

From Necessity to Opportunity

Lessons for Integrating Phone and In-Person Data Collection for Agricultural Statistics in a Post-Pandemic World

Alberto Zezza

Kevin McGee

Philip Wollburg

Thomas Assefa

Sydney Gourlay



WORLD BANK GROUP

Development Economics

Development Policy Team

September 2022

Abstract

The COVID-19 pandemic has disrupted survey and data systems globally and especially in low- and middle-income countries. Lockdowns necessitated remote data collection as demand for data on the impacts of the pandemic surged. Phone surveys started being implemented at a national scale in many places that previously had limited experience with them. As in-person data collection resumes, the experience gained provides the grounds to reflect on how phone surveys may be incorporated into survey and data systems in low- and middle-income countries. This includes agricultural and rural surveys supported by international survey programs such as the World Bank's Living Standards Measurement Study—Integrated Surveys on Agriculture, the Food and Agriculture Organization's AGRISurvey, or the 50x2030 Initiative. Reviewing evidence and experiences

from before and during the pandemic, the paper analyzes and provides guidance on the scope of and considerations for using phone surveys for agricultural data collection. It addresses the domains of sampling and representativeness, post-survey adjustments, questionnaire design, respondent selection and behavior, interviewer effects, as well as cost considerations, all with an emphasis on the particularities of agricultural and rural surveys. Ultimately, the integration of phone interviews with in-person data collection offers a promising opportunity to leverage the benefits of phone surveys while addressing their limitations, including the depth of content constraints and potential coverage biases, which are especially challenging for agricultural and rural populations in low- and middle-income countries.

This paper is a product of the Development Policy Team, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at pwollburg@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

From Necessity to Opportunity: Lessons for Integrating Phone and In-Person Data Collection for Agricultural Statistics in a Post-Pandemic World

Alberto Zezza^{*}, Kevin McGee^{*}, Philip Wollburg^{*}, Thomas Assefa[‡], Sydney Gourlay^{*1}

JEL Classification: Q10 Agriculture; C83 Survey Methods, Sampling Methods; R2, Household analysis

Keywords: Phone surveys; Agriculture; Household surveys; Living standards; Data

¹ Author ordering randomized using American Economic Association's randomization tool (confirmation code: Yj0CGbPwsdua). ^{*} Living Standards Measurement Study, Development Data Group, World Bank. [‡] University of Georgia. Authors may be contacted at pwollburg@worldbank.org. This is a publication of the 50x2030 Initiative to Close the Agricultural Data Gap, a multi-partner program that seeks to bridge the global agricultural data gap by transforming data systems in 50 countries in Africa, Asia, the Middle East and Latin America by 2030. For more information on the Initiative, visit 50x2030.org. The authors thank Daniel Ali, Chiara Brunelli, Andrew Dillon, Neli Georgieva and Diego Zardetto for their valuable comments on an earlier draft, as well as participants in IFAD's 2022 conference on "Jobs, innovation and rural value chains in the context of climate transition: Bridging the gap between research and policy" for their useful feedback. The authors are solely responsible for the contents of the paper.

1. Introduction

In early 2020, the COVID-19 pandemic forced the suspension of in-person data collection around the world. As data is a critical means to better understand the impacts of the pandemic, many governments, national statistical offices (NSOs), international organizations, and researchers came to rely on phone surveys in lieu of in-person interviewing. Phone surveys offered advantages for rapid data collection in an emergency both in terms of cost and flexibility (United Nations and World Bank, 2020). While already commonly used tools in high-income countries (National Research Council, 2008; Slavec and Toninelli, 2015), prior to the COVID-19 emergency phone surveys were far less prevalent in low- and middle-income countries where limited phone penetration had in the past been an obstacle to their implementation (GSMA, 2018, 2020). Phone surveys had for instance been used for specific populations and purposes, such as in response to the 2014 Ebola outbreak (Etang and Himelein, 2020) and the 2017 drought and conflict crisis in the Republic of Yemen, Somalia, South Sudan, and Nigeria (Hoogeveen and Pape, 2020), but only in a handful of regional or national projects (Ballivian et al., 2015; Dabalén et al., 2016). The COVID-19 emergency, however, prompted a rapid and wide-scale uptake of phone surveys in these countries, supported not only by the urgent need for real-time data but also by recent expansions in mobile network coverage (Tomlinson et al., 2009; Dillon, 2012; Demombynes et al., 2013; Ballivian et al., 2015; Larmarange et al., 2016; Lau et al., 2019).

This surge in phone surveys provided an unprecedented opportunity for NSOs and international organizations operating in low- and middle-income countries to build experience and the necessary infrastructure to implement such surveys at scale. Gourlay et al. (2021) provide an overview of how this acceleration materialized, drawing on the experience of the Living Standards Measurement Study (LSMS) survey program as well as other survey initiatives. In this paper, we approach the problem specifically from an agricultural and rural angle with a view to informing data collection activities going forward.

Agriculture, especially smallholder farming, comes with several specific challenges for data collection whether conducted in-person or remotely. Among other factors, agricultural production processes are seasonal and irregular, agricultural outcome variables have different timeframes, non-standard units are very common, respondents are more often illiterate, and access to mobile phones is persistently and comparatively low in rural areas (Carletto et al., 2021). To address some of these challenges, survey programs such as the World Bank's LSMS – Integrated Surveys on Agriculture (LSMS-ISA), FAO's AGRISurvey, and the 50x2030 Initiative often plan multiple field visits to improve data quality by reducing the length of recall period, but this can increase the burden for survey respondents and add to the cost of survey operations. Despite improvements brought about by many recent methodological developments, agricultural outcome variables continue to be plagued by measurement error related to respondent recall (Beegle et al., 2012a; Carletto et al., 2013; Deininger et al., 2012; Gourlay et al., 2019; Wollburg et al., 2020). Phone surveys, if adequately integrated with traditional survey operations, offer an opportunity to enrich the toolbox of agricultural statisticians in low- and middle-income countries, by providing an additional mode to collect data in a cost-effective and timely manner, taking some of the burden off the in-person interviews. Far from being a panacea, phone data collection has its own challenges.

Phone surveys cannot be expected to be a substitute for in-person surveys when it comes to administering long and complex survey instruments and can face serious issues with coverage bias when phone ownership and network coverage are less than complete in the population of interest. It is critical for data

producers and researchers to consider when and how phone surveys should be used in the future; when and for what purposes it is advantageous to employ phone surveys either to replace (parts of) in-person data collection or to complement it; how integration or supplementation of in-person and phone surveys should be conducted and what are the best practices for doing so.

In this paper, we review experiences with phone surveys in low- and middle-income countries against the conceptual framework provided by the Total Survey Error (TSE) paradigm (Biemer, 2010; Groves and Lyberg, 2010), drawing implications for the design and implementation of future survey practice. While there is a substantial amount of literature on the topic for high-income countries, particularly the US (AAPOR, 2010; Lavrakas et al., 2017; Groves et al., 1988) and Europe (Jäckle, Roberts and Lynn, 2006; Häder, Häder, and Kühne, 2012; Lynn and Kaminska, 2013; Villar and Fitzgerald, 2017), the trajectory in the adoption and diffusion of survey modes in those countries has been quite different than it is in many low-income countries. Landlines in the US and Europe were at one point nearly universal and are now being complemented and partially supplanted by mobile phone and web connections. Therefore, while the experience accumulated in the US and Europe is very relevant, many of the issues that matter to survey practitioners appear with several different features and nuances in low-income settings, particularly in rural areas.²

Against this backdrop and drawing on an extensive review of the literature and analysis of primary data, we conclude that there is great potential for the combination of phone and in-person surveys for agricultural data collection. To reach that conclusion our review takes into consideration the specific constraints around using phones for agricultural and rural surveys, including the nature of the questions, abbreviated interview duration, and the need for post-survey adjustments to account for uneven mobile phone coverage, as well as the advantages of phone surveys, namely the flexibility to schedule interviews appropriately around the agricultural season and the reduced costs and simplified logistics compared to in-person interviews.

The paper is organized as follows. In Section 2, we present the conceptual framework (based on the TSE paradigm) and adapt it to evaluate the role phone surveys can play in agricultural and rural survey systems in the coming years. We also introduce the phone survey programs and experiences on which we draw throughout this review. In Section 3 we discuss issues of coverage and nonresponse bias that emerge when designing and implementing phone surveys on agriculture, focusing on sampling and representativeness (3.1), respondent selection (3.2), and post-survey adjustments (3.3). Section 4 discusses issues of measurement including questionnaire design and survey timing (4.1), the role of the respondent (4.2), and the role of the interviewer (4.3). Section 5 discusses cost considerations. We conclude in Section 6.

² The literature in high-income countries distinguishes at times between fixed and mobile phone surveys. In low-income settings, and particularly in rural areas, mobile phones are by far the most prevalent. In what follows, we therefore just use the term phone survey, having in mind that in most instances the means of contact will be a mobile phone.

2. Conceptual framework and sources

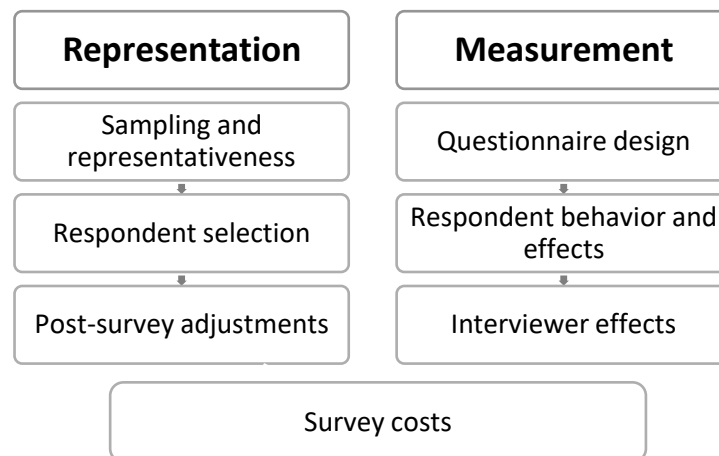
2.1 Conceptual framework

To frame our review, we lean on the Total Survey Error (TSE) paradigm as articulated in the survey methodology literature (Biemer, 2010; Groves et al., 2011; Groves and Lyberg, 2010) and applied to agricultural data collection among others by Carletto et al. (2021). Total survey error refers to the sum of errors, that is, bias and variance, arising in the survey lifecycle. Each stage of the survey lifecycle and related survey design choices are associated with different sources of survey error. The objective for survey designers is to make survey design choices that reduce the total survey error, subject to a budget constraint.

In their useful systematization of the survey lifecycle, Groves et al. (2011) distinguish the dimensions of *measurement* and of *representation*, as depicted in Figure 1. The *measurement* dimension broadly comprises what data a survey collects and how. The *representation* dimension refers to the target population a survey describes.³ Survey errors associated with the representation dimension include coverage, sampling, and non-response errors. Measurement error in its various forms as well as data processing errors are related to the measurement dimension.

In keeping with this framework, we analyze phone survey programs for agricultural data collection in terms of both the representation and the measurement dimensions, each with its associated sources of error. On the representation side, covered in Section 3, we tackle (i) sampling and representatives, (ii) respondent selection, and (iii) post-survey adjustments. On the measurement side, covered in Section 4, we tackle (iv) questionnaire design and look at the roles of (v) respondents and (vi) interviewers.

Figure 1. Conceptual framework



Source: Authors' adaptation based on Groves et al. (2011)

³ On the *measurement* side, Groves et al. (2011) distinguish 'construct', 'measurement', 'response', and 'edited response' as the stages in the survey lifecycle at which survey design choices to reduce error are made. On the *representation* side, Groves et al. (2011) distinguish 'target population', 'sampling frame', 'sample', 'respondents', and 'post-survey adjustments'.

We are also cognizant that there are trade-offs between reducing survey errors or increasing survey quality, and survey costs (Dillon et al., 2020). The most accurate measurement methods can be expensive and can come into direct competition with other survey objectives for example with covering a larger sample or collecting more information. Cost considerations are an important aspect in assessing how phone surveys can best complement or replace parts of in-person agricultural data collection. We discuss cost considerations in Section 5, highlighting the interactions and trade-offs between survey errors and survey costs.

2.2. Recent experience of agricultural and rural phone surveys in low-income settings

We draw from the experiences of several phone survey efforts implemented in rural areas, some of which were focused explicitly on agriculture, including: the World Bank-supported High Frequency Phone Surveys (HFPS), with an emphasis on the seven HFPS survey programs supported by the LSMS team in Burkina Faso, Ethiopia, Malawi, Mali, Nigeria, Tanzania, and Uganda (Living Standards Measurement Study, 2022); the Innovation for Poverty Action (IPA) RECOVR survey⁴; the World Food Programme's mVAM Project (World Food Programme, 2020); the Young Lives at Work program (Young Lives, 2022); the International Food Policy Research Institute (IFPRI) phone surveys (Alvi et al., 2021; Hirvonen et al., 2021a; Hirvonen et al., 2021b; Minten et al., 2020); World Bank's Listening to Africa and Listening to Latin America and the Caribbean initiatives (Ballivian et al., 2015; Dabalen et al., 2016); a World Bank Gender Innovation Lab survey of women in agricultural households in Western Uganda (Sharma et al., 2021); and the recent Georgia Survey of Agricultural Holdings (50x2030 Initiative, 2020; FAO, 2018). Appendix I includes a brief description of these surveys.

