples from countries with higher mobile penetration rates more closely resembled the actual population. After weighting on demographic variables, sample imprecision was a challenge in the two lower feasibility countries (Ethiopia and Mozambique) with a sampling error of +/- 5 to 7 percent, while Zimbabwe's estimates were more precise (sampling error of +/- 2.8 percent). Surveys performed reasonably well in reaching poor demographics, especially in Afghanistan and Zimbabwe. Rural women were consistently under-represented in the country samples, especially in Afghanistan and Ethiopia. Countries' linguistic fractionalization may influence the ability to obtain nationally representative samples, although a material effect was difficult to discern through penetration rates and market composition. Although the experimentation design of the incentive compensation plan was compromised in Ethiopia and Zimbabwe, it seems that offering compensation for survey completion mitigated attrition rates in several of the pilot countries while not reducing overall costs. These effects varied across countries and cultural settings.

JEL Codes: O33, O14, O55

Keywords: mobile phone surveys, mobile phones, zimbabwe, ethiopia, mozambique, afghanistan

Center for Global Development
www.cgdev.org

Working Paper 398
April 2015

Do Mobile Phone Surveys Work in Poor Countries?

Ben Leo
Center for Global Development

Robert Morello
Center for Global Development

Jonathan Mellon
University of Oxford,
Global Governance Practice - World Bank

Tiago Peixoto
Global Governance Practice - World Bank

Stephen Davenport
Global Governance Practice - World Bank

Ben Leo is a Senior Fellow at the Center for Global Development. Robert Morello is a Research Assistant at the Center for Global Development. Jonathan Mellon is the British Election Study Research Fellow at Nuffield College, University of Oxford and a consultant at the Global Governance Practice at the World Bank. Tiago Peixoto and Stephen Davenport are in the Global Governance Practice at the World Bank Group. This work was co-funded and co-implemented by the World Bank's Governance Partnership Facility as part of activities of the Digital Engagement Evaluation Team (DEET).

CGD is grateful for contributions from its board and funders, including the William and Flora Hewlett Foundation, in support of this work.

Ben Leo, Robert Morello, Jonathan Mellon, Tiago Peixoto, and Stephen Davenport. 2015. "Do Mobile Phone Surveys Work in Poor Countries?" CGD Working Paper 398. Washington, DC: Center for Global Development.
http://www.cgdev.org/publication/do-mobile-phone-surveys-work-poor-countries-working-paper-398

Center for Global Development
2055 L Street NW
Washington, DC 20036

202.416.4000
(f) 202.416.4050

www.cgdev.org

Contents

I. Introduction

The exponential rise in mobile phone penetration rates throughout developing countries has provided a powerful platform for engaging citizens, whether for political, commercial, or social purposes. There is a growing proliferation of technology-based applications that have the power to gather large amounts of data quickly from thousands of participants. These range from citizen accountability mechanisms (e.g., IPaidABribe.com) to crowdsourcing tools (e.g., Ushahidi) to citizen engagement platforms (e.g., UReport). For instance, the UReport system in Uganda leverages mobile phones to collect perspectives on a broad range of issues from its nearly 300,000 participating members. This information is then used to inform national political and policy debates.

At the same time, many development agencies have recognized an operational need for rapid, accurate, flexible, and cost-efficient tools for gathering information. This reflects a growing imperative to engage intended beneficiaries throughout a project life cycle, including: identification of citizen priorities, assessing related needs, monitoring project implementation, and conducting ex-poste evaluations. In this context, mobile phone based approaches may be able to augment traditional data collection techniques, such as household surveys or consultative roundtables. In some instances, such as identifying citizen priorities or assessing beneficiaries' views about completed projects, they may even be able to replace more time consuming and expensive approaches.

Despite this potential, there have been few studies that have rigorously assessed the feasibility of mobile phone based survey techniques. We seek to address this relative gap in the literature by conducting our own mobile phone surveys to analyze whether these techniques are a feasible and cost-effective approach for gathering statistically representative information in four low-income country contexts (Afghanistan, Ethiopia, Mozambique, and Zimbabwe). In this paper, we focus on the technical aspects of our four surveys, including implementation, sample composition, and other operational considerations.[1] Specifically, we focus on three primary research questions. First, can the mobile phone survey platform reach a nationally representative sample of the country's population in a range of local environments? If so, then how many people must be called to produce it? Second, to what

[1] These surveys had a substantive focus on respondents' national development priorities. This paper will not address the surveys' substantive focus or actual survey answers, beyond the demographic questions. A second paper will address these questions, including practical concerns related to these topics.

extent does linguistic fractionalization affect the ability to produce a representative sample? Third, how effectively does monetary compensation impact survey completion patterns?

The paper is organized as follows. Section two briefly examines several World Bank-sponsored pilots that use mobile phone platforms to collect information from citizens and project beneficiaries. Section three outlines our primary research questions in more detail. Section four describes our methodology for selecting the four pilot countries. Section five describes the survey implementation modalities, including dialing techniques, demographic targeting, and language coverage. Section six outlines our survey collection results and assesses performance against our primary research questions. Section seven concludes with a discussion of lessons learned and areas for further testing and research.

II. Mobile Phone Survey Studies in Low-Income Countries

While an expanding number of organizations are utilizing mobile phones to collect information and conduct surveys in developing countries, it remains a nascent field within the policy research community. We identified four rigorous research projects that used mobile phone surveys to collect high frequency longitudinal data from respondents. The World Bank was involved in each of these projects, either as a sponsor or implementer.[2]

A. Tanzania Cotton Farmer Project

The first study, which was entitled Research on Expectations in Agricultural Production (REAP), focused on rural Tanzanian cotton farmers' expectations, production, and income levels.[3] Call center-based enumerators contacted 200 cotton farmers every three weeks on a rolling schedule among the 15 selected communities. The study began with an in-person survey during which phones were also distributed to participants. While in the enumeration area, researchers made arrangements with a charging station to pay for survey cellphone charges before the interviews. After the completion of the interview, a small amount of credit was transferred to the participants' survey phones. On average, each interview lasted approximately 19 minutes.

[2] The World Bank Institute sponsored this study as well. There have been mobile phone based projects by other development and research institutions; however, we were unable to find detailed analyses of their results and/or lessons.

[3] Dillon, 2010.

Dillon (2010) briefly addresses the choice to rely upon live enumerators in a call center. He chose not to use SMS due to concerns about literacy and technological experiences amongst a substantial proportion of targeted respondents. Also, the live enumeration allowed real-time feedback of compensation and follow up from the enumerators. As opposed to face-to-face enumerated surveys, he found significantly improved supervision factors as well. Fewer supervisors were able to more closely oversee call center enumerators.

The enumerator's inability to control the respondents' environment was one unexpected complication. In face-to-face surveys, enumerators typically interview respondents in private in order to avoid biasing responses from other people. They also report any intimidation or other bias that may have occurred during the survey. Although attentive phone interviewers may be able to detect other people in the immediate area and encourage respondents to answer privately, they have significantly less control over the situation. Taken together, Dillon concludes that for similar studies mobile phone based surveys offer a superior way to gather data.

B. South Sudan Regional Capital Household Project

The second study is the South Sudan Experimental Phone Survey (SSEPS).[4] In this project, researchers used mobile phones to monitor households in South Sudanese regional capitals. Demombynes, Gubbins, and Romeo (DGR) chose to focus on regional capitals because of the presence of mobile phone coverage in these areas. During a face-to-face baseline survey, enumerators distributed mobile phones to participants. Using a call center in Nairobi, Kenya, they conducted monthly interviews that lasted between 15 and 20 minutes.

The researchers experimented with the type of telephones as well as remuneration levels provided for survey completion. Contrary to their expectations, they found a slightly higher rate of attrition for respondents receiving higher compensation and no correlation between type of phone and response attrition over time.

Overall, DGR make three basic conclusions: (1) mobile phones are a viable means of data collection; (2) using participant's own phones is preferable to distributing phones; and (3) researchers should be especially concerned with selective non-response in mobile phone-based longitudinal studies.

[4] Demombynes, Gubbins Romeo, 2013

C. Liberia Economic Impact of Ebola Crisis

World Bank researchers, partnering with Gallup and the Liberia Institute of Statistics and Geo-Information Services (LISGIS), used a high frequency mobile phone survey to gather the first information on the socio-economic impacts of Ebola in Liberia.[5] The researchers leveraged a partially completed Household Income and Expenditure Survey (HIES) to identify a sample frame, which had collected phone numbers from approximately half of the respondents (all urban households and a third of rural households). Although a debated decision, the survey used live enumerators in Gallup's Nebraska call center because LISGIS lacked the technical capacity to implement the project. The target interview length was 15 minutes.

In the first round of the survey, the researchers found a lower than expected response rate of 30 percent of the available numbers. In an attempt to increase response rates, they used text messages to explain the survey and a monetary incentive ($1). These efforts only slightly increased the number of responses. They found that only 61 percent of the phone numbers in their sampling frame were ever turned on during the week of calls. They hypothesize that, during a crisis, respondents chose not to charge their phones either because they had less income or lacked the capability to charge.

The researchers acknowledge that the low response rates as well as an urban bias in the phone survey hinder the sample's representativeness. They used a weighting process to conform the sample to population parameters. However, the author acknowledges that the results are "more informative than representative."[6]

D. Listening to Dar Project

The last study, which is entitled *Listening to Dar*, was originally initiated by Twaweza, a Tanzanian non-governmental organization. The study made high frequency telephone calls, typically on a weekly basis, to residents of Dar es Salaam.[7] A portion of the survey questions changed weekly to reflect current events and research interests, while other questions tracked respondents' standard of living over time. In the first phase of the surveys, the researchers did not exclude baseline survey participants who did not own a mobile phone. Eighteen months after the start of the survey, researchers distributed phones to baseline survey

[5] Himelein, 2014.

[6] *Ibid.*

[7] Croke et al, 2010.

participants who did not own them. They used live enumerators in a call center to conduct the surveys. On average, the *Listening to Dar* interviews lasted between 20 and 30 minutes. This project remains ongoing.[8]

E. Cross-Project Lessons and Implications

All four of these studies used live enumerated calls. In a separate study by MIT researchers, they tested three types of mobile phone-based data collection approaches: 1) electronic forms with menus and numeric choices; (2) SMS; and (3) voice coupled with a live enumerator. Despite a very small sample, they concluded that voice reporting had the lowest error rate. *Listening to Dar* has experimented with other mobile-based survey approaches, including interactive voice recognition (IVR). The World Bank researchers found these other approaches technologically ill-suited to their operating environment.[9]

DGR (2013) compare each mobile data collection approach's relative strengths and limitations. When comparing call centers and IVR, they find that both require a reliable network signal and that both work well in settings with low rates of literacy. While call centers allow enumerators to build relationships with and clarify questions for respondents, they also require significant additional management, quality control, and budgetary resources. IVR simulates a voice call in a locally recognizable language and accent, but does not have the benefit of real-time interaction. DGR also note how keyboard navigation could be potentially cumbersome.[10]

III. Primary Research Questions

In this study, we aim to contribute research that examines whether mobile phone surveys can provide representative, cost-effective, and timely information as an input into policy and political discussions. In this context, we focus on three specific methodological areas: (1) prospective reach; (2) linguistic fractionalization; and (3) respondent attrition rates.

A. Prospective Reach

The credibility and applicability of a given survey ultimately depends upon, amongst other factors, producing a representative sample. This is important for using the results as a way of

[8] Current information and results can be found at http://listeningtodar.org/.

[9] Croke et al, 2010.

[10] See Demombynes, Gubbins, and Romeo's (2013) for a thorough discussion of the strengths and limitations of all approaches, including SMS, USSD, and GPRS.

feeding citizen views – thereby augmenting any existing formal and informal mechanisms – into discussions about resource allocation, project implementation, or ex-post evaluations. In this manner, we systematically sought to assess:

- ✓ *Research Question*: Can the mobile survey platform reach a nationally representative sample of the country's population in a range of local environments? If so, how many people must be called to produce it?

- ✓ *Test*: We tested this research question by focusing on both countries with high and low mobile phone penetration rates. [Our country selection methodology is outlined in section IV.]

- ✓ *Hypothesis*: We predicted that the mobile survey technology would reach a nationally representative sample in countries with high cellphone penetration rates. We also expected that in countries with low penetration rates it would require significantly more calls to produce a nationally representative sample.

B. Linguistic Fractionalization

Many developing countries have significant linguistic fractionalization, which may complicate citizen consultation efforts and serve to reduce civic voice and influence within political processes.[11] Therefore, it is important to demonstrate that a given survey can be representative, conducted in multiple languages, and broadly reflect different ethno-linguistic groupings.

- ✓ *Research Question*: To what extent does linguistic fractionalization affect the ability to produce a nationally representative sample?

- ✓ *Test*: We chose pilot countries with high and low linguistic fractionalization. In the post-analysis phase, we considered the language of the survey to identify any affects among language options utilized as well as any translation effects.

[11] The literature examining the impact of linguistic and ethnic fractionalization on democracy, economic growth, and conflict is long and mixed. In this project, we are primarily focused on assessing whether there is an impact on survey response rates in countries with high degrees of linguistic fractionalization.

- ✓ *Hypothesis*: We expected countries with higher linguistic fractionalization to have higher levels of attrition and non-completion, thereby necessitating a greater number of calls.

C. Respondent Attrition Rates

The existing literature suggests that mobile phone surveys can be associated with high respondent drop-off rates.[12] Given this, researchers and practitioners typically utilize some type of response incentive, such as providing free airtime or emphasizing that the responses will be used for a specific, worthwhile purpose. Ensuring that respondents complete all, or at least most, of the questionnaire is important for both cost and survey method purposes.

- ✓ *Research Question*: How effectively does monetary compensation affect respondent behavior and survey completion patterns?

- ✓ *Test*: We experimentally used monetary compensation (airtime credits) to test its impact on respondents' willingness to complete the questionnaire.

    - ✓ All respondents received a pro-social or intrinsic incentive. The introduction included a sentence explaining the survey's purpose and how the data could be used in the future. One-third of respondents only received this intrinsic incentive, which served as our control or comparison group.

    - ✓ One-third of respondents were informed that, upon completing the survey, they would become eligible for an extrinsic benefit (airtime raffle). Two of these respondents per country received two hours of airtime.

    - ✓ One-third of respondents were informed that, upon completing the survey, they would receive 4 minutes of airtime.[13] Compensating for the same amount of airtime adjusts for different income environments and ties the compensation to the mobile technology used for the survey.

[12] For example, see http://www.votomobile.org/files/VOTO_SurveyingLessonsLearned.pdf.

[13] The size of this airtime credit mirrored the expected time to complete the survey instrument.

- ✓ *Hypothesis*: We predicted that the extrinsic reward options would be a more effective retention tool. We also expected that the pro-social control would have a modest impact on retention rates. However, we were not able to test that effect given the lack of another control group.

