

# Agricultural Survey Design

## Lessons from the LSMS-ISA and Beyond

Andrew Dillon, Gero Carletto, Sydney Gourlay, Philip Wollburg, Alberto Zezza

2021-07-21



# Contents

<b>Acknowledgments</b>	<b>5</b>
<b>Acronyms and Abbreviations</b>	<b>7</b>
<b>1 Chapter 1: Introduction</b>	<b>9</b>
<b>2 Chapter 2: Survey Methodology for Agricultural Data Collection</b>	<b>13</b>
2.1 Innovations in Research on Survey Design Choices . . . . .	15
2.2 Survey Design Choices: Units of analysis for agricultural measurement . . . . .	23
2.3 Survey Design Choices: Respondents . . . . .	27
2.4 Survey Design Choices: Recall periods . . . . .	28
2.5 Survey Design Choices: Minimizing attrition in longitudinal surveys	29
<b>3 Chapter 3: Production</b>	<b>31</b>
3.1 Land Measurement . . . . .	37
3.2 Field Crop Production and Yield Measurement: Self-reported, remote-sensing, and crop-cut production measures . . . . .	49
3.3 Agricultural Production Data: Design Features of the Reference Questionnaire . . . . .	57
<b>4 Chapter 4: Agricultural Inputs</b>	<b>61</b>
4.1 Agricultural labor . . . . .	61
4.2 Seed, fertilizer, pesticides and herbicides . . . . .	63
4.3 Machinery and Farm Implements . . . . .	65

4.4	Agricultural Inputs: Design Features of the Reference Questionnaires . . . . .	69
<b>5</b>	<b>Chapter 5: Livestock</b>	<b>71</b>
5.1	Measuring Stocks . . . . .	71
5.2	Costs of Production . . . . .	72
5.3	Income sources . . . . .	73
5.4	Challenges to Livestock Measurement . . . . .	74
5.5	Livestock Data: Design Features of the Reference Questionnaires	78
<b>6</b>	<b>Chapter 6: Field Implementation</b>	<b>81</b>
6.1	Fieldwork Implementation Tradeoffs and Survey Design Choices .	82
6.2	Piloting for Improved Survey Design . . . . .	84
6.3	Mode of Data Collection . . . . .	85
6.4	Fieldwork Organization and Logistics . . . . .	86
6.5	Plot Visits and Georeferencing . . . . .	89
<b>7</b>	<b>Chapter 7: Conclusions and Future Directions for Agricultural Survey Design</b>	<b>91</b>
7.1	Data Integration . . . . .	92
7.2	Measuring Theoretical Concepts More Precisely . . . . .	93
7.3	Open Measurement Questions for Validation Research . . . . .	95
<b>8</b>	<b>References</b>	<b>97</b>
	<b>Appendix I: Agricultural Reference Questionnaire</b>	<b>109</b>
	<b>Appendix II: Livestock Reference Questionnaire</b>	<b>111</b>
	<b>Appendix III: Glossary</b>	<b>113</b>

# Acknowledgments

The authors are grateful to Didier Yelognisse Alia, Leigh Anderson, and Paul Winters for their valuable peer review of this document, as well as the continued efforts and dedication of the NSOs involved in the Living Standards Measurement Study – Integrated Surveys on Agriculture program. Much of the research and fieldwork underpinning this document was made possible thanks to the generous financial contributions of the Bill and Melinda Gates Foundation, the United Kingdom’s Foreign, Commonwealth & Development Office (formerly FDID), and the United States Agency for International Development.



# Acronyms and Abbreviations

APIs Application programming interfaces

CAPI Computer assisted personal interviewing

CR Compass and rope

CSPro The Census and Survey Processing System

GPS Global Positioning System

ICLS International Conference of Labour Statisticians

LSMS-ISA Living Standards Measurement Study - Integrated Surveys on Agriculture

NGO Non-governmental organization

NSO National Statistical Office

ODK Open Data Kit

PAPI Paper assisted personal interviewing





# Chapter 1

## Chapter 1: Introduction

Household surveys are the primary tool through which international development goals are monitored, and policy questions are informed based on empirical research. Accurate agricultural data are predicated on careful survey design and data collection, though practices for survey design and implementation of data collection exercises vary considerably across national statistical offices (NSOs), international organizations that generate international data public goods, NGOs, and researchers. While the importance of agricultural statistics is widely acknowledged as essential to evidence-based policymaking, tracking progress in the agricultural sector, and conducting innovative research, best practices in data collection and the research foundations of these recommendations are often ambiguous. Data collection methods may not be completely generalizable and need to be adapted to the local context, but the lessons learned across contexts and the lessons that we can generalize, can be available to the community of researchers and practitioners who collect and use agricultural data. It is the purpose of this volume to aggregate innovations in agricultural survey design from the experience of the Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA), academia, and other survey operations to provide recommendations for researchers, survey practitioners and policy analysts on lessons learned in survey design, while also highlighting topics for which more research is required to provide a conclusive survey design recommendation. From these agricultural survey design ‘lessons learned’, we highlight tradeoffs between household survey design choices, measurement error and data use.

In their synthesis of household survey practices, Grosh and Glewwe (2000) outlined lessons learned from 15 years of the World Bank’s Living Standards Measurement Study (LSMS) surveys. The genesis of this investment in the design and implementation of household surveys was motivated by the realization that progress towards reducing poverty could only be achieved if the goal could be measured, along with determinants of poverty to facilitate policy

analysis by NSOs and relevant national ministries and policymakers. Concurrently, an academic interest in understanding household consumption in low- and middle-income countries led to research foundations for measuring household consumption aggregates (Deaton and Zaidi, 2002) and the canonization of the agricultural household model (Singh et al., 1986) as the relevant unit of observation for understanding smallholder production and consumption decisions. These insights provided a basis for survey design for multi-topic household surveys focusing on linkages between productivity changes (human capital, labor productivity and later agricultural productivity) and welfare which are essential to policy analysis. A focus on household production also facilitates supply and demand comparisons for food security and agriculture sector analysis.

Out of the lessons of the past 18 years since the Grosh and Glewwe volumes, innovations in methods and measurement have improved our understanding of best practices in multi-topic household survey design. A widely noted limitation of the previous volumes and multi-topic household surveys generally has been agricultural module recommendations. Though acknowledged as a focal livelihood practice of the world's poor, multi-topic household surveys designed to measure poverty have often been limited by cursory agricultural data in part due to the burden of measuring consumption aggregates.

Rozelle (1991) described three differing approaches to agricultural survey design including a production function approach, an income statement approach and a balance sheet approach. A production function approach places emphasis on capturing how inputs create outputs for an agricultural household unit with an aim to potentially estimate the returns to different inputs on farm productivity. An income statement approach measures farm revenues and expenses with an objective of measuring farm profits. A balance sheet approach is related to the income statement approach but differs in valuing farm assets and liabilities as well as the sources of all inputs and outputs. Agricultural sector specific surveys often take alternative approaches to using the farm or area sampling as the unit of analysis rather than the household (Benedetti et al., 2010).

A recent innovation in survey design has been the LSMS-Integrated Surveys on Agriculture which have redesigned agricultural modules by recognizing first, that the unit of observation for agricultural production is often not the household, but the plot that may be managed by differing household members, and second, that improved agricultural statistics may be collected and respondent survey burden reduced by timing agricultural data collection to the agricultural production seasons in a post-planting and post-harvest questionnaire. Multiple time periods facilitate agricultural input and production recall, while also lowering the per interview time for a multi-topic household questionnaire. Repeated survey visits open the possibility of different panel types where the unit of analysis could be either the household, plot or farmer. The LSMS-ISA surveys track households and plot managers over time, but not plots due to high plot visit costs and challenges related to plot recall over time.

Improved agricultural data opens a wide range of policy analysis for both re-

search and policy. NSOs are common implementers of agricultural surveys and require high quality agricultural data to meet a number of national policy objectives, including annual or seasonal estimation of crop production and monitoring change in agricultural production over time. Agricultural surveys, if carefully designed, can meet these objectives while also enabling additional research on the drivers of agricultural production and productivity, for example, which can inform policies aimed at improving agricultural production and rural livelihoods.

This volume is organized to assess innovations in agricultural survey design, provide practical recommendations on the design and tradeoffs inherent in agricultural survey modules, and lastly to provide lessons learned on the organization of fieldwork and the linkage of agricultural household data to other data sources. In the next chapter, Chapter 2, we assess recent methodological research conducted by the LSMS team and the academic community to inform current best practice recommendations. We discuss topics including questionnaire design, respondent selection, units of analysis, timing of data collection to understand seasonality, recall periods, and innovations in the measurement of production, land, and agricultural labor.

In Chapter 3, we discuss the design of agricultural production modules, particularly as they relate to the organization, unit of analysis, and differences in production modules by agricultural system (field crops, root crops, and agroforestry). A key innovation in the design of agricultural surveys, which is discussed in Chapter 3, is the enumeration of land holdings of the household by agricultural plot and plot manager. This survey design choice improves the measurement and analysis of agricultural yields and generates sex-disaggregated agricultural data but creates survey design tradeoffs in the respondent's ability to recall allocation of inputs to plots which may be purchased in bulk to distribute across plots (fertilizer) or be difficult to retrospectively recall (household labor hours or days per plot).

Chapter 4 addresses the design of modules to measure agricultural factors of production given plot-level reporting including labor, capital, inputs (fertilizer and seed), and water management. Chapter 5 covers livestock production for which previous multi-topic survey design attention has been limited. As low- and middle-income countries' economies have grown, animal-based sources of food production independent of cereal and tree crop production are increasingly important. Motivation for measuring livestock or animal holdings in previous multi-topic surveys has often focused on livestock as an asset stock, yet pastoralists rely on animal rearing not only as a stock of wealth, but also as a flow of income. Measuring revenues and costs from animal production has been challenging given the frequent and often small costs associated with animal rearing and the imperfectly measured effort and time that are often important determinants of animal health and quality.

In Chapter 6, we provide practical guidance for the collection of agricultural data in a multi-topic household survey including field work logistics, some perspectives of implementers including NSOs, NGOs, and researchers. To conclude,

we summarize key lessons learned and highlight future methodological research which may be promising to advance agricultural survey design methodology in Chapter 7. In the Appendices, we provide a reference questionnaire, which we use as a basis for discussing agricultural survey design choices throughout this volume.

## Chapter 2

# Chapter 2: Survey Methodology for Agricultural Data Collection

This chapter reviews the general insights from the survey design literature and recent innovations in agricultural data collection which inform agricultural module design choices. It is important to remember that survey design for multi-topic household surveys is a relatively new and interdisciplinary endeavor. De Heer et al. (1999) discuss the history of survey design, documenting how innovations in sampling methods, data collection techniques, and statistical methods for data analysis were developed in the 1930s. The early field of survey research emphasized limiting measurement error and nonresponse through protocols developed in surveys conducted by Bowley and Burnett-Hurst (1926) to measure poverty among the working class in England. Agricultural measurement and related issues of non-random measurement error have been discussed in some of the earliest work by Fisher (1926) and Working (1925),<sup>1</sup> but also have important overlap with survey methodology as discussed in the statistics, applied econometrics, development economics, and labor economics fields.

The survey methodology literature has focused on minimizing bias relative to the cost of improved data collection in framing the questionnaire design problem as minimizing total survey error (Gideon, 2012). Total survey error can be disaggregated into sampling error and non-sampling error, where this volume focuses primarily on the latter. The questionnaire design affects non-sampling

---

<sup>1</sup>For a deeper discussion of the evolution of methods and measurement in agricultural economics, see Fox (1986) and Herberich et al. (2009).

error, which can be further disaggregated into non-response errors (for example refusals, missing data, respondent not knowing) and response errors (such as question framing, question sequencing, and social desirability bias). While this framework is useful in considering sources of sampling and measurement error in survey response, a researcher faces additional decisions and tradeoffs in the research design process. Minimizing total survey error with expensive measurement methods ignores the research design cost-variance tradeoff and the full set of research design choice variables. For example, in the case where the researcher completely conceives of the research design, a researcher chooses an identification strategy along with the questionnaire design and a sample size using a sampling strategy given their budget constraint. A researcher may be willing to accept some measurement error if reducing such error reduces the statistical power of the research design. If a researcher is implementing a randomized controlled trial, measurement error that is not correlated with treatment status may not bias estimates, whereas in a non-experimental design measurement error might bias parameter estimates.

To conceptualize more carefully these research design tradeoffs and to emphasize the point that questionnaire design is one set of choices among many in the research design, which include an identification strategy, sampling strategy and field implementation protocol, Dillon et al. (2020) introduce the idea of the data quality production function. The researcher's objective is to maximize knowledge or evidence generated from the research project. The researcher makes choices about the identification strategy, statistical power and external validity that are subject to budget constraints and the data quality production function. The data quality production function includes questionnaire design choices, but also interacts with the other choice variables such as sampling, empirical approach and field implementation protocols and constraints. From the perspective of NSOs designing surveys for both policy and potential research purposes, additional constraints need to be considered when designing agricultural surveys, including financial resource availability, personnel capacity, and competing demands on and mandates of the agency. Considerations for designing surveys to meet the needs and constraints of NSOs is further discussed in Chapter 6.

The history of the Living Standard Measurement Study surveys began in the 1980s as part of a concerted effort at the World Bank to provide global poverty measures and support the statistical capacity of developing country governments to collect household data. The Grosh and Glewwe (2000) manuals on questionnaire design were an important consolidation of field knowledge and research on questionnaire design that resulted from the Living Standard Measurement Study survey program. An early innovation in the LSMS approach and in papers such as Deaton and Zaidi (2002) was the clear discussion between the theoretical motivation for estimating an empirical relationship and its linkage to questionnaire design in a multi-topic household survey. Since these seminal works, research on questionnaire design has continued to improve the internal validity of empirical estimates through improved measurement and data quality

that reduces bias from non-random measurement error.

In the late 2000s, earlier surveys with limited agricultural modules used household level reporting but did not establish a clear linkage between agricultural inputs and outputs. The inherent tradeoff in a multi-topic household survey is how to maximize potential empirical analysis within and across modules given survey budget and respondent fatigue constraints. Carletto et al. (2010, 2015) outlined the methodological innovations in designing integrated agricultural modules in a multi-topic household survey. If surveys were going to inform food price policy, poverty reduction in rural areas, or agricultural policy to improve farm productivity, agricultural modules needed to better measure the relationship between inputs and outputs at the plot level, measure changes in agricultural livelihoods over time and account for seasonality. These innovations raised a host of survey design considerations and the potential to integrate cost-effective technologies through computer assisted personal interviewing (CAPI) and GPS devices both of which necessitated a research program using survey design experiments to better inform the tradeoffs and potential biases from survey design choices. Below, we review what has been learned from these survey experiments and the broader literature since the Grosh and Glewwe (2000) agricultural chapter.

We first present a framework for thinking about research design and questionnaire design where the objective is to generate as much knowledge as possible. We then focus on questionnaire design choices that researchers make including the units of analysis of agricultural measurement, respondent choice, recall periods, module sequencing, question phrasing, and panel data design. We outline the theoretical motivation for some of the design choices and how innovations in questionnaire design have addressed non-classical measurement error. In our review of specific modules in subsequent chapters, we emphasize questionnaire design choices that may reduce measurement error, but also highlight where choices represent tradeoffs in research design whose objective may not be to solely reduce measurement error if identifying a specific empirical objective is compromised due to cost tradeoffs.

## 2.1 Innovations in Research on Survey Design Choices

Given the multitude of potential questionnaire design choices, how does a researcher or NSO decide which choices are most important? While field testing and piloting strategies are often used to improve survey design, an emerging literature uses survey experiments or validation studies to inform either the relative or absolute biases of survey design choices. De Weerd et al. (2019) provide a recent summary of the survey experiment and validation literature. Survey experiments randomly assign survey design choices to assess their *relative* causal effect. Random assignment provides strong internal validity which

is important due to multiple design choices contained in a multi-topic questionnaire. However, the strength of this approach is also a potential limitation. In particular, the measurement error associated with each choice is not measured against ‘objective truth’. We can know the tradeoffs between design choices and their magnitude, but both design choices could be measured with error.

Validation studies rely on comparing alternative designs to a predefined truth such as administrative data, an objective measure or repeated observations over time. The advantage of this survey methodology ‘methodology’ is that the causal effect of survey design choice is estimated relative to this benchmark, quantifying the measurement error associated with each choice. The limitation is that ‘objective truth’ might not be measured in administrative data if it too is reported with error or the objective measure itself may not have been validated. When validation studies rely on repeated observations to assess consistency of responses, assuming that convergence reflects the true survey response, the effect of repeated interviewing may bias estimates. In the remainder of this document, we focus on methodological innovations that have been validated either through survey design experiments or validation studies with an objective measure that has either been deemed the gold-standard approach or validated against the relevant gold-standard.

A survey design literature outlines the importance of module sequencing, question ordering and question phrasing (Iarossi, 2006) in household survey design that is also relevant, but not specific to, agricultural survey design. Surveys should follow a logical sequence, collect relevant information from the best-informed respondent, and be coherent to the respondent. Little methodological research has been implemented on module sequencing as it is presumed feedback from piloting that detailing individual level outcomes within the household via the household roster, education and labor modules, for example, should precede household level outcomes such as consumption, non-food consumption, and household enterprises.

Manski and Molinari (2008) framed the choice of skip sequencing in a module as a formal design process. When including skip sequences, a questionnaire design makes a survey design choice that may increase item nonresponse and response errors leading to non-random measurement error. Manski and Molinari (2008) illustrate that the tradeoff in using skip sequences which reduce respondent burden depends on the likelihood and the cost of item nonresponse and response errors that increase bias in the survey data. If estimates of item nonresponse and response errors could be known from a pilot, a questionnaire design could make a comparison between the bias induced by skip sequences and the survey time cost.