We also draw on experiences from phone surveys used in regions affected by conflicts or natural disasters (Hoogeveen and Pape, 2020) and during the 2014-2016 Ebola epidemic in West Africa (Etang and Himelein, 2020; Himelein et al., 2015; Maffioli, 2020; World Bank, 2014; Zafar et al., 2016). Several survey experiments involving the use of phone surveys aimed at measuring agricultural labor and crop production (Arthi et al., 2018; Gaddis et al., 2021; Kilic et al., 2021) also provide valuable insights for the administration of agricultural surveys via Computer Assisted Telephone Interviewing (CATI). Finally, we draw lessons from several review articles on phone surveys conducted in low- and middle-income countries, including Dabalen et al., 2016; Dillon, 2012; Etang and Himelein, 2020; Glazerman et al., 2020; Gourlay et al., 2021; Henderson and Rosenbaum, 2020.

3. Issues of coverage and nonresponse

3.1. Sampling and representativeness

In phone surveys focusing on agriculture, the survey design problem when it comes to sampling and representativeness is reaching a representative sample of the target population of agricultural households, holdings, or individual farmers over the phone. Many of the challenges are common to any survey, but a few are specific to the phone survey mode (phone ownership or access, the existence or ability to generate a list of phone numbers, the difficulties of securing respondents' consent at a distance) and to the fact that agricultural subjects (and in particular a non-random portion of them) are often less reachable by phone than the general population. In this section we illustrate how some of these issues play out in practice and the way and extent to which recent survey efforts have been able to overcome these issues to produce representative national estimates.

⁴ For more on IPA's RECOVR Survey, visit: <https://www.poverty-action.org/recovr/recovr-survey>.

The nature and prevalence of these issues of representativeness is often highly dependent on the type of sampling frame used. This choice affects the extent to which a representative sample can be achieved, and post-survey adjustments can improve representativeness ex-post. While in-person surveys in low- and middle-income countries often follow the common two-stage sampling approach, phone surveys adopt a more varied set of approaches. There are three predominant approaches to establishing a sampling frame and drawing a sample for a phone survey, each with distinct advantages and disadvantages when it comes to both implementation as well as representativeness (Lepkowski et al. 2007; Glazerman et al., 2020; Himelein et al., 2020; Gourlay et al., 2021):

1. **Recontact:** Respondents are drawn from a list of phone numbers from previous surveys or programs or from a recent listing of households/holdings conducted within selected EAs (or other groups).
2. **Random Digit Dialing (RDD):** A sample of randomly drawn numbers corresponding to the construction of phone numbers in the country is drawn and called.
3. **Telecom list:** Respondents are drawn from a list of active phone numbers obtained in most cases from network service providers.

When it comes to agricultural surveys, these three approaches will sometimes require specific recommendations and considerations, which will be noted in the sections to follow.

Mobile phone ownership and coverage

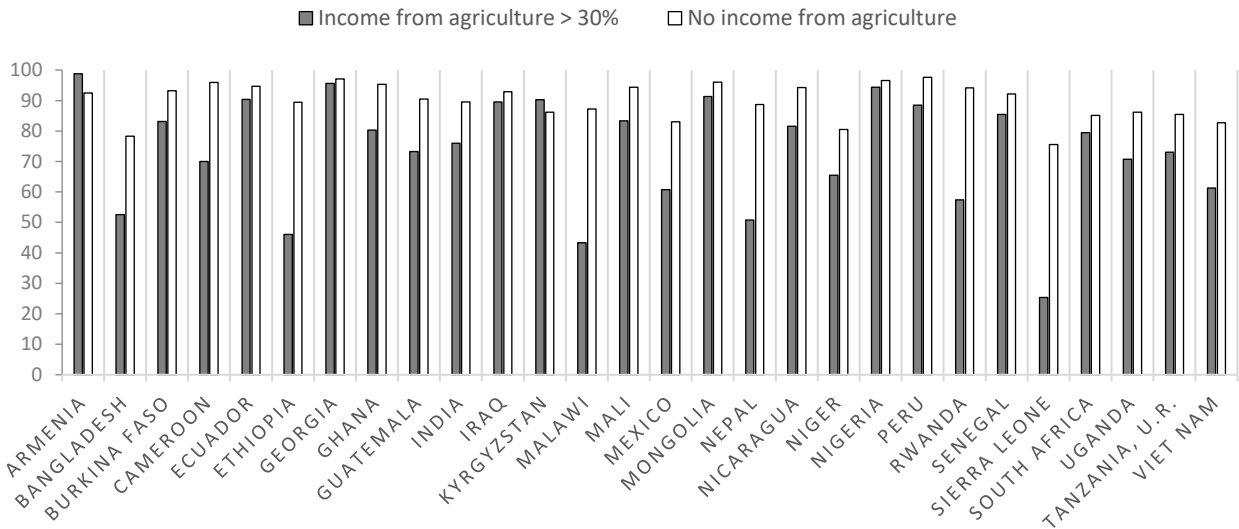
One of the challenges in order for mobile phone surveys to yield data representative of the population of interest is at the first step tied to the level of mobile phone penetration in the population (Leo et al., 2015). Mobile phone ownership is not universal in low- and middle-income contexts: mobile penetration rates were at 67 percent of the general population globally in 2019, and at 45 percent in Sub-Saharan Africa (GSMA, 2020), though penetration continues to increase at a significant rate. Moreover, agricultural households are less likely to own phones than the general population (Ambel et al., 2021; Himelein et al., 2015; Leo et al., 2015) and adoption of mobile phones in low- and middle-income countries has been found to be associated with wealth, gender, remoteness, and education (Henderson and Rosenbaum, 2020; Himelein et al., 2020). Drawing on household survey evidence from 28 countries, **Error! Reference source not found.** Figure 2 illustrates the discrepancy in access to mobile phones, with access defined as mobile phone ownership within the household, for the population with no household income from agriculture and those with over 30% of household income coming from agriculture.⁵ The majority of countries exhibit a pattern in which households that are reliant on agriculture, i.e., those with over 30% of their total income coming from agriculture, have lower rates of mobile phone access.⁶ This is especially pronounced in Ethiopia, Malawi, and Sierra Leone where the difference between the share of

⁵ Households earning between 0% and 30% of total income from agriculture are excluded for presentation purposes. The rates of phone access for this group of households generally falls between the rate of those with no income from agriculture and those earning more than 30% of income from agriculture.

⁶ Only countries with data on access to mobile phones in 2010 or later are included. The year of data collection by country is as follows: Armenia - 2013; Bangladesh - 2010; Burkina Faso - 2014; Cameroon - 2014; Ecuador - 2014; Ethiopia - 2016; Georgia - 2015; Ghana - 2013; Guatemala - 2014; India - 2012; Iraq - 2012; Kyrgyzstan - 2013; Malawi - 2017; Mali - 2017; Mexico - 2014; Mongolia - 2014; Nepal - 2011; Nicaragua - 2014; Niger - 2014; Nigeria - 2019; Peru - 2019; Rwanda - 2014; Senegal - 2011; Sierra Leone - 2011; South Africa - 2015; Tanzania - 2015; Uganda - 2016; Vietnam - 2010. For more details on the country surveys and the RuLIS methodology, visit: <https://www.fao.org/in-action/rural-livelihoods-dataset-rulis/en/>.

the population with mobile phone access between the non-agricultural population and the most dependent on agriculture is more than 40 percentage points.

Figure 2. Share of population with mobile phone access, by share of total household income from agriculture.



Source: Authors' calculation based on the FAO Rural Livelihoods Information System (RuLIS) database

A similar pattern in mobile phone access between agricultural and nonagricultural households is observed in the countries where the World Bank HFPS were implemented. Using HFPS and LSMS-ISA data from Ethiopia, Malawi, Nigeria, and Uganda, Ambel et al. (2021) demonstrate that the sample of households who own or have access to mobile phones are considerably different from the sample without access to a mobile phone. Households without access to a mobile phone tend to be poorer and have lower living standards (across multiple dimensions) relative to the sample of households with access to a mobile phone (Ambel et al., 2021).

This issue of coverage is common to all three of the predominant phone survey sampling approaches. In what follows we discuss possible approaches to mitigate this risk during different stages of survey implementation, and in Section 3.3 we discuss post-survey adjustments that attempt to manage this problem.

Nonresponse

While the issue of coverage is critically important for the representativeness of phone surveys, another important factor is the pattern of nonresponse. In phone surveys in particular, nonresponse can be linked not only to refusals, but to failure of the interviewers to establish contact. Just like refusals, this type of nonresponse can be systematic and therefore introduce bias. For example, in a low-income country setting, nonresponse tends to be higher in areas with unreliable network or electricity which makes it more difficult to successfully reach respondents. Areas with poor mobile network or electricity are likely to be rural and poorer, so higher nonresponse in these areas will bias the sample toward urban and richer

areas. Since agricultural households predominantly reside in rural areas where poor reliability of the mobile network and electricity is more common, they likewise may be underrepresented.

The level and nature of nonresponse can vary substantially depending on the context as well as the sampling approach adopted for the phone survey. Recontact surveys generally have higher response rates relative to RDD and telecom list-based phone surveys. One reason for this is that sampled units in recontact phone surveys have already been previously successfully contacted and interviewed and have an established relationship with the surveying agency as well as familiarity with being interviewed (relative to RDD and telecom list-based respondents). Furthermore, with RDD generated numbers are often unassigned or non-working, and survey practitioners are unable to tell in advance whether a number belongs to a household or a firm. For example, in an Ebola survey in Liberia, 214,823 calls to RDD-generated numbers were made with only 24,000 of the numbers connecting; while in Ghana over 1 million RDD numbers were needed to get 16,003 connections (a connection rate of only 1.5%) (Himelein et al., 2020; Maffioli, 2020). Once a frame of phone numbers known to be assigned and working is established, response rates in RDD-based phone surveys are much higher, though usually still lower than recontact surveys (Henderson and Rosenbaum, 2020). IPA's Recovr surveys, which relied on random digit dialing of phone numbers achieved response rates of between 4 percent in Mexico and 59 percent in Burkina Faso, with an average response rate of 28 percent (IPA, 2020). By contrast, some of the World Bank HFPS (recontact) surveys relied on phone numbers from previous LSMS-ISA surveys and achieved response rates ranging from 60 to 93 percent at an average of 74 percent (Gourlay et al., 2021; Table 1).

Table 1. Response rates in round 1 of IPA Recovr surveys based on RDD (top panel) and HFPS surveys based on contact information from previous surveys (bottom panel)

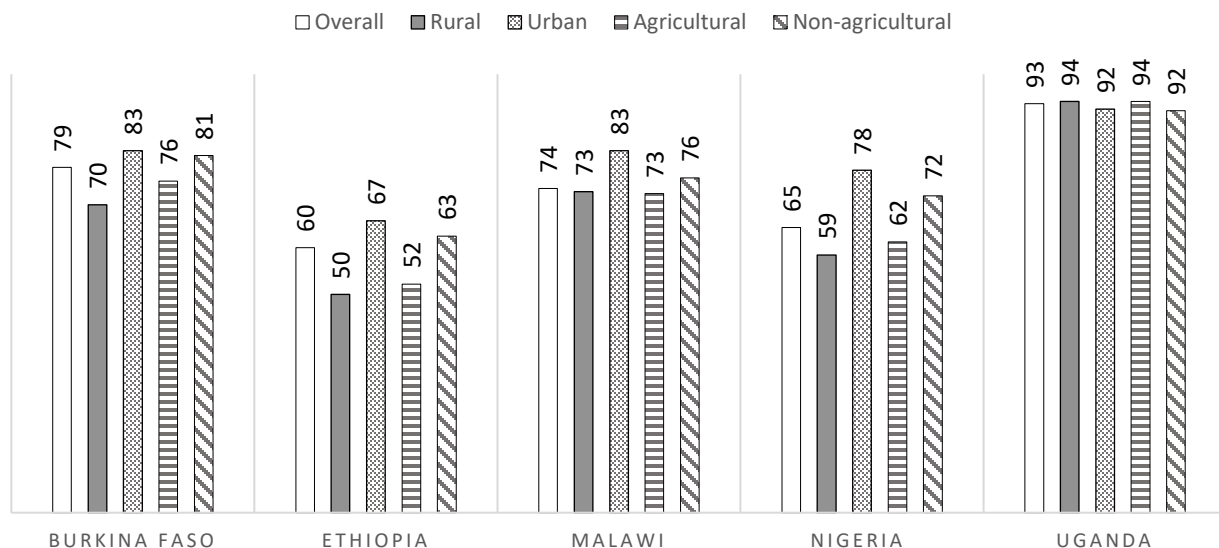
Country	Respondents	Attempted contacts	Response rate
Recovr - Random Digit Dialing (RDD)			
Burkina Faso	1,356	2,284	59%
Colombia	1,507	6,984	22%
Côte d'Ivoire	1,329	3,022	44%
Ghana	1,357	10,781	13%
Mexico	1,330	29,876	4%
Philippines	1,389	8,378	17%
Rwanda	1,482	4,234	35%
Sierra Leone	1,304	3,831	34%
HFPS - Existing phone contact from previous survey			
Burkina Faso	1,968	2,500	79%
Ethiopia	3,249	5,374	60%
Malawi	1,729	2,337	74%
Nigeria	1,950	3,000	65%
Uganda	2,227	2,386	93%

Sources: Innovation for Poverty Action (IPA, 2020) and Gourlay et al. (2021). Based on survey round 1 in each country. The IPA Recovr surveys used random digit dialing of a nationally representative sample of phone numbers.