IV. Country Selection Methodology

A. Development Need and Public Spending Filters

Our focus is confined to low-income countries (LICs) and lower-middle income countries (LMICs), as defined by the World Bank.[14] Due to the surveys' substantive focus (e.g. citizens' development-related priorities), and recognizing that they could be relevant for both national government and development assistance spending priorities, we applied three initial filters as proxies for: development needs, the relative importance of development assistance, and public sector resource availability.[15]

- ✓ *Development Need*: The UN Human Development Index (HDI) is used as a proxy for overall development levels within a country. The HDI ranks countries into four tiers of human development based upon a composite score encompassing life expectancy, education, and income levels. Countries with a HDI score below the LIC/LMIC median were considered for the pilot project.

- ✓ *Official Development Assistance*: We focused on those countries that receive significant official development assistance (ODA) relative to income levels. Specifically, our pilot is limited to those countries that received ODA totaling at least 5 percent of gross national income (GNI).[16]

- ✓ *Public Sector Resources*: While all governments must make difficult budget allocation decisions, the tradeoffs can be particularly challenging in countries with very limited public sector resources.[17] We focused on those countries with government spending per capita below the LIC/LMIC country median.[18] Alternative measures could have

[14] The World Bank classifications include: (1) low-income countries (gross national income per capita ≤ $1,045); and (2) lower-middle income countries (gross national income per capita between $1,046 and $4,125).

[15] Although we do not address the surveys' substantive focus in this paper, we include this section to explain the motivation behind our country selection.

[16] Alternative measures could have focused on ODA as a percentage of government expenditures.

[17] See Pattillo et al (2010) and Gupta et al (2002).

[18] In 2013, LICs and LMIC public spending per capita was about $400. Source: International Monetary Fund, World Economic Outlook Database, October 2013 and authors' calculations.

focused on total government spending or government spending as a percentage of gross domestic product (GDP).

There were 30 prospective pilot countries based on these three initial filters.[19] In 2012, these countries had a combined total of $100 billion in government expenditures[20] and received nearly $39 billion in development assistance.[21] They also account for 40 percent of total ODA to LICs and LMICs over the last five years.

B. Mobile Phone Penetration and Linguistic Fractionalization Filters

Within this universe, we considered feasibility of mobile phone surveys in selecting pilot countries. Feasibility largely depends upon two factors: (1) cell phone penetration rates; and (2) linguistic fractionalization. However, there are other important considerations, such as cultural norms and cell phone ownership and usage patterns. We did not account for these factors during the project design phase, which led to a number of lessons learned within several specific countries. We explore these issues in section VI of the paper.

- ✓ *Cell Phone Penetration Rates*: Large polling organizations, such as Gallop, have traditionally used in-person interviews unless landline or mobile phone penetration rates exceed 80 percent of the total population. However, only six of the 30 prospective countries have penetration rates at this level (Benin, Cote d'Ivoire, the Gambia, Mali, Senegal, and Zimbabwe). Despite this, we tested whether it was possible to generate a representative sample in countries with lower rates.[22] To do so, we divided our sample into two groups based on those countries with mobile penetration rates above the group median and those below it (see figure 1 below), identifying pilot countries from within each.

[19] These countries include: *Afghanistan*, Benin, Burkina Faso, Burundi, Central African Republic, Chad, Comoros, Democratic Republic of Congo, Cote d'Ivoire, Eritrea, Ethiopia, Gambia, Guinea, Guinea-Bissau, *Haiti*, Liberia, Madagascar, Malawi, Mali, Mozambique, *Myanmar*, *Nepal*, Niger, Rwanda, Senegal, Sierra Leone, Tanzania, Togo, Uganda, and Zimbabwe. [Non-African countries are presented in italics.]

[20] Source: IMF *World Economic Outlook* database, October 2013.

[21] Source: OECD DAC database. Figures reflect total gross official development assistance minus debt relief. World Bank assistance to these countries totaled nearly $4 billion in 2012, and nearly $20 billion over the last five years.

[22] The data measures mobile cellular subscriptions (per 100 people).

Figure 1 – Cell Phone Penetration Rates[23]

Above Median		Below Median	
Mali	129.1	Tanzania	55.7
Gambia	100.0	Mozambique	48.0
Zimbabwe	96.3	Comoros	47.3
Côte d'Ivoire	95.4	Sierra Leone	44.1
Benin	93.3	Uganda	44.1
Senegal	92.9	Congo (Dem. Rep.)	43.7
Guinea-Bissau	74.1	Niger	39.3
Nepal	71.5	Madagascar	36.1
Afghanistan	70.0	Chad	35.6
Haiti	69.4	Malawi	32.3
Burkina Faso	66.4	Central African Rep.	29.5
Guinea	63.3	Ethiopia	27.3
Togo	62.5	Burundi	25.0
Liberia	59.5	Myanmar	12.8
Rwanda	56.8	Eritrea	5.6

Source: International Telecommunication Union

- ✓ *Linguistic Fractionalization*: We anticipated that a country's linguistic fractionalization would affect the feasibility of an automated survey.[24] To our knowledge, polling organizations have not applied a general rule concerning linguistic fractionalization. Instead, they typically have attempted to address any challenges through local interpreters on a region-by-region basis.[25] We divided our sample between countries with rates of linguistic fractionalization below the median, where we expect the feasibility to be higher and those with fractionalization rates above the median (see figure 2 below).

[23] These figures represent the number of mobile subscriptions per 100 inhabitants in 2013.

[24] Alesina et al, 2003.

[25] Alesina (2003) approximates the 'differentness' of the peoples that compose a country's population. Since it does not consider fluency in a second language, it is an imprecise measure of the difficulty to execute a survey. In the 17 countries for which we have Afrobarometer surveys and Alesina data, we compared the Alesina score to the percentage of respondents who completed the Afrobarometer survey in the most popular language for a particular country. We find that Alesina roughly approximates this measure of linguistic fractionalization from Afrobarometer, except for in the case of Tanzania where 100 percent of the respondents completed the survey in Kiswahili but has a high fractionalization score of 0.8983. We chose to use Alesina in order to have one equivalent measure across all of the countries.

Figure 2 – Linguistic Fractionalization

Above Median		Below Median	
Uganda	0.923	Guinea	0.773
Liberia	0.904	Sierra Leone	0.763
Tanzania	0.898	Burkina Faso	0.723
Togo	0.898	Nepal	0.717
Congo, Dem. Rep.	0.871	Senegal	0.708
Chad	0.864	Eritrea	0.653
Mali	0.839	Niger	0.652
Central African Rep.	0.833	Afghanistan	0.614
Guinea-Bissau	0.814	Malawi	0.602
Mozambique	0.813	Myanmar	0.507
Gambia	0.808	Zimbabwe	0.4472
Ethiopia	0.807	Burundi	0.298
Benin	0.791	Madagascar	0.02
Cote d'Ivoire	0.784	Comoros	0.01

*Source: Alesina et al (2003). **Note – The sample median is 0.7835.*

We then divided the 30 prospective pilot countries into four groups. Countries are predicted to have higher survey implementation feasibility if their linguistic fractionalization was below the group median and mobile penetration was above it.[26] Lower feasibility countries have linguistic fractionalization above the group median and mobile penetration rates below the median. A country was predicted to have moderate feasibility if either the linguistic fractionalization was below or mobile penetration was above the median.

[26] The median LIC/LMIC mobile phone penetration rate is 70.6 subscriptions per 100 inhabitants. The median linguistic fractionalization score was 0.55. For the 30 countries, the median mobile phone penetration rate is 49.3 subscriptions per 100 inhabitants, and the median linguistic fractionalization score was 0.77.

Figure 3 – Mobile Survey Feasibility

Low Feasibility	Moderate Feasibility		High Feasibility
	Linguistic Fractionalization Below Median	**Mobile Penetration Above Median**	
Central African Rep.	Burundi	Benin	Afghanistan
Chad	Comoros	Cote d'Ivoire	Burkina Faso
Congo, Dem. Rep.	Eritrea	Gambia	Guinea
Ethiopia	Madagascar	Guinea-Bissau	Haiti
Mozambique	Malawi	Liberia	Nepal
Tanzania	Myanmar	Togo	Senegal
Uganda	Niger	Mali	Zimbabwe
	Rwanda		
	Sierra Leone		

C. Final Prospective Pilot Countries

Based upon these factors, we selected four countries to serve as representative test cases. Budgetary restrictions forced us to choose only a sub-set of countries for the pilot study. Ideally, we would have included all 14 of the high and low feasibility countries. Within our sub-sample, several countries were excluded due to domestic stability concerns, such as the Central African Republic. Since the universe only includes four non-African nations, we chose to include at least one non-African country in the pilot phase (Afghanistan).

- ✓ *High Feasibility*: Afghanistan and Zimbabwe
- ✓ *Low Feasibility*: Ethiopia and Mozambique

Overall, this methodology allowed us to test survey feasibility and citizen response patterns across a range of different environments. Moreover, they provided reasonable representation of a number of additional factors, such as: (i) primary official languages (Anglophone, Lusophone, Pashto, Dari, and Amharic); (ii) geographic region (African sub-regions and other developing regions); (iii) levels of fragility and conflict or post-conflict[27]; (iv) GDP and income levels; (v) population size; and (vi) national government budget transparency levels. Collectively, these countries have a combined GDP of $84 billion and a population of 160 million people. In 2012, government expenditures totaled $40 billion and they received ODA totaling nearly $20 billion.

[27] Based upon the World Bank's FY14 Harmonized List of Fragile Situations.

Figure 4 – Pilot Country Characteristics, Select Indicators

Country	Region	Fragile	GDP (US$ billions)	Population (mln)
Afghanistan	Non-African	X	$18.0	29.8
Ethiopia	East Africa		$41.6	91.7
Mozambique	Southern Africa		$14.6	25.2
Zimbabwe	Southern Africa	X	$9.8	13.7
Total	**-**	**-**	**84.0**	**160.4**

Source: IMF WEO Database and World Bank 2014 WDI Database

Figure 5 – Pilot Country Characteristics, Budget Openness and Public Sector Governance

	Open Budget Index (out of 100)			PRSP	Public Sector Governance Quality (2012 CPIA, Range = 1-6)	
Pilot Country	2008	2010	2012	Date	Quality of Public Admin	Transparency, Corruption, Accountability
Afghanistan	8	21	59	2008	2.5	2.0
Ethiopia	-	-	-	2011	3.5	3.0
Mozambique	-	28	47	2011	3.0	3.0
Zimbabwe	-	-	20	-	2.0	1.5

Source: International Budget Partnership, IMF, and World Bank

D. Demographics of Mobile Phone Ownership in Pilot Countries

We utilized Demographic and Health Surveys (DHS) to estimate the ownership of mobile phones by demographic groups in the four pilot countries. DHS are known for their highly reliable and accurate household surveys. Recent DHS survey data (either 2010 or 2011) is available for all four of the countries, which include variables for cell phone ownership at the household level. Although information on individual-level ownership and usage would be preferable, we use this household data since it is the only available cellphone ownership data for all of our countries.

Figure 6 – DHS Survey Coverage

Country	Year	Sample Size
Afghanistan	2010	180,676
Ethiopia	2011	77,744
Mozambique	2011	62,750
Zimbabwe	2010-2011	41,946

Source: DHS Program

We examined cell phone ownership across a range of demographic cross-tabulations. As expected, rural households are less likely to have a mobile phone. This trend is substantially more prevalent in Ethiopia, Mozambique, and Zimbabwe. Across all countries, households with more educated members are more likely to own a mobile phone. Finally, poorer people are substantially less likely to own phones, especially in countries with low rates of ownership. In Ethiopia and Mozambique, these low levels present significant feasibility concerns for obtaining a nationally representative sample. Appendix III includes greater detail on this information.

This initial desk research provided some sense of a probable sample that would be obtained through mobile phones surveys. However, we found that these probable samples, at best, only weakly approximated the actual samples. The main driver for this weak approximation likely rests with the DHS household ownership question, which fails to account for gender-based ownership or usage patterns. Also, since the DHS data is three to four years old, there seems to have been substantial growth in mobile phone ownership, particularly among poorer segments of the population.

V. Pilot Survey Implementation

A. Interactive Voice Recognition

We chose to use interactive voice recognition (IVR) in these surveys. This technology utilizes a recording to ask the survey questions, and the survey participants respond by pressing a number on the phone dial pad. The system then records the selected number through dial tone sensing (DTMF) technology.

There are a variety of technical solutions that use mobile phones to remotely gather answers, including SMS, USSD, and live call centers. Each of these technologies has its own

advantages and disadvantages.[28] Briefly, we chose IVR because we hypothesized that it would be more user friendly for low-literacy populations than SMS-based techniques, but would still retain some cost-savings over using a live call center. We recognize the need to more fully understand the advantages and disadvantages of these technologies, including their effect on the information gathered. We encourage further research in this area.

B. Random Number Generation

The survey operator, Voto Mobile, uses a random number generation system to obtain a sample of mobile users. More specifically, it randomly generates a list of numbers that conform to the target countries' mobile number formations. In cases where geographic codes or telecom number batch information is available, the system's random generation reflects these numeric combinations. It calls these numbers in turn, moving to the next number when the dialed number is found to be inactive. The system continues to call down the list until the desired number of responses is obtained. Respondents receive a call from an international phone number.[29]

This system does not use national or telecom user registries. Voto Mobile's experience has shown that these directories are often very difficult and expensive to obtain and/or are out of date. These registries rarely contain any demographic information related to the numbers' owners. Also, it is difficult to obtain registries from every telecom operating within a country. Thus, using only a sub-set of registries can systematically exclude entire sets of customers. Due to telecom provider's varied subscriber profiles (e.g., regional, ethnic, and economic status factors), excluding any provider could introduce systemic bias into the survey approach.

Therefore, the list of all possible telephone numbers essentially serves as the project's sampling frame. Since we are unaware of any demographic characteristics associated with the number until after the survey, it is not possible to stratify the sample. We use simple random selection from this list, and attempt to connect with the numbers. Successful connections that lead to completed surveys compile our sample.

---

[28] See Demombynes, Gubbins, and Romeo's (2013) for a thorough discussion of the strengths and limitations of all approaches, including SMS, USSD, and GPRS.

[29] For these surveys, all respondents received a call from the same international number which had a 001 country code and a 202 area code. Voto Mobile is able to mask the incoming number as any number in the 001 country code. We chose to keep the number constant for all respondents in order to avoid an additional element of variation in the study.

C. Demographic Data Collection Methods

Data collection for most demographic variables is straightforward. Our survey instrument used categorical multiple-choice questions to measure the respondents' age, gender, educational attainment, and geographic location (urban or rural). Compared to open response questions that ask respondents to type their age or years of schooling, categorical responses seem to have lower rates of attrition and produce more reliable data.

We explored collecting more detailed information on place of residence, such as including selection options for major cities and/or sub-national administrative units. This would have required multiple survey questions or lengthy cascading selection options, which would have extended the call length and increased respondent attrition levels. This was due to the presence of many major urban centers in several countries, particularly in Ethiopia. In addition, there were concerns about how familiar some respondents may be with sub-national administrative units or official names. For these reasons, we decided to proceed with a simple standardized binary choice of urban or rural for all of the pilot countries. For future projects, we would give further consideration to these options. Especially in countries with low cell phone penetration rates and where the political and/or commercial capital city accounts for a significant share of the national population (e.g., Mozambique).