Methodological research has focused on question phrasing including ordering screening questions to make ambiguous concepts clearer or using different word choices. An earlier survey design literature established the importance of question phrasing (Payne 1980; Sudman and Bradburn 1973; Converse and Presser 1986; Fowler 1995; Gideon 2012). General recommendations from this liter-



ature focused on the way respondents understand questions in a face-to-face interview. Misunderstanding questions from the respondent’s perspective can result from ambiguity, framing, technical wording, or hypothetical construction. Bautista (2012) describes a theoretical model of the survey response process in four steps: the respondent’s comprehension of the question, the cognitive process of information retrieval from memory, judgement of the appropriate answer, and communication of the answer. A respondent’s answer to questions can also be misreported by enumerators if response categories or the question’s intention is not clear to the enumerator, which may result in incorrect administration of the questionnaire.

We summarize key agricultural survey design choices in Table 1, describing these design choices in detailed subsections.

Table 1. Survey Design Choices

Survey design choice	Considerations	Key knowl- edge gaps	References
Sampling unit: household vs agricultural holding <sup>a</sup>	<b>Household:</b> representative estimate of household welfare, household-run farms, but not of agricultural sector; no direct link to household welfare. <b>Holding:</b> representative estimate of agricultural sector; no direct link to household welfare. A further consideration is the availability of recent sampling frames, that is, lists of households or holdings. Household frames are usually based on the most recent population and housing census and tend to be more widely available than up-to-date list frames of agricultural holdings which may be based on agricultural registers or the most recent agricultural census.	Integration of household and house- hold frames	2010; Global Strat- egy to im- prove Agri- cultural and Rural Statis- tics, 2017; FAO 2015a; FAO 2015b.

Survey design choice	Considerations	Key knowl- edge gaps	References
Definitions of the household: production-based vs consumption-based; wording of definition	<p><b>Consumption:</b> individuals must regularly share meals to be members of the same household.</p> <p><b>Production:</b> Adult individuals co-mingle revenues from agricultural production and non-farm enterprise for consumption. Issues such as (seasonal) migration and polygamy may complicate the concept of household and varying definitions may affect the measured size and composition of households and have an impact on recorded agricultural production, among others. To produce comparable and reliable agricultural data, surveys need to use comparable and suitable household definitions.</p>	Impact of household definition agricultural land holding and production	Beaman and Dillon, 2012; Guirkingner and Plateau (2014); Kazianga and Wahhaj (2013)

Survey design choice	Considerations	Key knowl- edge gaps	References
Unit of observation for crop production measurement: Plot/parcel vs household/holding	<p><b><i>Plot/parcel<sup>b</sup></i></b>: production technologies, ownership, and productivity differ across plots within a household, so that the plot is a more appropriate unit of observation to match outputs with inputs, and for understanding dynamics within the household; labor and fertilizer input allocation at plot-level may be more burdensome to recall, while land input, output, and tenure is likely measured more reliably at plot-level.</p> <p>Plots may change season-on-season, so that tracking plots over time is difficult, while parcels are expected to remain constant.</p> <p><b><i>Household/holding</i></b>: easier to recall when inputs are purchased collectively rather than separately for each plot/parcel; potentially reduces questionnaire length;</p>	Quantified impact of measuring at plot vs household level on reported inputs and outputs, and how they compare to an objective measure	Just and Pope, 2001; Carletto et al., 2017; Carletto et al., 2010

Survey design choice	Considerations	Key knowl- edge gaps	References
Respondent selection: self vs proxy respondents; group interviews vs individual interviews	<p><b><i>Self/individual:</i></b> individual reporting considered more reliable than proxy reporting, especially when it comes to capturing gender dynamics; interviewing each respondent individually allows for truthful reporting; may lead to inconsistent reports between different respondents;</p> <p><b><i>Proxy reporting:</i></b> less time intensive and methodologically challenging; may lead to overlooking individuals, their contributions and vulnerabilities</p>	Impact of proxy-reporting	Ambler et al., 2017; Bardasi et al., 2011; Coates et al., 2010; Dillon et al., 2012; Dillon et al., 2021; Doss et al., 2019; Jacobs and Kes, 2015; Janzen, 2018; Kilic et al., 2020b; Kilic and Moylan, 2016; Palacios-Lopez et al., 2017; Serneels et al., 2017

Survey design choice	Considerations	Key knowl- edge gaps	References
Recall periods: length vs number of field visits	<p><b><i>Length of recall period:</i></b> important to match recall period with activity appropriate time horizon; longer recall periods increases cognitive burden of recalling events, especially when these are less salient; very short recall periods may lead to double counting;</p> <p><b><i>Number of field visits:</i></b> more visits imply shorter recall periods for outputs and inputs, as do shorter implementation periods; more visits increase fieldwork costs.</p>	Effect of repeated visits on re-sponses (such as through learning, experience)	Beegle et al., 2012; Gaddis et al., 2019; Kilic et al., 2018; Wollburg et al., 2020

Survey design choice	Considerations	Key knowl- edge gaps	References
Longitudinal surveys: Panel unit; attrition, representativeness, and tracking; conditioning	<p><b><i>Choice of panel unit:</i></b> panel may be at the level of the enumeration area, household/farm, parcel, plot, each allowing for more sophisticated analysis but also increasing logistical difficulty and potentially costs; plot is expected to change season-on-season and may therefore not be a suitable panel unit. With any choice of panel, attrition is a concern as it leads to attrition bias; tracking split-off households is advisable but increases costs and logistical challenges. A representative sample is representative at the time it is drawn, but changes to populations can lead panel samples to lose representativeness over time, which may be aided through refreshing the sample. There are also potential response effects from re-visiting the same households, panel conditioning.</p>	Effects of long-term panels on sample of farmers and their characteristics	Garlick et al., 2020; Bevis and Barrett, 2020; Schündeln, 2018; Zwane et al., 2011

Survey design choice	Considerations	Key knowl- edge gaps	References
<p><sup>a</sup> According to the FAO World Program of the Census of Agriculture (WCA), the agricultural holding is defined as an “economic unit of agricultural production under single management comprising all livestock kept and all land used wholly or partly for agricultural production purposes, without regard to title, legal form or size.”</p> <p><sup>b</sup> A parcel is any piece of land of one land tenure type entirely surrounded by other land, water, road, forest or other features not forming part of the holding, or forming part of the holding under a different land tenure type. A plot is defined as a continuous piece of land on which a specific crop or a mixture of crops is grown or which is fallow or waiting to be planted.</p>			

## 2.2 Survey Design Choices: Units of analysis for agricultural measurement

The unit of analysis is critical for the validity and comparability of empirical estimates. In multi-topic household surveys that include welfare measures based on household consumption as well as production, several issues related to the

units of analysis must be addressed in the context of a nationally representative survey.

### *Households*

In defining the population frame, is the relevant unit of analysis the household, farm, holding, or enterprise? A national survey of the agricultural, manufacturing or industrial sector would not necessarily focus on the household as the relevant unit of analysis. Agricultural censuses use a population frame defined by agricultural holdings, acknowledging that both small-scale farms operated by the majority of rural farmers and agri-industrial farms of larger scale contribute to the agricultural sector. Using a household population frame would inherently exclude larger scale enterprises and farms that are not operated by households. As the analytical objective of the LSMS surveys has focused on poverty measurement and the determinants of changes in household welfare, the household is the relevant unit of analysis which can still inform how policy changes agricultural productivity, poverty, and other facets of rural livelihoods. As a nationally representative household survey, the household as the unit of analysis is inherently microeconomic in approach as opposed to providing a sector-representative sample. By using the household as the unit of analysis, we do not estimate agricultural sector indicators, as the sample will exclude farm enterprises that operate in the private sector as opposed to within agricultural households (Table 1).

How exactly is the household defined in the context of extended and nuclear family overlapping responsibilities in production, income sharing, and social connectedness? Household definitions differ widely across surveys and have important measurement implications in part because social and economic definitions of the household diverge (Beaman and Dillon, 2012), especially in communities with extended farming families with land inheritance claims, common production of family lands, or complicated use-rights. Key words in the definition of the household in multi-topic household surveys vary depending on their survey purpose and can emphasize either a family structure such as the nuclear family, consumption requirements such as eating from a ‘common pot’, or production based definitions that include requirements that adults co-mingle revenues from agricultural production and non-farm enterprise for mutual consumption. The household definition matters from an agricultural perspective as who is included in the household will be included in modules on agricultural land holdings, labor, assets, and marketing decisions. In polygamous households, multiple adult farmers, competing land use rights, and collective agriculture for some plots may be key features of the agricultural production system that nuclear definitions of the household may exclude. Empirical studies of polygamy and agriculture have incorporated wider definitions of the household to measure the complexities introduced in production (Guiringer and Platteau, 2014; Kazianga and Wahhaj, 2013). Definitions of the household that exclude agricultural producers lower reported agricultural production. Residency requirements also complicate measurement of pastoralist activities when transhumant pastoralism is used by the



household.

Beaman and Dillon (2012) show that consumption and production key words in the definition of the household do not alter household size estimates, but do affect household composition. Hess et al. (2001) describe the results of the Census Bureau's Questionnaire Design Experimental Research Survey where individual- and household-level measures were collected. They find that with the exception of asset data and health insurance prevalence reporting that household level enumeration of survey questions provided equivalent estimates to that of individual level data for demographic and government transfer reporting. Given the costs of collecting individual level data, they conclude that household level data are preferable, but the type of variable likely affects data quality and differences in reporting biases attributable to the unit of analysis (Table 1).

### *Plots*

In the design of agricultural modules, household level measurement of agricultural inputs and outputs did not correspond well with production technologies and reasonable assumptions about these technologies as represented in the literature on production and profit function estimation (Chambers 1988; Chambers and Quiggin 2000; Just and Pope 2001). An important survey design choice has been selecting a better unit of analysis than the household to measure agricultural production. Measuring productivity and input use at the plot is consistent with the production literature, but multiple cropping, seasonality, and distinctions between plot owners and plot managers create new measurement challenges in attributing inputs to agricultural outputs.

Tradeoffs in measurement error related to different units of analysis also imply different empirical applications. Nationally representative surveys such as FAO's Agricultural Census provide useful agricultural information which may be preferred to a household or plot level analysis if the researcher wants to make sector specific analysis related to farm size when a country has large scale commercial farms, infer changes in agricultural labor markets, or analyze changes in agricultural businesses. LSMS survey design is motivated by empirical analysis that can be tested using an agricultural household model (Singh et al., 1986).

Across different agricultural systems, the vocabulary associated with an agricultural landholding may also differ. Farmers use different words to indicate their farm, parcel and plot, often with contradictory meanings. It is important that agricultural survey design reflect a clear conception of the hierarchy of units consistent with the agricultural system that is being measured. Carletto et al. (2016) provide an overview of land area measurement survey design issues, noting differences in units of land measurement as well as variation across LSMS-ISA surveys in land reporting units. The holding, parcel, field and plot have internationally accepted definitions, though their interpretations by both academics, NSOs, and policy makers often lead to ambiguity.

The definition of the agricultural holding is the primary unit of analysis in agricultural surveys, whereas the household is the primary unit of analysis in

household surveys. FAO (2015a) defines the agricultural holding as, “economic unit of agricultural production under single management comprising all livestock kept and all land used wholly or partly for agricultural production purposes, without regard to title, legal form or size...” Holdings can be divided into parcels and the FAO notes that the difference between the holdings land sub-units are distinguished such that “A field is a piece of land in a parcel separated from the rest of the parcel by easily recognizable demarcation lines.... A field may consist of one or more plots, where a plot is a part or whole of a field on which a specific crop or crop mixture is cultivated, or which is fallow or waiting to be planted.” (FAO, 2015a). When designing and implementing an agricultural survey, practitioners should confirm the tiers and definitions used by the national statistical office as they may not coincide with the FAO definitions. Note that the reference questionnaire included in Appendix I utilizes only the parcel and plot tiers, as this is common in many contexts.

Recording agricultural information at the household level inherently aggregates individual production and imposes a linearity assumption across plots for input utilization and asset use. The main tradeoff in recording agricultural information at the plot level is that farmers must recall input allocation at the plot level which requires more cognitive effort and response time. These recall biases may be compounded by proxy response bias as plot level self-reporting is time consuming in the field and may be not feasible for all survey responses. Proxy respondents may have incomplete information on plots managed by other household members. For farmers who purchase inputs collectively with their family for multiple plots, it may be difficult to accurately assess how much fertilizer, seed, or other input was applied to a particular plot. Consistent with time use data, it may also be difficult for a farmer to recall individual household labor allocations to particular plots over an agricultural season or with respect to particular agricultural tasks. While more research is needed to understand the measurement implications of the disaggregation of input and production data to the plot level from the household level, the known analytical advantages of doing so, such as analysis of male-managed plots vis-à-vis female-managed plots, likely outweighs the unknown risk of aggregation in many surveys, including the LSMS-ISA.

Due to variation in land tenure status and land use rights, it is also important to account for seasonality in production on plots and changes in plot management when considering the unit of analysis. Depending on the agricultural season, a plot may be cultivated by a different member of the household and use different levels of inputs along with different cropping choices. Researchers have often cited asymmetries in crop type and input use, and therefore productivity and earnings, by plot manager gender. O’Sullivan et al. (2014) estimate that, after controlling for plot size and region, productivity differences across male and female-managed plots in Africa range from 23-66 percent. In order to appropriately account for plot-level production, and to enable analysis of the timing of production and gender asymmetries, both season and plot manager should be considered. The plot manager may differ from one season to the next,

depending often on gender-based norms.

## 2.3 Survey Design Choices: Respondents

A larger discussion within the survey design literature is estimating the potential bias from self-reporting versus proxy reporting (Moore, 1988; Krosnick, 1999). In labor modules that collect earnings data, it is presumed that self-reported data may be the most accurate as respondents have full information regarding their earnings which may not be fully disclosed to other household members. This assumes that self-respondents do not have incentives to mis-report information if they think that lower earnings increase their probability of future program participation or transfer. Bardasi et al. (2011) found in the context of labor market modules that using proxy responses as opposed to individual self-reports had no effect on female labor statistics, but that proxy reports did lead to under-reporting of men's participation rates in agricultural activities. Kilic et al. (2020a) also find significant impacts of respondent strategy in the collection of labor data in their analysis of two concurrent surveys in Malawi. They find that the use of the common approach which allows proxy and non-private interviewing results in significant underreporting of employment relative to measurement via private, self-respondent interviews, with stronger effects for women (Kilic et al., 2020a).

Kilic et al. (2020b) find that the common practice of proxy land reporting results in different findings for land assets than with self-response reporting. The use of proxy reporting by the "most knowledgeable household member" results in higher rates of exclusive reported and economic ownership of agricultural land among men, and lower rates of joint reported and economic ownership among women as compared to individual, self-respondent interviews.<sup>2</sup>

Dillon et al. (2021) also look directly at the effects of proxy response on agricultural statistics including gender dimensions of proxy response using a survey experiment with household heads, random proxies and self-reports. They find no effects of respondent type on total landholdings reported for the household, but statistically significant effects of area cultivated by random proxy reports relative to self-reported land data (11 percent of the standard deviation). Effect sizes are much larger on land reported by household heads and random proxies relative to self-reports for field crops and pasture land. Household heads also over-report production of cereals, cash crops and crop diversity as computed by crop diversity scores, relative to self-reports. Household heads (+ 18 percent of a standard deviation) and random proxies (- 37 percent of a standard deviation)

---

<sup>2</sup>The gold standard approach of individual, self-respondent interviews for the measurement of asset ownership and control in household surveys has been recently researched and supported by the United Nations Statistical Commission through the United Nations Evidence for Gender Equality (EDGE) Initiative. The EDGE Initiative led to the publication of the United Nations Guidelines for Producing Statistics on Asset Ownership from a Gender Perspective (UNSD, 2019).

also provide different biases relative to self-reported agricultural labor. Female proxies report lower levels of fertilizer for the household and higher frequencies of crops such as legumes and vegetables that women traditionally produce in Burkina Faso.

## 2.4 Survey Design Choices: Recall periods

Recall periods for agricultural statistics require consideration of the different frequencies and seasonal timing of input decisions and harvest periods. The number and timing of survey visits will have immediate implications on the recall periods employed in the survey. Survey visits may be employed once per year (with recall for the full year), in two visits per agricultural season (one at the post-planting stage to collect data on preparation and planting and a second visit after the harvest, to inquire about harvest and sales activities), or more than two visits. The more frequent the survey visits, the less recall the survey requires. We recommend the implementation of a two-visit approach where feasible, as this offers a balance of practical implementation feasibility and limited recall bias. The reference questionnaire in Appendix I assumes a two-visit approach and therefore includes separate post-planting and post-harvest questionnaires. If a two-visit approach is not feasible due to budgetary or time constraints, a one-visit option may be employed. However, one-visit surveys often bear a heavy respondent burden as data for the full season, or multiple seasons within the year, as asked about in one sitting.

In the context of an LSMS survey in particular, or any agriculture focused survey that is also measuring household consumption, a tradeoff in survey implementation arises when sampling designs spread survey implementation over a 12 month period to balance consumption seasonality. This survey implementation choice then creates variation in the recall period for production and input information. Recall bias depends on the salience of the event reported, its frequency and the time period over which information must be recalled. For different types of agricultural activities, salience can vary. For example, labor inputs from household members may be frequent during the planting period and difficult to recall while fertilizer costs incurred during the planting season may be easier to recall as a large expenditure would be significant to farmers. Survey designers have some discretion over the length of the recall period by choosing the timing and number of field visits during the agricultural year.

