Economic Premise

Collecting High-Frequency Data Using Mobile Phones: Do Timely Data Lead to Accountability?

74668

Kevin Croke, Andrew Dabalen, Gabriel Demombynes, Marcelo Giugale, and Johannes Hoogeveen

As mobile phone ownership rates have risen dramatically in Africa, there has been increased interest in using mobile telephones as a data collection platform. This note draws on two largely successful pilot projects in Tanzania and South Sudan that used mobile phones for high-frequency data collection. Data were collected on a wide range of topics and in a manner that was cost-effective, flexible, and rapid. Once households were included in the survey, they tended to stick with it: respondent fatigue has not been a major issue. While attrition and nonresponse have been challenges in the Tanzania survey, these were due to design flaws in that particular survey, challenges that can be avoided in future similar projects. Ensuring use of the data to demand better service delivery and policy decisions turned out to be as challenging as collecting the high-quality data. Experiences in Tanzania suggest that good data can be translated into public accountability, but also demonstrate that just putting data out in the public domain is not enough. This note discusses lessons learned and offers suggestions for future applications of mobile phone surveys in developing countries, such as those planned for the World Bank's "Listening to Africa" initiative.

Timely, high-quality information about socioeconomic outcomes related to well-being, service delivery, income, security, health, and many other topics is not readily available in Africa. This is because such data are typically collected by nationally representative, face-to-face household surveys. Such surveys are expensive and time consuming and are, for these reasons, not implemented very frequently. There is strong (latent) demand for more timely welfare and monitoring information—decision makers need information that allows them to monitor the situation in their country in "real time." Program managers, researchers, and citizens could all benefit from such information.

Because mobile phone networks now reach most areas of Africa, this note proposes a mobile phone survey model that combines a standard baseline survey with the distribution of

mobile phones to respondents at the time of baseline, who then are called regularly (weekly or every two weeks) with follow-up questions. These questions can be comparable to questions asked earlier, to allow for tracking of changes over time, or can address new issues that have arisen since the implementation of the baseline survey. This approach enables high-frequency, low-cost collection of a wide range of data related to household welfare.

Two African countries, South Sudan and Tanzania, have already piloted mobile phone surveys. The survey in Tanzania has been running longest (33 rounds to date), while the survey in South Sudan operates under more difficult conditions. These are not the first examples of using a mobile phone platform to collect high-quality panel data in Africa. Dillon (2012), for instance, used mobile phones to track how farm-

ers' expectations of the next harvest changed over time. This note differs from Dillon's work in two ways: it proposes the creation of nationally representative (panel) surveys that ask questions via mobile phones every two weeks, and then proposes the data be publicly released (in anonymity) within weeks of collection. This note also discusses how to translate the data into policy in the public domain.

Mobile Phone Surveys

Conducting surveys by phone is standard practice in developed countries, but typically has not been considered feasible in poor countries, because phone ownership rates are too low (especially in the premobile phone era). In Tanzania, for example, just 1 percent of households own a landline phone (DHS 2010). However, the rapid rise of mobile phones in Africa has changed this. Again in Tanzania, mobile phone ownership increased from 9 percent of all households according to the 2004–5 Demographic and Health Survey (DHS), to 28 percent by 2007–8. By 2010, this number had almost doubled again, to 46 percent of households. Even higher rates of phone ownership have been recorded in urban areas: phone ownership was 28 percent for urban households in the 2004–5 DHS, and more than doubled to 61 percent in 2007–8, and reached 78 percent by 2010. In the Tanzania baseline survey in Dar es Salaam, mobile ownership was found to be 83 percent. In Kenya, the sub-Saharan country that is leading in terms of mobile phone ownership, the Afrobarometer survey of November 2011 shows that 80 percent of Kenyan adults have their own mobile phone; 81 percent use their mobile phone to make a call at least once a day; 61 percent send or receive a text message at least once a day; and a remarkable 23 percent send or receive money or pay a bill via mobile phone at least once a day.

With such high rates of mobile phone ownership, representative household surveys using mobile phones start to become a feasible option. For example, only 80 percent of U.S. households own landlines, but political polling still typically uses landline samples only, corrected by reweighting (Blumberg and Luke 2009). This suggests that phone ownership in Kenya or in urban Tanzania is already high enough for reasonable inference to be made from surveys that exclusively rely on mobile phones.