Measuring survey respondents' economic status was the most challenging or complex component of the demographic data collection process. The difficulties of measuring economic status in the developing world are well recorded – particularly since many people lack a regular cash income. By illustration, in Afrobarometer's round five surveys of 33 African countries (2011-13), 43 percent of respondents reported "always" or "many times" lacking a cash income. Therefore, researchers are unable to directly and reliably ask respondents about typical monthly earnings. In response to these constraints, large nationally representative household surveys utilize a series of asset or consumption questions to construct a wealth factor index.[30]

The inherent brevity of mobile phone IVR surveys precluded us from asking several asset or consumption questions in order to construct an index. Given these constraints, we used two questions to proxy respondents' economic status. Using recent DHS data, we identified two assets that correlate closely with the survey's 'wealth index' quintiles. By referencing survey respondents' asset ownership characteristics, we were able to assess whether the pilot survey

[30] For examples, see Filmer and Pritchett (2001), Schellenberg et al. (2003), and McKenzie (2003).

adequately reached all income level segments within each of the countries, with a particular emphasis on the bottom two quintiles.

Figure 7 – Asset Ownership Characteristics by Wealth Quintile

Country	Asset Type	1st Quintile "Poorest"	2nd Quintile "Poorer"	3rd Quintile "Middle"	4th Quintile "Richer"	5th Quintile "Richest"
Afghanistan	Radio	26%	63%	73%	82%	86%
	Finished Floor	1%	10%	38%	53%	73%
	Both	0%	3%	24%	42%	62%
Ethiopia	Finished Roof	1%	19%	43%	70%	96%
	Chairs	10%	28%	42%	62%	76%
	Both	0%	1%	13%	43%	72%
Mozambique	Finished Floor	14%	26%	34%	55%	93%
	No Toilet Facility	94%	51%	30%	20%	2%
	Both	1%	5%	23%	45%	92%
Zimbabwe	Radio	16%	33%	36%	49%	64%
	Finished Floor	1%	27%	83%	96%	84%
	Both	0%	11%	32%	48%	63%

Source: DHS and authors' calculations

D. Targeting Demographic Quotas

Voto Mobile's IVR system can utilize the demographic questions at the beginning of the survey to obtain targeted numbers of specific demographics. For example, once the predetermined quota for urban male respondents has been reached, the survey can conclude automatically for any other urban male respondent as identified from the first questions. This approach could reduce the costs of unnecessarily over sampling certain demographics.

E. Translation and Language Selection

Due to financial constraints, we were unable to conduct the survey in all official and local languages. Instead, we first identified the spoken languages in the four pilot countries from a variety of sources.[31] Then, we sought to maximize the percent of the population covered within a few constraints. First, the total number of languages was capped at five per country.

[31] Sources included the World Factbook from the Central Intelligence Agency, Ethnologue: Languages of the World, and the most recent Afrobarometer survey in the country.

Second, we recognized the low marginal returns of incorporating an additional language, especially after 80 percent of the population was covered.[32] Taken together, this process produced 15 different languages across the four pilot countries. With only one exception, the survey instrument was translated from English into the local languages through a double blind technique. In Ethiopia, the survey implementer was able to identify only one Sidamo translator for this project.

Figure 8 – Survey Languages

Afghanistan	Ethiopia	Mozambique	Zimbabwe
Dari	Amharic	Changana	English
Pashto	Oromo	Chuabo	Shona
	Sidamo	Makua	SiNdebele
	Somali	Portuguese	
	Tigrayan	Sena	

VI. Survey Results and Analysis

A. Demographic Filtering

As data collection began, the demographics of our initial samples differed substantially from the population estimates.[33] In three countries (Afghanistan, Ethiopia, and Zimbabwe), we experimented in a limited way with demographic filtering in order to better reflect the population in the sample. Unsurprisingly, we found that filtering respondents significantly increased our costs per completed response. In many cases, it doubled the cost per completed response. Although the price per completed response increased, we ultimately limited the number of overrepresented demographics in the final sample.

In Afghanistan, we experimented with demographic filtering on the fifth, sixth, and seventh days of implementation.[34] On the fifth day, urban males were filtered. On days six and seven, all males were filtered. Overall, 1,786 respondents were precluded from finishing the survey; 632 were rural males, and 1,154 were urban males. In Ethiopia, we intentionally filtered respondents on the sixth day of the survey.[35] However, due to a miscommunication with the

[32] For example, translating the instrument into Urdu and Pashto covers 85 percent of the Afghan population, while we would need to add Uzbek and Turkmen to cover additional 11 percent.

[33] Please see this section's subsection C for a full comparison between the completed sample and population demographics.

[34] Days 5-7 were July 20-22, 2014.

[35] Day 6 was Oct. 16, 2014.

implementing partner, unintentional filtering began again on the twelfth day and continued unevenly across survey versions through the end of data collection.[36] In total, 579 respondents were filtered in Ethiopia. The vast majority of filtered respondents were urban males (475), the remaining were urban females (90) and rural males (14). In Zimbabwe, all urban respondents were filtered on the seventh and eighth days of data collection.[37] In total, 507 urbanites were precluded from finishing the survey.

Demographic filtering merits further consideration and research. During this pilot, we only began to explore this technique. One area for future research may be the temporal implications of filtering entire demographics. If response patterns are tied to time-sensitive events, then excluding a demographic from an entire period of data collection could bias the survey. Future research may also need to explore the full methodological implications of filtering. Once filtering begins, the sample no longer reflects the mobile phone owning population; however, it may more accurately reflect the national population.

B. Required Number of Calls

In order to obtain a completed survey, calls passed through several stages. Each one of these stages reflects financial implications. First, the survey operator randomly generated a large batch of numbers. Only a fraction of these numbers actually connected. Numbers did not connect for several reasons, such as: (i) the phone number was not assigned; (ii) the number was assigned but was not active at the time of the call[38]; (iii) the phone rang but the owner was unable to answer; or (iv) the phone was answered by an automated message or voicemail system and the survey operator terminated the call.[39] Specifically in Mozambique, we believe that undetected voicemails may have inflated the connection rate.

[36] Filtering began again on day 12 of the survey (Oct. 22, 2014). No respondents were filtered on day 13 of survey data collection. Then, filtering occurred on the last four days on the survey (Oct. 24-28). Urban males were filtered out of the closed surveys with the transfer incentive and no incentive on day 14 (Oct. 24). For the open surveys, the unintentional filtering happened on the final four days of data collection (Oct. 22-27). On open surveys either without a material incentive or with a transfer incentive, all demographics except rural females were filtered. All urbanites were filtered out on the open survey with a raffle incentive.

[37] The seventh and eighth days of data collection were July 22 and 23, 2014.

[38] For instance, the SIM card was not in the phone or phone was out of coverage at the time of the call.

[39] We acknowledge the possibility that an undetected voicemail could have resulted in a call being counted as a connected.

We also look beyond the national trends to the specific carriers. We find that publically available estimates for carrier market share largely correspond with the connection rates.[40] Corresponding to the carrier market share, we found a large variance in the number of attempts to result in a connected call. For example, for Afghan Telecom, a small carrier with about one percent of market share, Voto Mobile needed to attempt more than 21 calls to produce a single connected dial. This is due to the large size of inactive cell phone numbers within Afghan Telecom's allotted prefixes. In contrast, we attempted only one or two calls to produce a connected call amongst Econet subscribers, a Zimbabwean carrier with 63 percent market share.

The successful connection rate depends on the density of assigned phone numbers out of all possible numbers. National density stems first from mobile penetration rates, but also from the nature of the mobile market. For example, both Mozambique and Ethiopia have connection rates of 20 percent, even though Mozambique has double the mobile penetration of Ethiopia. The latter has a single carrier, while the former has three. Afghanistan, which has a much higher mobile penetration rate than these two countries, but has a lower connection rate at 15 percent. The Afghan telecom market has five carriers, including one carrier with a small share of the market.

This observation has at least two implications for future mobile phone surveys. First, if researchers or policymakers choose to use all carriers within a market, they may encounter increased costs in markets with multiple carriers, especially in countries with many small carriers. Second, in light of this first implication, researchers may choose not to use all carriers when using random digit dialing techniques. However, we intentionally included all carriers to prevent against systematic exclusion of population segments. This risk is particularly acute if the type of carrier is correlated to demographic factors. This could be a material concern for future research attempting to obtain nationally representative samples.

---

[40] Market share sources: Afghanistan: USAID. *The State of Telecommunications and the Internet in Afghanistan.* March 2012. http://www.internews.org/sites/default/files/resources/Internews TelecomInternet_Afghanistan_2012-04.pdf.

Mozambique: Gareth van Zyl. IT Web Africa. *Door-to-door sales help Movitel win Mozambican mobile subscribers.* Aug. 25, 2014. http://www.itwebafrica.com/mobile/429-mozambique/233419-door-to-door-sales-help-movitel-win-mozambican-mobile-subscribers.

Zimbabwe: Postal and Telecommunications Regulatory Authority of Zimbabwe (POTRAZ). *Postal and Telecommunications Sector Performance Report.* Fourth Quarter 2013. http://www.potraz.gov.zw/images/files/stats/Sector_Perfomance_4th_Quarter%202013.pdf.

Figure 9 – Attempted versus Connected Numbers, by Country and Carrier

Country/Carrier	Calls Attempted	Calls Connected	Percent Connected	Average Attempts Per Connected	Carrier Market Share	Mobile Penetration Rate
Afghanistan	**191,845**	**28,549**	**15%**	**6.7**	**-**	**59.1**
AWCC	49,033	5,657	12%	8.7	20%	
Afghan Telecom	49,265	2,300	5%	21.4	1%	
MTN	49,346	10,198	21%	4.8	26%	
Etisalat	19,749	3,934	20%	5.0	20%	
Roshan	24,452	6,460	26%	3.8	32%	
Ethiopia	**280,820**	**54,806**	**20%**	**5.1**	**-**	**16.8**
ETHMTN	280,820	54,806	20%	5.1	100%	
Mozambique	**314,389**	**62,620**	**20%**	**5.0**	**-**	**39.9**
Movitel	104,904	14,173	14%	7.4	22%	
Vodacom	104,779	34,563	33%	3.0	34%	
mCel	104,706	13,884	13%	7.5	44%	
Zimbabwe	**180,450**	**56,605**	**31%**	**3.2**	**-**	**79.4**
NetOne	52,103	3,542	7%	14.7	18%	
TelCel	55,047	8,048	15%	6.8	19%	
Econet	73,300	45,015	61%	1.6	63%	

Second, after the call is connected, the respondent hears the language selector question. We encountered some survey attrition at this phase. We hypothesize that these calls could have ended for one of five reasons: (i) the call inadvertently connected to an automated message[41]; (ii) the respondent attempted to select a language but the dial tone function did not operate properly; (iii) the respondents chose not to participate upon hearing the language selector; (iv) the respondents desired to participate but was unable to navigate the keyboard functionality; or (v) respondents were unable to find their language. Unfortunately, we are unable to identify which one of the five reasons led to survey respondents' failure to select a language.

Within this stage, we looked for any clustered carrier effects. If the effect is concentrated on a single carrier, it is more likely that a technical reason explains the respondents' failure to answer the language selector. For example, only 16 respondents answered the language

[41] Automated messages include an undetected voicemail or carrier messages, such as the subscriber being outside coverage area or the system temporarily not functioning.

selector out of 2300 calls connected on Afghan Telecom. We surmise that Voto Mobile's system to sense dial tones (DTMF) failed to operate effectively on this carrier.[42]

Figure 10 – Connected versus Answered Language Selector, by Country and Carrier

Country/Carrier	Calls Connected	Answered Language Selector	Percent of Connected Calls that Answer Language Selector	Connected Calls per Language Selected
Afghanistan	**28,549**	**8,864**	**31%**	**3.2**
AWCC	5,657	2,202	39%	2.6
Afghan Telecom	2,300	16	1%	143.8
MTN	10,198	3,798	37%	2.7
Etisalat	3,934	1,317	33%	3.0
Roshan	6,460	1,531	24%	4.2
Ethiopia	**54,806**	**10,517**	**19%**	**5.2**
ETHMTN	54,806	10,517	19%	5.2
Mozambique	**62,620**	**5,820**	**9%**	**10.8**
Movitel	14,173	1,070	8%	13.2
Vodacom	34,563	3,241	9%	10.7
mCel	13,884	1,509	11%	9.2
Zimbabwe	**56,605**	**4,781**	**8%**	**11.8**
NetOne	3,542	1,134	32%	3.1
TelCel	8,048	841	10%	9.6
Econet	45,015	2,806	6%	16.0

Once a respondent selects a language, we are sure that the call has resulted in a real human respondent with functioning technology. At this point, the respondent hears the introduction to the survey in the selected language as well as the randomized incentive treatment. After listening to the introduction, respondents answer up to ten or eleven questions to complete the survey. For operational use, we defined an open survey complete when a respondent answered the open-ended question (i.e., the seventh question). Open surveys included an additional four closed ended questions, for a total of eleven questions. Respondents completed the closed survey when they answered all ten questions. While

[42] We acknowledge concerns that the technological failure with Afghan Telecom may bias the sample. The extent to which Afghan Telecom users have different observed and unobserved demographics from other carriers could bias our sample. Our concern is somewhat mitigated by the observation that Afghan Telecom has only 1 percent of market coverage and accounted for only 8 percent of all connected calls.

response rates varied across countries, we find fairly stable rates across carriers within a country with the exception of Afghan Telecom.[43]

Figure 11 – Answered Language Selector versus Complete Survey, by Country and Carrier

Country/Carrier	Answered Language Selector	Completed Survey	Survey Completion Rate	Percent of Connected Calls that Answer Language Selector	Connected Calls per Language Selected
Afghanistan	**7,078**	**2,123**		**30%**	**3.3**
AWCC	1,733	543	31%	31%	3.2
Afghan Telecom	15	0	0%	0%	
MTN	3,076	928	30%	30%	3.3
Etisalat	1,035	348	34%	34%	3.0
Roshan	1,219	304	25%	25%	4.0
Ethiopia	**9,938**	**2,258**		**23%**	**4.4**
ETHMTN	9,938	2258	23%	23%	4.4
Mozambique	**5,820**	**2,229**		**38%**	**2.6**
Movitel	1,070	402	38%	38%	2.7
Vodacom	3,241	1,310	40%	40%	2.5
mCel	1,509	517	34%	34%	2.9
Zimbabwe	**4,274**	**2,192**		**51%**	**1.9**
NetOne	992	518	52%	52%	1.9
TelCel	751	375	50%	50%	2.0
Econet	2,531	1,299	51%	51%	1.9

We now turn to answer the second part of our first research question: how many numbers must be called to produce our sample? As expected, we find that the number varies widely depending on mobile penetration rates and the nature of the telecom market.[44] In Afghanistan, which has an estimated penetration rate of nearly 60 percent, we attempted approximately 190 thousand numbers. Yet in Mozambique, which has a penetration rate of 40 percent, we attempted 70 percent more numbers, or about 314 thousand. In Ethiopia, which has the lowest mobile penetration rate of all the countries, we attempted less calls

[43] This trend does not hold for Afghan Telecom. Please see the preceding paragraph in this section and footnote.