Gaddis et al. (2019) and Arthi et al. (2018) use methodological validation studies to analyze the impact of recall on the measurement of agricultural labor in Ghana and Tanzania, respectively. In Tanzania, seasonal recall of agricultural labor at the person-plot level, a common method employed, resulted in reported labor up to four times the number of hours reported through weekly interviews (Arthi et al., 2018). Similarly, Gaddis et al. (2019) find that agricultural labor is overestimated by approximately ten percent through the seasonal recall ap-

proach vis-à-vis weekly interviews in Ghana. Both findings, while different in magnitude, suggest that a seasonal recall approach to agricultural labor measurement may result in underestimated labor productivity.

Beegle et al. (2012) estimate potential biases due to differences in recall periods in surveys conducted in Malawi, Kenya and Rwanda. Due to the design of these surveys which were implemented over a 12-month period with clusters randomly assigned for interview across regions, recall of agricultural input choices and harvest period outcomes were de facto randomly assigned depending on when the household was interviewed. The authors find no evidence of bias in harvested quantities across the three countries, examining both staple and cash crops. Malawian tobacco farmers did over-report the cash value of their harvest, while also under-estimating fertilizer use as recall periods increased. Households did not exhibit recall bias with respect to fertilizer use for staple crops, except for female-headed households that cultivated maize. Longer recall periods lowered reported fertilizer use for female-headed maize crop cultivators. Recall bias was important in hired labor reporting, but the direction of these biases varied by country. Beegle et al. (2012) results suggest over-reporting in Kenya and under-reporting of hired labor in Malawi as recall periods increase. Recall bias tended to differ by subgroup, even if overall trends were not statistically significant. While there were no significant recall effects for Rwandan sorghum and coffee farmers, coffee farmers with small landholdings did under-report hired labor as recall periods increased. Using a similar identification strategy, Wollburg et al. (2020) find the recall period length consistently affects farmer reporting of plots listed, labor and other inputs, and maize production in Malawi and Tanzania, an effect that carries over to common measures of agricultural productivity. This is related to the design choice of number and timing of field visits, as more visits and shorter roll-out periods reduce the length of the recall period.

## 2.5 Survey Design Choices: Minimizing attrition in longitudinal surveys

Agricultural surveys may be administered either cross-sectionally or longitudinally. Longitudinal surveys, or *panel* surveys, offer additional benefits in terms of monitoring change over time for the same households. In panel surveys such as the LSMS-ISA, households (or other units of analysis) are tracked from survey round to survey round. This often requires specialized fieldwork teams that are responsible for using the information collected in previous rounds and at the suspected place of residence, to locate and interview (i) households that have moved from their previous residence and (ii) households that have “split-off” from the original survey household, such as when household members leave the original survey households to create their own household. If households are not tracked successfully, or they refuse to participate in future survey rounds, this results in attrition. While panel surveys offer analytical advantages, attrition

poses a risk to the representativeness of the panel sample over time.

Survey design choices can reduce attrition by facilitating tracking with high quality data, but it can also capture information on potential reasons for household attrition that might be used to quantify and assess attrition bias. The literature on attrition bias has primarily focused on household and individual attrition. After reviewing 13 national panel surveys in developing countries, Hill (2004) found that mobility was the greatest source of attrition bias. Fitzgerald et al (1998) and Witoelar (2011) provide a general review of the attrition issues in longitudinal surveys while Thomas et al. (2012) discuss low cost design protocols that limited attrition in the context of the Indonesia Family Life Survey.

In even the most remote villages, a variety of information based on how households and individuals are connected to each other can help reduce attrition by finding respondents, but also understanding the sources of attrition. The principle in survey design to facilitate tracking individuals and their households is to build in triangulated sources of information on a particular individual or household in case one source of information does not lead to success in tracking the observation. Mobile phones are frequent in many rural contexts and are useful sources of information about the individual or household itself as well as friends who might know how to contact the individual or household.

In the context of agricultural modules specifically, the survey design choice depends on the researcher's preference for panel or repeated cross-sections of a household's land portfolio. The literature on repeated cross-sections in estimating poverty dynamics is well-established with empirical strategies that are replicable in the context of agricultural surveys (see for example, Deaton (1985), Antman and McKenzie (2007), Dang et al. (2014)). In the context of agricultural modules, identification of parcels and plots over time is greatly improved with the use of GPS coordinates that can help identify locations across surveys. However, interview costs increase considerably if each parcel and plot must be visited in each survey round. While parcel or plot panels may facilitate the estimation of dynamic agricultural modules, recall of parcels and plots across survey rounds is nontrivial. LSMS-ISA surveys have implemented parcel panels in select countries with success, but have not employed plot panels primarily due to the fluid nature of plot dimensions within a given parcel over time. Approaches to address recall bias include naming parcels and plots, drawing maps, or taking plot characteristics from previous survey rounds into subsequent rounds, but little research has been done outside of piloting to provide an estimate of bias reduction due to different recall methods.

## Chapter 3

# Chapter 3: Production

This chapter reviews agricultural survey design innovations and their methodological validation studies that inform survey design choices related to crop production. We focus on methodological innovations that have been validated either through survey design experiments or validation studies with an ‘objective’ measure. We first discuss measurement issues related to production before summarizing survey design choices and practical implementation in the reference questionnaire provided in the appendix of this volume.

LSMS surveys have primarily taken a production function design objective with complementary questions related to farm revenues and costs without extensively providing a method to calculate farm profits. In early versions of the LSMS surveys and many other multi-topic household surveys, agricultural activities were not detailed at the plot level, as this requires a higher respondent burden relative to household level recall. One innovation of recent LSMS-ISA surveys has been to use the plot as the unit of analysis for field crop data collection. The advantage of this approach is that it more accurately measures the relationship between input and outputs in agricultural production when the household head may not manage all production. Production heterogeneity results from differences in crops cultivated that require different levels of inputs and management by several plot managers. The level of detail required in subsequent modules does not make this survey design choice trivial. Plot level data collection requires not only measuring land area at the plot level, but levels of production, labor, capital, chemical inputs, and land management techniques disaggregated at the plot level.

Choosing to measure production at the plot level requires a wider series of choices in identifying the unit of analysis which has design and empirical implications. First, agricultural production is seasonal and may involve multiple crop cycles on a given plot that need to be measured. Second, plots are not always associated with a single crop, as multi-cropping or inter-cropping is an important land management practice to increase yield and land quality. Third,

property rights are not necessarily well established in many rural contexts, and multiple family members may work on the plot or cultivate the plot in different seasons. The land owner may be different than the person making decisions to cultivate the land, as share-cropping, land leasing or land lending may mean that land owners are not making agricultural decisions. Holden et al. (2016) describe survey design choices in modules used to describe land tenure. Depending on the empirical applications of data collected, information on sources of land acquisition including inheritance and legal status, land transactions, formal and informal property rights, land conflicts, perceptions of tenure security and trust in land-related institutions may all be important additional questions complementary to the land roster. This is especially true if the survey aims to monitor Sustainable Development Goals related to land tenure, particularly SDG 5.a.1 and 1.4.2. Recognizing the importance of land tenure security as it relates to investment and the control of and access to other assets such as credit, both of these SDG Indicators seek to measure and monitor specific aspects of land tenure at the *individual* level, rather than at the household level.<sup>1</sup>

In summary, the key innovation in conducting plot-level production analysis is not to simply measure inputs and outputs at the plot level, but to distinguish the unit of analysis as plot-crop-season-manager. This unit of analysis facilitates comprehensive measurement of household production, allowing multiple analytical strategies from seasonal, crop and gender perspectives, but also has some limitations particularly in the context of a panel survey. Tracking plots is difficult between survey years which is one of the reasons that the LSMS-ISA surveys are conducted as a household panel (or household-parcel panel in select countries) with repeated cross-sections of plot, production, and input information. Another difficulty is related to plot characteristics that are relatively time-invariant within the survey period (such as soil characteristics, water source, plot topography) and time-varying characteristics (including labor, inputs, and capital utilization). We devote the next two chapters to the measurement of these time-invariant and time-variant characteristics but note that allocating these inputs and labor to the plot-crop-season-manager unit is often challenging.

<b>Post-Planting Questionnaire</b>		
<b>Questionnaire module</b>	<b>Key design choice</b>	<b>Motivation/Considerations</b>

<sup>1</sup>For more detailed guidance on data collection for monitoring SDGs 5.a.1 and 1.4.2, on land tenure, refer to the guidance document developed by the custodian agencies (FAO, The World Bank, and UN Habitat, 2019); available here: <http://documents.worldbank.org/curated/en/145891539095619258/pdf/Measuring-Individuals-Rights-to-Land-An-Integrated-Approach-to-Data-Collection-for-SDG-Indicators-1-4-2-and-5-a-1.pdf>



<b>PP 1 - Parcel Roster</b>	Land area measurement with GPS	Farmer self-reported land area has been found to be systematically biased so that farmer-reported land area measure is considered unreliable. Farmer-reported land area is still collected as it can be used to impute the area of parcels for which the GPS area has not been measured.
	Non-standard units of measurement for farmer self-reported land area.	Reporting in standard units (square meters or hectares) may pose difficulties for respondents in some cases and can worsen reporting error. Non-standard units should be country-specific. In some cases, conversion factors for non-standard units may vary by region, district, province, etc. Conversion factors must be available (either through existing data or supplemental conversion factor development work) in order to convert all land areas to a common unit in the analysis phase.
<b>PP 2 - Parcel Details</b>	Individual-level ownership determined	Asset ownership, particularly of land in agrarian settings, is a critical factor of individual wealth and security. Individual-level ownership of land allows for gender-based analysis and monitoring of select SDG Indicators when coupled with other questions on land.
	Network identification	Network identification allows respondents to identify co-ownership of parcels with individuals outside of the household.
	Land area measurement with GPS	Farmer self-reported land area has been found to be systematically biased so that farmer-reported land area measure is considered unreliable. Farmer-reported land area is still collected as it can be used to impute the area of plots for which the GPS area has not been measured.

<b>PP 3 - Plot Roster</b>	Non-standard units of measurement for farmer self-reported land area.	Reporting in standard units (for example, square meters or hectares) may pose difficulties for respondents in some cases and can worsen reporting error. Non-standard units should be country-specific. In some cases, conversion factors for non-standard units may vary by region, district, province, etc. Conversion factors must be available (either through existing data or supplemental conversion factor development work) in order to convert all land areas to a common unit in the analysis phase.
	Plot as primary unit of observation for land preparation and crop planting information	Plot-level data collection is designed to improve respondent recall and makes possible analysis of the data at the plot-level.
<b>PP 4 - Plot Details</b>	Recording of plot manager	Plot manager characteristics matter for plot level outcomes such as harvest. Identification of plot manager also allows for gender-specific productivity analysis, analysis of input allocation, gender-based crop cultivation patterns, etc.
	Respondent selection/recording by plot	- It is recommended that the plot manager, being the most knowledgeable person concerning the plot, respond to the questions about plot characteristics. Plot managers may differ from plot to plot which means that each plot may have a different respondent. - For data analysis, recording the ID of the respondent allows controlling for the effect that the respondent's characteristics such as education on gender have on the data at the same level of observation of the data.
	Plot-crop as unit of observation for crops planted	Allows for detailed accounting of each of the crops planted and enables plot-level analysis of production.

<b>PP 6 - Crop Roster</b>	Permanent and temporary crops in the same roster but with differentiated questions.	The plot may contain both permanent and temporary crops so both need to be reported in the same module, using the plot to anchor recall on planting. Permanent crops, such as trees, are likely not planted in the ongoing agricultural season, so year of planting instead of month of planting is recorded.
	Reporting of seed quantity in non-standard units	Reporting in standard units (e.g. kilograms) poses difficulties for respondents and can worsen reporting error. Non-standard units should be country-specific.
	Reference period for seed acquisition is 'reference agricultural season'	The reference period aligns with the actual production cycle aiding respondent recall and improving data reliability.
<b>Post-Harvest Questionnaire</b>		
<b>Questionnaire module</b>	<b>Key design choice</b>	<b>Motivation/Considerations</b>
	Respondent selection/recording by plot-crop	- It is recommended that the plot manager, being the most knowledgeable person concerning the plot, respond to the questions about harvest. Plot managers may differ from plot to plot which means that each plot may have a different respondent. - For data analysis, recording the ID of the respondent allows controlling for the effect that the respondent's characteristics such as education on gender have on the harvest data.
	Reporting of harvest quantity in non-standard units	Reporting in standard units (e.g. kilograms) poses difficulties for respondents and can worsen reporting error. Non-standard units should be country-specific.

<b>PH 7 - Field Crop Production</b>	Reporting of harvest quantity in different conditions	The weight and value of harvest may vary depending on the condition in which it was harvested and/or weighed.
	Decision-maker concerning use of harvest recorded	This allows for understanding household decision-making over the production process, which is especially relevant from a gender perspective.
<b>PH 8 - Field Crop Disposition</b>	Respondent selection/recording by crop	Decision maker may vary from crop to crop, data considered most reliable if each decision-maker can respond in person rather than proxy response.
	Reporting of harvest sales quantity in non-standard units	Reporting in standard units (e.g. kilograms) poses difficulties for respondents and can worsen reporting error. Non-standard units should be country-specific.
	Reporting of harvest sales quantity in different conditions.	The weight and value of harvest may vary depending on the condition in which it was harvested and/or weighed.
	Decision-maker concerning earnings from sold harvest recorded	This allows for understanding household decision-making over the production process, which is especially relevant from a gender perspective.
	Reference period for seed acquisition is 'reference agricultural season'	The reference period aligns with the actual production cycle aiding respondent recall and improving data reliability.

## 3.1 Land Measurement

We briefly review the land measurement literature and methodological guidance that has emerged from a recent set of land measurement experiments, primarily focused in sub-Saharan African countries. Much of this literature has been summarized in Carletto et al. (2016), but this literature continues to evolve, particularly as most empirical evidence has been focused on sub-Saharan African production systems. The extent of possible land measurement error is highly dependent on the method used to measure land size. Three common measurement methods include compass and rope, plot measurement with GPS devices, and farmer self-reported land size. Remotely sensed plot measurement is also feasible if the GPS coordinates of the plot and its boundaries are identified. A longstanding approach to collect land size is the compass and rope (CR) method which uses poles, ropes and a compass to carry out a systematic measure of land size (FAO 1982). With a compass, measuring tape, ranging poles, a programmable calculator or other computational tool, and two to three persons, the CR method can measure the area of a plot with significantly more accuracy than by subjective estimates. GPS and CR measures require careful implementation by a field team, though GPS measurements may also be affected by the GPS device's calibration error which can be affected by environmental factors. GPS measurements may be subject to more error than CR since the coordinates measured by the GPS device are not exact.<sup>2</sup> When carefully implemented, CR provides precise estimates, and it can therefore be considered as a benchmark against which to assess the precision of other methods. However, it should be noted that the compass and rope method is not immune to measurement error of its own. In their analysis of three methodological validation studies, Carletto et al. (2016 and 2017) illustrate various cases of closing error in compass and rope measurements as well as provide anecdotal evidence of the challenges of this form of measurement, including significant difficulty in the reading of compasses due to the poor eyesight of many enumerators.

Although the CR method is highly accurate when performed properly, it is time consuming and costly due to the necessity of careful training, monitoring, transportation costs to individual farmer plots, and enumerator time spent measuring the plot. The use of GPS can on average require as little as 28 percent of the time needed for compass and rope (Keita and Carfagna, 2009; Schøning et al., 2005). Keita and Carfagna (2009) find that on small plots CR can take up to 17 times longer than GPS, though GPS measurement still requires survey team relocation to the plot and survey team time to trace the field's perimeter. Similarly, Carletto et al. (2017) find that CR measurement takes approximately four times as long as GPS measurement on average, in methodological validation studies in Ethiopia and Tanzania. Because of this, many surveys rely on

---

<sup>2</sup>Standard GPS devices have reported position accuracy of approximately ten meters depending on the model's satellite calibration algorithm (<http://www8.garmin.com/aboutGPS/>). GPS receivers using WAAS (Wide Area Augmentation System) can have accuracy of three meters by correcting for atmospheric conditions.

a farmer's own estimate of land size which avoids the time and cost of actual measurement. However, self-reported estimates could be subject to greater error given that first, many farmers in developing countries acquire land through informal means where record keeping (and thus information on plot dimensions) is limited and second, farmers are more likely to give rounded and inexact size estimates.

Most land measurements require the farmer to first identify the plot correctly for the interviewer. Plot measurements are then linked to the unique plot identifier, but misclassification is entirely possible. Because of the intensity of this field activity, monitoring systems are paramount to ensure that plots enumerated are actually visited and plot measures are properly linked to household survey data, particularly if the land measurement teams work independently of household interview teams.

Donaldson and Storeygard (2016) review the integration of satellite data for economic analysis, noting that remotely sensed data were used as early as the 1930s to study US farmland production and land use practices<sup>3</sup>. Agricultural statistics derived from remotely sensed data have been used primarily to compare land use information and changes over time, though applications to yields (Lobell et al., 2009) and economy-wide outcomes such as food security are becoming more common (Grace et al., 2014). Lobell et al. (2009), Grace et al. (2012, 2014), and Husak and Grace (2016) apply remote sensing to the estimation of yields, land use, and total agricultural production for a region, but are limited by their ability to link this data to specific households and other inputs in the agricultural production process.