For the HFPS surveys, we can further diagnose the observed patterns of nonresponse. Figure 3 displays response rates from five countries broken down by urban/rural and agricultural/nonagricultural status. The response rates in Uganda, which are very high, benefitted from the relatively recent pre-COVID in-person survey which allowed for an updating of contact information shortly before implementation of the

HFPS. In Burkina Faso, Ethiopia, Malawi, and Nigeria response rates are lower (sometimes substantially) for rural and agricultural households, suggesting potential for nonresponse bias in these segments. However, as for the case of coverage bias, there are post-survey adjustments (described in Section 3.3) that can be implemented to at least partially correct for nonresponse bias.

Figure 3: HFPS Response Rates



Source: Authors' calculation based on HFPS data.

Sampling methods for phone surveys

When considering the sampling approach to adopt for a survey, there are substantial trade-offs to each in terms of representativeness and methods available for correcting bias. Among the three common approaches (i.e., recontact, RDD, or telecom list-based), recontact surveys that use phone numbers from previous in-person, representative surveys may be considered the preferred sampling method and has also been the most commonly used approach in low- and middle-income settings (Ceballos et al., 2020; Dabalen et al., 2016; Glazerman et al., 2020; Himelein et al., 2020). The main advantage of this approach is that there exists a well-defined (often representative) sample selected for a previous in-person survey and for which there is a profile of characteristics available for the sample units (i.e., households or individuals). This permits a direct assessment of the representativeness of the selected and successfully interviewed phone survey sample and facilitates post-survey adjustments to reduce selection biases (Ambel et al., 2021; Gourlay et al., 2021; Himelein et al., 2020). The existing information is also critical for identifying agricultural households.

Response rates are typically far lower for RDD surveys than recontact surveys, thus requiring more attempts and a larger selected sample to achieve the same number of completed interviews. Furthermore, under RDD, response rates can be expected to worsen if data collection targets agricultural or rural subjects, as they will need to be identified. Effectively, this will add a step to the sample selection process to verify with respondents if they belong to the target population. With contacts from existing

surveys, prior survey information may help in identifying households that are active in agriculture or reside in rural areas (with some limitations, as these features are not time invariant), whereas in the case of RDD no prior information is available to inform the selection prior to attempting to contact the respondent.

There are drawbacks and limitations to using contact information from previous surveys as sampling frames for agricultural phone surveys, however. The sample size of the previous in-person survey inevitably constrains the sample size of the phone survey, which may be problematic especially when contact phone numbers are available only for a smaller share of in-person households. Moreover, depending on the time passed since the previous in-person survey, the list of phone numbers may not be current. This could mean that many phone numbers have been disconnected in the meantime which could also mean the list of households is no longer representative of the target population. Another drawback of selecting a phone survey sample from previous in-person surveys is that there is some loss in sampling efficiency due to the multi-stage, clustered design of in-person survey samples. These design effects from clustering the in-person sample will be imported into the phone survey sample and therefore increase the variance of estimates obtained from the phone survey sample relative to a simple random sample of the same size. An additional consideration when using phone numbers from previous in-person surveys is the purpose and target population of the previous in-person survey. It is important to make sure that the population of interest of the phone survey aligns with the previous in-person survey (Glazerman et al., 2020) or contains the population of interest. For example, an in-person survey whose population of interest was farming households would not be a suitable frame for a phone survey targeting the general population.

A sampling strategy based on a list of phone numbers generated by random digit dialing (RDD) or obtained from a mobile network provider is an alternative (or the only feasible option) especially when no previous contact information is available. These strategies also eliminate the need for sample clustering, which is generally required for cost-efficiency in in-person survey operations, and do not constrain *ex ante* the size of the sample that can potentially be contacted. However, both strategies have drawbacks in terms of sample representativeness. The lack of *a priori* information on the potential respondents associated with each phone number impedes the analysis of non-responders and limits post-survey reweighting to counteract selection biases (Himelein et al., 2020). Further, a list obtained from a mobile network operator is unlikely to represent even the phone-owning population if it is from one specific operator in a country where there are multiple operators.

A mixture of these methods could also be implemented to arrive at a representative sample. For example, in a situation where a recent previous survey is available, but the sample of phone numbers is small, survey designers can supplement the list with additional phone numbers generated through RDD. This was the approach taken in the Kenya high-frequency phone survey on the socio-economic impacts of COVID-19 in which the World Bank is collaborating with the Kenyan National Bureau of Statistics (Pape et al., 2020).

Increasing coverage, reducing non-response, and improving representativeness

Given the above issues, survey designers can draw on several approaches to attempt to reduce the level and impact of these sources of bias in phone survey samples. In-person survey operations should invest in the first place in the collection of phone contact information of surveyed households and individuals

and subsequently keep these lists updated. Collecting phone numbers from more than one adult household member increases the chances of successfully re-contacting the household for any future survey. Updating and maintenance of the list of phone numbers may be done through brief follow-up calls or surveys though one needs to be mindful of inducing increased attrition via too many contacts. When possible, working with local officials or community leaders to update and maintain lists may be beneficial. This approach proved successful in increasing coverage ex-post in a recent World Bank Africa Gender Innovation Lab survey of rural agricultural households in Western Uganda (Sharma et al., 2021). An important additional consideration is to ensure that respondents can give informed consent to their phone numbers remaining on file for follow-up interviews (Glazerman et al., 2020).

In a situation where no recent previous phone contact information is available, survey designers can create a new sampling frame of phone numbers through in-person visits, listing households, their phone numbers, and some basic demographic information in target enumeration areas much like listing is done in traditional in-person surveys. The newly created lists can then be used for follow-up phone surveys.

Coverage at the household level may be increased by collecting contact information from a reference person outside the household, such as a friend, relative, or neighbor, particularly in situations when none of the household members own a phone. In the HFPS that drew upon the LSMS-ISA surveys, some of which used this approach, the availability of reference person contact information helped not only to retain households that do not have phone numbers of their own but also facilitated contact with households that could not be reached on their own phone(s) but were reached through a reference person's. The share of HFPS respondents who were ultimately reached through reference person contact information ranged from 7 percent in Burkina Faso to 20 percent in Malawi. In all cases, the share of respondents reached through reference contacts was higher for rural and agricultural households with a substantial 22 and 23 percent of agricultural households reached through reference contacts in Ethiopia and Malawi, respectively.

Another strategy to increase coverage and response rates is by providing free phones to respondents. This strategy was used by the World Bank for its Listening to Africa (L2A) series of evaluations as well as by other researchers for individual experiments or impact evaluations (Dabalén et al., 2016; Dillon, 2012; Gaddis et al., 2021; McCullough et al., 2020). Beyond the cost implications of this strategy, providing mobile phones may meaningfully affect respondent behavior which the survey is seeking to measure and therefore interfere with the survey objectives or experimental design. Survey designers and researchers should carefully consider the implications of providing phones to respondents. An approach to limit this effect is to hand out phones to certain reference contacts in each community which can then connect selected respondents with interviewers.

One method to improve the chances of successful contact and interview with a respondent is to arrange some automated pre-contact to the sample units, typically and perhaps easiest through SMS message. This pre-contact can help to inform respondents of the survey as well as when they can expect a follow-up call for the interview. With this information, respondents may be more willing to answer a call from an interviewer and more closely monitor their phone if the call is expected. Pre-contact SMS messages have been shown to reduce nonresponse and enhance cooperation among phone survey respondents (Dal Grande et al., 2016).

In general, it is good practice to implement a clear protocol for when to contact respondents and how many call attempts to make. Ensuring variation in the day of the week and time of day that call attempts are made can help in reaching a respondent at a day and time where they are available to answer and be interviewed. Providing small incentives to respondents may also help to ensure cooperation and reduce nonresponse, though the incentive has to be carefully calibrated to not contaminate responses from the respondent. One critical principle is to make all efforts to limit respondent fatigue through keeping the interview length manageable and the questions asked straightforward and (wherever possible) avoiding asking sensitive or uncomfortable questions. These and other approaches are further discussed in Section 4.2 below.

3.2. Respondent selection

The choice of respondent is a key survey design decision for any survey. The survey objectives, and, in particular, the desired unit of observation, inform this decision. Survey respondents need to have knowledge of the issues in which the survey is interested, and the choice of survey respondent should reflect that. This requirement is not unique to phone surveys, but it can be substantially more difficult in practice to reach a desired respondent through a phone survey. Phone ownership within the household and among the general population is often skewed towards household heads, men, and older members (Brubaker et al., 2021), so interviewers may need to ask for phone owners to pass the phone along to the desired respondent, provided they are available at the time of the call (or schedule an appointment for another call). If survey questions are sensitive and require respondent privacy, phone surveys may not be a suitable tool (Section 4.2). Interviewing more than one household member, for instance all plot managers in an agricultural household, increases the response and time burden on the household and may carry the risk of greater attrition.

When it comes to collecting individual-level data, for instance on nutrition or individual food security, selecting a sample of individuals representative of the population of individuals is preferable. For phone surveys, the challenge for individual-level data collection again comes down to the pattern of phone ownership whereby some segments of the population are less likely to own or have access to a mobile phone and thus will not be adequately represented in phone survey samples. Such issues can be magnified in recontact surveys that use the household (rather than the individual) as their unit of observation. Brubaker et al. (2021) show that the main respondents in High Frequency Phone Surveys on COVID-19 in four African countries are predominantly household heads, and also better educated and more likely to own a non-farm enterprise than adults overall, so that individual-level estimates are not representative of the population at large. In turn, the scope for a ‘proxy respondent’ answering on behalf of other household members depends critically on the kind of information collected and has been linked to measurement error (Kilic et al., 2020). For these reasons, a randomized (probability-based) respondent selection protocol—whereby a targeted respondent within the roster of eligible household members is randomly selected for interview—is the preferred strategy to achieve unbiased individual-level data.

3.3. Post-survey adjustments

When mobile phone ownership is not complete, the phone-owning population and the general population that the phone survey seeks to represent are different, often resulting in coverage bias. Similarly, if nonresponse is systematic, then the responding sample will not be representative of the population of interest. There are a range of post-survey reweighting techniques that can be used to attempt to counteract these differences and improve the representativeness of the estimates obtained (Valliant et

al, 2013). These techniques are particularly critical for agricultural phone survey samples where issues of coverage and nonresponse are likely to be magnified given a lower share of mobile phone ownership and likely less reliable mobile networks and electricity in rural areas. The availability and success of these different adjustment methods will also vary depending on the sampling method for the survey and the amount of reference information available on the general population.

One commonly used reweighting approach is weighting class adjustments in which the sample is divided into cells based on the characteristics of sampling units (households, farms, individuals, etc.) such as location, size, or sex that are demonstrated or assumed to be associated with coverage or nonresponse bias. Then, the sampling units in each cell are weighted proportional to the general population total of each cell (Little, 1986).

A second reweighting technique is response propensity adjustment (Valliant et al, 2013). This technique models the probability of response (i.e., selection and successful interview) for each sampling unit given a range of its observable characteristics. Then the inverse predicted probability of selection derived from this response probability model is used to adjust sampling weights and counteract the bias (see Himelein et al. (2020a) for an in-depth discussion). Response propensity adjustment techniques can accommodate many variables, allowing reweighting to be more fine-grained, while weighting class adjustment can only accommodate a handful of characteristics as otherwise the number of different cells becomes too large.

In addition, it is good practice to calibrate the adjusted sampling weights to match known population totals (e.g., number of households by urban, rural) which can help to reduce overall standard errors and also help to reduce bias due to coverage or nonresponse. There are several calibration methods that can be implemented, depending on the level of detail of information available for calibration (Valliant et al, 2013; Lundström and Särndal, 1999; Andersson and Särndal, 2016).