[44] This analysis includes the filtered responses. 1786 filtered responses in Afghanistan. 507 filtered responses in Zimbabwe.

than in Mozambique. This is mainly due to differences in the telecom market, as noted previously. The results are highly variable by carriers as well.[45]

Figure 12 – Survey Completion versus Attempted Calls, by Country and Carrier

Country/Carrier	Calls Attempted	Complete	Percent of Attempted Calls that Complete Survey	Number of Attempted Calls Required to Obtain One Complete Survey
Afghanistan	**191,845**	**2,123**	**1.11%**	**90**
AWCC	49,033	543	1.11%	90
Afghan Telecom	49,265	0	0.00%	*N/A*
MTN	49,346	928	1.88%	53
Etisalat	19,749	348	1.76%	57
Roshan	24,452	304	1.24%	80
Ethiopia	**280,820**	**2,258**	**0.80%**	**124**
ETHMTN	280,820	2258	0.80%	124
Mozambique	**314,389**	**2,229**	**0.71%**	**141**
Movitel	104,904	402	0.38%	261
Vodacom	104,779	1,310	1.25%	80
mCel	104,706	517	0.49%	203
Zimbabwe	**180,450**	**2,192**	**1.21%**	**82**
NetOne	52,103	518	0.99%	101
TelCel	55,047	375	0.68%	147
Econet	73,300	1,299	1.77%	56

In terms of survey length, respondents generally spent equal amounts of time, averaging between 2 and 3 minutes across countries.[46] On a per question basis, respondents spent about 30 to 40 seconds per question answered. However, many of these respondents did not complete the survey. Respondents who completed the survey spent 4 to 5 minutes on average total, and slightly under 30 seconds per question. This faster time likely results from greater familiarity with the technology as the survey progressed.

[45] For example, 53 numbers were attempted on Afghanistan's MTN network to yield a complete survey, while more than five times as many numbers were attempted on Mozambique's Movitel to yield a complete response.

[46] This is the average for calls in which a respondent engaged with the survey, by responding to the language selector. The average for all calls is much smaller at about 40 seconds, probably because undetected voicemails or respondents refused to answer the language selector. See subsection C of this section for more details.

C. Obtaining a Nationally Representative Sample

Next, we explore whether the surveys were able to obtain nationally representative samples. We consider each country in turn to assess two aspects: raw demographics and population parameter adjustment weights.

To adjust our sample to reflect population demographics, we used an iterative proportional fitting algorithm (also known as raking).[47] The weights for Afghanistan, Mozambique, and Zimbabwe converged in less than 10 iterations.[48] The weights for Ethiopia did not converge within 500 iterations. The maximum deviation between the sample and the population demographics remained at 0.105 percent.[49]

Next, we measured how much sample weighting was required to conform it to population parameters. We first look at the largest and smallest weight applied to any observation. Also, we use the common measure of design effect, using Kish (1992) approximation $DEFF^2_{KISH} = 1 + CV(\mathrm{w_i})^2$, where $CV(\mathrm{w_i})^2$ is the coefficient of variation of the weights $\mathrm{w_i}$.[50] Assuming that equal weights are optimal for our surveys, the design effect represents the variability in the weights, and subsequent loss of precision. An intuitive way to present loss precision is through a reduced sample size. Dividing the current sample size by the design effect approximates the effective sample size reduction. The result is a hypothetical un-weighted sample that could produce an equivalent level of precision. For an additional intuitive way to measure loss of precision, we also approximate a sample error for this hypothetical sample.[51]

---

[47] Deming and Stephan (1940) first proposed this concept. More recently, Battaglia et al. (2011) discuss practical considerations for raking survey data. We performed stepwise adjustment, using Bergmann (2011) *ipfweight* statistical package for Stata, to conform the raw survey data to known population margins identified from the country's most recent DHS. We did not weight the sample on language demographics. See part D in this section for further discussion. For Afghanistan, the most DHS does not include information on education. We use information from the Asia Foundation's Survey of the Afghan People as a population standard in this instance.

[48] Afghanistan weights converged in 7 iterations. Mozambique converged in 8 iterations. And, Zimbabwe converged in 6 iterations.

[49] This deviation occurs on the "no formal education" demographic.

[50] Kish originally developed this technique to measure the loss of precision in stratified and clustering sampling techniques. Kalton, Brick, and Lê (2005) note that this measure is a poor approximation for design effect when the weights are created from population totals. In these cases, the equation indicates loss in precision. We choose to use this measure as a helpful benchmark in order to compare across surveyed countries. We caution against directly comparing this approximated design effect to those calculated for household surveys with stratified and clustered designs.

[51] We base the sample error on a 95 percent confidence interval and 50 percent sample proportion to create a conservative estimate for the entire survey.

Survey practitioners use several methods to limit increases in variability and losses of precision when using population parameter weights.[52] For this analysis, we used the most conservative and straightforward weighting approach in order to judge how well our samples represented the population at large. Among these widely used techniques is trimming, when practitioners limit the maximum weight given to any observation. DeBell et al (2009) suggest limiting any observation's weight to 5. While we did not limit the weights, we report the number of observations with weights greater than five. In this analysis, we use a variety of measures to assess the representativeness of each pilot country sample.

Afghanistan (High Feasibility Country)

We find several notable differences and similarities between the population and our sample demographics. Overall, we find an absolute average difference of 11 percentage points from the population. Two specific demographic groups were under-represented in our sample. First, while rural respondents accounted for nearly 60 percent of the sample, they were nationally under-represented by 20 percentage points. Second, the female population was under-represented by 28 percentage points. More specifically, we find that fewer rural female respondents were an important driver of these results.[53]

However, economic status measures closely tracked the national population. Household radio ownership was within 1 percentage point of DHS estimates. In addition, survey respondents with a finished floor were within 6 percentage points of national estimates. Although mobile phone ownership is often viewed as highly correlated with economic status, we found little evidence of this relationship with these two wealth proxies in Afghanistan.

[52] See Kalton and Flores Cervantes (2003) for a discussion of weight creation and methods to reduce variability and loss of precision.

[53] From the population, we would have expected 828 rural women observations (39 percent of the population). The actual sample only included 233 rural females, or roughly 11 percent of total observations. We note that in person enumerated polling may also underrepresent Afghani women. For example, the Asia Foundation's *2013 Survey of the Afghan People* interviewed 9,260 people, of which, only 38 percent were women.

Figure 13 – Afghanistan Sample versus Population Estimates

	Demographic Variable	Population Over 15	Un-Weighted Sample	*Difference*
Place of Residence	Rural	79%	59%	*-20%*
	Urban	21%	41%	*20%*
Gender	Male	50%	77%	*28%*
	Female	50%	23%	*-28%*
Wealth Measure #1	Finished Floor	38%	33%	*-6%*
	Unfinished Floor	62%	67%	*6%*
Wealth Measure #2	Radio	67%	66%	*-1%*
	Radio no	33%	34%	*1%*
Age	Age 15-24	40%	61%	*21%*
	Age 25-34	21%	30%	*9%*
	Age 35-54	26%	8%	*-18%*
	Age 55+	12%	1%	*-11%*
Education	No formal	58%	46%	*-3%*
	Primary some	9%	20%	*-12%*
	Primary completed	5%	8%	*11%*
	Secondary some	5%	7%	*2%*
	Secondary complete	3%	3%	*0%*
	Secondary plus	20%	17%	*3%*
Language	Dari	50%	63%	
	Pashto	35%	37%	
	Total Observations	-	**2,123**	-

Next, we applied the weighting technique to conform the sample to the population parameters.[54] From these weights, we approximate a design effect of 6.3. This would suggest an effective sample size of 337, or an 84 percent reduction compared to the raw sample. The effective sample size indicates a margin of error equaling slightly more than +/- 5 percent. The design effect is primarily driven by a few observations receiving large weights. Forty-seven observations receive a weight greater than 5.

Ethiopia (Low Feasibility Country)

Among the four surveyed countries, the Ethiopia sample differed the most from population demographics. This was an expected outcome based on Ethiopia's low mobile phone

[54] The minimum weight created is 0.052, which was applied to 31 urban males with some primary education who have a radio and unfinished floors. The maximum weight is 62.067 applied to a single respondent who is rural, female, over the age of 55, with more than a secondary education, a radio, and a finished floor. Similarly large weights (38.478) are applied to two respondents who are rural females over the age 55 with radios but without formal education and finished floors.

penetration rates and the concentration of phone ownership amongst wealthier urban residents. Overall, there is an average absolute difference of 19 percentage points between the sample and population demographic measures. As expected, we find a large difference in the place of residence (urban or rural), with urban respondents being over-represented by 43 percentage points. We also find that female respondents are 27 percentage points under-represented. Moreover, survey respondents tended to be more educated than the general population, specifically those with secondary education or higher.

We find mixed results in terms of the economic status measures. Our raw sample perfectly matched household ownership of chairs within the national population. However, we find a marked difference in terms of finished roofs. This suggests that the possession of a finished roof is more highly correlated with mobile phone ownership. Based on this, we caution against relying on a single asset as a measure of socio-economic status. In addition, future researchers may want to include more than two assets as proxies for respondents' economic status.

Figure 14 – Ethiopia Sample versus Population Estimates

	Demographic Variable	**Population Over 15**	**Un-Weighted Sample**	***Difference***
Place of Residence	Rural	78%	35%	*-43%*
	Urban	22%	65%	*43%*
Gender	Male	48%	74%	*27%*
	Female	52%	26%	*-27%*
Wealth Measure #1	Finished Roof	49%	80%	*31%*
	Unfinished Roof	51%	20%	*-31%*
Wealth Measure #2	Chair	44%	44%	*0%*
	No Chair	56%	56%	*0%*
Age	Age 15-24	34%	38%	*4%*
	Age 25-34	24%	41%	*17%*
	Age 35-54	26%	18%	*-8%*
	Age 55+	15%	3%	*-13%*
Education	No formal	50%	8%	*-42%*
	Primary some	34%	27%	*-7%*
	Primary completed	4%	11%	*7%*
	Secondary some	6%	8%	*2%*
	Secondary complete	1%	11%	*10%*
	Secondary plus	5%	35%	*31%*
Language	Amharic	29%	72%	
	Oromo	34%	28%	
	Sidamo	4%	0%	
	Somali	6%	0%	
	Tigrayan	6%	0%	
	Total Observations	-	**2,258**	-

For the Ethiopian sample, the statistical package was unable to converge the weights[55], leaving the maximum deviation at 0.105 percent on the "no formal education" demographic. The sample weights suggest a design effect of 11.6, the largest of all of the surveys. This would suggest an effective sample size of 194, or a 91 percent reduction compared to the raw sample. The effective sample size indicates a margin of error of +/- 7 percent. Eighty-one observations receive a weight greater than 5.

Mozambique (Low Feasibility Country)

The Mozambique sample differs less from the population than Ethiopia, but more than the other two surveyed countries. The average absolute difference between the sample and population demographic measures is 18 percentage points. Once again, we find that rural and female populations are significantly under-represented in the sample, 28 percentage points and 19 percentage points respectively. In contrast to Afghanistan and Zimbabwe, we also find that the sample differs significantly in measures of economic status. Our survey respondents have greater access to toilet facilities and finished floors. Lastly, the survey sample tends to be younger and better educated than the broader Mozambican population.

Next, we applied the weighting technique to conform the sample to the population parameters.[56] The sample weights suggest a design effect of 5.2, less than Afghanistan and half of Ethiopia. This would suggest an effective sample size of 430, or an 81 percent reduction compared to the raw sample. The effective sample size indicates a margin of error of +/- 4.7 percent. Sixty-three observations receive a weight greater than 5.

[55] The smallest weight is 0.00016 on 17 urban, male respondents ages 15-24 with a complete secondary education, finished roof, but without a chair in the household. The largest weight is 51.38 on single rural female respondent without a formal education or finished roof, but with a chair in the household.

[56] The minimum weight is 0.028 applied to 18 urban male respondents in the age bracket 15-24 who have finished floors, toilet facilities, and more than a secondary education. The maximum weight is 27.048 applied to two rural female respondents in the 55 plus age group without toilet facilities, finished floors, or formal education.

Figure 15 – Mozambique Sample versus Population Estimates

	Demographic Variable	Population Over 15	Un-Weighted Sample	*Difference*
Place of Residence	Rural	68%	41%	*-28%*
	Urban	32%	59%	*28%*
Gender	Male	48%	67%	*19%*
	Female	52%	33%	*-19%*
Wealth Measure #1	Finished Floor	44%	61%	*16%*
	Unfinished Floor	56%	39%	*-16%*
Wealth Measure #2	Toilet Facility	61%	74%	*14%*
	No Toilet Facility	39%	26%	*-14%*
Age	Age 15-24	32%	52%	*20%*
	Age 25-34	24%	35%	*11%*
	Age 35-54	29%	11%	*-19%*
	Age 55+	15%	2%	*-12%*
Education	No formal	41%	12%	*-29%*
	Primary some	44%	24%	*-20%*
	Primary completed	5%	10%	*5%*
	Secondary some	7%	23%	*16%*
	Secondary complete	1%	21%	*20%*
	Secondary plus	1%	10%	*10%*
Language	Portuguese	15%	85%	
	Changana	11%	8%	
	Makua	26%	5%	
	Chuabo	7%	2%	
	Sena	8%	0%	
	Total Observations	-	**2,232**	-

Zimbabwe (High Feasibility Country)

While the raw Zimbabwe survey sample differs from the national population in a few important ways, it more closely reflects the population than the other three surveyed countries. The average absolute difference between the sample and population demographic measures is only 7 percentage points. Most notably, the sample under-represents the rural and female populations. Although to lesser degree than in the three other countries, we again encounter a missing rural female population of roughly 14 percentage points.[57] The sample also seems to over-represent respondents with more than a secondary education.

[57] The sample includes 480 rural females or 22 percent of respondents, while the population parameters would suggest a sample of 789 rural females, or 36 percent of all respondents.

As with Afghanistan, our sample does *not* appear skewed toward wealthier individuals. The first economic status measure (finished floors) perfectly matched the DHS estimates. However, the radio ownership variable is 22 percentage points off compared to the population parameters. This likely reflects the Zimbabwean government's ban on short wave radios, which was imposed in early 2013 after the DHS survey was completed.[58] Thus, ownership of radios could have changed, or some respondents may have declined to acknowledge owning potentially banned radios. We failed to account for this local dynamic during the project design phase. Given this material shortcoming, we excluded the radio ownership variable when creating sample weights.[59]

Figure 16 – Zimbabwe Sample versus Population Estimates

	Demographic Variable	**Population Over 15**	**Un-Weighted Sample**	***Difference***
Place of Residence	Rural	65%	53%	*-12%*
	Urban	35%	47%	*12%*
Gender	Male	46%	58%	*12%*
	Female	54%	42%	*-12%*
Wealth Measure #1	Finished Floor	73%	73%	*0%*
	Unfinished Floor	27%	27%	*0%*
Wealth Measure #2	Radio	41%	63%	*22%*
	No Radio	59%	37%	*-22%*
Age	Age 15-24	33%	32%	*-1%*
	Age 25-34	25%	40%	*15%*
	Age 35-54	26%	24%	*-2%*
	Age 55+	16%	3%	*-13%*
Education	No formal	8%	6%	*-2%*
	Primary some	15%	13%	*-2%*
	Primary completed	16%	10%	*-6%*
	Secondary some	23%	19%	*-4%*
	Secondary complete	33%	36%	*3%*
	Secondary plus	5%	17%	*12%*
Language	English	40%*	13%	
	Shona	20%	77%	
	Ndebele	40%	10%	
	Total Observations	-	**2,192**	-

***Includes second language speakers

[58] Source: IRIN, "Zimbabwe police ban radios, crack down on NGOs", February 26, 2013. http://www.irinnews.org/report/97549/zimbabwe-police-ban-radios-crack-down-on-ngos

[59] Although in the section on Ethiopia we cautioned against using a single asset to approximate economic status, the contextual evidence that the radio ownership rate is an unreliable measure leads us to discount it in Zimbabwe.