Listening to Africa

Mobile surveys, such as the proposed Listening to Africa project, will not rely exclusively on interviews over the phone, but will employ a face-to-face baseline survey where household and respondent characteristics will be collected, and during which the respondent for the mobile phone part of the survey will be selected. Listening to Africa aims to collect population data, requiring that an (adult) respondent be randomly selected from the household. The baseline survey

is also when phones are distributed, and in locations with limited access to electricity, solar chargers. Alternatively, a village kiosk owner who provides phone-charging services could be contracted to offer free phone charging to participating households.

Once the baseline has been completed, respondents receive calls from a call center at regular intervals during which additional questions are asked. In this way, a high-frequency panel is created. Answers from respondents can be collected in various ways because mobile phones offer a multitude of opportunities to obtain feedback, including through SMS, WAP, IVR, and USSD. When this note mentions mobile surveys, it refers to surveys carried out by call centers where operators call respondents, interview them, and enter the responses into a database using a CATI (computer-aided telephone interview) system.

The decision to rely on call centers for mobile phone data collection was informed by experiences with WAP, IVR, and USSD in the early stages of the Tanzania mobile survey. The flexibility that call centers offer, particularly when asking complex questions requiring explanation, for accommodating illiterate respondents or low-end phones without Internet connectivity makes voice the most feasible technology for sub-Saharan Africa. Call centers are able to deal with respondents who speak different native languages, good enumerators can build a rapport between respondent and phone operator, and call centers offer the opportunity to ask in-depth (qualitative) questions.

## Results from the Tanzania and South Sudan Mobile Surveys

The mobile survey in South Sudan revisited 1,000 respondents in 2010 in 10 urban areas covered in South Sudan's 2009 National Baseline Household Survey. During this revisit, enumerators selected respondents, handed out mobile phones (half of them with integrated solar chargers), and then called respondents on a monthly basis from a call center in Nairobi, using interviewers capable of speaking South Sudan's main languages. Respondents who successfully completed an interview were rewarded with an amount varying from $2 to $4.

The survey in Tanzania visited 550 households in Dar es Salaam in August 2010, administered a baseline survey, randomly selected an adult respondent for the mobile survey, and called respondents on a weekly basis (25 rounds), and later (8 rounds) every two weeks. The survey in Dar es Salaam did not initially distribute phones, but recently, after round 33, some phones have been distributed to respondents who had never participated in the survey before. Both surveys are still running.

The mobile survey interview format does not appear to pose major limitations on what can be asked, except that

the length of an interview should probably not be more than 20–30 minutes. So an elaborate consumption module, for example, or a detailed health module with birth histories, is less suited to this type of survey. The mobile surveys in South Sudan and Tanzania collected information on a wide variety of issues including health, education, water, security, nutrition, travel times, prices, electricity, and governance. The surveys have been used to ask perception questions on topics varying from what respondents considered the most pressing problems to be addressed by the city government to people's opinion about the draft constitution. They have also been used to collect baseline information for large-scale programs on food fortification.

Figure 1. In the Last Month, How Often, If Ever, Have You or a Member of Your Household Gone Without Enough Food to Eat?
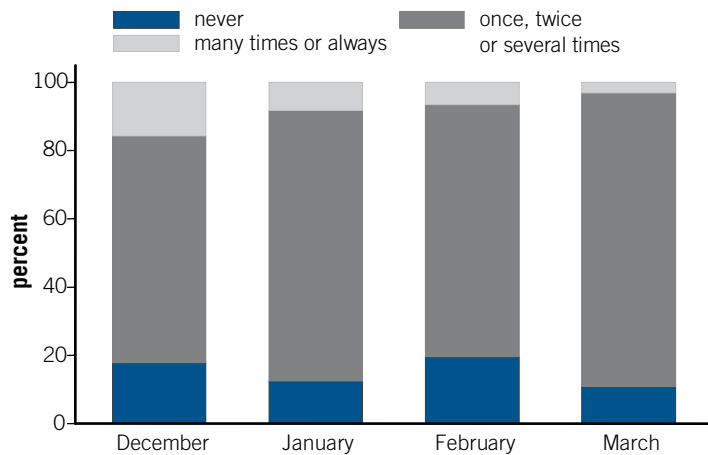
Source: Croke et al. 2012.