With a previous in-person survey as a sampling frame (i.e., recontact surveys), there is often a wealth of information associated with each household (or individual or farm), which readily allows modeling the response probability. This was the approach taken in the LSMS-supported HFPS, for example (Gourlay et al., 2021). With sampling frames based on RDD or lists provided by mobile network operators, however, there is usually no or limited information available in addition to telephone numbers. Auxiliary data sets are required for post-survey adjustments to reduce selection biases. A recent census or a recent nationally representative survey may serve this function. The phone survey data and the auxiliary data need to have in common a set of variables, such as demographic and location variables, for weighting adjustments to be possible (Lepkowski et al. 2007; Himelein et al., 2020).

Post-survey adjustments such as reweighting can never fully correct coverage and non-response biases in phone survey data. Their relative success in reducing biases depends on how much of the population is left out, how different the selected sample is from the general population, and how much information is available for bias reduction methods. Ambel et al. (2021) examine how well response propensity adjustments perform at reducing coverage and non-response biases in LSMS-supported HFPS in four African countries. They find biases to be substantially reduced but not fully eradicated in all four surveys. In contrast to Ambel et al. (2021), Brubaker et al. (2021) investigate respondent selection biases in individual-level data from the same surveys. Selection biases are found to be more pronounced at the individual level than at the household level, such that reweighting is relatively less successful in overcoming these often more substantial biases. Obtaining more representative individual-level data would require a probability-based/randomized respondent selection protocol.

4. Issues of measurement

4.1. Questionnaire design, seasonality, and survey implementation

Agricultural data collection often requires complex instruments reflecting the nature of agricultural production, with seasonality and a high prevalence of shocks as well as different input decisions taken with varying frequencies and at different times. Integrated agricultural surveys, like the LSMS-ISA, 50x2030 Initiative, and FAO's AGRISurvey, collect data ranging from livestock production and asset ownership to agricultural inputs and labor use, and crop harvest quantity and value, among other issues. In some cases, input and output data are collected at the plot-level and multiple crops and seasons are covered in a single questionnaire (Dillon et al., 2021). The basic principles of questionnaire design apply as much to phone surveys as to other modes of agricultural data collection. However, phone surveys are considered more constrained than in-person surveys in terms of questionnaire length and complexity.

As we argue in this section, it is not realistic to administer such complex and extensive survey questionnaires over the phone in full. These operations will continue to require in-person interviewing. However, shorter special purpose survey questionnaires can be (and have been) administered in full by phone. Under certain conditions, phone surveys can complement in-person interviewing even for extensive agricultural surveys: some survey modules can benefit from the flexibility in timing that phone interviews offer, and some survey modules lend themselves more readily to data collection over the phone. Phone surveys can also provide an opportunity to collect additional information of interest.

Questionnaire length and complexity

There is consensus that phone interviews should not be too long and recent research documents how response fatigue decreases data quality as the duration of phone surveys gets longer (Abay et al., 2021; Ballivian et al., 2015; Ceballos et al., 2020). Acceptable survey length varies depending on the topic, the country context, whether there is a previous relationship with the respondents, and on the respondents themselves. Glazerman et al. (2020) recommend interviews should not exceed 20 minutes while Mathur (2020a) find that the ideal survey length in that context was roughly 10–15 minutes. The LSMS-supported HFPS were designed to be around 20 minutes on average, though there was some variation across rounds and between countries, some of which averaged around 30 minutes (Gourlay et al., 2021). Survey designers will need to pilot questionnaires to determine what survey length is feasible without jeopardizing data quality and nonresponse.

In addition to length, questionnaire complexity is a limiting factor in phone surveys, which is thought to increase response fatigue and, in turn, affect data quality and attrition. Excessively complex questions should therefore be avoided in surveys of any mode. However, key concepts in agricultural statistics are inherently complex. Some complexity is necessary for survey questionnaires to reflect this adequately as oversimplification comes at a loss of information that can reduce the value of the data. In in-person surveys, interviewers rely on visual aids to make difficult questions more palatable. This is not possible with phone surveys.⁷

Given these limitations in terms of length and complexity, certain topics of agricultural data cannot be collected over the phone. There are, however, some experiences with simplifying questionnaires to make

⁷ An alternative to be explored for the future may be the use of smartphone applications that provide visual aids and other survey support. An early example, Daum et al. (2019), used a pictorial smart-phone app to collect time use and nutrition data in rural Zambia.

them suitable for phone surveys but these should be considered carefully in light of the objectives of a given survey. One approach is asking questions at a more aggregated level of observation, such as the farm or the household instead of the plot or the individual. In this approach, instead of enumerating a plot roster and eliciting inputs and outputs at the plot-level, questions are asked in aggregate, e.g. *“What is the total area of land, summing all parcels, you operate?”*. This approach was used in some of the LSMS-supported HFPS. A similar kind of aggregation was done in the IPA’s RECOVR surveys (Innovations for Poverty Action, 2020) but with respect to household composition and demographic characteristics. This kind of aggregation may save time, though some aggregation questions, such as total area of land summing all parcels, may require respondents to do computations which increases respondent burden and introduces potential for error. Also, aggregation of individual- or plot-level outcomes at the household-level necessarily entails a loss of information.

Another strategy to simplify complex questions, such as those involving land or harvest quantifications, is posing such questions in relative terms, as compared to a previous period. This makes responses easier because farmers only need to roughly compare two periods, without the need of an exact quantification. For example, IFPRI’s vegetable value chain phone surveys in Ethiopia ask questions like *“Do you use more or less fertilizer currently than normally?”* with responses *“1. The same; 2. Lower; 3. Higher”* (Hirvonen et al., 2021b; Minten et al., 2020). This simplification strategy is useful for before-and-after comparisons when no baseline quantification exists for comparison, as was the case in many of the COVID-19 phone surveys. However, responses may be biased positively or negatively depending on the sources of bias at play (for example, social desirability bias or perceived benefits to indicating poor outcomes). Moreover, the loss of information is significant, so that this strategy is not suitable for many types of agricultural surveys. For instance, it would not be feasible for a large-scale crop production survey seeking to generate national statistics.

A third approach is to retain some detailed and disaggregated data collection but reduce the response burden by asking fewer questions or covering only a sub-sample of crops, plots, or individuals. IFPRI’s vegetable value chain phone surveys in Ethiopia asked detailed questions focusing on the most important vegetable household’s grow (Hirvonen et al., 2021b; Minten et al., 2020). The LSMS-supported HFPS followed the same approach, focusing on the ‘main crop’ in the reference agricultural season. In contrast, the World Bank’s Africa Gender Innovation Lab used phone calls to collect input data for one selected parcel in a survey of rural agricultural households in Western Uganda (Sharma et al., 2021). In the context of their survey experiments, Arthi et al. (2018), Gaddis et al. (2021), Kilic et al. (2021) successfully collected plot-level input and output data over the phone, focusing on one specific topic (labor inputs and extended-harvest crop production). In all three settings, the plot-level data collected through phone interview was deemed reliable and even superior to the standard end-of-season labor module administer in a one-off in-person visit. These examples illustrate that it is possible to retain a certain level of complexity if focusing on a narrow set of issues or a sub-sample of plots, crops, or individuals, thus allowing for short phone interviews. Sub-sampling has implications for the representativeness of the data so this approach may not be suitable to replace an extensive agricultural sample survey in full but may be used to complement in-person data collection.

Survey timing and seasonality

In addition to questionnaire design, survey timing is a critical design choice in agricultural data collection, particularly because of the highly seasonal nature of agricultural production and the varying frequency

and intensity with which agricultural activities of interest are performed. Some outcomes in agriculture (labor inputs, harvesting of some crops) that are dynamic over the agricultural season would benefit from more frequent data collection, some happen only during certain periods, while others are static within a given crop growing season and may well be collected only at one point during the season. Centered around the main crop growing season, agricultural surveys often visit farms at the end of the season. This approach can lead to respondents having to remember activities and outcomes many months in the past, and recall decay has been found to affect data quality in agricultural surveys (Beegle et al., 2012a; Wollburg et al., 2020). Increasing the frequency of data collection has been shown empirically to improve data quality in several domains of interest (Arthi et al., 2018; Gaddis et al., 2021; Kilic et al., 2021; Knippenberg et al., 2019).

Phone surveys offer the flexibility and cost advantage to conduct surveys at the appropriate time and more frequently. Agricultural labor inputs are used frequently and often in a manner that is difficult to track. Arthi et al. (2018) in Tanzania and Gaddis et al. (2021) in Ghana show experimentally that weekly phone surveys improve the quality of labor data relative to the standard end of agricultural season in-person modules. Weekly phone surveys have also shown to yield better harvest data quality for extended harvest crops, like tomatoes, roots and tubers (Ceballos et al., 2020; Kilic et al., 2021), which are often harvested over the course of the season rendering recall difficult. The Georgia Survey of Agricultural Holdings fielded an AGRIS production survey over the phone relying on quarterly data collection to shorten the recall period (50x2030 Initiative, 2020). Higher frequency data collection over the phone can also be suitable for agricultural input and output prices (Hermosilla, 2017; Hoogeveen et al., 2014), though community- rather than household-level data collection is likely appropriate. Recall bias has also been documented in the livestock sector, with Zezza et al. (2016) providing evidence on recall bias in the measurement of milk off-take in Niger. The Global Strategy for Improving Agricultural and Rural Statistics successfully used phone surveys in the collection of livestock productivity data, also in Niger (Bako, 2018).

There are other agricultural variables to which similar reasoning applies, and which would likely benefit from higher frequency data collection by phone, though this has not been validated empirically. These include non-labor inputs such as fertilizers and pesticides as well as weather shocks such as droughts, flooding, extreme heat, or insect infestations, and the damage these shocks cause (Knippenberg et al., 2019). Some agricultural variables are generally static over a crop growing season, such as access to irrigation, land ownership, stocks of large ruminants, and more frequent data collection is unlikely to improve the quality of these variables, diminishing the possible benefits of collecting such information via phone.

Food consumption (expenditure) data could in principle benefit from more frequent data collection, particularly to temper the excess variation that may be associated with collecting data over a short time interval (Gibson and Alimi, 2020). Full-fledged consumption modules are however too time intensive to be suitable for implementation over the phone (Abate et al., 2021). More synthetic measures of consumption like dietary diversity indicators or food frequency questions can be adapted and collected

over the phone (Aggarwal et al., 2021; Picchioni et al., 2021). Higher frequency data collection may also be preferable for monitoring food security (Knippenberg et al., 2019).⁸

Finally, for some agricultural variables in-person visits remain strictly preferable. This is the case for those variables for which plot-level information or physical measurements are sought, such as inventories of agricultural parcels/plots and crops planted as well as land area owned or cultivated, for which measurement with GPS devices is best practice in low-income, smallholder settings (Carletto et al., 2017; Dillon et al., 2021). GPS land area measurement relies on trained interviewers being on-site. Even when land area measurement is based only on farmer-reporting, taking a full inventory of parcels/plots and the crops planted on them is complex and important enough to warrant an in-person visit, where possible. Given the problem of long recall periods, these variables are probably best collected in a pre- or post-planting visit. Similarly, in the case of seasonal crops harvest in-person harvest or post-harvest visit will be necessary if crop-cutting is included in the survey. Even without crop cutting, the level of complexity and length of best-practice harvest modules makes them suitable for in-person data collection and not appropriate for data collection over the phone. Phone surveys may however be used to support harvest diaries such as was done by Kilic et al. (2021) and Deininger et al. (2012). A summary of survey topics and their estimated suitability for implementation via phone survey is presented in Table 2.

Table 2. Potential for data collection over the phone by survey topics and data types

Topic/Data Type	Suitable for phone surveys	Comments	References
Food security	Yes	Higher frequency data collection, including over the phone, may be beneficial. The metadata for the Food Consumption Score (FCS) recommends implementation every 1 to 6 months if the objective is to monitor food security, and food frequency data are collected over the phone. For FIES, FAO recommends to randomly rotate the sample among rounds.	Knippenberg et al. (2019); Amankwah and Gourlay (2021); Horjus (2010); Picchioni et al. (2021); personal communication with FAO FIES Team ⁸
Consumption	Not for full-fledged, expenditure module	Phone surveys unlikely to be suitable to administer either weekly diaries or a full-fledged 7-day recall module. Even the 24-hour recall has not performed well in tests.	Abate et al. (2021); Aggarwal et al. (2021)
Land inventory, plot characteristics, and land area	No	Plot level interviews too burdensome for phone interviewing. Best-practice GPS land area measurement requires in-person visit.	Dillon et al. (2021)

⁸ The Food Insecurity Experience Scale (FIES) proved to be a valid tool when applied in repeated surveys. More specifically, in case of surveys repeated in short intervals and more importantly in case of panels, FAO recommends to randomly rotate the sample among rounds, avoiding asking the same questions to the same people after a short period of time, thus minimizing dropouts and adaptation strategies. In these cases, FAO also advises to use shorter recall periods that do not overlap among rounds (FAO FIES Team, personal communication, March 2022).