While a primary advantage of remote sensing data for land measurement and land characteristics is the ability to measure over large geospatial variation and easily construct panel measures of such variables, limitations of remotely sensed data are also described by Willis et al. (2015), Donaldson and Storeygard (2016), and Jain (2020). While remotely sensed data are not subject to respondent recall biases, satellite images are linked to GPS coordinates of farmer plots reported by farmers and images must be interpreted using either subjective evaluations or machine learning algorithms, both of which could be biased relative to actual plot size. Jain (2020) reviews errors in preprocessing imagery including sensor characteristics, satellite angle and atmospheric conditions. Validating classifications of plantation, agricultural and forestry land uses, Jain (2020) find that the error structure is non-random for remotely sensed data with an overall accuracy of 93.9percent of land classifications. Carletto et al. (2016) also note remotely sensed data used to measure land size may be more biased for plots with larger slope gradients than relatively flat plots due to interaction with sensor angles. The data generating process to use remotely classified data requires respondent input, measurement from remote sensors, and the processing of this information,

---

<sup>3</sup>In addition to land use information, Donaldson and Storeygard (2016) note remote sensing data sources available to produce data related to mineral deposits, elevation, terrain, and land cover, as well as airborne pollution, fish abundance and electricity use.

linking data sources for interpretation and analysis. Remote sensing is not without potential implementation biases. Remotely sensed images must be relevant to the validation period's season and year. Imagery must be of high resolution to facilitate respondent identification of plots. Respondents must be capable of identifying their plot on an image which may be complicated by changing plot boundaries or ambiguous plot boundaries. For these reasons, despite the advantages of remotely sensed data, empirical applications using remotely sensed data should also be careful of potential biases.

Land *quality*, in addition to quantity, is essential for productivity measurement yet challenging to measure accurately in a household survey context. Household survey instruments, including the one included in Appendix I, typically collect information on soil quality and health through a series of subjective farmer assessments. Findings from methodological research, such as that by Carletto et al. (2017a) and Gourlay et al. (2017), illustrate the limitations of such subjective soil data, highlighting the limited variability observed in such data and the weak relationship of certain subjective questions, particularly overall soil quality, with laboratory-based soil quality indicators. This finding is echoed by Berazneva et al. (2018), who find in Kenya that farmer-reported soil type is reasonably correlated with objective measures of soil fertility while farmer-reported soil quality is not. The alternatives to subjective farmer assessment of soil in household surveys are the integration of laboratory testing, which is both logistically complex and costly, the use of in-situ tools for objective or pseudo-objective measures, and the integration of survey data with available geospatial soil products. While integration with publicly available geospatial soil products is likely the most cost-effective approach, recent research has found that current offerings inadequately capture variation in soil properties at the local level. Berazneva et al. (2018) and Gourlay et al. (2017), for example, both suggest that the Africa Soil Information Service (AfSIS) 250m soil map (Hengl et al., 2015) fail to sufficiently capture local-level variation vis-à-vis ground-based objective measures. The future of soil measurement in household surveys likely falls at the intersection of subjective, geospatial, and in-situ tools. Use of a sample of ground-based soil tests to calibrate geospatial soil products to better fit local conditions and farm management practices may result in improved soil parameters at a national level, for example. That, coupled with the growing availability and reliability of soil-assessment tools, including mobile applications like the Land Potential Knowledge System (LandPKS, [landpotential.org](http://landpotential.org)) and handheld spectrometer technology suggests improved soil data in household surveys is soon to be realized, though additional methodological research is needed.

#### *Methodological studies on land area measurement bias*

The spread of GPS measured land size has led to several studies that examine the differences between the GPS and both CR and self-reported measures including the potential impact of measurement error resulting from inaccurate land size measurements. In two examinations of GPS versus CR measures in sub-Saharan Africa, Schøning et al. (2005) and Keita and Carfagna (2009) find that GPS

based measurements are generally lower than CR measures though neither study provides a technical explanation for this difference. Both studies find that the GPS-CR difference is statistically significant for smaller plots but not for larger plots<sup>4</sup>. It appears that for smaller plots, the fixed margin of error associated with determining GPS coordinates is large relative to the size of the plot and thus results in a greater measurement error. Since most plots in developing countries are relatively small, the potential for land mismeasurement using GPS could be significant. However, rounding of very small plot areas in Schoning et al. (all plots less than 0.01) clouds the reliability of their findings for smaller plots.

A cross-country study by Carletto et al. (2015) showed that the error in farmer-estimated area is both significant in magnitude and systematic in nature. Data pooled from nationally-representative surveys in Niger, Malawi, Tanzania, and Uganda suggests that self-reported areas are underestimated by 2.5 percent relative to GPS measures on average, but that respondent estimation on the smallest plots is overestimated by more than 100 percent on average while respondents underestimate the area of larger plots (Carletto et al., 2015). Findings from methodological studies in Ethiopia, Tanzania, and Nigeria illustrate a similar trend, with self-respondent area estimation significantly overestimated on the smallest plots, with overestimation decreasing with plot size (Carletto et al., 2017). Findings from pooled data from methodological studies in Ethiopia, Tanzania, and Nigeria, however, suggest that on average GPS plot area measurements are only 1 percent different than the CR measurements (Carletto et al., 2017). On the smallest plots, those under 0.05 acres, the mean GPS and CR measures are not statistically different (Carletto et al., 2017). Similarly, Desiere (2015), in an analysis of more than 50,000 agricultural plots in Burundi, finds support for implementation of GPS measurement on small plots. The study finds that, for plots smaller than 500m<sup>2</sup>, bivariate regressions of GPS on CR explained approximately 97 percent of the variance, though they do cite increased precision of GPS measurements as plot size increases (Desiere, 2015).

Several studies have found inconsistencies between GPS and self-reported land sizes (Goldstein and Udry 1999; Carletto, Savastano and Zezza, 2013; Carletto, Gourlay and Winters, 2015; Dillon et al., 2019; Dillon et al., 2021). Although ancillary to their main analysis, Goldstein and Udry (1999) found that the correlation between GPS and self-reported land size was only 0.15 in their dataset from Ghana. The authors do point out that, historically, field measurements in the region were based on length and not area and that this could partially explain the lack of a strong correspondence between farmer and GPS estimates. However, Carletto, Savastano, and Zezza (2013) also find a sizeable difference in Uganda though one that varies with plot size. The authors find that the bias is wider for smaller (less than 1.45 acres) and larger (greater than 3.58

---

<sup>4</sup>Schoning et al. (2005) find a statistically significant difference for plots under 0.5 hectares but not for plots equal to or greater than 0.5 hectares. Keita and Carfagna (2009) separate their sample into five clusters based on plot size. They find a statistically significant difference for the clusters with the lowest land size but not for the clusters with larger plots. However, they do not specify the plot size range for each cluster.



acres) plots while self-reported measures are a reasonable estimate for medium size plots. In a similar analysis, Carletto, Gourlay, and Winters (2015) find inconsistencies between GPS and self-reported measurements in Malawi, Uganda, Tanzania, and Niger. They find that in all four countries self-reported plot sizes were higher than GPS for smaller plots but lower for larger plots, suggesting that farmers over-report the area of small plots and under-estimate the area of large plots. Carletto, Savastano, and Zezza (2013) find that in addition to plot size, the rounding of self-reported measurements, the age of the household head, and whether the plot was in a dispute with relatives were all positively associated with a greater GPS-self-report difference in land size in Uganda. Carletto, Gourlay, and Winters (2015) similarly find that rounding of farmer estimates are consistently associated with a difference between GPS and self-reported plot size in all four countries. Dillon et al. (2019) also find that GPS measurements of land area are similar to compass-and-rope (CR) estimates and more reliable than farmer estimates, where self-reported measurement bias leads to overreporting land sizes of small plots by 83 percent and underreporting of large plots by 21 percent of the compass and rope estimate. Their study in Nigeria finds that the error observed across land measurement methods is nonlinear, is not resolved by trimming outliers, and results in biased estimates of the inverse land size–productivity relationship. A key econometric advance in this paper is the ability to control for plot fixed effects which may bias parameter estimates. They also investigate input demand functions that rely on self-reported land measures and find that these measures significantly underestimate the effect of land on input utilization including fertilizer and household labor.

Dillon et al. (2021) estimate measurement differences between self-reported, GPS and remotely sensed land size measures. In their comparison of results from Lao People’s Democratic Republic, Philippines, Thailand and Viet Nam, they find significant differences between GPS and remotely sensed data only in Viet Nam, where plot sizes are small relative to the other countries. The magnitude of farmers’ self-reporting bias relative to GPS measures is nonlinear and varies across countries, with the largest magnitude of self-reporting bias of 130 percent of a standard deviation (2.2-hectare bias) in the Lao People’s Democratic Republic relative to Viet Nam, which has 13.3 percent of a standard deviation (0.008-hectare bias). While the land measurement literature has expanded largely due to LSMS-ISA collaborations in Africa, variation across cropping systems in other continents can yield different types of land measurement biases in self-reported land data.

Table 3. Research on Land Area Measurement

Paper and mea- sure- ment meth- ods	Country/Dataset	Key insights
Goldman and self- Udryreport (1999)GPS	Eastern Ghana / Agricultural Innovation and Resource Management in Ghanaian Households survey	<ul style="list-style-type: none"> <li>• Correlation between self-reported and GPS measured plot size is only 0.15</li> <li>• Proposed interpretation: agricultural history in Eastern Ghana in which local field measurements were often based on length rather than area, making it hard for respondents to report in square meters</li> </ul>

Paper and Country/ mea- sure- ment meth- ods	Key insights
De Farm Southern Mali / Field experiment data Groote self- collected by Farming Systems Research team and report in Mali Traofcom- (2005) pass and Rope	<ul style="list-style-type: none"> <li>• Relative to compass and rope, farmer self-reported area (supported by enumerators) over-estimated on small plots, under-estimated on large plots, but under-reported by 11 percent on average.</li> <li>• Measurement error smaller for cotton fields than for cereals</li> <li>• Difference between GPS and Compass and Rope is small and significant on small plots (&lt;0.5ha) but not on larger plots.</li> <li>• Farmer self-report unreliable.</li> <li>• Compass and Rope takes 3.5 times as long as GPS.</li> </ul>
Schon Uganda / 2003 Uganda Pilot Census of Farm self- Agriculture report; GPS; Com- pass and Rope	

Paper and Country/Dataset mea- sure- ment meth- ods	Key insights
Keita GPS; Cameroon, Madagascar, Niger, Senegal / and Com-sample of 207 purposively selected plots Carfagna from Cameroon, Madagascar, Niger, Senegal (2009) and Rope	<ul style="list-style-type: none"> <li>• GPS tends to underestimate plot size relative to Compass and Rope, but the error is small.</li> <li>• Accuracy of GPS appears to be higher without cloud cover.</li> <li>• No correlation between plot size and difference between GPS and Compass and Rope measurements</li> <li>• Compass and rope takes on average takes 3.8 times longer than GPS, and this is worse for medium and small plots.</li> </ul>

Paper	Land Country/Dataset	Key insights
Cardoza et al. (2013)	Uganda / Uganda National Household self-Survey 2005/06 GPS	<ul style="list-style-type: none"><li>• Relative to GPS, land area over-estimated on small plots, under-estimated on large plots.</li><li>• Other determinants of measurement error: rounding, boundary delineation of plot, tenure status of plot</li></ul>

Paper and Country/Dataset mea- sure- ment meth- ods	Key insights
Cardon et al. (2015) Malawi, Niger, Tanzania, Uganda / self- LSMS-ISA Malawi Integrated Household Survey 2010/11, Uganda National Panel GPS Survey 2009/10, Tanzania National Panel Survey 2010/11, Niger ECVM/A 2011	<ul style="list-style-type: none"> <li>• Relative to GPS, self-reported area of small plots (less than 0.5 acres) systematically over-estimated.</li> <li>• Degree of over-estimation varies, but in all countries the mean self-reported area is overestimated by at least 90 percent of the mean GPS area of plots on small plots.</li> <li>• Area of large plots under-estimated</li> </ul>

Paper Land Country/Dataset mea- sure- ment meth- ods	Key insights
Desiree GPS;Burundi / Enquete nationale agricole du (2013)Com-Burundi 2011/12 pass and Rope	<ul style="list-style-type: none"> <li>• GPS slightly over-estimates area relative to Compass and Rope area but the under-estimation is negligible</li> </ul>
Carleton Ethiopia, Nigeria, Tanzania (Zanzibar) / et self- Ethiopia Land and Soil Experimental al. (2016) Research 2013, Nigeria Area Measurement and GPS;Validation Study, Measuring Cassava 2017)Com-Productivity study Tanzania (Zanzibar) pass and Rope	<ul style="list-style-type: none"> <li>• GPS is reliable even on smaller plots (less than 0.5 hectares), with reliability increasing rapidly for plots larger than 0.1 hectares.</li> <li>• Error in self-reported area depends on plot area, with small plots over-estimated, large plots under-estimated.</li> <li>• In comparison to Compass and Rope, GPS is generally very reliable.</li> </ul>

Paper and Country/Dataset mea- sure- ment meth- ods	Key insights
KilicFameUganda, Tanzania / Uganda National Panel et self- Survey 2009/10, Tanzania National Panel al. (2017) Survey 2010/11 GPS	<ul style="list-style-type: none"> <li>• Not all plots are measured with GPS because some are too far away or otherwise inaccessible, which may lead to bias.</li> <li>• Farmer self-reports are predictive of GPS measured area so that Multiple Imputation techniques can be used to impute area of plots where GPS is missing.</li> </ul>
DilloFarmIrao PDR, Philippines, Thailand, Vietnam / and self- land measurement experiment among rice Rao report producers (2021) GPS; satel- lite data	<ul style="list-style-type: none"> <li>• Remotely sensed plot area differs from GPS area estimate only in Viet Nam, where plot sizes are small.</li> <li>• Farmer self-reported plot area estimates are most overstated for small plots.</li> </ul>



3.2. FIELD CROP PRODUCTION AND YIELD MEASUREMENT: SELF-REPORTED, REMOTE-SENSING, AND CROP-CUT PRODUCTION MEASURES

Paper	Country/Dataset	Key insights
Dilloo et al. (2019)	Nigeria / Nigeria General Household Survey self-report; GPS; Compass and Rope	<ul style="list-style-type: none"><li>• GPS more reliable than self-reporting, relative to Compass and Rope</li><li>• Farmers underestimate land area on large plots and overestimate it on small plots.</li><li>• GPS error is relatively small relative to Compass and Rope.</li></ul>

3.2 Field Crop Production and Yield Measurement: Self-reported, remote-sensing, and crop-cut production measures

Accurate measures of production likely rely on reducing the cognitive burden of reporting for the respondent. As applied to the measurement of production, plot-level recall by plot managers may actually be cognitively less burdensome than asking a respondent, potentially the head of household, to aggregate across all household plots to report production. Plot level production reporting also allows women’s production to be more accurately measured if intrahousehold gender asymmetries exist. In facilitating respondents’ accurate reporting, a principle of agricultural survey design has been to let respondents report in the units in which they have measured their own production. This principle opens the possibility that respondents may not use standard units of measures, such as kilograms to report production.

Oseni et al. (2017) present a comprehensive guide for measuring conversions

from non-standard units to standard units for consumption and agricultural production data. Conversions require not only question response category flexibility to allow respondents to report quantities in non-standard units, but inter-linked data collected through market surveys to measure systematically conversions of non-standard units by crop or product into standardized units (kilograms or liters). Alternative approaches to collecting non-standard unit conversions formally in the market survey is to design a module to ask respondents directly for conversions, but a respondent's estimate of the conversion factor may not be made based on objective standards, creating bias in measures across respondents that is directly correlated with their conversion estimation skill. A significant benefit of CAPI implemented interviews is the possibility to document in advance non-standard measures which provide easier reference for respondents, as well as a pre-loaded conversion factor. A classic example is bunches of bananas for which standardizing even this non-standard measure is particularly challenging.

Alternatives to self-reported production include more intensive field approaches such as crop-cutting or integrated survey techniques such as first geo-referencing plots and then estimating production using remotely sensed data, or remotely sensed data in combination with ground-based data. It is important to underscore that using either crop-cutting or remotely sensed data to measure production involves *estimation* of production quantities. While these approaches may be more reliable than self-reported production by farmers, they are not without bias themselves. Researchers should avoid labeling these measures as objective or direct measures unless the full plot is crop-cut because plot samples are used to then extrapolate total plot production.

Crop-cutting methods of plot production estimation and yield calculation can be undertaken using several different measurement methods. Plots are defined by field teams and respondents to delineate the plot's size. Plot size is measured often using either compass and rope or GPS techniques. The plot area is demarcated, and a crop-cutting rule is implemented where the field area is sampled, the sampled plot areas are clear cut, production is dried, weighed, and then based on the sampling fraction of the plot samples, total production is estimated. Crop-cuts are estimates of crop production, potentially more reliable than self-reports, but subject to biases from the measurement process. First, budgets will determine the sampling fraction of plot subsamples taken from the plot from which the plot production estimate will be calculated. The higher the sampling fraction, the more reliable the plot estimate. Second, crops must then be dried, threshed and weighed which often takes additional days of fieldwork, sample tracking protocols, and standardization of drying and weighing techniques. Finally, the plot sub-sample production weights are then used to extrapolate the total plot production measure. Crop production may vary within plot due to planting techniques, plot slope that may pool water differentially within the plot, uniformity of plot fertilization and weeding, or soil quality. Animal infestation or weather damage may also not be uniform within the plot. Ferment and Benson (2011) review the sources of bias and history of crop cut-

### 3.2. FIELD CROP PRODUCTION AND YIELD MEASUREMENT: SELF-REPORTED, REMOTE-SENSING, AND

ting in Ugandan agricultural surveys, noting the widespread variation in crop cutting measurement methods and yield estimates using various crop-cutting measures.