Figure 2. Questions Asked to Primary School Children about Teacher Presence and Use of Books while in School
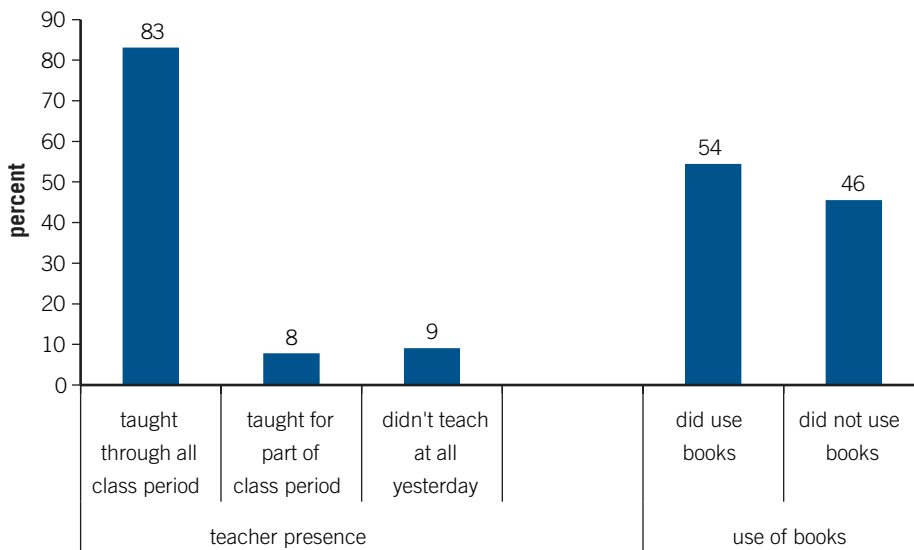
Source: Croke et al. 2012.

The data can be used to report on a single issue, but become of greater interest when the same information is tracked over time (figure 1).

Because questions can be changed every round of the survey, it is possible to accommodate new data requests or to respond to emerging issues. For instance, relevant questions were included in the Dar es Salaam survey after major floods hit the city in December 2011 to get a sense of the number of people that had been affected. Mobile surveys can also ask the main respondent to pass the phone on to someone else in the household to get certain questions answered, if necessary. Figure 2, for instance, presents responses to questions asked to children attending primary school about the presence of their teacher and the use of books. Finally, though this has not been tried yet, the mobile survey can be used to field screening questions to identify respondents who qualify for in-depth interviews.

Nonresponse and Attrition

A key challenge for high-frequency mobile surveys is nonresponse (when a respondent participates in some but not all rounds) and attrition (when a respondent drops out of the survey completely). What does the Tanzania experience tell us about attrition and nonresponse in mobile phone surveys? On the negative side, there was a large initial burst of attrition. This can largely be attributed to the fact that the survey team did not hand out phones to those who did not already own one. When the team initially visited and administered the baseline survey to 550 respondents, it was found that 418 owned their own phone. Once the mobile survey began in earnest, an average of 304 respondents, or 66 percent, participated during the first 19 rounds of the survey. Later, once the survey was put under World Bank management and oversight was tightened (after a four-month gap in interviews!), the number of respondents increased to 343 respondents on average (75 percent of the sample). So after 33 rounds of mobile interviews, the overall response rate is 75 percent from the 458 households in the sample that had access to phones (62 percent of the 550 households in the baseline survey). The rate of attrition, narrowly defined as those who did not respond at all to the mobile survey is much lower: only 4 percent of 18 out of the 458 households never responded to a mobile survey, while 66 percent responded to at least two out of every three surveys.

Is the Tanzania Mobile Survey Representative?

Attrition and nonresponse are particularly problematic when they occur in a nonrandom manner. The working paper from which this note is drawn includes regression analysis of the determinants of attrition. In these regressions, the dependent variable is the number of rounds (out of 25) in which the household participated. When attrition among all 550 initially visited households is analyzed, results show—perhaps unsurprisingly for a survey that did not distribute mobile phones—that wealth is correlated with survey participation. When analysis restricts the regression to households that were identified as reachable in the first full round of the survey (that is, households that own phones and that gave a working phone number to the survey team), the impact of wealth is no longer significant, while residence in the rural part of Dar es Salaam and using the premium mobile phone provider (Vodacom) remain significant variables.