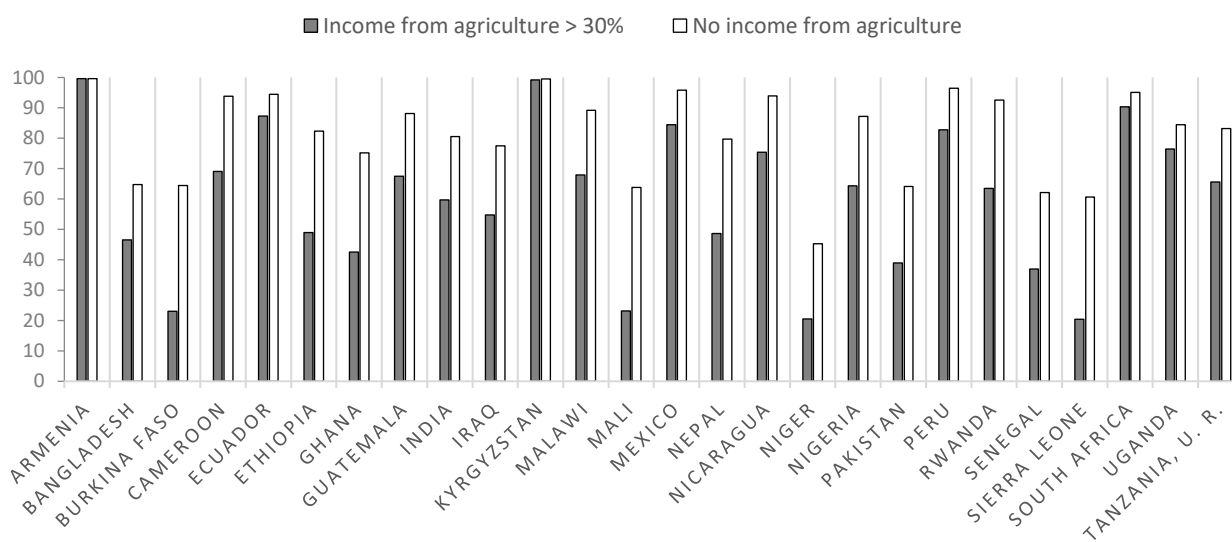
Agricultural assets	Yes	List of basic agricultural assets likely possible over the phone. Unlikely to change over the season so no specific comparative advantage for phone surveys. However, information on certain assets may be sensitive complicating data collection over the phone.	
Agricultural labor inputs	Yes	Evidence suggests phone surveys can improve data quality. Auxiliary diaries may be beneficial for data collection.	Arthi et al. (2018); Gaddis et al. (2021)
Non-labor inputs	Potentially	Phone surveys can be used to reduce recall length which may improve quality of the data. Auxiliary diaries may be beneficial for data collection.	
Livestock production	Yes	Phone survey was tested successfully in Niger. Higher frequency data collection considered advantageous for extended production livestock products (milk, dairy). Auxiliary diaries may be beneficial for data collection. Phones may also be used to track or contact semi-nomadic herders as they move, provided sufficient penetration, signal, energy access is available.	Bako (2018); Zezza et al. (2016); Hiraga, Uochi and Doyle (2020); Karamba and Salcher (2020); Teickner et al. (2020)
Harvest of extended harvest crops	Yes	Evidence suggests phone surveys can improve data quality. Auxiliary diaries preferable.	Ceballos et al. (2020); Kilic et al. (2021)
Crop harvest	No	Standard crop harvest module at crop-plot level likely too heavy for phone surveys in most countries. Crop-level harvest module implemented in e.g. Georgia. Harvest diaries have also been tested but phone surveys unlikely to be suitable to administer such diaries. Crop cutting requires in-person visit.	Deininger et al. (2012)
Market prices	Yes	Market price data collection appropriate at the community or market level but farm-gate prices at household/farm level likely difficult.	Hermosilla (2017); Hoogeveen et al. (2014)
Shocks and damage caused by shocks	Potentially	Phone surveys can be used to reduce recall length which may improve quality of the data. Auxiliary diaries may be beneficial for data collection.	Knippenberg et al. (2019)

4.2. Respondent behavior and effects

The characteristics of respondents in agricultural surveys may differ from national populations in ways that are material for designing and implementing phone surveys for agricultural data collection.

Agricultural and rural households in low-income settings often have higher illiteracy rates than the general population, for example. This is illustrated in Figure 4, which presents the literacy rates for households with no income from agriculture vis-à-vis those earning more than 30% of total income from agriculture for 26 low- and middle-income countries.⁹ If a meaningful share of respondents can be expected to have literacy difficulties, text message-based and Interactive Voice Response (IVR) phone survey modes are likely unsuitable as these usually require the respondents to read and/or write.

Figure 4. Adult literacy rate (ages 15+), by share of total income from agriculture.



Source: Authors' calculation based on the FAO Rural Livelihoods Information System (RuLIS) database

Response fatigue and incentivization

Fatigued respondents may cease to answer truthfully, refuse to answer questions, or stop participating in a survey over time. Various factors can contribute to the respondent burden, including survey complexity and length. For instance, Ambler et al. (2021), through an experiment in Ghana, find that the roster position of a household member has implications for activities reported in the labor module, with the movement of an individual down in the roster by one position reducing the number of their reported productive activities by 2.2%. In a recent experimental study, Abay et al. (2021) show that delaying a module on dietary diversity by 15 minutes led to underestimation of the dietary diversity score due to respondents' fatigue in a phone survey in Ethiopia.

Recall length also contributes to respondent burden and has been found to affect data quality (Beegle et al., 2012a; Wollburg et al., 2020). As discussed in the previous section, phone surveys can be used to shorten recall periods through higher frequency data collection (Arthi et al., 2018; Gaddis et al., 2021). Shorter lags between survey rounds tend to also lower the chances that respondents attrite (Gourlay et

⁹ Households earning between 0% and 30% of income from agriculture are excluded for presentation purposes. Only countries in the RuLIS database with data on adult literacy rates since 2010 are included. The year of data collection by country is as follows: Armenia - 2013; Bangladesh - 2010; Burkina Faso - 2014; Cameroon - 2014; Ecuador - 2014; Ethiopia - 2016; Ghana - 2013; Guatemala - 2014; India - 2012; Iraq - 2012; Kyrgyzstan - 2013; Malawi - 2017; Mali - 2017; Mexico - 2014; Nepal - 2011; Nicaragua - 2014; Niger - 2014; Nigeria - 2019; Pakistan - 2014; Peru - 2019; Rwanda - 2014; Senegal - 2011; Sierra Leone - 2011; South Africa - 2015; Tanzania - 2015; Uganda - 2016.

al., 2021). At the other extreme, respondents may also feel burdened if they are contacted too frequently. Garlick et al. (2020) found that phone interviews at a weekly frequency did not induce persistent changes in respondent reporting nor increase permanent attrition but did increase the incidence of missed interviews among microenterprise owners in South Africa. In contrast, Schündeln (2018) finds panel conditioning effects, whereby reported consumption levels are correlated to the number of interviews, for households visited up to 10 times per month during the Ghana Living Standards Measurement Study.

Incentives are key to increase participation rates in phone surveys and were provided in the vast majority of phone surveys discussed in this paper, including those implemented at a national scale. There is a common understanding that monetary payment is important to compensate participants for their opportunity cost of time as well as the cost of airtime and electricity that a phone conversation entails (Glazerman et al., 2020). Furthermore, incentives are provided to encourage engagement in future interviews and are found to increase cooperation and response rates (Gibson et al., 2019; Glazerman et al., 2020; Gourlay et al., 2021; Singer and Ye, 2013). LSMS-supported HFPS provided airtime-based incentives of between \$0.70 and \$1.83. Gibson et al. (2019) found little importance of the exact amount of the incentive, though Singer and Ye (2013) suggest incentives exhibit diminishing marginal returns as the amount increases. According to a number of studies, survey incentives, ranging in size from \$0.10 to \$3.00, increase response rate by around 5 percentage points, although this effect ranges from 2 to 10 percentage points depending on context and incentive size (Ballivian et al., 2015; Gibson et al., 2019; Glazerman et al., 2020).

Respondent contact protocols and privacy

Carefully planning how and when to contact survey respondents is another way to reduce burden, increase cooperation and counteract attrition. Özler and Cuevas (2019) significantly increased recontact rates among a sample of refugees in Turkey by increasing the number of call attempts, spacing the call attempts appropriately, and reaching respondents in the evenings and weekends. Mathur (2020b) finds similar results when attempting to reach households in rural India. For agricultural data collection, deciding the dates and times to contact respondents requires that survey designers have information on the work schedule over several months of a typical agricultural household or holding. This informs the design decision of when to ask about specific farming activities, while also helping to understand when farmers will have time to participate in a survey.

Ensuring respondent privacy over the phone is very difficult (Glazerman et al., 2020). Calls to agricultural households can attract attention from other household members or non-household members, threatening respondent privacy and potentially biasing responses. This issue is more acute when sensitive questions are part of the survey, which may include questions on land rights and tenure. Highly sensitive questions, such as on violence against women and girls, cannot be collected in a safe fashion remotely and require a different approach (UN Women and World Health Organization, 2020).

4.3. Interviewer effects

Interviewers can also affect respondent behavior and data quality regardless of survey mode. Evidence of interviewer effects in surveys has been widely documented (West and Blom, 2016). Randall et al. (2013) show in an African context that interviewer behavior and characteristics can have an impact on the pattern of nonresponse. This is because interviewers often need to convince respondents to participate which requires considerable social skills. Interviewer behavior and characteristics can also affect, and bias, survey responses. Such effects appear to be meaningful for types of questions deemed sensitive though

they appear to be minor for less sensitive survey questions (Di Maio and Fiala, 2020; Garlick et al., 2020; Himelein, 2016). The social similarities of the interviewer and the respondent, or lack thereof, are a key factor in how much interviewers affect both responses and participation in the survey, with language playing a crucial role (Di Maio and Fiala, 2020; Randall et al., 2013).

As with in-person surveys, choosing, training, and supervising interviewers well is an important aspect of phone survey design. In addition to the skills that are needed specifically for phone surveys, such as desirable phone etiquette and, preferably, experience with the CATI application that is being used to record data (or other selected phone survey implementation mode), familiarity with the content of the questionnaire can be of great value. In cases of recontact surveys, the interviewers employed in the baseline survey may hold valuable knowledge of the questionnaire content, making training more efficient, as well as some developed relationships with respondents, which can also help reduce non-response.

Training and supervision

To limit the impact of interviewer effects, thorough and consistent training of both interviewers and supervisors is essential. Training for phone surveys, when feasible, can be conducted in person similar to the manner in which traditional in-person surveys are administered. In cases where in-person training is not a possibility, such as in the case of the LSMS-supported HFPS surveys, Amankwah et al. (2020) and Glazerman et al. (2020) recommend the use of various remote learning tools, including videos, web-based tutorials, and quizzes.

In the case of CATI implementation, quality control and supervision may extend beyond the team supervisors themselves. The use of audio recordings of phone-based interviews, potentially on a random sample of interviews and/or random segments of interviews, can allow survey managers at headquarters to review interviewer skills, explanations, and interactions with respondents, enabling the identification of behaviors that are outside the norm or are not as intended by survey management. These audio audits can also allow cross-referencing of data points in the recorded data and the audio exchange, as a further quality control measure. In the LSMS-supported HFPS surveys, audio audits were employed in Ethiopia and Nigeria and were found to be effective in identifying weakly performing interviewers, to which additional supervisory efforts were directed (Gourlay et al., 2021).

5. Cost considerations

Cost considerations are central to every survey operation. Perhaps the most appealing feature of phone-based survey implementation in low-income contexts is the reduced cost relative to traditional in-person data collection. Data for the World Bank's LSMS-ISA and High-Frequency Phone Surveys on COVID-19 offer a useful cost comparison of in-person and phone survey modes, for the same countries, at about the same time and on overlapping samples (Table 3).

When comparing *costs per completed interview*, in-person surveys are about 30 times as expensive compared to phone surveys. In this sample of countries, the cost per in-person interview ranges from 230 USD per household in Burkina Faso, to 409 USD in Tanzania.¹⁰ For the same countries, the range of per interview costs for phone surveys is from 8 USD in Tanzania to 12 USD in Malawi. Comparing the *cost per question* substantially reduces these cost differences, with phone surveys generally appearing to be 3-4

¹⁰ See Kilic et al. (2017) for a more comprehensive review of the cost of in-person household surveys.

times less expensive per question asked.¹¹ Cost per question ranges from 10 cents in Malawi to 25 cents in Tanzania for in-person interviews, and from 3 cents in Uganda to 6 cents in Malawi for phone surveys (Table 3).