A large body of literature has focused on comparisons of crop-cut measures relative to self-reported measures, particularly focused on the empirical relationship between yield and land size. Desiere and Joliffe (2018) find that self-reports over-estimate yield on small plots and underestimate yield on large plots relative to the crop cut measures which they use as their validation measure. Their production results largely corroborate evidence from the land measurement literature that find nonlinear plot reporting effects, particularly for smaller plots. Gourlay et al. (2019) find similar results in the analysis of maize in Uganda, with self-reported yields being significantly overestimated on the smallest plots (by more than 3,500kg/ha on average in the smallest plot area quintile). A divergence between self-reported and crop-cut-based yields is also observed by Lobell et al. (2020) who find that, although the means are similar, self-reported estimates of sorghum yields in Mali are only weakly correlated with yields measured via crop-cutting (correlation coefficient of 0.33). Bevis and Barrett (2020) argue that agricultural labor intensity increases at a plot's edge and thus drives erroneous production self-reports and the inverse land size relationship in their paper which also uses crop-cuts to measure production. Abay et al. (2019) estimate the magnitude of bias on the inverse land size relationship describing potential econometric implications of correcting both land size and production measures. An important insight from econometric applications that they apply to agricultural data is that correcting one measurement error in either land or production estimates may increase bias when measurement errors are correlated. They find that reducing bias by using crop-cutting production measures with compass and rope land measures causes the inverse land size relationship to disappear in their Ugandan sample.

The evidence on remote sensing to measure production is an evolving scientific field. Carfagna and Gallego (2005) argue that remotely sensed data are particularly useful for area frame measurement to define sampling units but hesitate to recommend remotely sensed images to estimate crop production because the classification of pixels assigned to specific crops is often strongly biased. This is one of the reasons remote sensing-based estimation of yields in intercropped areas is particularly challenging, as discussed in Lobell et al. (2019). This field continues to evolve and identification algorithms linked to biological science are improving. It is important to remember that measurement of plot production is based on models of interpreting visual images that correspond to biological plant growth. Brown and Pervez (2014) provide evidence in the US context that some models of production prediction are increasingly accurate to measure land use and production, calibrating and validating their models with USDA data. In Uganda, Lobell et al. (2019) estimate maize plot yields using (i) full plot crop cuts, (ii) partial, sub-plot crop-cutting, and (iii) remotely sensed images of plot yields. To estimate yields via remote sensing imagery, Lobell et al. (2019) estimate two models, one calibrated on ground-based, crop-cutting

data and one without calibration of ground data. They find that calibrated remotely sensed yield estimates captured half the variability of yield estimates in comparison to full-plot crop cut production measures for plot sizes above 0.1 hectare. Uncalibrated remotely sensed yield estimates were one ton per hectare higher than crop-cut measures. These results are both incredibly encouraging and require a further research agenda to better understand best practices for remotely sensed yield estimates that extend to a wider set of crops.

*Challenges to production measurement in intercropped systems*

Much of the literature on production measurement has considered the case of a single field crop, but farming systems are often more complex and not all agricultural production is based on field crops. We review measurement approaches when planting techniques on a plot use mixed, multiple crops or inter-cropping techniques as well as cases of roots, tubers and tree crops whose harvest period may be multiple rather than at a singular period of time in the production process. The LSMS-Integrated Surveys on Agriculture incorporate seasonal planting and harvest for annual crops in its biannual interview structure, but concern over recall bias for root, tuber and tree crops that do not overlap with the main agricultural season varies across agro-ecological zones and across national surveys. Production surveys focused on root, tuber and tree crops are often timed to the main harvest periods for these crops, but supplement production reports with higher frequency surveys. Kilic et al. (2018) conduct a survey experiment in Malawi over a 12-month period where variations of diary and farmer recall are used to produce cassava production estimates. Their production reporting treatments vary the recall period (daily diary recalls (supervised in person or via telephone), six-month and annual recalls). The survey experiment indicates that annual recall underestimates annual production by 21 percent relative to the in-person supervised diary visit. Daily diary reporting supervised by a telephone visit increased annual production reports by 28 percent compared to the in-person supervised diary visits.

We have also sidestepped an important complication in crop production measurement, specifically the case when production systems use mixed or inter-cropping planting techniques. LSMS-ISA data from Tanzania, for example, indicates that approximately 64 percent of cultivated plots are cultivated with more than one crop (Wineman et al., 2019). Estimating both land area apportioned to a particular crop and its production is particularly challenging. Measuring yields on these intercropped plots can take many forms, through varying methods of estimating the denominator, including: (i) using the full plot area for each crop; (ii) using the share of plot area under a given crop, identified by seeding rate, plant density, or area estimation, such that the total cultivated area sums to the total plot area; (iii) allocating total plot area equally across crops cultivated on the plot such that the total cultivated area sums to the total plot area; or (iv), imputing the area which a crop would occupy if it were mono-cropped (where the total cultivated area summed across crops may exceed the total plot area). Most household surveys acknowledge the compli-

3.2. FIELD CROP PRODUCTION AND YIELD MEASUREMENT: SELF-REPORTED, REMOTE-SENSING, AND

cations of production and input estimates on inter-cropped plots by identifying these plots and apportioning the area planted to divide plot-level input reports to production reported by crop. Proportional input attribution implies crop input demands including fertilizer, weeding, and harvest time are similar by crop.

The Global Strategy to Improve Agricultural and Rural Statistics (2018) provides methodological guidance on implementing the above methods to measuring the area under a given crop in intercropped systems. However, a best practice recommendation supported by a methodological survey experiment is not currently available. Remote sensing or crop-cut production estimates are possible alternatives, but these measures are also challenging to implement. Lobell et al. (2019) report lower accuracy of remotely sensed production estimates compared to crop-cut production estimates for maize intercropped plots in Uganda.

Table 4. Research on Production Measurement

Paper	Production measurement methods	Crop/Country/Dataset	Key insights
Fernandez-Beneito (2011)	Farmer self-report vs crop cutting	Various, Various/Various, focus on Uganda / Various production datasets, Agricultural Censuses from 1965, 1990-91, 2008-09; Annual Survey from 1967-68, Uganda National Household Survey 1999/2000 and 2005/06	<ul style="list-style-type: none"><li>• Crop cutting associated with over-estimation of production in many contexts</li><li>• Farmer estimation of production also exhibits significant biases; appear to be closer to objective measures in several cases.</li><li>• Lack of systematic evidence on sources and effects of biases make recommendation impossible</li></ul>

Paper	Product	Crop/Country/Dataset	Key insights
Deininger et al. (2012)	Figure 3.1: self-report; harvest diaries	Various/Uganda / Uganda National Household Survey 2005-06	<ul style="list-style-type: none"> <li>• End-of-season recall diverges significantly from harvest diaries, which are deemed more reliable.</li> <li>• Findings vary by crop type: for most crops, end-of-season recall associated with under-reporting, especially for extended-harvest crops such as cassava or banana; cash crop production recorded as significantly higher in end-of-season recall module.</li> </ul>

### 3.2. FIELD CROP PRODUCTION AND YIELD MEASUREMENT: SELF-REPORTED, REMOTE-SENSING, AND

Paper	Product/Crop/Country/Dataset	Key insights
Kilic et al. (2018); Farmer self-report; harvest diaries; crop cutting	Cassava/Malawi / Methodological Experiment on Measuring Cassava Production, Productivity, and Variety Identification 2015-16	<ul style="list-style-type: none"> <li>• Comparison of weekly harvest diaries to two different recall methods for cassava production in Malawi: a single visit with a 12-month reference period and two visits with 6-month reference periods.</li> <li>• Relative to crop cutting, the recall methods lead to under-reporting, especially with 12-month reference period.</li> <li>• Crop cutting is an upper bound for extended harvest crop cassava.</li> <li>• Harvest diaries, especially with the support of mobile phones, perform well to capture cassava production.</li> </ul>
Desideri and Joliffe (2018); Farmer self-report; maize crop cutting	Maize/Ethiopia / Ethiopia Socioeconomic Survey 2011-12 and 2013-14	<ul style="list-style-type: none"> <li>• Relative to the crop cut benchmark, respondents over-report maize production on small plots and under-report it on large plots.</li> </ul>

Paper	Product	Crop/Country/Dataset	Key insights
	mea- sure- ment meth- ods		
Gounie et al. (2019)	Farmer self-report; partial and full plot crop cutting	Maize/Uganda / Methodological Experiment on Measuring Maize Productivity, Soil Fertility and Variety 2015 and 2016	<ul style="list-style-type: none"> <li>• Relative to the crop cut benchmark, respondents significantly over-report maize production on smaller plots.</li> <li>• Over-reporting is related to rounding/heaping</li> </ul>
Abay et al. (2019)	Farmer self-report; crop cutting	Wheat/Ethiopia / Randomized controlled trial of 482 wheat farmers in Ethiopia 2013-14	<ul style="list-style-type: none"> <li>• Relative to crop cutting, self-reporting overestimates production, with small plots overestimating more than large plots.</li> </ul>
Bevis and Barrett (2020)	Farmer self-report	Various/Uganda / IFPRI, UBOS, and National Science Foundation panel survey of 972 households, 2002-3 and 2013.	<ul style="list-style-type: none"> <li>• The edges of plots are more productive than the inside, which may explain why farmers overstate yield on small plots and understate yield on large plots and may also affect reliability of crop cuts.</li> </ul>



Paper	Product	Crop/Country/Dataset	Key insights
Lobe et al. (2019)	Remote sensing; self-report; crop cutting; full plot crop cutting	Maize/Uganda/ Methodological Experiment on Measuring Maize Productivity, Soil Fertility and Variety 2015 and 2016	<ul style="list-style-type: none"><li>• Self-reported performed very poorly relative to full-plot crop cut.</li><li>• On average, difference between partial crop cut and full-plot crop cut not significant.</li><li>• Crop cut yield estimates captured one-quarter of full crop cut yield variability.</li><li>• Both calibrated and uncalibrated satellite yield estimates captured half of full crop cut yield variability on pure stand plots above 0.10 hectare.</li><li>• Uncalibrated yield estimates were consistently one ton per hectare higher than full-plot and partial crop cut.</li></ul>

### 3.3 Agricultural Production Data: Design Features of the Reference Questionnaire

We explain the questionnaire structure in the post-planting and post-harvest reference questionnaires that relate to measuring production that are reviewed in this chapter and earlier summarized in Table 2 “Survey decisions that Inform Production Measurement.” To measure production and farm productivity using

yields, the reference questionnaire first enumerates the household's parcels in the post-planting questionnaire using the parcel roster and parcel details which includes the parcel's size and tenure status. A parcel is defined as a piece of land exploited by one or more persons as a single farming unit. A parcel may be bounded by natural boundaries and may comprise one or more plots. After parcels are listed, plots within parcels and their characteristics are recorded in the plot roster and plot details. A plot is defined as a continuous piece of land on which a unique crop or a mixture of crops is grown, under a uniform, consistent crop management system. All outputs and inputs are recorded at the plot level.

The primary differences in the parcel and plot details modules of the questionnaire relate to the acquisition of the parcel and its tenure status in the parcel details module, and the management and soil characteristics of the plot including water sources and soil type in the plot details module. The crop roster in module 6 documents at the parcel-plot-crop level the list of crops cultivated on the plot during the reference agricultural season and seed varieties. The module also includes a question about productivity expectations which can be used to compare with actual productivity recorded in the post-harvest questionnaire.

In the post-harvest questionnaire, the parcel and plot roster can be updated in modules 1 and 2. The plot-crop roster in module 6 validates the crops that were reported to have been planted from the post-planting questionnaire and records information on why crops may not have been harvested. The primary sections of the questionnaire related to production are found in modules seven to ten. Modules seven and eight record information on production and the disposition of field crops, while modules nine and ten capture production and disposition of permanent tree crops.

The field crop production module includes information on field crop harvest including when the harvest of a particular parcel-plot-crop unit began, what percentage of the crop was harvested, and self-reported production information. If the harvest has not been completed, information is collected on when the harvest is expected to be completed and how much of the harvest had been completed to date. Finally, information on who controls the disposition of the crop harvest is recorded by respondent ID. The disposition of field crops is focused on capturing unprocessed crops sales including the quantity and total value of such sales. The volume of transactions related to the total sold and the transportation costs associated with the crop sales are also recorded. Other dispositions of field crops include auto-consumption, gifting and in-kind payments. Lastly, the storage of crops is recorded including the quantity and method of storing. The intended use of the storage for later sale, future consumption, use as seed or animal feed are asked in this module.

We make special note in this section of the questionnaire on the calculation of output prices from the sales data recorded in these modules. Revenues recorded provide the basis for the calculation of output prices, but special note of the location of sale is important analytically. In theory, farmgate prices should reflect the output price from sale on the farm or local market, yet crop sales often occur

### 3.3. AGRICULTURAL PRODUCTION DATA: DESIGN FEATURES OF THE REFERENCE QUESTIONNAIRE

in markets outside of the farmer's village. Revenue from sales in larger markets outside the farmer's village may include transportation cost premiums above the actual farmgate price. The timing of sale may also affect imputed output prices from revenue data. A large literature explores the seasonality of crop prices which may reflect risk or storage premiums in addition to the farmgate price. From this discussion, it is also clear that not all revenue sales represent the 'market' price as a market survey might collect. In the literature on food prices imputed from expenditure surveys, Deaton and Zaidi (2002) recommend taking the village or enumeration area median price to impute consumption per capita to account for variations in food prices. In a similar sense, estimating output prices from agricultural modules ideally would account for sale location as well as the timing of sale.

The tree and permanent crop production module is organized in a different manner than the field crop module. The primary motivation for this difference is that the number of trees per plot is necessary to estimate yield for tree crops. The last harvest date is recorded as well as the quantity of production and any losses on the parcel-plot-tree/permanent crop unit of analysis. The tree and permanent crop disposition module is similar in structure to the field crop disposition module, but uses a 12-month recall period.



## Chapter 4

# Chapter 4: Agricultural Inputs

The production function approach to questionnaire design requires mapping inputs to the outputs produced at plot level. This chapter reviews the measurement of chemical and organic inputs, agricultural labor and agricultural capital allocated to the plot including farm implements and machinery. The primary measurement challenge in designing an agricultural questionnaire to capture seasonal plot level production is precisely the attribution of inputs to the plot. Agricultural labor is drawn from the household's members, exchange labor, or hired labor from the community. In the case of exchange and hired labor, recall at the plot level may be facilitated by supervision requirements, the timing of the labor demand during the production process (planting, harvest, etc.) and the relative infrequency of these types of labor on plots. Household agricultural labor recall may suffer from attribution of household labor across multiple plots and over relatively longer recall periods. These measurement challenges are closely related to general challenges in time use modules. Fertilizers, herbicides and pesticides are normally purchased in bulk and distributed across the household or farmer's plots. Measuring machinery or farm implement use at the plot-level is related to the attribution challenge of non-labor inputs. Machinery and farm implements are capital goods that are used over multiple agricultural seasons. Quantifying how much they are used on a particular plot and the associated cost depends on the type of machinery and agricultural production process.

### 4.1 Agricultural labor

A large body of literature is devoted to agricultural labor measurement to understand national employment, labor productivity and worker earnings. The

primary motivation for agricultural labor measurement is to better understand smallholder agricultural productivity and its determinants. One of the most important of these determinants being is labor for which poor households have relatively larger endowments compared to capital or other inputs. Researchers and survey practitioners interested in measuring participation rates in work and employment are encouraged to consult the 19<sup>th</sup> International Conference of Labour Statisticians (ICLS) standards and the pilot studies conducted by the International Labour Organization related to the measurement of those concepts (Benes and Walsh, 2018a, 2018b). The focus in this section is exclusively on agricultural labor conducted on the household farm, including that by household members and hired and exchange laborers.<sup>1</sup>

Sagesaka et al. (2019) summarizes much of the current literature on measuring work in agricultural household surveys. Dixon-Mueller and Anker (1988) highlight how the classification of workers based on their main activity leads to underestimating the number of economically active women, primarily because women contribute significantly to both household and agricultural work. A significant number of survey experiments have also addressed survey design issues in agricultural labor measurement. A key set of survey design choices in agricultural labor measurement have been how to frame questions regarding farm work and whether labor information needs to be self-reported or permissive of proxy response. Bardasi et al. (2011) estimate biases in labor force participation, hours worked and income by gender and sector of employment due to questionnaire design such as screening questions and proxy response. In their survey experiment from Tanzania, they find no difference in female labor statistics due to proxy response, but lower male employment rates due to underreporting agricultural activity by proxy respondents. In Malawi, Kilic et al. (2020) find that proxy reporting increases underreporting of employment when recall periods increase and when women are the subject of proxy reporting. Arthi et al. (2018) estimate the effects of recall period, either weekly agricultural labor reporting or end of the season reporting. End of season recall increases by four times the hours reported by individuals at the plot level relative to those reporting weekly. They note aggregation to household level reporting causes the reported hour differences between the weekly and end of season recall periods to disappear. Arthi et al. (2018) interpret these findings as evidence that recall bias is driven by mental burdens of reporting such disaggregated time use patterns as well as memory. Given that these biases may differ depending on the level of aggregation at which they are used, the Tanzanian experiment suggest that agricultural labor productivity would be understated which are similar to findings by Gaddis et al. (2019) in Ghana.

Given the difficulties of measuring labor productivity, due to attributing output to a particular worker and the unit of time in which they produced such output, Akogun et al. (2020) measure the physical activity of sugarcane cutters which

---

<sup>1</sup>For guidance on the collection of data for labor more broadly, see the LSMS “Guidebook on Labor: Work and Employment in Multi-Topic Household Surveys” (forthcoming).

is a direct measure of effort in their piece rate wage setting. They find a high correlation between administrative data on output per worker recorded by the firm and the worker's physical activity, as well as large changes in the intensity of such activity in response to malaria testing and treatment. Integrating physical activity measures into national surveys may be possible using a subsample to calibrate biases in reported time use as well as predict effort-based measures of agricultural labor productivity.