Next, we weighted the sample to match population parameters.[60] The sample weights suggest a design effect of only 1.8. This would suggest an effective sample size of 1,196, or a 45 percent reduction compared to the raw sample. This hypothetical sample would have a sample error of +/- 2.8 percent. There are 31 observations with a weight greater than 5.

Summary Findings

These four country-level experiences suggest two key sets of conclusions. First, looking at correlations, mobile penetration rates appear to have a substantial effect on the ability to produce nationally representative samples. With appropriate caveats, we are confident that the Zimbabwe survey sample, with modest weighting, broadly reflects the national population. In sharp contrast, we have little confidence that the Ethiopia results are nationally representative. In Mozambique and Afghanistan, the results are more mixed and nuanced. Mozambique's *raw* sample is almost as non-representative as Ethiopia's, but the design effect (following weighting) is significantly lower than Afghanistan.

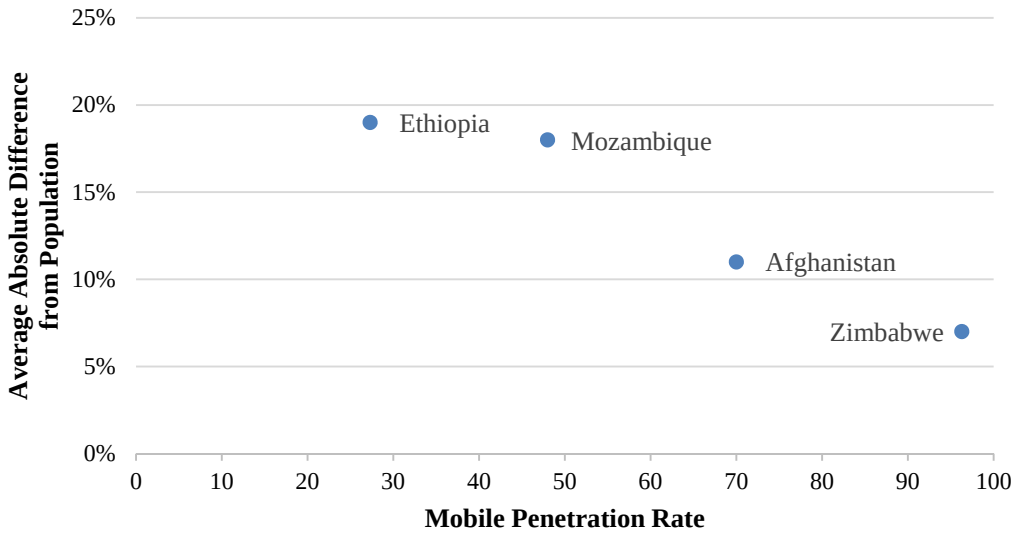
Second, while penetration rates are clearly important, who adopts mobile technology matters as well. Afghanistan's mobile penetration rate is 22 percentage points higher than Mozambique. However, cultural norms have slowed the adoption and usage of mobile technology by rural Afghani women.[61] This factor inflates the weights required to reflect the broader population, which lowers the degree of sample precision. Although we found a bias against rural women across countries, it was particularly acute in Afghanistan. These findings raise important questions about whether mobile phone surveys are well suited for capturing viewpoints from rural women. Moreover, there is a need to test additional techniques and methods aimed at better reaching this important demographic group.[62]

[60] The minimum weight is 0.125, which is applied to 7 respondents who are urban males in the 25 to 34 age bracket with unfinished floors and more than a secondary education. The maximum weight is 9.352, applied to three rural females over 55 years old with finished floors and a complete primary education.

[61] See GSMA (2012).

[62] By illustration, future surveys could be recorded using a local, female voice instead of a male voice.

Figure 17 – Mobile Penetration Rate versus Sample and Population Difference



Recognizing that the effective sample size increases sample imprecision, we naturally ask how many observations would be needed to obtain an effective sample size with commonly accepted levels of precision. This information is presented in Figure 18 below.[63] Countries with smaller design effects unsurprisingly require few observations to reach acceptably precise effective sample sizes. Fewer observations are associated with lower costs, although price information is highly variable by country. These estimated costs are extrapolated from the executed surveys' costs, including variable costs such as airtime, incentives (see part E of this section for further details), transcription of a voice recorded question, and our experimentation with demographic filtering (see part A of this section for further details).

Figure 18 – Required Observations to Create Effective Sample Size with Estimated Cost

Effective Sample Size	Afghanistan	Ethiopia	Mozambique	Zimbabwe
1000 (+/- 3.1 percent)	6,300	11,600	5,200	1,830
	($23,783)	($65,165)	($26,431)	($14,343)
2000 (+/- 2.19 percent)	12,600	23,200	10,400	3,660
	($47,216)	($129,615)	($51,672)	($28,087)

[63] These costs reflect our experiences and are broadly illustrative of this technique. However, they do not reflect pricing quotes.

D. Linguistic Fractionalization and Nationally Representativeness

Linguistic fractionalization appears to play a role in obtaining a nationally representative sample. In this section, we outline several operational challenges while working in two linguistically fractionalized countries (Ethiopia and Mozambique). Then, we observe how linguistically fractionalized environments may have influenced the samples' ability to represent the national population in both the language selection phase and through respondents' survey attrition patterns.

Operationally, translating the survey into multiple languages was more difficult than anticipated. The survey was translated in 15 languages. The survey operator experienced challenges in identifying translators for several of the less common local languages. After finding acceptable translators, these languages' translations seemed to take longer, encountering multiple and extended delays. For example, the Sidamo translation delayed the launch of the Ethiopian survey by roughly one month. The low response rates for these minor local languages made the lengthy delays particularly problematic.[64]

Beyond these operational considerations, we first address the implications of linguistic fractionalization at the language selector phase of the survey, and then address how it may have influenced attrition rates. At the language selector stage, we observed major differences between the population and sample demographics in terms of language.[65] In Zimbabwe, an estimated 40 percent of the population speaks Shona, while 76 percent of the sample selected it. In Ethiopia, we found another example when 85 percent of the sample selected Amharic, but only an estimated 29 percent of Ethiopians speak it as their primary language. We observe this trend in both linguistically fractured and less fractionalized countries.

[64] For example, only 28 Ethiopian respondents chose Somali from the language selector, and none of these went on to complete the survey. Only 38 respondents selected Sidamo at the beginning of the survey, and only one of them completed the survey. For Tigrayan, 242 respondents chose it, and only 7 of them completed the survey. Overall, nearly 100 percent of Ethiopians completed the survey in the two primary languages (Amharic and Oromo). In Mozambique, we found a similar trend with Sena, where 42 respondents chose the language and none of them completed the survey. While we erred on the side of linguistic inclusiveness, future researchers may wish to consider these experiences when selecting the survey instrument languages.

[65] We caution against heavy comparison between the estimated proportion and population. As discussed in this section, respondents' answers to the language selector may not reflect the respondents' true first language. Also, the estimates are compiled from multiple sources, including the most recent Afrobarometer (in applicable countries), the CIA World Factbook, and www.ethnologue.com. Although we feel confident about using these composite figures as rough estimates, they may not reflect exact population proportions. For this reason, we chose not to weight on this variable and urge appropriate caution when making analytical comparisons.

However, these trends are particularly prevalent in Ethiopia and Mozambique, where several minority languages appear severely under-represented in our sample.[66]

At least two factors may explain the disparities between expected population demographics and respondents' language selection decisions. First, when the languages were listed, the order remained unchanged, placing the most widely spoken first and the least widespread language last. Respondents may have chosen the first language in which they felt comfortable, particularly in Mozambique and Ethiopia, where respondents had up to five language choices.[67] Moreover, if respondents exhibited this type of selection behavior, then there may have been varying levels of fluency in the selected language. In the future, this potential bias could be mitigated by randomly changing the order of languages included in the selector question.

We acknowledge an alternative explanation that may more fundamentally bias the sample. Minority language speakers may have lower mobile penetration rates than majority speakers because the adoption of mobile phones may not expand evenly across the population. To the extent that the language spoken is correlated with outcomes of interests, this alternative explanation could raise more fundamental questions about mobile surveys' ability to produce nationally representative surveys.

Second, we find different rates of survey completion across the selected languages. This evidence confirms our hypothesis that countries with greater rates of linguistic fractionalization would exhibit higher rates of survey attrition. We find that some minority languages have much higher attrition rates or even no completed surveys at all. This trend may be partially driven by small numbers of minority language respondents starting the survey in the first place.

Again, we identify two possible explanations that may explain this observation. First, the translation quality may have varied across languages. Although measures were taken to ensure that each translation was high quality in terms of content and audio comprehensibility, the possibility remains that the survey was more difficult to complete in

[66] For example, the estimated percent versus selection of language is under-represented for Sidamo (3.6 percentage points), Somali (5.7 percentage points), and Tigrayan (4 percentage points) in Ethiopia, and for Sena (7 percentage points) in Mozambique.

[67] By illustration, in Mozambique, until recently all formal schooling occurred in Portuguese. Although only an estimated 15 percent of the population speaks Portuguese as their first language, a much greater proportion may use Portuguese as a second language.

some languages. Another possible explanation is the selected language correlates with other unobserved factors, which may reflect a reduced capacity to complete the survey. For example, speaking minority languages may also correlate with lower education or less familiarity with technology. Thus, we would observe higher attrition rates with these languages. Since DHS questionnaires do not standardly cover linguistic issues, we were unable to mitigate this potential phenomenon during the project design stage or monitor its impact during the analysis stage.

Figure 19 – Completion Rates by Language

	Estimated Population	Percent Answered Language Selector	Percent of Completed Sample	Percent Completing By Language Selected
Afghanistan				**24%** (*national avg*)
Dari	50%	61%	62%	25%
Pashto	35%	39%	37%	23%
Ethiopia				**21%** (*national avg*)
Amharic	29%	85%	72%	18%
Oromo	34%	12%	28%	48%
Sidamo	4%	0.4%	0.04%	3%
Somali	6%	0.3%	0%	0%
Tigrayan	6%	2%	0.3%	3%
Mozambique				**38%** (*national avg*)
Changana	11%	9%	8%	32%
Chuabo	7%	9%	2%	9%
Makua	26%	12%	4%	14%
Portuguese	15%	68%	86%	48%
Sena	8%	1%	0%	0%
Zimbabwe				**46%** (*national avg*)
English	~40%	14%	12%	39%
Ndebele	20%	10%	8%	37%
Shona	40%	76%	80%	48%

Overall, our findings suggest that working in linguistically fractionalized countries may influence the ability to obtain a nationally representative sample. However, fully understanding how these contexts influence the sample is still an area for further research. These four countries only start to illuminate the interplay between technological constraints and contextual realities. Many of the practical technological challenges may be resolved in future surveys, while researchers may only be able to mitigate other on-the-ground realities. In this manner, many of these concerns may also need to be resolved on a country-by-country basis through further pilots.

E. Impact of Response Incentives

Improving survey response and completion rates is a primary challenge for researchers and practitioners. One common solution is to provide monetary compensation to respondents. For instance, in the field of mobile surveys, the provision of extrinsic incentives such as airtime rewards has been seen as a quick fix to deal with the non-response issue. Yet, there are significant reasons to further examine whether monetary incentives are effective at improving mobile response rates.

More generally, the literature on incentives to promote specific pro-social behaviors has shown a number of cases where the opposite effect is achieved, thereby leading to a crowding out of incentivized behavior.[68] With regard to surveys (e.g. postal, interviewer-mediated), most studies point to a positive effect of incentives on response rates across most related modalities, including web-based ones, cross-sectional, and panel studies.[69] Yet, the evidence is mixed when examining specifically the case of mobile phone-based surveys, with some cases reporting no effects of monetary compensations.[70]

Thus, to further test the effect of incentives on mobile surveys, we randomized the introduction that respondents received when they started the survey, testing two treatment conditions.

- ✓ All respondents received an introduction with a sentence explaining the survey's purpose and how the data could be used in the future. One-third of respondents only received this message (control group).

- ✓ The first treatment (transfer condition) also informed respondents that, upon completing the survey, they would receive 4 minutes of mobile airtime.

- ✓ The second treatment (raffle condition) also informed individuals that, upon completing the survey, they would be entered into a raffle to win a substantial amount of mobile airtime.[71]

Before presenting our results, we briefly outline the adjustments done prior to the data analysis. To begin, we included only the first contact attempt that was made to a given

[68] For an overview of when and why incentives may work, see Gneezy et al. 2011.

[69] For a review of the literature on the effects of incentives on surveys, see Singer & Ye 2013.

[70] See for instance, Oldendick & Lambries 2013.

[71] One percent of respondents received two hours of airtime.

number. Many of the phones were called multiple times before someone answered. Multiple calls are a more efficient way to get responses, but they complicate experimental analysis. Consequently, we excluded subsequent attempts since receiving previous call attempts might affect how likely someone is to complete the survey. The second adjustment restricts the dataset to just those respondents who got as far as the introduction and excludes any dropped calls or subjects who hung up before hearing the introduction. In order to look only at the effect of the message as received by the respondent, we calculate response rates only for those individuals who select a language to hear the message in. This excludes the vast majority of calls made in the dataset, as most subjects either hung up before making a language selection or did not pick up the call at all (see sub-section B in this section). However, this adjustment rests on the assumption that there are no differences between the experimental groups prior to receiving the treatment. We examine this assumption in more detail further below.

Results and Discussion of Findings

Pooling the data from all four surveys together shows that both the raffle and transfer conditions have significantly higher completion rates than the control conditions (see figure 20 below). However, the pooled analysis does not indicate that either treatment condition was more effective than the other.

However, before presenting these findings in detail, a major caveat should be noted: the results rest on the assumption that random assignment was done correctly and hence that there are no differences between the experimental groups prior to treatment. Unfortunately, our analysis (see appendix IV) suggests a failure in random assignment for Zimbabwe and Ethiopia. The data show that there are significant differences in survey start rates across experimental groups in these two cases. This observation indicates an unintended selection effect caused by which respondents are selected into each treatment. We, therefore, cannot exclude the possibility that apparent treatment effects may instead actually be selection effects. Consequently, the observed effects in these two countries may well be caused by factors other than the incentives themselves. In other words, from an experimental perspective, we are unable to make any definitive claims on the effects of incentives beyond the cases of Mozambique and Afghanistan.[72]

[72] While this limitation is disappointing, this exercise highlights a valid lesson for researchers and practitioners in the field of mobile phone surveys. One of the advantages of mobile surveys is the ability to

Recognizing these shortcomings, we find substantial variation in the effects across the different country contexts. Afghanistan shows no significant differences between any of the experimental conditions. Both treatment conditions are significant in Ethiopia but there is no observed difference between the effect of giving respondents a raffle reward and giving them a direct mobile credit transfer.[73] The raffle reward gives a significantly higher completion rate in Mozambique when compared with the control group. Taking these results together, we can see that the raffle is the most consistently effective treatment across three of the countries.