The lesson learned from this regression is that with phone distribution to poor households and more careful collection of multiple, verified phone numbers from each respondent, nonresponse could have been significantly reduced. The second lesson is that this survey, in order to remain representative, would have to be reweighted ex post. Figure 3 illustrates how reweighting is able to restore the survey's representativeness by showing how the changing composition of the sample affects the percentage of households allocated to different wealth quintiles. The first column presents the survey baseline (550 respondents); the sample is then divided equally among the five wealth quintiles: each quintile has exactly 20 percent of the sample. When looking

at the breakdown across the 458 respondents that were included in the mobile phone survey, one notes that poor households are underrepresented. The distribution becomes more skewed toward wealthier households in round 26 (341 respondents). The final set shows what the final distribution looks like once the mobile phone sample has been reweighted using information about the determinants of attrition. It shows that the original distribution is essentially restored.

Using Mobile Data to Enhance Accountability

Since management of the Tanzania survey was transferred to the World Bank, active efforts have been made to use the survey for accountability purposes. Questions are carefully identified for their potential use as accountability tools. Once data are collected, easy-to-understand, factual reports are prepared presenting the findings. These reports are disseminated through a dedicated Web site from which all survey data, baseline survey as well as data from the mobile survey rounds, can be downloaded (www.listeningtodar.org). The reports are also shared by email using a distribution list that includes journalists and other stakeholders.

So, does it work? While the Web site itself attracts relatively modest traffic, the project has gotten quite a bit of media traction, especially since May 2011, when it partnered with a local television journalist, who now holds a press conference to introduce each new survey report to the local media. Reports produced have been discussed on blogs and in academia, on television in Tanzania, and have led to various newspaper articles. It is harder to assess what happens once information is published, but there are indications that the information is "received" by those responsible for results. For instance, the managing director of Tanzania's electricity company felt compelled to explain to the media why so many households connected to electricity are experiencing power cuts and what his company is doing about them.[1] Once published, information travels fast and far. A brief about food price increases was used in a front page article in Tanzania's *Citizen* newspaper, which was picked up by others including the *Rwandan Times,*[2] and is cited in the World Bank's 2012 *Global Monitoring Report*. Information also tends to go in unexpected directions. A brief about the limited increase in water connections in Dar es Salaam, despite a large-scale investment program, got media attention because of the discrepancy between the data reported by the mobile survey and official government statistics.[3]

What lessons can be drawn thus far? One lesson is that providing citizens with relevant, timely, and accurate data about the actions of politicians, policy makers, and public service providers is not sufficient. For the data to have impact, they need to be accessible and disseminated widely, and in

Figure 3. Changing Wealth Composition of Sample
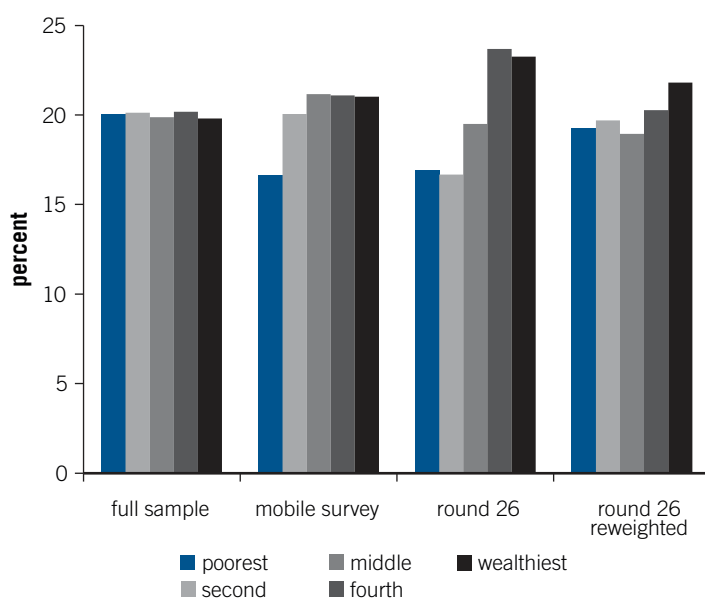
Source: Croke et al. 2012.

ways that allow them to be utilized by already existing institutions and actors.