An often overlooked agricultural labor measurement issue is the classification of children's time in agriculture. The ILO's International Conference of Labour Statisticians in 2008 focused on child labor measurement methodology and statistics. ILO (2008 and 2017) summarizes many of the measurement issues related to the classification of children's time in agriculture as work, child labor, or hazardous work. Definitions of child labor and hazardous work depend on the country legal and agricultural production context, despite efforts to standardize official reporting of child labor statistics. Guarcello et al. (2008) review 87 datasets from 35 countries that collect information on child labor. They conclude variation in child labor statistics varies more broadly within and across surveys relative to other children's activities such as schooling. While variation such as questionnaire design and sampling do explain some of this variation, Guarcello et al. (2008) note the importance of interviewer training and social stigma as potentially influential factors that may also create variation in reported child labor information.

Dillon et al. (2012) extend the analysis from Bardasi et al. (2011) in the context of child labor statistics and survey design, finding that in rural Tanzania proxy reporting from adults did not reduce child labor reporting relative to children's own self-reports. Rather, classification using less precise questions regarding work activities that should be classified as labor force participation reduced by 16 percentage points girl and boy reported participation in agriculture in comparison to survey designs that included screening questions listing specific work activities. The Tanzanian results are somewhat surprising in that the presumption of the international community is often that parents will want to avoid reporting child labor. However, in rural communities such as those in Tanzania, child work in agriculture is not uncommon and social stigma is low as children's work is viewed as human capital building. Careful training of interviewers to be mindful of cultural context may help improve reporting on (presumed) sensitive topics.

## 4.2 Seed, fertilizer, pesticides and herbicides

Measuring seed acquisition in input modules requires a questionnaire design that differs from recording other non-labor agricultural inputs. This is primarily due to the common informality of seed exchange among rural producers, but even in more formal seed markets, significant differences in seed brand names and seed

traits lead to potential misclassification. For some empirical applications, the researcher may want to link seed brand names to consumer choices. In many seed applications, the researcher may be interested in the complementary nature of seed choice and other input choices, or the return to specific seed traits such as drought tolerance or germination speed, in which case, the seed trait is the relevant seed characteristic. These classification issues, identifying both brand and traits of a specific input, are not exclusive to seeds. In the case of fertilizer, pesticides and herbicides, brand names are not directly aligned to chemical compositions of the input. Fertilizer, pesticide and herbicide packaging lists the chemical composition of the product which can be used to identify product traits along with the brand name. Non-standard unit reporting and planting technique or application method are also important non-labor agricultural input measurement issues. We have discussed non-standard units of reporting in the previous chapter. We note Oseni et al. (2017) provides extensive discussion of measurement techniques for non-standard units. We also highlight that planting techniques and application methods are critical to understand as complementary to the traits of the non-agricultural inputs that agronomically will affect the returns to these inputs. Planting and land-management choices such as micro-dosing, distance between crops, mounding or bunding, or inter-cropping will change the potential yield of seeds. Mixing techniques for herbicides and pesticides, as well as application methods affect chemical absorption. Similarly, fertilizer application methods such as broadcasting and micro-dosing will affect the return to these inputs.

An objective of measuring input use is to understand the returns to different agricultural investments, hence measuring the often-unobservable quality of inputs is an important characteristic of input investments. With many products, quality is not completely observable and quality perceptions matter. This holds true for seeds and other non-labor inputs. Improved seed varieties have been shown to produce greater, more resistance yields. However, due to several factors including seed recycling practices, informal or unregulated seed markets, counterfeit seed, and asymmetric information, among others, farmers are often uninformed or misinformed of the true variety they are cultivating. Given the numerous stages of the seed supply chain in which variety can be masked or modified, it is not surprising that farmer knowledge of seed variety is generally limited. Farmer identification of simply the type of seed, whether improved or local seed rather than the specific variety, is still challenging. Wineman et al. (2020) find, from a methodological study comparing farmer seed type identification against DNA fingerprinting, that farmers in Tanzania correctly identified maize seed type (improved or local) 70 percent of the time. Similarly, research in Ethiopia on the ability of farmers to identify the type and variety of sweet potato found that 20 percent of farmers falsely reported local varieties as improved, and 19 percent falsely reported improved varieties as local (Kosmowski et al., 2018). With respect to the quality of other inputs, Ashour et al. (2019) find that farmer's perceptions of herbicide quality are overstated, but that product quality is adulterated in their sample. Fifteen percent of the



herbicide samples they test are missing the active ingredient, while farmers believe that 41 percent of herbicides are counterfeit in their sample. Michaelson et al. (2018) report on farmer perceptions of input quality, comparing farmer perceptions with nutrient quality from laboratory-based measures. They find that in their sample from Tanzania fertilizers purchased in the market were not adulterated, yet farmers believed that they had been based on observable characteristics of the fertilizer sample which are not necessarily correlated with nutrient quality.

### 4.3 Machinery and Farm Implements

Agricultural capital in the form of machinery and farm implements can increase the productivity of smallholder farmers. Understanding how farm size and profitability change over time are linked to the mechanization of agriculture. While it is generally regarded as easy for farmers to recall agricultural capital within the household, the plot-level attribution and control of such capital are measurement challenges. Plot level attribution of machinery use is often avoided as it may be assumed by the survey designer that agricultural capital is shared equally in the household.

A large literature on women's empowerment in agriculture has focused on the correct measurement between women and men's ownership of assets versus their use-rights. Alkire et al. (2013) and Doss and Kieran (2014) provide a comprehensive review and guidelines for collecting sex-disaggregated asset data which apply generally to agricultural capital modules. They emphasize the importance of respondent and the method of collecting ownership and use-rights. Data from the machinery and farm implements modules can be linked to plot disaggregated production and other inputs modules to assess differences in intrahousehold allocation of inputs (Udry, 1996).

Recall periods for agricultural machinery and implements usually focuses on the availability of assets over the previous 12 months, differences in input use by crop-plot-season are important to capture, but may not be possible if the frequency of survey administration is annually, rather than seasonally. Machine age is usually collected with the intention that depreciation might be calculated, but much machinery depreciation depends on use frequency and maintenance which are more difficult to capture in household surveys.

[Table 5. Survey Decisions that inform input measurement](#)

---

*Post-Planting  
Questionnaire*

---

Module	Key Design Choices	Data quality Implications
<b>PP 5A - Household Member Labor Inputs</b>	Unit of observation for labor input is the individual-plot-level.	This unit of observation makes possible the analysis of labor input at the plot-level and allows gender-disaggregation of labor input
	Labor input data collected both in post-planting and in post-harvest visit.	This reduces the length of the recall period for the respondents thus reducing reporting error related to recall decay.
	Respondent selection/recording by individual-plot	In-person response considered more reliable than proxy-response, so respondent recording allowed to vary at plot-individual level.
<b>PP 5B - Hired and Exchange Labor Inputs</b>	Plot-person type as unit of observation for hired and exchange labor inputs	Makes possible analysis of labor inputs at the plot-level. Person types are women, men, children under 15, whose labor inputs are considered to have varying returns.
<b>PP 7 - Seed Acquisition</b>	Crop as unit of observation for seed acquisition	Allows analysis at plot-level for seeds which are key input in crop production.
	Reference period for seed acquisition is 'reference agricultural season'	The reference period aligns with the actual production cycle aiding respondent recall and improving data reliability.
	Reporting of seed quantity in non-standard units	Reporting in standard units (e.g. kilograms) poses difficulties for respondents and can worsen reporting error. Non-standard units should be country-specific.

---

*Post-Planting  
Questionnaire*

---

*Post-Harvest  
Questionnaire*

<b>PH 3 - Input Use</b>	Reference period for non-labor inputs is 'reference agricultural season'	The reference period aligns with the actual production cycle aiding respondent recall and improving data reliability. This module is administered in the post-harvest visit because non-labor inputs may be used throughout the agricultural season.
	Unit of observation for non-labor inputs is the plot-level.	Allows analysis at plot-level for non-labor inputs which are key input in crop production.
	Non-standard units of measurement for reporting of input quantities	Reporting in standard units (e.g. kilograms) poses difficulties for respondents and can worsen reporting error. Non-standard units should be country-specific.
	Respondent selection/recording by plot	Decision maker may vary from plot to plot, data considered most reliable if plot decision-maker can respond in person rather than proxy response.
<b>PH 4 - Input Roster</b>	Unit of observation for input acquisition of non-labor inputs is input type	Input is presumably acquired by type and not for each plot separately making this unit of observation easier to recall.
	Reference period for non-labor input acquisition is 'reference agricultural season'	The reference period aligns with the actual production cycle aiding respondent recall and improving data reliability. This module is administered in the post-harvest visit because non-labor inputs may be used throughout the agricultural season.

---

*Post-Planting  
Questionnaire*

---

<b>PH 5A - House-hold Member Labor Inputs</b>	Unit of observation for labor input is the individual-plot-level.	This unit of observation makes possible the analysis of labor input at the plot-level and allows gender-disaggregation of labor input
	Labor input data collected both in post-planting and in post-harvest visit.	This reduces the length of the recall period for the respondents thus reducing reporting error related to recall decay.
	Respondent selection/recording by individual-plot	Data are most reliable if each individual responds for him or herself personally, so respondent recording allowed to vary at plot-individual level.
<b>PH 5B - Hired and Exchange Labor Inputs</b>	Plot-person type as unit of observation for hired and exchange labor inputs	Makes possible analysis of labor inputs at the plot-level. Person types are: women, men, children under 15, whose labor inputs are considered to have varying returns.
<b>PH 11A - House-hold Members Post Harvest Labor</b>	Crop-individual as unit of observation	Post-harvest labor is not usually tied to plot as the natural unit of production. Instead, post-harvest labor activities, such as shelling or processing, may vary by crop.
	Respondent selection/recording by individual-crop	Data are most reliable if each individual responds for him or herself personally, so respondent recording allowed to vary at crop-individual level.

---

*Post-  
Planting  
Ques-  
tionnaire*

---

<b>PH 11B - Hired/Exchange Post Harvest Labor</b>	Crop-person type observation of observation for hired and exchange labor inputs	Makes possible analysis of labor inputs at the plot-level. Person types are: women, men, children under 15, whose labor inputs are considered to have varying returns.
<b>PH 12 - Farm Imple- ments, Machin- ery, Struc- tures</b>	Past 12 months reference period	Farm implements, machinery, structures in their lifetime are independent of the agricultural season so that 12 months is a more suitable length of reference period
<b>PH 13 - Exten- sion Services</b>	Past 12 months reference period	Extension services do not follow a seasonal pattern so that 12 months is a more suitable length of reference period

---

## 4.4 Agricultural Inputs: Design Features of the Reference Questionnaires

Input utilization is captured in both the post-planting and post-harvest questionnaires. In the post-planting questionnaire, labor used for land clearing and planting is recorded in module 5 as well as seed acquisition in module 7. In the post-harvest questionnaire, input use, labor inputs including household, hired and exchange labor, farm capital, and extension service use is documented in modules 2-5 and 11-13.

Labor in the post-planting questionnaire relates specifically to land preparation, planting, weeding, ridging or fertilizing, and supervision labor. Labor inputs are recorded at the parcel-plot-individual level in the household labor input module 5A. The number of days and the average number of hours provided on the plot are registered. In module 5B, hired and exchange labor is recorded at the parcel-plot by person type (for instance men, women, and children under 15). The number of person types used on the parcel-plot, the days, the hours per day, activities and remuneration are documented. With respect to remunera-

tion, the value of payments including those in-kind are asked. Exchange labor is also recorded at the parcel-plot level, but no estimation of remuneration is documented, only the number of person types, days, hours per day and activities are included in the post-planting labor modules. This survey design feature captures local context of reciprocal labor markets but makes calculating agricultural wages challenging due to the sampling of households which may have low labor demand. Alternative labor market surveys or supplemental agricultural labor market modules may be essential to accurately capture agricultural wage variation.

Information on seed acquisition is found in module 7 of the post-planting questionnaire. Seed acquisition information is linked to the crop codes, rather than alternative units of analysis such as the parcel-plot. As a survey design choice, collecting seed information at the plot level may be overly repetitive for the respondent, but the seed acquisition module does allow for multiple seed types to be recorded per crop. In this module, the source of seed acquisition is the organizing principle of the module. Seed can be left over from a previous harvest, acquired freely or purchased.

In the post-harvest questionnaire, input use related to production after planting is collected. Organic and inorganic fertilizer use, herbicide/pesticides use, and any equipment or machinery used for harvesting are populated in the input roster collected at the parcel-plot level in post-harvest questionnaire module 3. The input roster, module 4, is organized by input type and documents input purchases, quantities, and who financed inputs. Input acquisition such as being left over from the previous season, own-produced inputs or free inputs are included in accounting for quantities obtained by the household.

Labor during the harvest period is documented in module 5A for household labor and module 5B for hired and exchange labor. The questionnaire modules are similar to those of the post-planting labor module though post-harvest activities such as threshing/shelling and cleaning are expressly not recorded, so that this information can be collected within the post-harvest labor modules found in modules 12A (household) and 12B (hired/exchange labor). These post-harvest labor modules relate to agricultural labor activities such as threshing/shelling, drying, cleaning, and processing.

The last input-related modules in the post-harvest questionnaire include the agricultural capital and extension services modules (13 and 14, respectively). The agricultural capital module captures farm implements, machinery and structures that are inputs into agricultural production. The unit of analysis in this module is that of the item. The number, age, and sale value of the item is recorded. Any rental costs or rental revenues from such items are also recorded in the questionnaire to facilitate cost analysis of production. The extension services module documents information received, who received information, and the frequency of contact with an extensive list of agricultural extension sources. These include both formal sources such as government, NGO, private extension agents, or cooperatives, but also from peer farmers and media sources.

## Chapter 5

# Chapter 5: Livestock

Pastoral households rely on livestock assets and revenue, but livestock are also central to rural and urban households in LMICs. While livestock is a vital component of the agricultural sector, data on livestock, costs of production (veterinary services, labor, feed, and shelter costs), income sources, mortality and livestock sales are often limited in nationally representative data. Livestock modules are increasingly integrated into agricultural surveys to capture the multiple sources of livestock use. In this chapter, we cover questionnaire design choices in livestock modules highlighting measurement and empirical reasons for attention to measurement error. Livestock measurement challenges relate to capturing the reproductive and production cycles of livestock systems, noting that the salience of different farmer choices and continuous nature of production create recall challenges in the context of annual multi-topic household surveys. Detailed livestock data require attention to both the stock and flow variables in such a production system. Anagol et al. (2017) note in the context of India that median returns for cows were -7 percent with a 17 percent median return for buffalos when they assumed household labor was valued at zero. Despite mixed median returns, 51 percent and 45 percent of cow and buffalo owners reported negative returns. Measurement error may explain, in part, high fractions of negative returns, underscoring the survey design challenges of livestock measurement in household surveys. An LSMS guidebook on the livestock modules provides a more detailed review including variations in short, standard, and extended questionnaire modules (Zezza et al., 2016b).

### 5.1 Measuring Stocks

A primary objective of livestock modules is to register the number of animals that are owned by a household. Description of the animal breeds, age, and sex are important characteristics to quantify the household's livestock portfolio.

With the information on animal breeds and counts, livestock indices such as tropical livestock units may be constructed to compare herds across households (FAO, 2011). It is important to note that the unit of analysis for most livestock modules in multi-topic household surveys is the household herd. Ownership of livestock as with land modules has congruent measurement issues related to ownership and management of animals. As with land modules, the livestock owner may not actually be the same person who manages the livestock. This presents a potential opportunity for sex-disaggregated analysis, linking to other sections of the multi-topic household questionnaire to infer effects on welfare if the owner and livestock manager reside in the same household.

In many pastoralist communities, sedentary households may own livestock that are managed by a nomadic herder or a local herder who manages several herds for households in a village. This raises multiple measurement issues. First, the characteristics of the livestock manager may not be available for analysis because the manager does not reside in the household and is not listed as part of the household roster. Second, the manager may derive compensation from animal products such as the milk of the animals. In a production function approach, revenues from animal sourced products should be listed as revenues from the livestock and payment in-kind to managers should be listed as a cost of production to the owners. In principle, an economic definition of pastoral income would also include the net livestock value change, but in the context of multi-topic household surveys and due to the challenges of valuing livestock value changes, we compare revenues and consumption to costs related to the ‘flow’ variables rather than changes in the value of the animal stock. If animals are herded remotely, then the unobservability of the use of animal products creates measurement issues. Third, changes in animal stocks such as births, deaths or theft may also not be observable to the owner until herds are brought closer to the sedentary household villages which usually occurs seasonally.

## 5.2 Costs of Production

As in much agricultural survey design, a production function approach to questionnaire design requires agricultural outputs to relate to inputs. Measuring livestock inputs and the costs of livestock rearing are difficult for many of the same reasons that seasonal crop input measurement is challenging. Some costs of livestock production are seasonal and many are irregular and involve high up-front costs. This is particularly true of veterinary costs which also illustrate how the unit of analysis can matter in survey design. From a veterinary perspective, immunizations and animal care costs relate to a specific animal’s health and are critical for different animals at different ages. In livestock modules collected for multi-topic household surveys, a household’s herd is recorded and disaggregated by animal breed rather than at the animal level. Veterinary services may be recorded either generally for the herd or by breed. In practice, reporting veterinary services is often difficult for farmers due to longer recall periods and the



often specificity of veterinary services required of individual animals rather than for an entire herd or breed. This has empirical implications on whether an analyst can estimate a herd production function or a livestock specific production function. Empirical objectives to measure either herd or livestock-specific production functions need to be balanced against whether it is realistic for farmers to provide the necessary detail for livestock inputs and outputs at the chosen unit of analysis. Other costs of production such as food costs are critical to understanding the return to livestock, but many rural animals may graze freely while other animals such as chickens or pigs may be confined to pens for their safety. In principle, the consumption of grazing lands by the herd should be valued, but this analysis would require a land-based sampling frame rather than a population-based sampling frame as in household surveys.