Figure 20 – Impact of Incentives on Survey Completions

Country	No Reward	Raffle	Transfer	Sig. Difference Between Treatments
Afghanistan	30.4	30.2	32.2	
Ethiopia	18.4	20.8*	20.9*	
Mozambique	33.4	37.6*	34.9	†
Zimbabwe	43.1	52.1*	54.7*	*
Combined	*28.4*	*31.7**	*32.1**	

Treatment is significantly different from no reward using a chi-squared test

** p<0.05 † p <0.1*

Another question refers to the cost-effectiveness of incentives. All connected calls are charged for a minimum of 30 seconds. Therefore, respondents who do not complete the survey represent sizable costs. If incentives raise the completion rate enough, then the reduced airtime costs could more than offset the cost of incentive payments. However, paying incentives also increases the fixed cost per completion. Since many of the respondents who receive incentives complete the survey anyway, the incentive paid to these individuals does not increase the number of responses to the survey.

Figure 21 below demonstrates the cost breakdown for the different methods. All countries illustrate lower airtime costs per completed survey in the transfer and raffle conditions. However, when accounting for the additional cost of incentives, the transfer condition is

conduct multiple experiments at reduced costs and on an ongoing basis. These experiments, for their part, may provide valuable information on how to maximize the cost-effectiveness of surveys. Nevertheless, as our experience shows, the technical challenges for implementing true experimental designs are still present. From a practical standpoint, the effort required for data cleaning, verification and validation – particularly with regards randomization – should not be underestimated.

[73] Bearing in mind that these results should be interpreted cautiously in the cases of Mozambique and Afghanistan given the problems noted with randomization.

much more expensive per completed survey than the control condition (e.g., no reward). The raffle condition costs roughly the same per completed survey as the control condition. This means that the slightly higher costs from providing financial rewards cancels out the slightly lower airtime costs. Overall, these results suggest that transfer payments are not a cost effective way of recruiting respondents.[74]

Figure 21 – Cost per Completed Response, Incentive Treatments

Country	Group	Total Cost Before Incentive	Total cost	Total Completions	Airtime Cost per Completion	Total Cost per Completion
Mozambique	No reward	$976	$976	670	$1.46	$1.46
Mozambique	Raffle	$944	$981	673	$1.40	$1.46
Mozambique	Transfer	$931	$1,342	664	$1.40	$2.02
Afghanistan	No reward	$826	$826	594	$1.39	$1.39
Afghanistan	Raffle	$849	$876	626	$1.36	$1.40
Afghanistan	Transfer	$829	$1,095	605	$1.37	$1.81
Ethiopia	No reward	$1,303	$1,303	639	$2.04	$2.04
Ethiopia	Raffle	$1,266	$1,340	659	$1.92	$2.03
Ethiopia	Transfer	$1,221	$2,031	658	$1.86	$3.09
Zimbabwe	No reward	$1,271	$1,271	599	$2.12	$2.12
Zimbabwe	Raffle	$1,441	$1,477	702	$2.05	$2.10
Zimbabwe	Transfer	$1,476	$1,914	730	$2.02	$2.62

It should be noted, however, that this analysis only considers the average cost of obtaining a completed survey. This ignores the benefit of having an overall higher response rate on the survey. Whether transfer payments are a cost effective method of increasing response rates comes down to each practitioner's willingness to pay for a higher response rate. Nevertheless, the increase in cost per respondent is substantial between the no reward and transfer conditions, so there may be cheaper ways to increase response rates.

Another question is whether different socio-economic groups are more affected than others by incentives. Due to challenges with the randomization (see discussion below) we limit our analysis here to Mozambique and Afghanistan. A likely hypothesis is that – all other things equal – poorer individuals are more responsive to monetary compensation. Yet, surprisingly, individuals across different income groups seem to be equally affected by incentives. In a similar vein, we found no difference on age, gender or education groups with regard to their response to the incentives treatment. The only exception found was in Afghanistan, with urban dwellers more likely to react to monetary incentives than rural ones.

[74] Idem.

Bearing these considerations in mind, a number of questions remain to be answered. The first one refers to the external validity of our findings. That is, the extent to which they can be generalized to other contexts. The second one refers to whether or not an optimal amount of incentives can be identified and, if so, what these amounts are. A third issue refers to the impact of monetary compensation on the quality of responses provided: are respondents who receive incentives equally likely to provide accurate answers? A fourth question concerns the medium and long-term effects of extrinsic incentives on respondents, and the extent to which they may crowd-out the willingness to respond to future surveys in the absence of rewards. Finally, the cost-effectiveness of incentives when compared to other efforts to increase response rates and non-response bias remains an open issue.[75] These questions, we contend, are essentially empirical ones and cannot be answered without further research.

VI. Summary and Conclusion

The exponential growth in mobile phone technology throughout the developing world has provided a powerful platform for engaging citizens for a range of purposes, including quickly gathering data from thousands of people. At the same time, many development agencies have recognized an operational need for rapid, accurate, flexible, and cost-efficient tools for gathering information from, and engaging with intended beneficiaries, throughout the project life cycle. In this context, mobile phone based approaches may be able to augment traditional data collection techniques, such as household surveys or consultative roundtables. Despite this potential, few studies have rigorously assessed the feasibility of mobile phone based survey techniques.

We have attempted to address some of these gaps in the policy research literature. Through our pilot project, there are a number of key analytical findings and lessons learned in terms of survey design and implementation. These include:

- ✓ Including the Poor: The survey performed reasonably well at reaching poor inhabitants in the focus countries, especially in Afghanistan and Zimbabwe.

- ✓ Reaching Rural Women: Rural women were consistently under-represented in the four countries' surveys, especially in Afghanistan and Ethiopia. Cultural norms and

[75] Singer and Ye, 2013.

mobile phone ownership patterns likely are the most important drivers.

- ✓ Higher Phone Penetration Impact: Samples from countries with higher mobile penetration rates differed less from the actual population, as expected.

- ✓ Sample Precision: After weighting on demographic variables, sample imprecision was a challenge in the two lower feasibility countries (Ethiopia and Mozambique) with a sampling error of +/- 5 to 7 percent. In Zimbabwe, the sample's estimates were more precise (sampling error of +/- 2.8 percent).

- ✓ Linguistic Fractionalization: Countries' linguistic fractionalization may influence the ability to obtain nationally representative samples, although a material effect on survey response or completion patterns was difficult to discern through other factors, such as penetration rates and market composition. Operationally, translating the survey into multiple languages was more time consuming than anticipated, which contributed to project implementation delays.

- ✓ Survey Completion Incentives: Compensating for survey completion, either through an airtime raffle or transfer, mitigated attrition rates in several of the pilot countries. However, the effects varied across countries and cultural settings.

Although our findings only start to rigorously assess the feasibility of mobile surveys, they suggest that mobile phone based mechanisms may be a promising approach for promoting representative engagement with citizens. Future research should continue to explore their feasibility across varying mobile penetration and linguistic environments. From our results, we have identified several key implications that merit further consideration, testing, and refinement.

- ✓ Already Reaching Urban Populations: Even in contexts with low mobile penetration, urban populations across demographics categories seem to have already adopted mobile technology. If future research aims to survey urban populations, it may already be feasible in low mobile penetration contexts, while nationally representative samples in these countries would be less feasible.

- ✓ Better Reaching Rural Women and Other Under-Represented Groups: There was a systematic challenge in reaching rural women in our four survey samples. This

illustrates the need to test a variety of additional approaches to reach under-represented demographic groups while still maintaining rigorous statistical methods. For instance, researchers could experiment with calling at different times of the day, using a female voice for the recorded survey, and/or using push notices.

- ✓ Piloting Methods to Increase Response and Completion Rates: While there are country-specific dynamics, the largest overall driver of mobile phone survey costs are: (1) non-responses; and (2) incomplete responses. Low response and completion rates can lead to significant cost overruns. Given this, researchers may wish to test additional methods for increasing response and completion rates, such as SMS notices ahead of survey calls.

Appendix I

Survey Instrument

I. Language Selection

If you would like to continue in [language], press 1.
If you would like to continue in [language], press 2.
If you would like to continue in [language], press 3.

II. Introduction

Hello. You have been randomly selected to participate in a survey representing [countrymen]. We're an independent research institution, and your answers will help inform important decisions for your country. The survey is only [ten] questions long and should only take 4 minutes to complete.

[*Insert incentive: 33 percent no incentive sentence, 33 percent guaranteed airtime; 33 percent lottery*]

[If you complete the whole survey, we will put 4 minutes of airtime on this mobile.]
[If you complete the whole survey, you will have a chance to win two hours of airtime on this mobile.]

Please know that your answers will be kept strictly confidential.

III. Targeting Information

(1) Do you live in a village or in a city?

If you live in a rural area or village, press one
If you live in an urban area or city, press two.

(2-3) Asset Questions Country Specific

[*Afghanistan*]
(2 – Afghanistan) What best describes your home's type of floor?

If your floor is finished, for example a tile or cement floor, press one.
If your floor is not finished, for example a sand or dirt floor, press two.

(3 – Afghanistan) Do you or anyone at home own a radio?

If you or anyone at home owns a radio, press one.
If you or anyone at home does not own a radio, press two.

[*Ethiopia*]
(2 – Ethiopia) What best describes your home's type of roof?

If your roof is finished, for example a metal, tile, or cement roof, press one.
If your roof is unfinished, for example a thatch or mud roof, press two.

(3 – Ethiopia) Do you or anyone at home own a chair?

If you or anyone at home owns a chair, press one.
If you or anyone at home does not own a chair, press two.

[*Mozambique*]
(2 – Mozambique) What best describes your home's type of floor?

If your floor is finished, for example a tile or cement floor, press one.
If your floor is not finished, for example a sand or dirt floor, press two.

(3 – Mozambique) What best describes the toilet at your house?

If you have somewhere you only use as a toilet, like a pit latrine or flush toilet, press one.
If you do not have a place you only use as a toilet, press two.

[*Zimbabwe*]
(2 – Zimbabwe) What best describes the type of floor you have at your house?

If your floor is finished, for example tile or cement floor, press one.
If your floor is not finished, for example a sand or dirt floor, press two.

(3 – Zimbabwe) Do you or anyone at home own a radio?

If you or anyone at home owns a radio, press one.
If you or anyone at home does not own a radio, press two.

IV. Other Demographic Information

(4) Are you a man or a woman?

If you are a man, press one.
If you are a woman, press two.

(5) How old are you?

If you are 15-24 years old, press one.
If you are 25-34 years old, press two.
If you are 35-55 years old, press three.
If you are more than 55 years old, press four.

(6) How much schooling did you attend?

If you attended no formal schooling, press one.
If you attended some primary school, press two.
If you finished primary school, press three.
If you attended some secondary school, press four.
If you completed secondary school, press five.
If you attended more than secondary school, press six.

V. Citizen Preference and Information

[Option A (no timeframe) 50 percent / Option B (long-term timeframe) 50 percent]

Survey Option A – No timeframe *[Randomize order of questions 7 and 8]*

(7/8A) In your opinion, of the following options, what could the government do to most improve your standard of living?

(7/8A) In your opinion, of the following options, what could international actors, like foreign aid agencies and NGOs, do to most improve your standard of living?

[Note: Survey Option A1 – Open ended, 5 percent]

[*2.5 percent*] In your opinion, what could the government do to most improve your standard of living?

[*2.5 percent*] In your opinion, what could international actors, like foreign aid agencies and NGOs, do to most improve your standard of living?

Survey Option B – Long-term timeframe *[Randomize order of questions 7 and 8]*

(7/8B) In your opinion, of the following options, what could the government do to most improve your standard of living over the next five years?

(7/8B) In your opinion, of the following options, what could international actors, like foreign aid agencies and NGOs, do to most improve your standard of living over the next five years?

[*Note: Survey Option B1 – Open ended, 5 percent*]

[*2.5 percent*] In your opinion, what could the government do to most improve your standard of living over the next five years?

[*2.5 percent*] In your opinion, what could international actors, like foreign aid agencies and NGOs, do to most improve your standard of living over the next five years?

[Options part 1 for 7 and 8]

Economic issues, such as unemployment and cost of goods, press one.
Infrastructure, such as roads, electricity, and water, press two.
Crime and security, such as personal safety and discrimination, press three.
Social services, such as health and education, press four.
Land use and protecting the environment, such as rivers, forests, and farms, press five.

[Options part 2 for 7 and 8]

[If economic issues] Thinking of specific economic issues, please select the area that you think should be focused on most.

Unemployment, press one.
Cost of goods, press two.
Access to credit, press three.
Wages, press four.
Taxes, press five.

[If infrastructure] Thinking of specific infrastructure, please select the area that you think should be focused on most.

Roads, press one.
Electricity, press two.
Water, press three.
Housing, press four.

[If crime and security] Thinking of crime and security, please select the area that you think should be focused on most.

Street crime, press one.
Domestic violence, press two.
Instability, press three.
Discrimination, press four.
Bribes/corruption, press five.
Ability to get justice through courts, press six.

[If social services] Thinking of specific social services, please select the area that you think should be focused on most.

Health, press one.
Education, press two.
Support for orphans, street children, and the elderly, press three.

[If land use and environment] Thinking of land use and the environment, please select the area that you think should be focused on most.

Droughts, press one.
Food shortages, press two.
Land ownership, press three.
Deforestation, press four.
Pollution, press five.