One should note, however, that the scope of data collection in the in-person surveys is far more extensive than that of the HFPS surveys, so these figures should be considered as illustrative of the cost differential with acknowledgment that the surveys are not comparable in terms of data collected. A truly meaningful comparison needs to qualitatively discount these numbers by also looking at the information that can be collected, and the discussion in earlier sections showed how some key objectives of agricultural sample surveys cannot be fulfilled via phone.

Table 3. Comparison of in-person and phone surveys (US\$)

By country				
<i>Country</i>	<i>Survey</i>	<i>Mode: phone vs. in-person</i>	<i>Cost per interview</i>	<i>Cost per question</i>
Burkina	EHCVM-1 2018/19	In-person	230	0.20
	HFPS	Phone	9	0.05
Ethiopia	ESS4 2018/19	In-person	245	0.17
	HFPS	Phone	10	0.04
Malawi	IHPS 2019	In-person	269	0.10
	HFPS	Phone	12	0.06
Nigeria	GHS-Panel 2018/19	In-person	351	0.21
	HFPS	Phone	12	0.05
Tanzania	NPS 2021	In-person	409	0.25
	HFPS	Phone	8	0.04
Uganda	UHS 2021/22	In-person	383	0.19
	HFPS	phone	9	0.03
Overall				
		<i>Min</i>	<i>Max</i>	<i>Mean</i>
<i>Cost per interview</i>				
In-person		230	409	314
HFPS (phone)		8	12	10
<i>Cost per question per interview</i>				
In-person		0.10	0.25	0.19
HFPS (phone)		0.03	0.06	0.05

Source: Authors' calculation based on personal communication with World Bank staff supporting the implementation of the surveys.

The above comparisons are based on surveys conducted via CATI, but phone survey costs can vary significantly by approach. Based on IPA's review of phone surveys using 27 studies in 18 countries, on a cost per interview basis, IVR is the cheapest mode (\$4.86 on average), followed by automated SMS (\$7.75), and CATI (\$11.97) (Glazerman et al., 2020). Henderson and Rosenbaum (2020) add to the comparison a manual SMS survey in Nepal, where interviewers instead of an automated system were used to send text messages to respondents, with a cost per interview higher than CATI at \$17.41 per

¹¹ The calculations in Table 3 are based on the number of questions in the questionnaires, not on the questions actually asked. The total number of questions actually asked to respondents will always be lower, as not all questions will be applicable to all households.

interview.¹² Lau et al. (2019) report slightly different conclusions, with IVR at 43 percent of the cost of CATI and SMS at 24 percent of the cost of CATI. Each of these studies suggests that phone surveys are substantially less costly than in-person survey implementation, and that CATI is among the most expensive modes among the options for phone-based survey implementation. However, with a suggested length of 5 to 10 items and a maximum length of 15 items, IVR and SMS surveys cannot accommodate the more complex interviews that are made possible in a 20 to 30 minutes conversation in CATI (Glazerman et al., 2020). In sum, given literacy and non-response considerations discussed previously, and the complexity of most agricultural surveys that cannot be easily accommodated by IVR and SMS surveys, CATI remains the preferred mode of phone surveys for agricultural data collection.

The structure of survey costs differs greatly between phone and in-person surveys, and specifically between CATI and CAPI. The main budget lines for in-person surveys are transportation, especially in agricultural surveys where teams travel to sparsely populated rural areas, and per diems for long fieldwork periods. These lines are mostly eliminated in phone surveys, for which total costs are heavily driven by fixed costs including personnel and the acquisition of tablets or computers.¹³ Costs for the establishment of call centers, where they do not already exist, such as rental fees for the space, need be considered. Variable costs for phone surveys, primarily in the form of airtime, are marginal.

For reference, in the LSMS-supported phone surveys, incentives made up between 6 and 17 percent of the per-interview cost. In addition to incentives, it may be necessary to provide mobile phones to respondents in the sample who do not already have one. This is not possible if using RDD, or using a register of existing numbers, but could be a possibility in phone surveys that are using an in-person survey as a baseline. This decision should be carefully considered, as discussed in Section 3.1. Kilic et al. (2021) provided respondents with a mobile phone worth \$15.71 and a solar charger worth \$12.86. This is roughly double the combined value of the cash and airtime incentives they provided (\$7.14 each). Given the low variable costs of phone surveys, they generally become even more cost effective when utilized for panel surveys, repeated cross sections, or when combining the use of a call center with other survey operations, thereby lowering the fixed costs on a per interview basis.

6. Discussion and conclusions

This paper explored how the lessons learned from the COVID-19 pandemic, as well as pre-pandemic, phone survey experiences can be used to inform the integration of phone-based data collection into the survey and data systems in low- and middle-income countries, with a specific focus on the surveys of interest to agricultural statisticians and economists. Several overarching findings emerged. First, given the remote nature of phone interviewing, respondent burden and attrition deserve particular attention, requiring simplified survey question formats as well as abbreviated interviewing durations. Second, limited and uneven mobile phone coverage remains a critical issue for rural and agricultural populations, which has the potential to bias survey estimates, making better coverage a priority and post-survey adjustments a necessity. Finally, phone surveys allow tailoring data collection to the timing of the agricultural season in a cost-efficient manner, which can improve the quality of agricultural data and

¹² For CATI Henderson and Rosenbaum (2020) caveat these values with the indication that a “portion of these estimates do not include fixed costs and underestimate total survey costs”. They also report large standard deviations for both IVR and CATI.

¹³ For example, in the Nigeria LSMS-ISA survey, direct transport costs represented approximately 23 percent of the overall survey budget while these costs are nonexistent in the Nigeria HFPS.

reduce recall burden, but with limited depth in questionnaire content. Taken together, these findings point to promising potential for the combination of phone surveys and in-person surveys for agricultural data collection going forward.

Phone surveys are cheaper on a per-interview and per-question basis than in-person surveys, but what are the cost implications of combining the two in a mixed-mode survey? The baseline scenario is complementing in-person surveys with phone interviews to improve data quality, while leaving the in-person survey component unchanged. In this case, the phone survey rounds add costs to the overall survey operation. Based on data from their survey experiment in Tanzania, Arthi et al. (2018) estimate that adding one phone interview round to a two-visit in-person survey increases the total cost by 6 percent, adding ten phone interview rounds increases total costs by 54 percent, while adding 20 phone interview rounds more than doubles the total cost (+108 percent). Similarly, Dillon (2012) estimates the added costs of one phone survey round to be 7 percent. For Malawi, Kilic et al. (2021) estimate that an in-person survey with subsequent semi-weekly phone calls over the course of 12 months was 77 percent more costly than the baseline of two in-person visits over the same 12-month period. The exact cost is likely to differ significantly depending on the implementation details of the phone surveys. For instance, the cost calculations in Arthi et al. (2018) and Kilic et al. (2021) include mobile phones distributed to respondents (as well as harvest diaries in the case of Kilic et al., 2021). Overall, the additional cost of phone surveys ranges from marginal for a small number of rounds of interviews to substantial if many phone survey rounds are added to an existing in-person component that is left unchanged.

However, phone surveys can also substitute parts of in-person surveys. For instance, if agricultural labor data are collected via phone, these data may no longer need to be collected in-person, reducing the time and burden of in-person interviews. Especially if phone interviews can substitute an entire in-person visit, a substantial cost reduction could be achieved. The combination of in-person and phone modes will also necessitate a series of decisions to be taken with respect to the organizational machinery behind a survey. With mixed mode surveys there will be economies of scale in the creation of a call center for data collection if it can be used for several surveys, for example. Similarly, the organization of the interviewers' training will have to adapt to the new structure and decisions will need to be made on whether to use the same interviewers for both modes or to use different interviewers for phone and in-person interviews.

As for which topics to include in the phone-based component of a mixed-mode survey, the discussion on survey length, response fatigue, data quality, and attrition indicates that phone interviews should not be too long, naturally limiting the number of topics that can be covered in any given phone call. Topical coverage could be increased by scheduling several phone survey rounds and spreading topics out across those rounds, a common strategy used in the World Bank's L2A surveys, the LSMS-supported HFPS, and IPA's RECOVR surveys (Etang and Himelein, 2020; Innovations for Poverty Action (IPA), 2020; Living Standards Measurement Study, 2022). Data items that require objective measures such as land area or crop cutting, or sensitive topics, are likely to be those where in-person remains the primary mode. Data that is more variable over time, less salient, or that can benefit from higher frequency data collection because they are more likely subject to recall error (agricultural labor inputs is one example) are on the contrary those where the promise for the use of phones appears to be higher.

A specific question regarding the adoption of phones for mixed mode surveys will be how to avoid or mitigate biases connected with phone access between and within households. Dealing with limited and uneven mobile phone coverage remains a design priority especially when targeting rural and agricultural

populations to avoid biasing survey estimates, specifically underrepresenting the poorest and most vulnerable households and individuals. While reweighting techniques function to counteract these biases and improve representativeness in important ways, they cannot be an all-around remedy for systematic under-coverage and nonresponse. One option is to provide households and individuals who would normally not have access to one with a mobile phone, but that requires considering not only cost implications but also how having a phone may alter respondents behavior in areas the survey is actually aiming to investigate, such as access to market information or extension services. And if phones are given to households participating in the surveys, protocols will have to be devised on how many phones per household to hand out and who receives them, as informational content can be enhanced, and biases reduced or introduced, with different decisions in this domain.

Given the remote nature of phone interviewing, respondent burden and attrition deserve particular attention. Survey designers should make use of the tools at their disposal, especially questionnaire design, survey timing, contact protocols, respondent selection, and incentives, to minimize burden, curb nonresponse, and avoid attrition, within the limits dictated by the survey objectives. As for the frequency, the nature of data collected and the survey budget are key factors in determining the desired frequency. Arthi et al. (2018) and Gaddis et al. (2021) use weekly phone calls to collect agricultural labor data. Kilic et al. (2021) even called twice a week for extended harvest crop production data. In contrast, the LSMS-supported HFPS on COVID-19 collected monthly data, and the Georgia Survey of Agricultural Holdings, which converted to phone-based implementation in light of COVID-19 restrictions on in-person interviewing, is a quarterly survey. In practice, respondent burden and fatigue need to be considered, which very high interview frequencies could conceivably increase, risking higher attrition rates (see discussion in Section 3.1). The extent to which a sequence of short but more frequent phone interviews are less burdensome on respondents than a longer in-person sitting remains an empirical question, as there is no ready prior on the number of in-person visits and telephone interviews to optimize the use of survey resources and maximize data quality.

Having an in-person baseline visit makes possible the use of diaries or journals that can be facilitated by phone interviews. Kilic et al. (2021) asked respondents to keep diaries to document cassava harvest and twice weekly phone calls served, among other things, to supervise and support the diary-keeping. The authors conclude that the diary-keeping yields better cassava harvest data than the common end-of-season recall module. Deininger et al. (2012) use harvest diaries for several different crops in Uganda, which is again deemed more reliable than end-of-season recall. Beegle et al. (2012b) compare personal diaries to several other common methods to measure household consumption and poverty, deeming diaries the most reliable method. These experiences indicate significant scope for improving data quality through diaries which can be readily combined with phone surveys, though survey designers need to consider cost and logistical implications of this method.

Finally, cost considerations and the possible improvements in data quality need to be also weighed against the loss of comparability stemming from possible mode effects. This is a preoccupation that concerns both the comparison with earlier survey years, but also the comparison of different interviews within the same survey round, to the extent that they are implemented using different data collection modes. Longitudinal surveys may attach a different weight to the loss of comparability than cross-sectional surveys (Cernat and Saskaug, 2021). Response burden and attrition patterns may also have features that vary across households with important characteristics of interest (e.g., income and opportunity cost of time, literacy).

With all the promise, the paper identified several open questions that need to be addressed to better assess the potential, limitations, and optimal features of a survey system that complements in-person surveys with phone-based data collection. Rigorous survey methods research will have to provide the basis to inform several of these open questions and trade-offs, but one should be realistic and pragmatic in acknowledging that pending conclusive evidence from methodological studies, survey design choices will have to be made based on incomplete evidence, and informed choices made based on reasonable priors, accumulated experience, and the effective boundaries provided by available human, financial and technical resources.