The cost of labor devoted to animal care and herding is also critical to understanding the costs of livestock production. Common issues with labor measurement, reviewed in the earlier chapter, apply to livestock as well. Time use is difficult to measure, particularly over recall periods longer than seven days as is attribution of labor time to certain animals.

For domestic animal production such as chickens, goats, or pigs, labor time devoted to animals may be interspersed in the households' daily activities. that includes production of household public goods. Chores such as feeding, care and minor animal enclosure maintenance would be considered as part of the household's labor market activities, while activities related to producing household public goods such as household cleaning or cooking would be considered domestic activities, not included in the labor market. Due to the interconnection of activities related to household activities and domestic animal care, women's role in livestock production may be undervalued in household surveys.

Children's work in livestock production both in terms of herding and domestic production is likely under-reported in agricultural surveys, though children play key roles in the livestock sector (FAO, 2013). While some of children's activities in the livestock sector may build their human capital, it may be considered child labor if the work includes hazardous activities or interferes with their formal education. Formal measurement of children's activities in livestock may be difficult in household-based surveys if child herders are not listed in the household roster or excluded due to household definitions that require minimum residency.

### 5.3 Income sources

The challenge of measuring revenue flows from stock variables requires carefully enumerating the multiple sources of revenue. For livestock, revenue flows from the sale of milk, eggs, animal traction and dung are all potential sources of revenue. Each is challenging as milk and egg production is irregular while demand

for animal traction and dung are seasonal. The survey experiment on measuring milk off-take by Zezza et al. (2016a), notes that imputing milk off-take is difficult because lactating females can be milked multiple times during the day, milking frequency varies seasonally, milk production varies by lactation stage, and reproductive and lactating females may not be milked. Milk production questions are often asked in a household survey by first soliciting the number of production months in the last twelve months and then the average production per month during production months. In addition to production-related measurement issues, valuing milk production without milk sales information requires analysts to make assumptions about sale frequency and milk prices which may vary daily or weekly depending on the milk market. In the Zezza et al. (2016a) survey experiment, they find that 12-month recall using the standard two-part household survey questions recorded similar production quantities relative to estimating a lactation curve which asks extended questions related to milk off-take at four different time periods during the annual recall period in an attempt to capture lactation dynamics. Apart from this study, there are few survey experiments to inform the design of livestock modules in agricultural surveys. Despite measurement issues in milk sales, estimates are used to inform important economic questions as in Casaburi and Macchiavello (2019) who use contract design experiments in milk sales markets to estimate time preferences for Kenyan dairy farmers and Hoddinott et al. (2015) who find a positive correlation in estimating the effects of cow ownership on children's milk consumption and stunting.

Seasonal sources of animal products such as animal traction or dung can be captured in agricultural surveys if surveys are implemented during periods of high demand for such animal products. In practice, animal traction services may be easier to value in active agricultural services markets, most likely in contexts with contract or irrigated agricultural schemes. Fully capturing dung sales is difficult as larger animals are often encouraged to freely roam in rural contexts to help fertilize the fields and eat crop residues, clearing fields for farmers. As this reciprocal arrangement between farmers and animal owners is beneficial for both parties, formal dung markets are rarer in rural contexts. Livestock holdings also serve to provide insurance against risk by providing non-crop sources of revenue from milk, egg or animal sales. Increased livestock holdings can improve food security through income diversification, but also providing a savings mechanism for rural households without access to formal financial services.

## 5.4 Challenges to Livestock Measurement

While Pica-Ciamarra et al. (2014) and Zezza et al. (2016b) review best practices in livestock module design, outstanding design issues are well documented in the survey design literature and remain challenges for agricultural survey design. Many of these challenges relate to the household unit of analysis at which

livestock data are collected. First, mortality and livestock sales are often under-reported in household surveys as rural households often maintain herds outside of their residence. Mortality may be under-reported if stigma is associated with animal loss or animal owners are particularly sensitive to difficult questions related to the death of animals. Livestock sales may not be frequent occurrences, so multiple sales and at which prices may be difficult for respondents to recall. Second, as noted above, many livestock sector outcomes should be recorded at the herd level rather than the household level, particularly for nomadic populations. Household surveys with a population-based sampling will never capture herd level outcomes for which area-based sampling measures may be a better unit of analysis. Lastly, FAO (2011) which outlines guidelines for preparing livestock sector reviews underscores environmental and animal welfare issues that are often outside the scope of standard household livestock modules designed to measure a herd production function. If environmental consequences or social issues related to animal herding are the objective of empirical analysis, area-based sampling may more accurately capture a representative sample of landscapes for which environmental consequences or conflict between pastoralists and farmers occur. Animal health and welfare are also not easily addressed when animal information is only captured at the animal breed level. Veterinary service expenditures may be capturing farmers engaged actively to help sick animals as opposed to evidence of animal welfare concerns.

Table . Design Decisions in Livestock Production Measurement

---

***Livestock  
Question-  
naire***

---

<b>LS SECTION A: Livestock Ownership</b>	Livestock breed as unit of observation and disaggregated by sex and age	Critical for accurate capture of herd size and composition information. Valuation of livestock differs by gender and age. Greater aggregation would place burden on respondent to sum up various different breed.
	Differentiation between local/indigenous vs improved/exotic breeds	Breed type has implication for valuation of livestock

---

**Livestock  
Question-  
naire**


---

	Respondent recommended to be recorded at the livestock group (large ruminants, small ruminants, poultry etc.) level	Management responsibilities likely vary by groups of livestock and so the most informed household member may also vary at that level
	Household member-level livestock ownership and management responsibility recorded	Recording who in the household owns livestock in key to understanding gender dynamics and gaps in asset ownership; additionally, ownership is differentiated from management, which is also likely to have a gender gradient
<b>LS SECTION B1: Changes in Stock over the Past 12 Months</b>	Recall period for small and large ruminants: 12 months	Events related to large and small ruminants are likely not highly frequent, e.g. gestation periods for large ruminants is 250-300 days and 140-160 days for small ruminants, making longer recall period best suited.
<b>LS SECTION B2: Changes in Stock over the Past 3 Months: Poultry</b>	Recall period for poultry: 3 months	Events related to poultry are more frequent, e.g. gestation periods are only a few weeks, and there is limited seasonality, so that a shorter recall period is better suited.
<b>LS SECTION C: Breeding, Housing, Water, Feeding, and Hired Labor</b>	Unit of observation is livestock group (large ruminant, small ruminants, poultry, etc.)	Management usually differs only between livestock groups (large ruminants, small ruminants, poultry etc.) but not between breed of the same group

---

**Livestock  
Question-  
naire**


---

	Recording of individual-level management responsibilities for different aspects of husbandry (breeding, water, housing, etc.)	Important for gender analysis
<b>LS SECTION D: Animal Health</b>	Unit of observation is livestock group (large ruminant, small ruminants, poultry, etc.)	Health-related practices usually differ only between livestock groups (large ruminants, small ruminants, poultry etc.) but not between breed of the same group
<b>LS SECTION E: Milk Production (Off-take)</b>	Recall period 12 months	Events related to large and small ruminants are likely not highly frequent, e.g. gestation periods for large ruminants is 250-300 days and 140-160 days for small ruminants, making longer recall period best suited
<b>LS SECTION F: Egg Production</b>	Recall period 3 months	Clutching period is usually a few weeks and seasonality limited so that shorter recall period better suited
<b>LS SECTION G: Animal Power</b>	Unit of observation is livestock group (large ruminant, small ruminants, poultry, etc.)	Limited variation between breeds within groups (large ruminants, small ruminants, poultry etc.).
<b>LS SECTION H: Dung</b>	Unit of observation is livestock group (large ruminant, small ruminants, poultry, etc.)	Limited variation between breeds within groups (large ruminants, small ruminants, poultry etc.).

---

## 5.5 Livestock Data: Design Features of the Reference Questionnaires

The reference livestock questionnaire, which is sourced from Zezza et al. (2016b), is found in Appendix II which also indicates a shorter base set of questions which are highlighted in green. For surveys which require comprehensive data on livestock activities, the full questionnaire covers livestock ownership; changes in stocks (using a 12-month-reference period for livestock and a three-month reference period for poultry), breeding, housing water, feeding and hired labor; animal health; milk production; egg production; animal power; and dung. The questionnaire modules, or *sections* as they are referred to in Zezza et al (2016b), and the related design decisions are summarized in Table 6.

The livestock ownership section of the questionnaire establishes which animals are owned by the household, the number of owned animals, who within the household owns or manages the animals, and the main reasons for holding animals. A key survey design choice is whether to limit questions to animal types and numbers within the household or to disaggregate livestock ownership and management responsibilities by person ID within the household. Disaggregated livestock ownership and management questions permit analysis by sex, age and other individual characteristics and by animal sourced revenues.

Sections B1 and B2 of the questionnaire classify the sources for changes in the stock of large and medium-sized animals over a 12-month reference period and changes in the stock of poultry over a three-month reference period. Differences in reference periods reflect assumptions concerning the reproductive cycles of animals, the liquidity of animal markets where smaller animals may be exchanged more frequently or more likely to have differential survival rates. Both modules pose questions related to the birth, purchase, gifting of animals (both those received and given), lost, sold or slaughtered.

Animal care practices are covered in Section C of the animal questionnaire. Animal care includes breeding practices, housing, watering, feed and hired animal care labor. Recording animal care practices documents investments in animal quality, but also provides associated cost information which are essential in estimating an animal profit function or the returns to livestock investment. Section D records animal health information specific to disease, curative care, and preventive practices such as vaccination. This subsection records disease incidence as well as costs associated with vaccines and treatment.

Sections E and F record animal-sourced revenues from milk and egg production. Milk production is recorded by animal where the average number of animals milked per month and the average amount of milk produced per day. It is possible to record the person ID of the household member responsible for milking and recommended to record the average amount and earnings from weekly milk sold. The egg production module follows a similar structure with deviations in the reference period. The eggs per clutching period, the number of eggs sold

#### 5.5. LIVESTOCK DATA: DESIGN FEATURES OF THE REFERENCE QUESTIONNAIRES 79

and earnings over the past three months are recorded for poultry.

Sections G and H record in-kind animal uses which increase agricultural productivity via animal power and dung production. Animal power is recorded related to both transportation and agricultural production. Earnings from animal services is recorded over the past 12 months. Animal dung uses are recorded for owned animals over the past 12 months as well. Animal dung uses include manure, fuel, construction material, feed to other animals, and sales. Any dung sales are recorded in local currency over the past 12 months.





## Chapter 6

# Chapter 6: Field Implementation

Agricultural surveys may be implemented to achieve various objectives, and these objectives ought to be considered when designing and fielding data collection efforts. NSOs, which are common implementing agencies for agricultural surveys, may seek to use agricultural surveys to produce estimates of crop production or monitor changes in agricultural production over time. These objectives will have implications on the survey design, as estimation of crop production may warrant a cross-sectional design with a larger sample in order to achieve estimates at the desired level of representativeness, while an objective of monitoring and understanding the dynamics of the agricultural sector warrants a panel approach. These objectives and the implications on survey design all need to be considered, while also balancing the often-binding resource constraints and wider mandate of the agency.

This chapter discusses practical considerations for implementing agricultural surveys, with consideration for the primary objectives and constraints faced by implementing agencies, building on the experience of the LSMS-ISA. This chapter primarily focuses on implementation by or in collaboration with NSOs, but many of the principles and challenges of collaboration are instructive for other national collaborations designed to produce data. For information on household surveys costs conducted in partnership with NSOs around the globe, Kilic et al. (2017) summarizes costs and lessons learned from such collaborations.

## 6.1 Fieldwork Implementation Tradeoffs and Survey Design Choices

Survey fieldwork is a costly and complex operation. Here we discuss various design and implementation decisions that are to be considered in order to achieve the survey objectives while balancing data quality with resource constraints that are common to NSOs and other implementing agencies. These decision points, in which tradeoffs between data quality and implementation cost are present, include but are not limited to the following:

- Number of survey visits
- Sample size
- Question translation
- Integration of objective measures
- Survey content and questionnaire design
- Mode of data collection
- Enumerator recruitment, training, and supervision

The reference questionnaire in Appendix I, as well as the discussion in previous chapters, promotes interviews being split up into two visits whereby households are visited once following planting and again following harvest. This approach is recommended as it shortens the recall period for planting activities and splits the respondent burden across multiple visits. However, the feasibility of this approach depends on two things: (i) the agricultural calendar of the given country, and (ii) the resource constraints and parallel survey operations of the implementing agency which may further strain resource availability. On the former, if a country has multiple cropping seasons per year, post-planting and post-harvest visits may not be realistic for every cropping season, and depending on the agricultural calendar, recall periods may be reduced in these multi-season years supporting a case for only a post-harvest visit for each season. Resource constraints in terms of funding and/or human resources may also prohibit the implementation of a two-visit survey structure, in which aggregation of data collection efforts into a single post-harvest visit would be the alternative. If resources allow, more frequent visits may be implemented to reduce recall bias.

The objectives of the survey will also inform the sample size required. For operations aimed at highly precise production estimates representative of small administrative areas, a large sample size will likely be needed, depending on the number of key crops and heterogeneity of the sample areas. For example, Ethiopia's 2015-16 Annual Agricultural Sample Survey included a sample of

more than 44,000 households, while the LSMS-ISA supported Ethiopia Socioeconomic Survey panel in the same year was implemented on a sample of roughly 5,000 households.<sup>1</sup> Tradeoffs will be necessary between the detail of the survey instrument and the sample size in order to accommodate resource constraints of the implementing agency. The reference questionnaire provided in Appendix I, for example, includes highly detailed information on the drivers of production that may not be feasible/realistic for implementation in large sample sizes given the time it takes to administer such a survey.

There is also a tradeoff between carefully worded questions and easily understandable questions. While survey designers may be inclined to phrase questions in a very precise manner to get at the exact data they are seeking, this specific phrasing may either be difficult to understand by respondents or lost in translation. Unless the questionnaire is originally designed in the language in which it will be administered, translation will be necessary to account for the language, terms, and concepts of the local context. Back translation, wherein the questionnaire is translated to the local language, and then back to the original language, is recommended to ensure proper translation. Piloting of the questionnaire will also serve to flag incorrect translations or context-specific concepts that are not appropriately reflected in the questionnaire.

Significant tradeoffs exist around the measurement methods employed for specific topics. As discussed previously, systematic bias is evident in the respondent reporting of key agricultural variables such as land area and crop production. These biases can be mitigated by the integration of objective measures into survey operations. These objective, or pseudo-objective, measures may include the use of GPS for area measurement, crop-cutting for measurement of crop production, soil sampling for soil quality, etc. These measures have been shown to dramatically improve data quality, though they do come at a cost. Costs include the procurement of relevant equipment, training on the measures, and additional fieldwork time. The timing of visits should also be considered when determining which objective measures to integrate into a survey, as some are time sensitive and cannot be completed in a single-visit survey (such as crop cutting). Where integration of these measurements on the full sample is not possible due to resource constraints, a combination of sub-sampling for objective measurement and imputation techniques can be used to improve data quality across the sample with limited financial investment. Kilic et al. (2017), for example, illustrate how multiple imputation techniques can be used to impute improved measures of plot area for observations where GPS measurement was not conducted.

A significant discussion with implementation partners is the survey content and questionnaire design choices which may increase field time and interview complexity. Our discussion throughout this volume was to document best prac-

---

<sup>1</sup>Sample size of the 2015-16 Annual Agricultural Survey extracted from the CSA's Report on Area and Production of Major Crops, available here. Sample size for the Ethiopia Socioeconomic Survey extracted from the World Bank's Microdata Catalog here.

tice where possible and provide questionnaire designers with information about survey design choices. We fully acknowledge that it may not be possible to implement best practice in each survey due to fieldwork constraints, budget or partner capacity, but these choices should be taken with as much information on tradeoffs as possible.

The reference questionnaire discussed and appended to this document is designed to provide a detailed understanding of agricultural activities. The thoroughness of this questionnaire, however, may not be necessary or desired in contexts where agriculture makes up only a marginal share of the overall economy. In such cases where agriculture is not the focus of the survey or where it makes up a small share of household income, it is likely still necessary to collect some data on agriculture, if only to allow computation of total household incomes. The World Bank's LSMS team has developed an agriculture questionnaire which is designed for these cases, including only the minimum set of agricultural data necessary for analyzing the role of farm activities in household livelihoods. This shortened agriculture questionnaire is provided along with a reference household questionnaire and related guidebook (Oseni et al., 2020).

While the shortened version of the agriculture questionnaire includes crop, livestock, fishing, and forestry production, it of course comes with tradeoffs in topical coverage and level of detail. Data on crop production is collected at the parcel level, rather than at the plot level, which shortens the questionnaire but eliminates the ability to conduct plot-level productivity analysis. By aggregating to the parcel level, the concept of the plot manager, which is often used to conduct gender-related analysis, is also omitted. Details on non-labor inputs are also significantly minimized, with binary questions asked on fertilizer use rather than quantity of inputs applied, for example. No information is collected on crop processing or extension services in the shorter module. Additionally, the shortened questionnaire is designed to be implemented in a single visit, in conjunction with the household survey, rather than split into post-planting and post-harvest visits.

The tradeoffs around mode of data collection and enumerators are discussed separately in the subsequent sections.