Appendix II

Demographics of Household Ownership of Mobile Phones

Demographic Group	Afghanistan	Ethiopia	Mozambique	Zimbabwe
Population	74%	24%	38%	63%
Male	74%	24%	38%	63%
Female	74%	23%	38%	63%
Urban	93%	67%	72%	92%
Rural	70%	14%	22%	50%
Male Rural	70%	15%	22%	51%
Female Rural	69%	13%	23%	49%
Male Urban	93%	68%	72%	92%
Female Urban	93%	67%	72%	92%
1st Wealth Quintile (Poorest)	34%	1%	0.10%	18%
2nd Wealth Quintile (Poorer)	72%	3%	3%	45%
3rd Wealth Quintile (Middle)	79%	11%	25%	63%
4th Wealth Quintile (Richer)	90%	30%	65%	89%
5th Wealth Quintile (Richest)	97%	74%	97%	98%
15-24 yrs	76%	30%	45%	69%
25-34 yrs	75%	30%	44%	71%
35-49 yrs	74%	24%	37%	66%
over 50 yrs	72%	20%	31%	55%
Male 15-24 yrs	77%	30%	46%	69%
Male 25-34 yrs	75%	33%	45%	71%
Male 35-49 yrs	74%	26%	37%	69%
Male over 50 yrs	72%	20%	33%	58%
Female 15-24 yrs	76%	30%	45%	68%
Female 25-34 yrs	74%	28%	44%	71%
Female 35-49 yrs	74%	22%	37%	64%
Female over 50 yrs	72%	19%	29%	53%
No formal	-	15%	25%	55%
Some Primary	-	26%	37%	53%
Primary	-	51%	63%	56%
Some Secondary	-	63%	80%	75%
Secondary	-	87%	96%	95%
Secondary Plus	-	87%	99%	96%

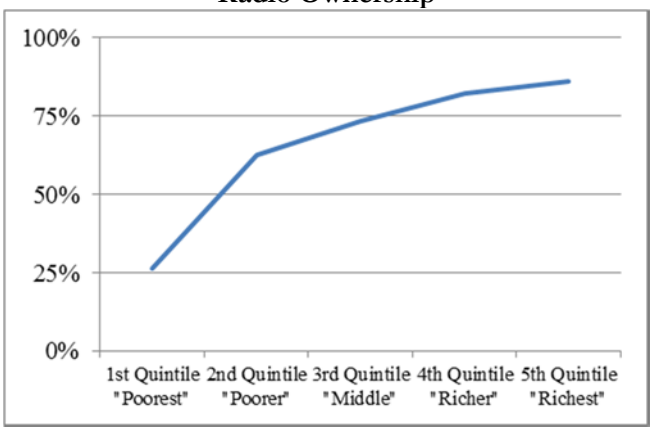
Appendix III

Ownership of Assets by Wealth Quintile

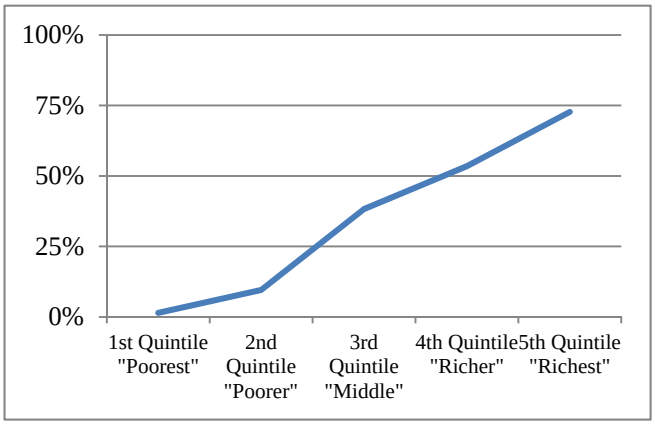
Afghanistan (2010)

Asset	1st Quintile "Poorest"	2nd Quintile "Poorer"	3rd Quintile "Middle"	4th Quintile "Richer"	5th Quintile "Richest"
Agriculture land	72%	70%	66%	62%	44%
Animal drawn cart	0%	0%	3%	3%	2%
Cabinet	7%	26%	44%	60%	84%
Bank Account	0%	3%	4%	5%	19%
Bicycle	4%	19%	35%	53%	56%
Car	0%	3%	7%	19%	37%
Cooking fuel: LPG	1%	6%	11%	20%	63%
Cooking fuel: wood	12%	27%	49%	58%	30%
Electricity	27%	35%	34%	35%	80%
Finished Floor	1%	10%	38%	53%	73%
Finished Roof	1%	10%	32%	45%	39%
Generator	1%	4%	7%	12%	36%
Mattress	92%	96%	96%	98%	99%
Motorcycle	16%	28%	23%	26%	29%
Toilet facility: none	46%	19%	22%	12%	1%
Toilet facility: pit latrine	15%	20%	14%	18%	26%
Radio	26%	63%	73%	82%	86%
Refrigerator	0%	0%	2%	4%	35%
Rickshaw	0%	1%	2%	2%	3%
Standing fan	0%	4%	12%	15%	63%
Television	5%	24%	30%	30%	79%
Video	1%	1%	3%	8%	8%
Watch	58%	83%	89%	96%	98%

Radio Ownership



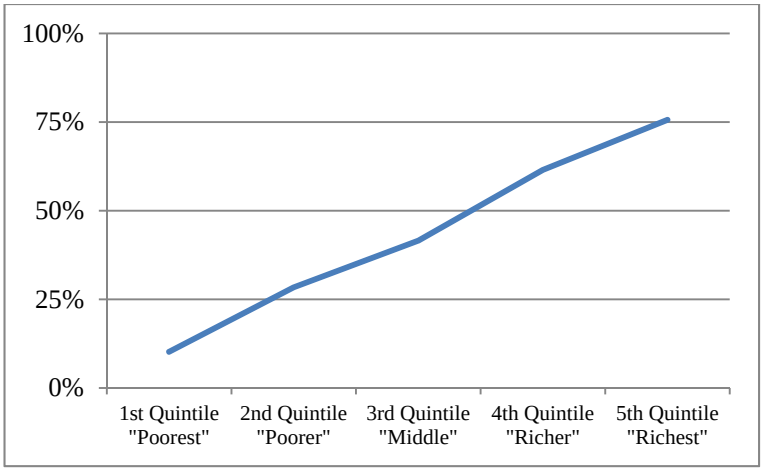
Finished Floors



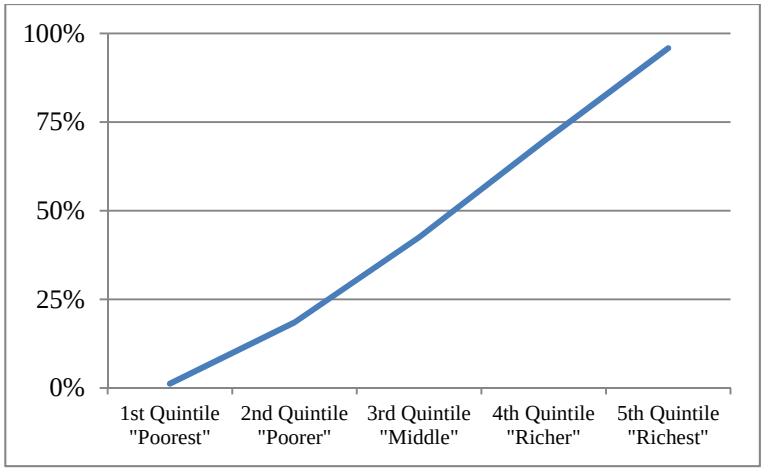
Ethiopia (2011)

Asset	1st Quintile "Poorest"	2nd Quintile "Poorer"	3rd Quintile "Middle"	4th Quintile "Richer"	5th Quintile "Richest"
Bank Account	1%	2%	5%	12%	30%
Bed	0%	3%	9%	31%	75%
Bed net	0%	0%	0%	0%	0%
Bicycle	0%	0%	0%	3%	9%
Car	0%	1%	0%	1%	6%
Chairs	10%	28%	42%	62%	76%
Cooking fuel: charcoal	0%	0%	0%	1%	28%
Cooking fuel: wood	93%	88%	85%	84%	54%
Electricity	0%	0%	0%	9%	84%
Finished Floor	0%	0%	0%	3%	57%
Finished Roof	1%	19%	43%	70%	96%
Lamp	8%	13%	16%	26%	19%
Motorcycle	0%	1%	0%	1%	2%
Toilet facility: none	78%	51%	32%	19%	12%
Radio	11%	29%	39%	60%	73%
Refrigerator	0%	0%	0%	0%	18%
Share toilet	1%	3%	22%	31%	55%
Table	4%	16%	31%	54%	75%
Telephone	0%	1%	0%	1%	6%
Television	0%	0%	0%	1%	47%
Water access on household premises	2%	1%	2%	2%	43%

Chairs



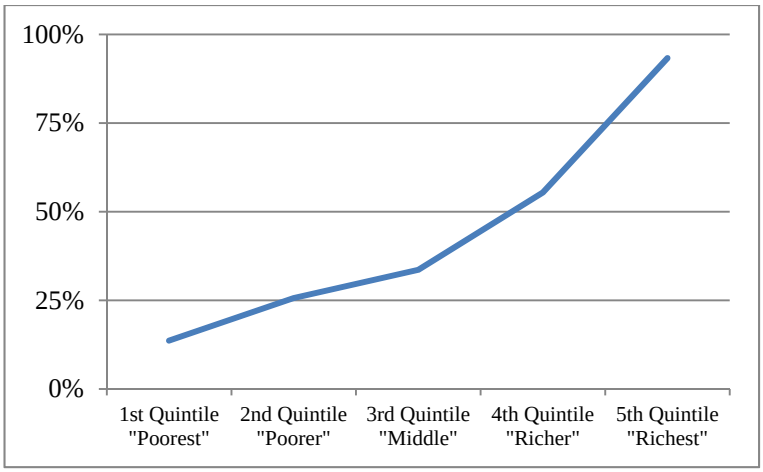
Finished Roof



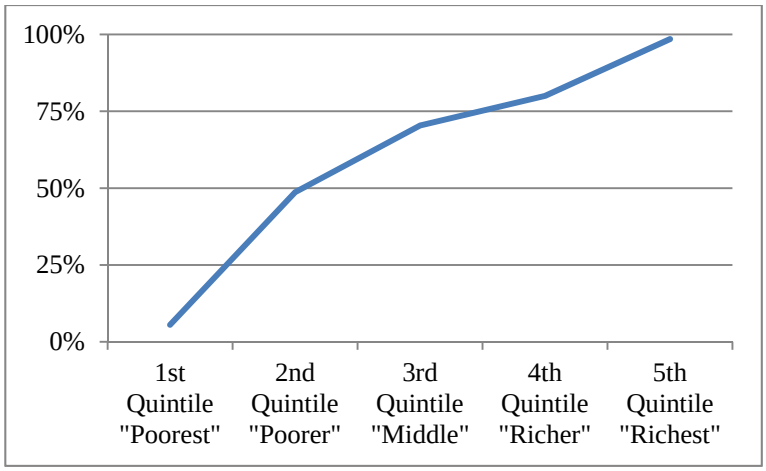
Mozambique (2011)

Asset	1st Quintile "Poorest"	2nd Quintile "Poorer"	3rd Quintile "Middle"	4th Quintile "Richer"	5th Quintile "Richest"
Agriculture land	99%	95%	88%	75%	37%
Animal drawn cart	0%	0%	2%	5%	2%
Bank Account	0%	0%	2%	15%	70%
Bed net	52%	55%	60%	64%	70%
Bicycle	43%	51%	54%	44%	22%
Car	0%	0%	0%	2%	21%
Cooking fuel: charcoal	0%	0%	2%	20%	57%
Chimney	0%	0%	0%	0%	0%
Cooking fuel: wood	100%	100%	97%	78%	22%
Electricity	0%	0%	2%	22%	90%
Finished Floor	14%	26%	34%	55%	93%
Finished Roof	0%	0%	8%	69%	89%
Motorboat	0%	0%	0%	0%	0%
Motorcycle	0%	2%	7%	12%	14%
Toilet facility	6%	49%	70%	80%	98%
Radio	30%	49%	60%	60%	75%
Refrigerator	0%	0%	0%	3%	64%
Share toilet	3%	10%	16%	17%	13%
Telephone	0%	0%	0%	2%	21%
Television	0%	0%	1%	21%	89%
Watch	3%	11%	23%	27%	38%
Water access on household premises	2%	3%	5%	11%	63%

Finished floors



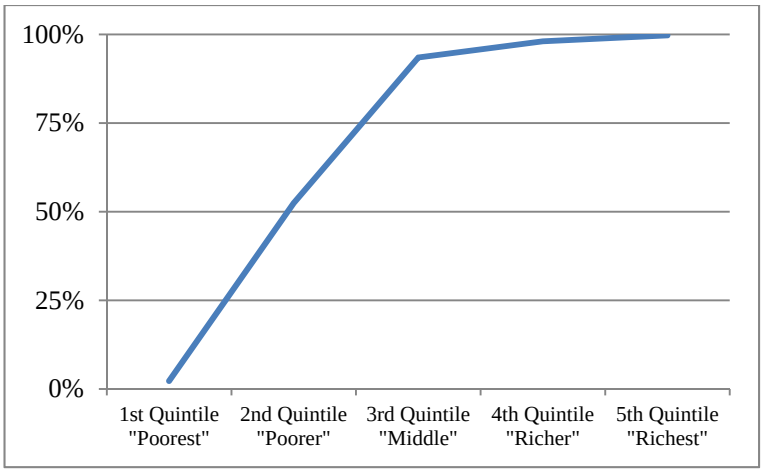
Toilet facility: none



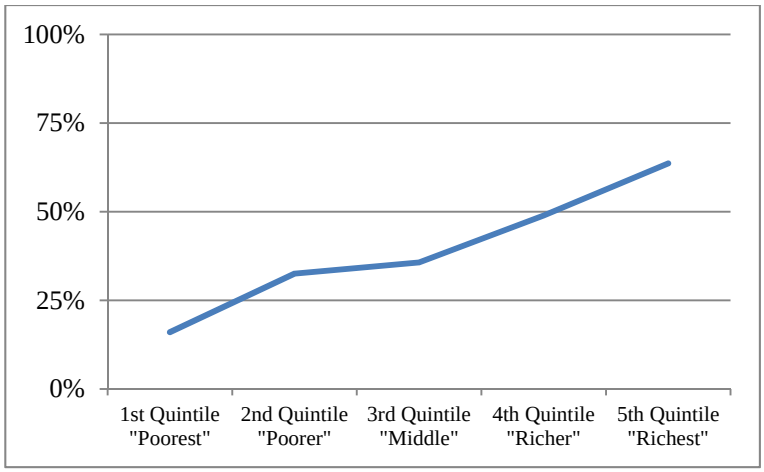
Zimbabwe (2010-2011)

Asset	1st Quintile "Poorest"	2nd Quintile "Poorer"	3rd Quintile "Middle"	4th Quintile "Richer"	5th Quintile "Richest"
Bank account	1%	5%	9%	26%	65%
Battery or generator	6%	13%	17%	20%	15%
Bed net	41%	37%	38%	42%	56%
Bicycle	14%	23%	30%	32%	28%
Car	0%	1%	2%	7%	31%
Cooking fuel: charcoal	0%	0%	0%	0%	0%
Computer	0%	0%	0%	1%	18%
Cooking fuel: wood	100%	100%	97%	58%	9%
Electricity	1%	3%	5%	59%	98%
Finished Floor	2%	52%	94%	98%	100%
Finished Roof	1%	27%	83%	96%	84%
Motorcycle	0%	0%	1%	1%	4%
Toilet facility: none	76%	44%	17%	3%	0%
Radio	16%	33%	36%	49%	64%
Refrigerator	0%	0%	0%	15%	76%
Share toilet	29%	35%	36%	55%	30%
Solar panel	13%	27%	36%	26%	5%
Telephone	0%	1%	2%	7%	31%
Television	1%	9%	15%	62%	94%
Tractor	0%	0%	0%	1%	4%
Water access on household premises	8%	15%	29%	60%	85%
Wheelbarrow	16%	32%	44%	35%	41%

Finished floor



Radio



Appendix IV

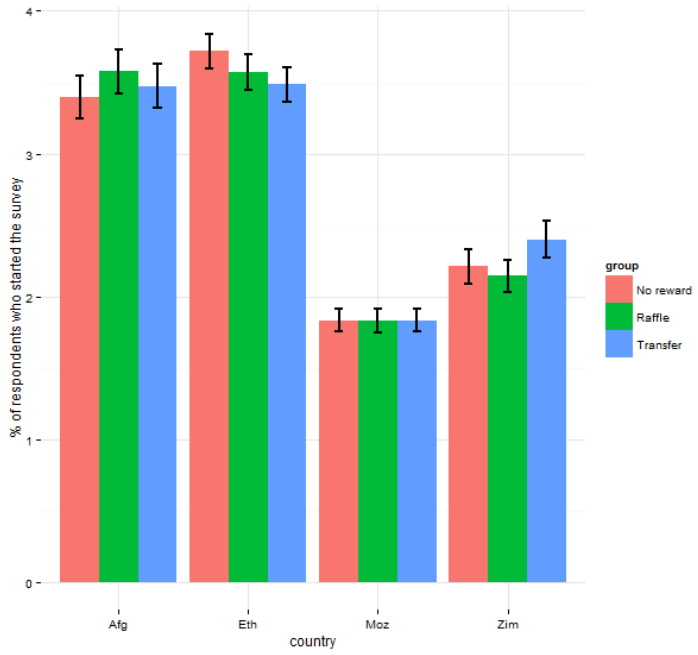
Impact of Survey Completion Incentives, Explanatory Note

The analysis of the impact of survey completion incentives rests on the assumption that random assignment was done correctly and hence that there are no differences between the experimental groups prior to treatment. Since we assign the treatment groups, there should be no way in which those groups differ other than the fact that they heard the different introduction messages. We can test whether this assumption holds by looking at whether the rates of respondents starting the survey vary across the different treatment groups.