References

- 50x2030 Initiative, 2020. 50x2030 on track to meet country engagement target for 2020 as Georgia begins planning | 50 by 2030 [WWW Document]. URL <https://www.50x2030.org/news/50x2030-track-meet-country-engagement-target-2020-georgia-begins-planning> (accessed 10.8.21).
- AAPOR Cell Phone Task Force. 2010. New considerations for survey researchers when planning and conducting RDD telephone surveys in the US with respondents reached via cell phone numbers. *Deerfield, IL: American Association for Public Opinion Research*.
- Abate, G.T., de Brauw, A., Hirvonen, K. and Wolle, A., 2021. *Measuring consumption over the phone: Evidence from a survey experiment in urban Ethiopia* (No. 2087). International Food Policy Research Institute (IFPRI).
- Abay, K.A., Berhane, G., Hoddinott, J.F. and Tafere, K., 2021. *Assessing response fatigue in phone surveys: Experimental evidence on dietary diversity in Ethiopia* (Vol. 2017). Intl Food Policy Res Inst.
- Aggarwal, S., Jeong, D., Kumar, N., Park, D.S., Robinson, J. and Spearot, A., 2020. Did COVID-19 market disruptions disrupt food security? Evidence from households in rural Liberia and Malawi (No. w27932). National Bureau of Economic Research.
- Alvi, M., Barooah, P., Gupta, S., Saini, S., Asia, S., 2021. Women's access to agriculture extension amidst COVID-19 : Insights from. *Agricultural Systems* 188, 103035. <https://doi.org/10.1016/j.agsy.2020.103035>
- Amankwah, A., Gourlay, S., 2021. Food Security in the Face of COVID-19 : Evidence from Africa. LSMS Integrated Surveys on Agriculture Washington, D.C. : World Bank Group.
- Amankwah, A., Kanyanda, S., Illukor, J., Radyakin, S., Sajaia, Z., Shaw, J.A., Wild, M., Yoshimura, K., 2020. High Frequency Mobile Phone Surveys of Households to Assess the Impacts of COVID-19 (Vol. 3) : Guidelines on CATI Implementation (English). World Bank, Washington, D.C.
- Ambel, A., McGee, K., Tsegay, A., 2021. Reducing Bias in Phone Survey Samples. Effectiveness of Reweighting Techniques Using Face-to-Face Surveys as Frames in Four African Countries (No. 9676), Policy Research Working Paper. World Bank, Washington D.C.
- Ambler, K., Herskowitz, S., Maredia, M.K., 2021. Are we done yet? Response fatigue and rural livelihoods. *Journal of Development Economics*, 153, 102736. <https://doi.org/10.1016/j.jdevec.2021.102736>
- Andersson, P. G., & Särndal, C. E. (2016). Calibration for nonresponse treatment using auxiliary information at different levels. In *The Fifth International Conference on Establishment Surveys (ICES-V)*, Geneva, Switzerland, June 20–23, 2016.
- Arthi, V., Beegle, K., Weerdt, J. De, Palacios-I, A., 2018. Not your average job : Measuring farm labor in Tanzania. *Journal of Development Economics*, 130, 160–172. <https://doi.org/10.1016/j.jdevec.2017.10.005>
- Bako, D., 2018. Test de collecte des données sur le cheptel par téléphone. Résultats et recommandations.
- Ballivian, A., Azevedo, J.P., Durbin, W., Rios, J., Godoy, J. and Borisova, C., 2015. Using mobile phones for high-frequency data collection. *Mobile Research Methods*, 21.
- Beegle, K., Carletto, C., Himelein, K., 2012a. Reliability of recall in agricultural data. *Journal of Development Economics*, 98, 34–41.
- Beegle, K., De Weerdt, J., Friedman, J., Gibson, J., 2012b. Methods of household consumption measurement through surveys: Experimental results from Tanzania. *Journal of Development Economics*, 1, 3–18.

- Biemer, P.P., 2010. Total Survey Error. Design, Implmentation, and Evaluation. *Public Opinion Quarterly* 74, 817–848.
- Brubaker, J., Kilic, T., Wollburg, P., 2021. Representativeness of Individual-Level Data in COVID-19 Phone Surveys Findings from Sub-Saharan Africa, World Bank Policy Research Working Paper No. 9660.
- Carletto, C., Dillon, A., Zezza, A., 2021. Agricultural Data Collection to Minimize Measurement Error and Maximize Coverage. In Christopher B. Barrett, David R. Just (eds.), *Handbook of Agricultural Economics*, Elsevier, Volume 5, Pages 4407-4480.
<https://doi.org/10.1016/bs.hesagr.2021.10.008>.
- Carletto, C., Gourlay, S., Murray, S., Zezza, A., 2017. Cheaper, Faster, and More Than Good Enough: Is GPS the New Gold Standard in Land Area Measurement? *Survey Research Methods*, 11, 235–265.
- Carletto, G., Savastano, S., Zezza, A., 2013. Fact or Artefact: The Impact of Measurement Errors on the Farm Size - Productivity Relationship. *Journal of Development Economics*, 103, 254–261.
<https://doi.org/10.1016/j.jdeveco.2013.03.004>
- Ceballos, F., Kannan, S., Kramer, B., 2020. Crop prices , farm incomes , and food security during the COVID-19 pandemic in India : Phone-based producer survey evidence from Haryana State. *Agricultural Economics*, 1–18. <https://doi.org/10.1111/agec.12633>
- Cernat, A. and Sakshaug, J.W. eds., 2021. *Measurement error in longitudinal data*. Oxford University Press.
- Dabalen, A., Etang, A., Hoogeveen, J., Mushi, E., Schipper, Y., Engelhardt, J. von, 2016. Mobile Phone Panel Surveys in Developing Countries: A Practical Guide for Microdata Collection. The World Bank, Washington D.C.
- Dal Grande, E., Chittleborough, C.R., Campostrini, S., Dollard, M. and Taylor, A.W., 2016. Pre-survey text messages (SMS) improve participation rate in an Australian Mobile telephone survey: an experimental study. *PLoS One*, 11(2), p.e0150231.
- Daum, T., Buchwald, H., Gerlicher, A., Birner, R., 2019. Times Have Changed: Using a Pictorial Smartphone App to Collect Time–Use Data in Rural Zambia. *Field Methods*, 31, 3–22.
- Deininger, K., Carletto, C., Savastano, S., Muwonge, J., 2012. Can diaries help in improving agricultural production statistics? Evidence from Uganda. *Journal of Development Economics*, 98, 42–50.
- Demombynes, G., Gubbins, P., Romeo, A. 2013. Challenges and Opportunities of Mobile Phone-based Data Collection: Evidence from South Sudan. World Bank Policy Research Working Paper No. 6321.
- Di Maio, M., Fiala, N., 2020. Be wary of those who ask: A randomized experiment on the size and determinants of the enumerator effect. *The World Bank Economic Review*, 34, 654–669.
- Dillon, A., Carletto, G., Gourlay, S., Wollburg, P., Zezza, A., 2021. Agricultural Survey Design: Lessons from the LSMS-ISA and Beyond, LSMS Guidebook. World Bank, Washington D.C.
- Dillon, A., Karlan, D., Udry, C., Zinman, J., 2020. Good identification, meet good data. *World Development*, 127, 104796. <https://doi.org/10.1016/j.worlddev.2019.104796>
- Dillon, B., 2012. Using mobile phones to collect panel data in developing countries. *Journal of International Development*, 24, 518–527. <https://doi.org/10.1002/jid.1771>
- Etang, A., Himelein, K., 2020. Monitoring the Ebola Crisis Using Mobile Phone Surveys, in: Hoogeveen, J., Pape, U. (Eds.), *Data Collection in Fragile States: Innovations from Africa and Beyond*. Palgrave Macmillan. <https://doi.org/10.1007/978-3-030-25120-8>
- FAO, 2018. Geostat Starts a New Survey | FAO in Georgia | Food and Agriculture Organization of the United Nations [WWW Document]. URL <https://www.fao.org/georgia/news/detail-events/en/c/1104050/> (accessed 1.31.22).

- Gaddis, I., Oseni, G., Palacios-Lopez, A., Pieters, J., 2021. Measuring Farm Labor: Survey Experimental Evidence from Ghana. *The World Bank Economic Review*, 35, 604–634.
<https://doi.org/10.1093/wber/lhaa012>
- Garlick, R., Orkin, K., Quinn, S., 2020. Call Me Maybe: Experimental Evidence on Frequency and Medium Effects in Microenterprise Surveys. *The World Bank Economic Review*, 34, 418–443.
<https://doi.org/10.1093/wber/lhz021>
- Gibson, J. and Alimi, O., 2020. Measuring poverty with noisy and corrected estimates of annual consumption: Evidence from Nigeria. *African Development Review*, 32(1), pp.96-107.
- Gibson, D.G., Wosu, A.C., Pariyo, G.W., Ahmed, S., Ali, J., Labrique, A.B., Khan, I.A., Rutebemberwa, E., Flora, M.S., Hyder, A.A., 2019. Effect of airtime incentives on response and cooperation rates in non-communicable disease interactive voice response surveys: randomised controlled trials in Bangladesh and Uganda. *BMJ Global Health*, 4, 1–11. <https://doi.org/10.1136/bmjgh-2019-001604>
- Glazerman, S., Rosenbaum, M., Sandino, R., Shaughnessy, L., 2020. Remote Surveying in a Pandemic: Handbook. *Innovation for Poverty Action*.
- Gourlay, S., Kilic, T., Lobell, D.B., 2019. A new spin on an old debate: Errors in farmer-reported production and their implications for inverse scale - Productivity relationship in Uganda. *Journal of Development Economics*, 141, 102376. <https://doi.org/10.1016/j.jdeveco.2019.102376>
- Gourlay, S., Kilic, T., Martuscelli, A., Wollburg, P., Zezza, A., 2021. High-Frequency Phone Surveys on COVID-19: Good Practices, Open Questions. *Food Policy*, 105, p.102153.
<https://doi.org/10.1016/j.foodpol.2021.102153>
- Groves, R.M., Biemer, P., Lyberg, L., Massey, J., Nicholls, W., Waksberg, J. 1988. Telephone Survey Methodology, New York: Wiley.
- Groves, R.M., Fowler Jr, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R., 2011. *Survey Methodology*. John Wiley & Sons.
- Groves, R.M., Lyberg, L., 2010. Total survey error: Past, present, and future. *Public opinion quarterly*, 74, 849–879.
- GSMA, 2020. The Mobile Economy 2020.
- GSMA, 2018. The Mobile Economy. GSMA Intelligence 53. <https://doi.org/10.5121/ijcsit.2015.7409>
- Häder, S., Häder, M., Kühne, M. 2012. Telephone Surveys in Europe. Research and Practice. Springer Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-25411-6>
- Henderson, S., Rosenbaum, M., 2020. Remote Surveying in a Pandemic: Research Synthesis. Innovation for Poverty Action.
- Hermosilla, R., 2017. Using web and mobile phone technologies to collect food market prices in Africa.
- Himelein, K., 2016. Interviewer Effects in Subjective Survey Questions: Evidence From Timor-Leste. *International Journal of Public Opinion Research*, 28, 511–533.
<https://doi.org/10.1093/ijpor/edv031>
- Himelein, K., Eckman, S., Kastelic, J., McGee, K., Wild, M., Yoshida, N., Hoogeveen, J., 2020. High Frequency Mobile Phone Surveys of Households to Assess the Impacts of COVID-19. Guidelines on Sampling Design. World Bank, Washington D.C.
- Himelein, K., Testaverde, M., Turay, A., Turay, S., 2015. The Socio-Economic Impacts of Ebola in Sierra Leone. World Bank Group 1–25.
- Hiraga, M., Uochi, I., Doyle, G., 2020. " Counting the uncounted – How the Mongolian nomadic survey is leaving no one behind." *Data Blog*, World Bank, 18 Aug 2020,
<https://blogs.worldbank.org/opendata/counting-uncounted-how-mongolian-nomadic-survey-leaving-no-one-behind>