Cost-Effectiveness of Mobile Surveys

How cost-effective are mobile surveys? World Bank data give a good sense of the marginal cost of the surveys. The call center in Tanzania was originally contracted to implement 12 survey rounds at a rate of $1,400 per round. If one adds the cost for consultants to maintain a Web site, supervise data collection and to analyze the data, the marginal cost per round increases to $2,500. Given that these rounds averaged 343 respondents, this comes to about $4.10–$7.30 per interview. Dillon (2012) notes a relatively similar marginal cost per survey: $6.98. In addition to these marginal costs, one needs to include the cost of a baseline, which will often be between $50 and $150 per interview, depending on the complexity of the survey and the distances that have to be covered. Whether this is cost-effective or not depends on the purpose. The ability to carry out an entire survey round and report on results for $2,500 is remarkably cost-effective. But if one keeps in mind that the typical round in the Tanzania survey asks 17 questions (with a maximum of 44), then the cost per question is about $0.42. This is relatively high, so if the intention is to ask many questions, it may turn out to be more cost-effective to opt for a face-to-face interview.

Conclusion

Evidence demonstrates that mobile surveys have great potential to provide rapid feedback and address existing data gaps at limited expense. Mobile surveys should not be considered substitutes for household surveys, but complements: mobile surveys may rely on an existing household survey to serve as baseline, and mobile surveys are not the right platform for lengthy interviews. When interviews are lengthy, face-to-face interviews are probably more cost-effective.

The evidence from the Tanzania and South Sudan surveys suggests that mobile surveys can collect quality data in a very timely manner that is of use to a wide range of data users: decision makers, program managers, statisticians, and researchers. The Tanzania survey highlighted the need for mechanisms that avoid attrition and nonresponse right from implementation of the baseline. Much of the attrition in the Tanzania survey can be explained by choices made in the organization of the survey (such as not to distribute mobile phones). Findings suggest that once households are included in the mobile survey, they are likely to remain in the survey: respondent fatigue was not found to be an issue. These pilot experiences also show that because of the high frequency of data collection, quality

control of mobile surveys is dynamic because issues identified in one round can be corrected in the next.

Finally, the use of the results from the surveys has been encouraging. Once results were disseminated systematically, civil society started to discuss them and media reported on them. At the same time, it is clear that if the objective is to use the results for accountability purposes, distinct efforts need to be made to ensure that the information reaches the actors in government and civil society who are most willing and able to make use of it.

About the Authors

Kevin Croke is a *Consultant* with the *World Bank's Africa Region. Andrew Dabalen* is a *Lead Poverty Specialist, Gabriel Demombynes* is a *Senior Economist, Marcelo Giugale* is *Director* for *Poverty Reduction and Economic Management*, and *Johannes Hoogeveen* is a *Senior Economist*, all with the *World Bank's Africa Region.*

Notes

1. http://www.ippmedia.com/frontend/index.php?l=39618 and http://www.ippmedia.com/frontend/index.php?l=39551.
2. http://www.newtimes.co.rw/news/index.php?i=14892&a=11334.
3. http://www.thecitizen.co.tz/sunday-citizen/-/20105-govt-figures-on-access-to-clean-water-inflated.

References

Baird, Sarah, and Berk Özler. 2011. "Examining the Reliability of Self-Reported Data on School Participation." *Journal of Development Economics* 98 (1): 89–93.

Blumberg, Stephen J., and Julian V. Luke. 2009. "Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July-December 2008." Centers for Disease Control, Atlanta, GA.

Croke, Kevin, Andrew Dabalen, Gabriel Demombynes, Marcelo Giugale, and Johannes Hoogeveen. 2012. "Collecting High-Frequency Panel Data in Africa Using Mobile Phone Interviews." World Bank Policy Research Working Paper 6097, Washington, DC.

Dillon, Brian. 2012. "Using Mobile Phones to Conduct Research in Developing Countries." *Journal of International Development* 24 (4): 518–27.

Lynn, Peter, and Olena Kaminsha. 2011. "The Impact of Mobile Phones on Survey Measurement Error." Working Paper No. 2011-7, Institute for Social and Economic Research, University of Essex.

Smith, G., I. MacAuslan, S. Butters, and M. Tromme. 2011. "New Technology Enhancing Humanitarian Cash and Voucher Programming." Research paper, Oxford Policy Management.