The figure below shows the difference in the percentage of subjects who started the survey across different experimental groups. Since starting the survey happens prior to receiving the different treatment messages, there should be no material difference across these groups.

The results for Mozambique show almost exactly the same start rate for all experimental groups. This suggests that treatment assignment was completely random. Contrast this to Ethiopia and Zimbabwe. In Ethiopia, respondents in the raffle and transfer conditions are both less likely to start the survey. This strongly suggests that the assignment of respondents in Ethiopia was not truly random. In Zimbabwe, a significantly higher proportion of subjects in the transfer condition start the survey compared to the raffle and control groups. The situation is less clear for Afghanistan. There are not significant differences between the different groups, but it is nowhere near as uniform as the results for Mozambique. Overall, we can say with some confidence that the subjects in Mozambique were truly randomly assigned to their groups and can tentatively say the same in Afghanistan. We also have confidence that this was not the case in Ethiopia and Zimbabwe.

Survey Start Rates for Subjects, Different Experimental Groups



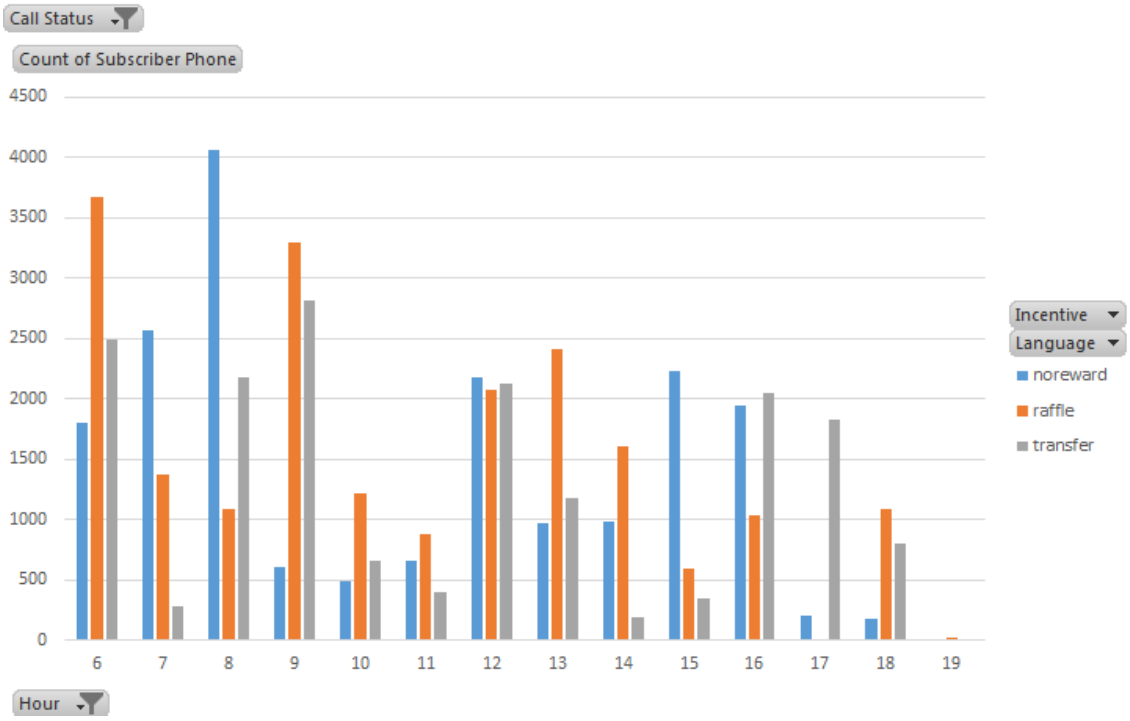
The survey operator, Voto Mobile, assigned the treatment subjects in blocks. These subjects were supposed to receive the calls at the same time.[76] However, in practice, it appears that these calls were not in fact *placed* at the same time. The figure below illustrates the call placement patterns in Zimbabwe. There are clear differences in the timing of the different groups. This likely explains the different start rates, as additional analysis has shown that the time of day that someone receives a call affects response rates. Similarly, if mobile networks are unreliable at times of peak demand, calls may be more likely to get through and to stay connected at certain times of day.

This non-random assignment is problematic because it opens up the possibility of selection effects. Suppose that there are two types of people: easy finishers (who will generally complete the survey if they start it) and difficult finishers (who will generally not complete the survey even if they start it). If a group has more easy finishers, it will end up with a higher completion rate than if it has more difficult finishers. If different proportions of these types of people end up in each experimental group, we won't be able to tell whether an apparent treatment effect is because of the treatment itself or because of the different composition of the groups. When randomization is done successfully, we don't have to

[76] Correspondence with Levi Goertz of Voto Mobile.

worry about composition effects such as these because any attributes (observable or unobservable) should be randomly distributed across the groups. However, once some form of non-random assignment happens, it is difficult to distinguish between the treatment effect and potential selection effects.

Survey Placement Time by Different Experimental Groups



One way to look at whether this non-random assignment is affecting the results is to use an intent-to-treat analysis, where we consider all phone numbers that were assigned a group as subjects rather than just those that heard the message. If selection effects are not present, then the effect directions and significance should not change depending on whether we use an intent-to-treat analysis or actually treated analysis. If the effects differ between the two analyses, it suggests that a selection effect may be happening. Using this analysis, the differences in the Ethiopian case are no longer significant. In Zimbabwe, the raffle condition is no longer significant when using an intent-to-treat framework, but the transfer condition does remain significant. The effects for Afghanistan and Mozambique remain the same when using an intent-to-treat framework, which further reinforces the impression that random assignment was done successfully in these cases.

Unfortunately, we were unable to address the selection effects among the treatment groups. We attempted using covariates to control for non-random assignment in some way. If the non-random assignment happened on a binary variable. For example, if one treatment group included more men than another, we could simply control for this variable and its interactions with a dummy variable. However, this proved very difficult because time is a continuous variable and we do not fully know how it affects completion rates. For instance, suppose that response rates increase linearly with time of day. If we put in dummy variables capturing each hour in the model, we would not have removed the effect of time because time would still have an effect within each hour. Similarly if we controlled for time linearly but if it had a curvilinear impact, we would still incorrectly control for its effect on the estimated treatment effects.

Appendix V

References

Acemoglu, Daron and Robinson, James. Why Nations Fail: The Origins of Power, Prosperity, and Poverty. Crown: 2012.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Replication data for: Fractionalization". Journal of Economic Growth 8 (June): 155-194 http://www.anderson.ucla.edu/faculty_pages/romain.wacziarg/papersum.html .

Anderson, Mary, Dayna Brown, and Isabella Jean. Time to Listen: Hearing People on the Receiving End of International Aid. CDA Collaborative Learning Projects.

Banerjee, Abhijit and Esther Duflo. Poor Economics, a radical rethinking of the way to fight global poverty, Public Affairs. 2011.

Barder, Owen. "It's the Politics, Stupid." Owen Abroad (blog). June 1, 2013.

Battaglia, Michael, David Izrael, David Hoaglin, and Martin Frankel. Nd. Practical Considerations in Raking Survey Data. Abt Associates Inc.

Bergmann, Michael. 2011. IPF-Algorithm to create adjustment survey weights. Stata Package. University of Mannheim.

Boakye, Kojo, Nigel Scott, and Claire Smyth. Keys Findings of Mobiles for Development, UNICEF. October 2010.

Bratton, Michael. Non-Governmental Organizations in Africa: Can They Influence Public Policy? Development and Change. Vol. 21: 87-118. 1990.

Bratton, Michael. Citizen Perceptions of Local Government Responsiveness in Sub-Saharan Africa, Afrobarometer Working Paper 119. 2010.

Brautigam, Deborah. The People's Budget? Politics, Participation and Pro-Poor Policy, Development Policy Review 22(6): 653-668.

Brinkerhoff, Derick. How Citizens Participate in Macroeconomic Policy: International Experience and Implications for Poverty Reduction, World Development 31(4): 685-701. 2003.

Calcott, Paul. New on Paternalism, Economics and Philosophy 2(10): 315-321. 2000.

Croke, Kevin, et al. Collecting High Frequency Panel Data in Africa Using Mobile Phone Interviews. World Bank Policy Research Working Paper. June 2012.

DeBell, Matthew, Jon Krosnick, Arthur Lupia, and Caroline Roberts. User's Guide to the Advance Release of the 2008-2009 ANES Panel Study. Palo Alto, CA and Ann Arbor, MI: Stanford University and University of Michigan. 2009.

Deming, W. Edwards and Frederick F. Stephan. On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known, in: The Annals of Mathematical Statistics 11 (4): 427-444. 1940.

Demombynes, Gabriel, Paul Gubbins, Alessandro Romeo. Challenges and Opportunities of Mobile Phone-Based Data Collection: Evidence from South Sudan, World Bank Policy Research Working Paper No. 6321. Jan. 2013.

Demographic and Health Surveys (various) [Datasets]. Calverton, Maryland: ICF International [Distributor], 2012.

Devas, Nick and Ursula Grant. Local Government Decision-Making: Citizen Participation and Local Accountability, Public Administration and Development 23(4): 307-316. 2003.

Dillon, Brian. Using Mobile Phones to Conduct Research in Developing Countries, Economic Development Initiatives Africa. Feb. 2010.

Driessen, Michael. Ethno-Linguistic Fractionalization: Dataset Review, APSA Comparative Politics. Winter 2008.

Filmer, Deon and Lant Pritchett. Estimating Wealth Effects Without Expenditure Data—Or Tears: An Application to Educational Enrollments in States of India, Demography 38(1): 115-132. 2001.

Gneezy, U., Meier, S., & Rey-Biel, P. When and why incentives (don't) work to modify behavior. The Journal of Economic Perspectives, 191-209. 2011.

GSMA. Women and Mobile: A Global Opportunity: A Study on the Mobile Phone Gender Gap in Low and Middle-Income Countries. 2012.

Gumboh, Stephen. Practicality of Stakeholder Participation in the Local Government Budgeting System in Zambia. 2012.

Gupta, Sanjeev, Benedict Clements, Emanuele Baldacci, and Carlos Mulas-Granados. Expenditure Composition, Fiscal Adjustment, and Growth in Low-Income Countries, IMF Working Paper 02/77. April 2002.

Himelein, Kristen. The Socio-Economic Impacts of Ebola in Liberia: Results from a High Frequency Cell Phone Survey. World Bank. Nov. 19, 2014.

Kalton, Graham, J. Michael Brick, and Thanh Lê. Estimating components of design effects for use in sample design. Household Sample Surveys in Developing and Transition Countries. United Nations Statistics Division. November 2005.

Kalton, Graham and Ismael Flores-Cervantes. Weighting Methods. Journal of Official Statistics, vol 19(2). 2003.

Krafchik, Warren. "Can civil society add value to budget decision-making? A description of the rise of civil society budget work." Citizen Participation and Pro-Poor Budgeting. 2005.

Leonard, Thomas C. Review of Richard Thaler and Cass Sunstein, Nudge: Improving Decisions about Health, Wealth and Happiness, Constitutional Political Economy 19(4): 356-360. 2008.

Leonard, Thomas C., Robert S. Goldfarb and Steven Suranovic, "Professor New on Paternalism and Public Policy." Economics and Philosophy 16(2): 323-331. 2000.

Matovu, George. Capacity Building for Participatory Planning and Budgeting in Africa: Initiatives and Strategic Perspectives, presented at the Pan African Conference of Ministers of Local Government held in Maseru, Lesotho. September 2006.

McKenzie, David. Measuring inequality with Asset Indicators. Journal of Population Economics 18(2): 229-260. 2005.

Merkel, Wolfgang and Brigitte Weiffin. Does Heterogeneity Hinder Democracy?, Comparative Sociology 11 (2012) 387–421.

New, Bill. Paternalism and public policy, Economics and Philosophy 15:63-83. 1999.

Oldendick, R. W., & Lambries, D. N. Incentives for Cell Phone Only Users: What Difference Do They Make? Survey Practice, 4(1). 2013.

Open Technology Development Alliance. Mobile Technology Empowers Citizens to Impact Budget Planning in Local Communities of Cameroon. http://odta.net/post/participatory-budgeting-cameroon.

Pattillo, Catherine, Michel Lazare, and Marco Cangiano. Budget Institutions and Fiscal Performance in Low-Income Countries, IMF Working Paper 10/80. March 2010.

Rivlin, Alice. Rationalism and Redemocratization: Time for a Truce, in Worst Things First? The Debate Over Risk-Based National Environmental Priorities. Resources for the Future press. 1994.

Robinson, Mark and Steven Friedman. Civil Society, Democratization, and Foreign Aid: Civic Engagement and Public Policy in South Africa and Uganda, Democratization 14(4): 643-668. 2007.

Schellenberg, Joanna Armstrong, et al. Inequities among the very poor: health care for children in rural southern Tanzania. Lancet 361: 561-6. Feb 2003.

Singer, E., & Ye, C. The use and effects of incentives in surveys. The ANNALS of the American Academy of Political and Social Science, 645(1), 112-141. 2013.

Svensson, Jakob. The institutional economics of foreign aid. Swedish Economic Policy Review 13: 115-137. 2006.

Yankelovich, Daniel and Will Friedman. Toward Wiser Public Judgment. Vanderbilt University Press. 2010.

Zamor, Garcia (edited by). Public participation in development planning and management: cases from Africa and Asia, Westview Press. 1985.