- Hirvonen, K., de Brauw, A., Abate, G.T., 2021a. Food Consumption and Food Security during the COVID-19 Pandemic in Addis Ababa. *American Journal of Agricultural Economics*, 103, 772–789. <https://doi.org/10.1111/ajae.12206>
- Hirvonen, K., Minten, B., Mohammed, B. and Tamru, S., 2021b. Food prices, marketing margins, and shocks: Evidence from vegetables and the COVID-19 pandemic in Ethiopia. *Agricultural Economics*, 52(3), pp.407-421.
- Hoogeveen, J., Croke, K., Dabalén, A., Demombynes, G., Giugale, M., 2014. Collecting high frequency panel data in Africa using mobile phone interviews. *Canadian Journal of Development Studies*, 35, 186–207. <https://doi.org/10.1080/02255189.2014.876390>
- Hoogeveen, J., Pape, U. (Eds.), 2020. Data Collection in Fragile States Innovations from Africa and Beyond. Palgrave Macmillan.
- Innovations for Poverty Action (IPA), 2020. IPA's RECOVR Survey [WWW Document]. Innovations for Poverty Action (IPA).
- Jäckle, A., Roberts, C., Lynn, P., 2006. Telephone versus face-to-face interviewing: mode effects on data quality and likely causes: report on phase II of the ESS-Gallup mixed mode methodology project (No. 2006-41). ISER Working Paper Series.
- Karamba, W., Salcher, I., 2021. Socioeconomic Impacts of COVID-19 on Households in Somalia : Results from Round 1 of the Somali High-Frequency Phone Survey. World Bank, Washington, DC.
- Kilic, T., Koolwal, G.B., Moylan, H.G., 2020. Are You Being Asked? Impacts of Respondent Selection on Measuring Employment. Impacts of Respondent Selection on Measuring Employment (February 18, 2020). World Bank Policy Research Working Paper No. 9152.
- Kilic, T., Moylan, H., Ilukor, J., Mtengula, C., Pangapanga-Phiri, I., 2021. Root for the tubers: Extended-harvest crop production and productivity measurement in surveys. *Food Policy*, 102033. <https://doi.org/10.1016/j.foodpol.2021.102033>
- Knippenberg, E., Jensen, N., Constatas, M., 2019. Quantifying household resilience with high frequency data : Temporal dynamics and methodological options. *World Development*, 121, 1–15. <https://doi.org/10.1016/j.worlddev.2019.04.010>
- Larmarange, J., Kassoum, O., Kakou, É., Fradier, Y., Sika, L. & Danel, C., 2016. Feasibility and Representativeness of a Random Sample Mobile Phone Survey in Côte d'Ivoire. *Population*, 71, 121-134. <https://doi.org/10.3917/popu.1601.0121>
- Lau, C.Q., Cronberg, A., Marks, L., Amaya, A., 2019. In Search of the Optimal Mode for Mobile Phone Surveys in Developing Countries . A Comparison of IVR , SMS , and CATI in Nigeria. *Survey Research Methods*, 13, 305–318.
- Lavrakas, P., Benson, G., Blumberg, S., Buskirk, T., Cervantes, I.F., Christian, L., Dutwin, D., Fahimi, M., Fienber, H., Guterbock, T. and Keeter, S. 2017. The future of US general population telephone survey research. *AAPOR Report*.
- Leo, B., Morello, R., Mellon, J., Peixoto, T., Davenport, S.T., 2015. Do Mobile Phone Surveys Work in Poor Countries? Center for Global Development Working Paper No. 398. Center for Global Development.
- Lepkowski, J. M., Tucker, N. C., Brick, J. M., De Leeuw, E. D., Japiec, L., Lavrakas, P. J., Brick, J. M., Link, M. W., De Leeuw, Edith, & Sangster, R. L., 2007. Advances in telephone survey methodology (Vol. 538). John Wiley & Sons.
- Living Standards Measurement Study, 2022. Living Standards Measurement Study [WWW Document]. World Bank. URL <https://www.worldbank.org/en/programs/lsmss/initiatives> (accessed 8.22.22).
- Little, R. J. 1986. Survey nonresponse adjustments for estimates of means. *International Statistical Review/Revue Internationale de Statistique*, 139-157.
- Lundström, S., & Särndal, C. E. 1999. Calibration as a standard method for treatment of nonresponse. *Journal of official statistics*, 15(2), 305.

- Lynn, P. and Kaminska, O., 2013. The impact of mobile phones on survey measurement error. *Public Opinion Quarterly*, 77(2), pp.586-605.
- Maffioli, E.M., 2020. Collecting data during an epidemic: A novel mobile phone research method. *Journal of International Development*, 32(8), pp.1231-1255.
- Mathur, M., 2020a. How to identify the best length and time for a phone survey [WWW Document]. IDinsight Blog. URL <https://medium.com/idinsight-blog/phone-survey-duration-and-timings-reaching-respondents-part-ii-b2c85627d576>
- Mathur, M., 2020b. 3 steps to designing an effective phone survey that reaches respondents. IDinsight Blog. URL <https://medium.com/idinsight-blog/three-steps-for-designing-a-phone-survey-reaching-respondents-part-i-fb968828ba4b> (accessed 11.5.20).
- McCullough, E., McGavock, T., Assefa, T., Mangan, N., Getahun, T., 2020. Her Time: a time use study of women participating in livelihoods programs in Ethiopia. AEA RCT Registry. June 04.
- Minten, B., Mohammed, B., Tamru, S., 2020. Emerging Medium - Scale Tenant Farming , Gig Economies , and the COVID - 19 Disruption : The Case of Commercial Vegetable Clusters in Ethiopia. *The European Journal of Development Research*, 32, 1402–1429. <https://doi.org/10.1057/s41287-020-00315-7>
- National Research Council, 2008. Understanding American Agriculture: Challenges for the Agricultural Resource Management Survey. Washington, DC: The National Academies Press, Washington, DC. <https://doi.org/10.17226/11990>
- Özler, B., Cuevas, P.F., 2019. Reducing attrition in phone surveys. World Bank Blogs. URL <https://blogs.worldbank.org/impactevaluations/reducing-attrition-phone-surveys> (accessed 11.6.20).
- Pape, U., Delius, A., Khandelwal, R., Gupta, R., 2020. The Socio-Economic Impacts of COVID-19 in Kenya. World Bank, Washington D.C.
- Picchioni, F., Goulao, L.F. and Roberfroid, D., 2021. The impact of COVID-19 on diet quality, food security and nutrition in low and middle income countries: A systematic review of the evidence. *Clinical Nutrition*.
- Randall, S., Coast, E., Compaore, N., Antoine, P., 2013. The power of the interviewer. *Demographic Research*, 28, 763–792.
- Schündeln, M., 2018. Multiple Visits and Data Quality in Household Surveys. *Oxford Bulletin of Economics and Statistics*, 80, 380–405.
- Sharma, A., Gruver, A., Montalvao, J., Sullivan, M.O., 2021. Coping with COVID-19 shocks in Western Uganda. World Bank, Washington, DC. <https://openknowledge.worldbank.org/handle/10986/36545>
- Singer, E., Ye, C., 2013. The Use and Effects of Incentives in. *The ANNALS of the American Academy of Political and Social Science* 645, 112–141. <https://doi.org/10.1177/0002716212458082>
- Slavec, A., Toninelli, D., 2015. An Overview of Mobile CATI Issues in Europe, in: Toninelli, D., Pinter, R., de Pedraza, P. (Eds.), *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*. Ubiquity Press, London, pp. 41–62.
- Teickner, H., Knoth, C., Bartoschek, T., Kraehnert, K., Vigh, M., Purevtseren, M., Sugar, M. and Pebesma, E., 2020. Patterns in Mongolian nomadic household movement derived from GPS trajectories. *Applied Geography*, 122, p.102270.
- Tomlinson, M., Solomon, W., Singh, Y., Doherty, T., Chopra, M., Ijumba, P., Tsai, A.C. and Jackson, D., 2009. The use of mobile phones as a data collection tool: a report from a household survey in South Africa. *BMC medical informatics and decision making*, 9(1), pp.1-8. UN Women and World Health Organization, 2020. Violence against women and girls data collection during COVID-19. UN Women and World Health Organization.

- United Nations, World Bank, 2020. Monitoring the State of Statistical Operations under the COVID-19 Pandemic : Highlights from the Second Round of a Global COVID-19 Survey of National Statistical Offices. World Bank Group, Washington D.C.
- Valliant, Richard, Jill A. Dever, and Frauke Kreuter. 2013. Practical tools for designing and weighting survey samples. New York: Springer.
- Villar, A. and Fitzgerald, R., 2017. Using mixed modes in survey research: evidence from six experiments in the ESS. In *Values and Identities in Europe* (pp. 299-336). Routledge.
- West, B.T., Blom, A.G., 2016. Explaining Interviewer Effects: A Research Synthesis. *Journal of Survey Statistics and Methodology*. smw024. <https://doi.org/10.1093/jssam/smw024>
- Wollburg, P., Tiberti, M., Zezza, A., 2020. Recall length and measurement error in agricultural surveys. *Food Policy*, 102003. <https://doi.org/10.1016/j.foodpol.2020.102003>
- World Bank, 2014. The Socio-Economic Impacts of Ebola in Liberia. The World Bank 1–21.
- World Food Programme, 2020. MVAM: THE BLOG [WWW Document]. WFP. URL <https://mvam.org/info/>
- Young Lives, 2022. Young Lives at Work. Available at: <https://www.younglives.org.uk/research-project/young-lives-work> (Accessed: 29 August 2022).
- Zafar, A., Batana, Y., Ndip, A.E., Mijiyawa, A., 2016. Republic of Guinea Socioeconomic Impact of Ebola using Mobile Survey. World Bank.
- Zeza, A., Federighi, G., Adamou Kalilou, A., Hiernaux, P., 2016. Milking the data: Measuring milk off-take in extensive livestock systems. Experimental evidence from Niger. *Food Policy*, 59, pp.174-186.

Appendix I: Recent agricultural and rural phone survey experience in low-income settings

This paper draws on the experiences of the following phone survey efforts which were implemented in rural areas, with some focused explicitly on agriculture, in addition to the survey experiments, previous phone surveys on climate and disaster, and review papers discussed in Section 2.2:

- **World Bank High Frequency Phone Surveys (HFPS):** The World Bank has supported phone surveys in 83 countries to monitor the impacts of COVID-19 on households and individuals.¹⁴ We draw most heavily on experience of the seven HFPS survey programs supported by the LSMS team, including those in Burkina Faso, Ethiopia, Malawi, Mali, Nigeria, Tanzania, and Uganda (Living Standards Measurement Study, 2022). These surveys are built on the nationally-representative, longitudinal LSMS-ISA program in those countries, with data collected on roughly a monthly basis for an initial period of 12 months and subsequently extended. The CATI surveys included questions aimed at measuring and understanding the effects of the pandemic on households' agricultural activities, among other topics. Sample sizes range between approximately 1,700 – 3,300 households per country.
- **RECOVR survey:** Innovation for Poverty Action (IPA) launched their RECOVR phone-based survey in 2020, with up to three rounds implemented in Burkina Faso, Colombia, Côte d'Ivoire, Ghana, Mexico, the Philippines, Rwanda, Sierra Leone, Uganda and Zambia. The emphasis of the survey was on the various impacts of COVID-19. Sampling was undertaken using RDD, with sample sizes varying by country (for example, 1,300 respondents in Côte d'Ivoire and Mexico, 1,600 in Ghana).¹⁵
- **World Food Programme's mVAM Project:** The World Food Programme has been conducting phone surveys in its mobile Vulnerability Analysis and Mapping (mVAM) project, which aims to monitor food security, nutrition, and food markets since 2013, with country coverage now up to 28 countries (World Food Programme, 2020). Data collection in the mVAM project includes, in addition to phone calls, the use of SMS and IVR approaches.
- **Young Lives:** The Young Lives at Work program adapted to the COVID-19 pandemic by conducting a five-round phone survey in Ethiopia, India, Peru, and Vietnam from 2020-2021, with a potential face-to-face follow-up survey planned for 2023. The questionnaires included, among other topics, questions aimed at measuring the impact of the pandemic on food security and consumption (Young Lives, 2022).
- **International Food Policy Research Institute (IFPRI) phone surveys:** IFPRI has conducted phone surveys in several locations, covering topics like food security (Hirvonen et al., 2021a), gender and agricultural extensions (Alvi et al., 2021), and value chains (Hirvonen et al., 2021b; Minten et al., 2020). In Addis Ababa, Ethiopia, for example, four rounds of a phone survey were conducted in 2020 building off of a previous in-person survey as part of a randomized controlled trial, with a

¹⁴ The information is accurate at the time of writing. Visit the COVID-19 Household Monitoring Dashboard for up-to-date information on the number of surveys supported and harmonized indicators:

<https://www.worldbank.org/en/data/interactive/2020/11/11/covid-19-high-frequency-monitoring-dashboard>.

¹⁵ For more on IPA's RECOVR Survey, visit: <https://www.poverty-action.org/recovr/recovr-survey>.

sample of approximately 600 households, focused primarily on food and nutrition (Hirvonen et al., 2021a).

- **Listening to Africa/LAC:** The World Bank’s Listening to Africa and Listening to Latin America and the Caribbean (LAC) initiatives utilize mobile phones to regularly collect data on living conditions, with phone calls following initial in-person surveys (Ballivian et al., 2015; Dabalen et al., 2016). Listening to Africa, which used a CATI approach, was implemented in Madagascar, Malawi, Senegal, Tanzania, Togo, and Mali, with sample sizes of approximately 1,500 to 2,000 urban and rural households in each country (with the exception of Mali, which interviewed 500 adult respondents).¹⁶ Listening to LAC was piloted in Peru and Honduras on 1,500 households each, and included CATI, SMS and IVR approaches.
- **World Bank’s Gender Innovation Lab survey:** The World Bank’s Gender Innovation Lab conducted a phone survey of women in agricultural households in Western Uganda (August-September 2020), as part of an ongoing study on land titling. The sample was made up of 1,289 households, with 89% of respondents being women farmers. The survey collected data on agriculture, as well as COVID-19-related shocks (Sharma et al., 2021).
- **Georgia Survey of Agricultural Holdings:** This quarterly FAO AGRISurvey-supported quarterly survey focuses on agricultural production among family farms and commercial holdings. In response to COVID-19 related restrictions on in-person interviewing, the survey of family farms was converted to a CATI phone survey. The survey of commercial farms is done through the internet (50x2030 Initiative, 2020; FAO, 2018).

¹⁶ For more on Listening to Africa, visit <https://www.worldbank.org/en/programs/listening-to-africa#1>.