## 6.2 Piloting for Improved Survey Design

Piloting surveys in the field before implementation is an important tool to assess survey design choices. We highlight the benefits of piloting particularly when a literature on a survey design choice is thin. Incorporating methodological experiments into the pilot phase of a project is an approach to maximize the benefit of field research for projects who have already budgeted for piloting. If survey design effects are large enough to potentially bias estimates, low cost methodological experiments should be able to detect survey design biases in relatively small sized experiments.

While integrating methodological experiments into piloting provides a direct estimate of biases in survey design choices, piloting can take many forms with informative results for survey design. Piloting even with a few respondents allows survey designers and national collaborators to assess a survey instrument's ease of implementation in the field, provide a method to evaluate interviewer's understanding of the questionnaire, and identify areas of questionnaire design improvement. Field protocols can be tested during pilots and a survey's approximate duration can be estimated. During piloting, it is good to evaluate survey duration using the pilot survey duration as an upper bound on survey duration during the whole of fieldwork as interviewers will become more efficient with more questionnaire experience.

Collaborations with national partners are enhanced by integrating survey designers with national stakeholders in the piloting process. National experts in complementary disciplines can also provide insights to better design surveys tailored to respondent thinking about variables of interest for researchers. Field partner collaboration extends beyond soliciting accurate response categories for questions and into how variations in production systems within countries are best represented clearly for respondents and translated into an agricultural questionnaire design. Question phrasing that better reflects cultural context will be easier for respondents to understand and accurately report their responses.

## 6.3 Mode of Data Collection

Agricultural survey data can be collected through paper-based questionnaires ("paper assisted personal interviewing" or PAPI) or tablet- or computer-based questionnaires ("computer assisted personal interviewing" or CAPI). Historically, PAPI has been the method of choice, largely for lack of alternative options. However, CAPI technologies have evolved and are now widely adopted. At the time of writing, all LSMS-ISA surveys have migrated from PAPI to CAPI implementation, using the World Bank's Survey Solutions CAPI software. Both modes of data collection come with pros and cons, though CAPI implementation is strongly recommended. PAPI implementation offers the advantage of relatively less upfront preparation time prior to fieldwork but requires significant effort in the data entry stage. In the case of PAPI implementation, it is recommended that a double-data entry approach be used in order to limit the influence of entry errors. Data entry, which can be done concurrently in the field by a designated data entry operator or at a centralized location, comes at a cost in terms of personnel and time required for data dissemination. Data quality controls in PAPI are also rather limited relative to CAPI operations. In PAPI, supervisors often review the paper questionnaire at the close of each day. If errors are found, this generally requires a re-visit to the household in order to confirm or rectify the relevant responses.

CAPI-based data collection benefits from real-time data quality controls as well

as trackable revisions and supervisor review. Research from various threads of development economics have cited a reduction in errors as a result of the use of CAPI implementation, for example the Asian Development Bank (2019) on labor force data and agricultural household survey data, Fafchamps et al. (2012) on microenterprise profit data, and Caeyers et al. (2012) on consumption data. Errors that are caught in real-time, through warning messages for pre-coded errors, prompt the enumerator to confirm or correct relevant inputs at the moment the question is asked, thereby increasing data quality and limiting the need to revisit the household. Data quality can be increased at the micro level, but also at the macro level through various survey management tools that allow office-based personnel to effectively monitor operations and data quality in real time. With relatively minimal investment, survey management tables and graphics can also be automated through application programming interfaces (APIs). An additional advantage of CAPI is the possibility of developing and administering interviews in different languages, allowing interviewers to toggle between different languages as necessary to accommodate respondents. Implementation via CAPI requires additional preparation time in order to test the application and program error checks, but it can significantly decrease the time from the close of fieldwork to data dissemination as data entry is not needed and data quality controls can be implemented continuously throughout fieldwork. Numerous open-source and private CAPI software options exist, common platforms include the World Bank's Survey Solutions software, SurveyCTO, Open Data Kit (ODK), CSPro and SurveyBee. The questionnaires found in Appendix I and II are also publicly available in Survey Solutions.

With the rate of mobile phone ownership growing globally, the potential for phone-based or phone-assisted surveys increases. While national-scale agricultural data collection through phone surveys has yet to be implemented, recent methodological research on the topic has shown promise for improving data quality in some areas, while the flurry of phone surveys implemented in the face of the COVID-19 pandemic shows promise for feasibility of implementation. Methodological research on agricultural labor data collection in Tanzania, for example, shows consistent data collected through weekly in-person visits, considered to be the gold standard, and weekly phone calls, suggesting that phone surveys could be a reasonable alternative (Arthi et al., 2018). Implementation of surveys via mobile phone will necessitate simplification of the questionnaire instrument and careful assessment of mobile phone coverage in the population of interest. Further research is needed in order to fully understand the risks and challenges of phone-based national-level agricultural survey data collection.

## 6.4 Fieldwork Organization and Logistics

Several features of fieldwork organization and logistics can impact data quality outcomes, including the composition, training, and structure of fieldwork teams. Fieldwork organization should include a supervisory hierarchy, with several field

teams deployed, each including a set of enumerators and a supervisor, and a separate headquarters-based supervisory team which conducts periodic field supervision and continuous data review. Where a panel approach is employed, a specialized tracking team can be deployed with the objective of tracking and interviewing respondents that have relocated between survey rounds and/or split off from previous households and created new *split-off* households.<sup>2</sup> In designing fieldwork structure, enumerators may be mobile (moving with teams from enumeration area to enumeration area) or resident (permanently residing in or close to the enumeration area for which s/he is responsible). The survey-structure and objectives of the survey should help to inform the fieldwork structure. Where multiple visits are necessary, such as when crop-cutting is to be implemented, resident enumerators may be more cost-efficient and effective.

At the base of the data generating process is the enumerator who manages the collection of data from respondents. A meta-analysis of the literature<sup>3</sup> establishes that enumerator behavioral traits and demographic characteristics influence survey responses and by extension data quality. Response rates and response biases are particularly influenced by specific enumerator characteristics (age, sex, ethnicity, experience, education, etc.), behaviors (formal versus conversational interview style), cognitive and non-cognitive skills (such as mathematical ability, reading, attention to detail, and empathy) and interviewer experience<sup>4,5,6</sup>. Recent research on enumerator effects by Di Maio and Fiala (2019), through a randomized experiment in Uganda, illustrate that enumerator characteristics and their differences from respondent characteristics affect survey responses and ultimately data quality, especially related to sensitive political topics. Responses in less sensitive topics were much less, or not at all, sensitive to enumerator characteristics. This is supported by additional research which suggests that salience and sensitivity of the questions at hand influence the nature and magnitude of enumerator effects (for example, Himelein, 2016; Laajaj and Macours, 2017). Beaman et al. (2018) find referral-based recruitment may disadvantage women in an enumerator recruitment experiment in Malawi. Marx et al. (2018) provide evidence on team composition and ethnic diversity on enumerator performance. Data on time use of field teams suggests horizon-

<sup>2</sup>Tracking protocols for the LSMS-ISA surveys are available as part of the documentation released with the survey data. Visit the World Bank's Microdata Library (<https://microdata.worldbank.org/>) to find all of the LSMS-ISA surveys. The tracking forms and protocols (included in the enumerator's manual) for the Malawi Integrated Household Panel Survey, for example, are available here: <https://microdata.worldbank.org/index.php/catalog/2939/related-materials>

<sup>3</sup>West, B.T. and Blom, A.G., 2017. Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, 5(2), pp.175-211.

<sup>4</sup>Cannell, C. F., and Laurent. (1977). A summary of studies of interviewing methodology. *Vital and Health Statistics: Series 2, Data Evaluation and Methods Research* (No. 69. DHEW Publication No. (HRA), 77-1343). Washington, DC: U.S. Government Printing Office.

<sup>5</sup>Fowler, F. J. Jr., and Mangione, T. W. (1985). *The value of interviewer training and supervision* (Final report to the National Center for Health services Research). Boston, MA: Center for Survey Research.

<sup>6</sup>Alcser K., Clemens J., Holland L., Guyer H., Hu M. 2016. *Interviewer recruitment, selection, and training. Cross-Cultural Survey Guidelines*. University of Michigan.

tally homogenous teams (enumerator teams) organized tasks more efficiently, while vertically homogenous teams (supervisors, enumerators, data monitors) had lower effort. Additional research is needed to gauge the sensitivity of agricultural data specifically to enumerator characteristics.

The capacity of enumerators should also be carefully assessed to ensure they possess the skills needed for the specific survey operation, with an eye for the data collection mode being employed. Some literature exists to suggest that enumerator recruitment and contract structure affect data quality and enumerator job performance. For agricultural surveys, knowledge of farming may prove to be an asset in terms of both understanding the questionnaire content and establishing rapport with respondents. This should be weighed against enumerator education, language skills, and familiarity with technological tools (in the case of CAPI implementation).

Effective survey implementation relies on a homogenous, thorough understanding of the questionnaire content and concepts. Training of field teams, as well as those involved in survey management at the headquarters level, is essential in developing this thorough understanding. Depending on the number of people involved and the structure of the NSO or other implementing agency, training can be conducted in a single centralized location or in several parallel decentralized training locations. Centralized trainings benefit from consistent messaging for all trainees, though may be ineffective if there is a large number of trainees. Decentralized trainings are often necessitated by large groups of trainees and/or decentralized statistical offices, though they come with the complication of ensuring discussions and questionnaire revisions are communicated across locations. Irrespective of the number of training locations, a training should be held prior to each survey visit (one training for the post-planting visit and one for the post-harvest visit, in the case of a two-visit survey). Prior to each training, a “training of trainers” is commonly held, especially in the case of decentralized trainings, in order to train the facilitators of the field team trainings and ensure the implementation and concepts are clear and harmonized prior to the full training. Training of trainers, and potentially fieldwork trainings, could also be held virtually in-part, if the context allows.

Fieldwork training should be inclusive of all facets of fieldwork, from protocols on conduct with respondents to questionnaire content to data review, revision, and submission. This generally requires both classroom-style training on relevant concepts, questionnaire content and layout, and CAPI software, if applicable. Classroom instruction should be complemented by mock interviews, both amongst the enumerators themselves and with pilot households, and training on objective measures that are being integrated into the survey. For example, training on the use of GPS devices for area measurement requires theoretical training in the classroom followed by repeated practice outside the classroom.<sup>7</sup> While the integration of objective measures is intended to reduce subjectivity

---

<sup>7</sup>For guidance on the implementation of GPS for area measurement, see the LSMS guidebook on Land Area Measurement in Household Surveys (Carletto et al., 2016).



in responses, training is still required to realize their maximum benefit.

Finally, enumerator activities should continue to be monitored throughout fieldwork. Supervision is particularly important at the onset of fieldwork, to identify and correct any misunderstandings, but is relevant for the full duration. Incoming data should be monitored both by supervisors in the field and the survey management team at the headquarters level, made quicker and more automated through the use of various CAPI tools. Additionally, supervisors from both the field and headquarters should conduct random spot checks or call backs to households to confirm data entered by enumerators matches that reported by the household.

## 6.5 Plot Visits and Georeferencing

An additional layer of consideration for the design and implementation of an agricultural survey is whether visitation to agricultural plots (or parcels) will be carried out. Visitation of agricultural plots, whereby an enumerator physically travels to plots typically with a member of the sampled household or farm, enables a wide range of additional survey operations and objective measures. Crop-cutting, which has been shown to significantly reduce measurement error in agricultural production estimates, as discussed in Chapter 3, for example, requires at least two visits to the selected plots (after planting and at time of harvest). Objective measures of agricultural land area, also discussed in Chapter 3, require physical presence on the land.

In addition to the benefits of potential objective measurements, visitation of agricultural plots allows for the *georeferencing* of plots (the collection of GPS coordinates of the plots and/or the identification of the boundaries of the plots), which can add significant value through the integration with other geospatial and remotely-sensed data sources such as data on soil properties, forest coverage, climate, and water quality, among many others. Integration of these properties with data on farming practices and outcomes can provide more complete insights into the agricultural landscape and barriers to increased productivity.

Integration of agricultural survey data with various types of spatial and Earth observation data, possible with the georeferencing of plots, has expanded the scope of analysis possible with the use of survey data, while also improving the performance prediction algorithms employed in spatial analyses. In recent years there has been growing demand for crop yield estimation via remotely sensed imagery, for example, for more frequent and rapid assessment of production estimates when agricultural surveys or censuses are not available. Recent research has cited improvements in yield estimation when estimation algorithms are calibrated on ground-based crop-cutting data collected through agricultural surveys. Lobell et al. (2019) find that calibration based on crop-cutting data from Uganda significantly limited the overestimation bias observed in the uncalibrated model of maize yields. Related research by Lobell et al. (2020) highlight the possibility

of sub-sampling of crop-cutting-based production measurement, whereby only a strategically selected subsample of plots are subject to crop-cutting rather than the full sample, could be sufficient to calibrate satellite-based estimation models. This would greatly improve the feasibility of scale-up in national survey operations.

While georeferencing of agricultural plots is strongly recommended when feasible, considerations need to be made to protect the confidentiality of survey respondents. GPS coordinates should not be disseminated to the public without an anonymization procedure.<sup>8</sup> Anonymization, or *offsetting*, of GPS coordinates will reduce the precision of the coordinates and therefore limit the value of integrated data sources through loss of variation at the local level. With raw GPS coordinates one is able to extract climate or soil conditions for that particular point, for example, whereas offset coordinates are commonly provided at the enumeration area level and thus the data extracted from geospatial data will be the same across the enumeration area. To maximize the value of integration with geospatial data sources while preserving respondent confidentiality, the LSMS-ISA prepares and disseminates a set of *geovariables* that are derived from the raw GPS coordinates, but does not release the raw coordinates themselves. Additional research is ongoing with respect to further ways of maximizing the use of georeferenced data while maintaining confidentiality. An active literature on creating synthetic data from confidential data may provide some insights to data integration challenges (see for example, Abowd and Schmutte 2019).

---

<sup>8</sup>An example of a GPS anonymization protocol employed by the DHS program can be found [here](#).

## Chapter 7

# Chapter 7: Conclusions and Future Directions for Agricultural Survey Design

In the 20 years that followed the Ghosh and Glewwe (2000), *Designing Household Survey Questionnaires for Developing Countries* volumes, innovations in agricultural survey design have expanded the set of answerable policy and research questions by filling data gaps, along with improving the quality of descriptive statistics that are critical to development monitoring and planning. This volume has focused on an approach to agricultural survey design that considers not only the objective of limiting measurement error through survey design, but also the tradeoffs that NSOs, researchers, policymakers and other stakeholders face when designing and using agricultural data. While the recommendations and design choices found in the reference questionnaires have evolved from the LSMS team's collaborations with NSOs, different studies and partnerships may find different data requirements are useful for achieving alternative survey objectives. Agricultural survey design must consider the real tradeoffs that field implementation and partnerships necessitate in concert rather than independently of the econometric consideration of minimizing measurement error to increase statistical precision and improve internal validity.

In this chapter, we take stock of future directions in agricultural survey design which may help guide a future methodological research agenda as well as stimulate debate about best practices. We highlight opportunities to expand the application of publicly collected data sets through data integration and the challenges of measuring fundamental agricultural concepts such as farmer preferences, labor, profitability and land.

## 7.1 Data Integration

The value of agricultural survey data can be increased through integration with other data sources. Integration of household survey and alternative data sources, such as telephone surveys, administrative records and geospatial data, can result in analytical gains from both sources and more precise measurement of variables when self-reports are not reliable. Household survey data, therefore, should be viewed as a complement to other data sources, not a substitute. Integrating supplementary data with household surveys is a trend for which increased investment by NSOs, researchers, international organizations and policymakers would yield a wider range of policy analysis. The integration of remotely sensed agricultural, ecological, and climate data is not uncommon in agricultural analysis, but the costs of such integration are often high. In Chapter 3, we highlight the potential of remotely sensed data. Here we also highlight limitations in terms of validation study requirements to calibrate remotely sensed data for ‘ground-truthing’ and the limitations of what can be imputed from remotely sensed crop images, particularly for tree and tuber crops and in intercropped areas. While these limitations are not trivial, remotely sensed plot data would permit a much leaner agricultural survey instrument which may improve self-reported variables on land use and production information. Without household survey integration, remotely sensed data itself does not satisfy most data requirements for policy analysis which require linking inputs and the use of production.

With careful design, agricultural survey data also has the potential to be integrated with administrative or census data. With the proper identifiers in the survey data to allow matching between the two sources, survey data can be merged, for example, with firm data linking employee- and employer-reported data, such as that done by de Nicola and Giné (2014) for fishing communities in India, or with microfinance or agricultural lender data to match access to credit with agricultural practices and outcomes. Agricultural administrative data are often under-utilized, in part, because extension data systems may not be well developed or data systems in Ministries of Agriculture may not be always linked to the National Statistical Office. Particularly in irrigated agricultural areas or commercial agricultural settings, administrative data on inputs such as water use, technology adoption, or crop yields can be of high quality. Where household surveys may also have the most promise in integrating with extension data sources are on topics of agricultural and environmental practices. A promising opportunity for integration may come with the expansion of digital, mobile and smart phone-based delivery of agricultural extension services, discussed for instance in Fabregas et al. (2019).

A promising innovation with potential applications for measuring physical activity, productivity and time use is wearable technology such as accelerometers worn either specifically for measurement purposes or embedded in mobile phones. Akogun et al. (2020) validate physical activity measures relative to individual farmer harvesting productivity, demonstrating that physical activity

measures are highly correlated with work effort. They also document potential uses and limitations of accelerometers for physical activity and productivity measures. Friedman et al. (forthcoming) illustrate a use case for physical activity data in understanding intra-household energy expenditure in rural settings. As the measurement of labor variables is one of the most challenging self-reported survey design problems, improved data integration using accelerometers and household surveys remains promising.

Lastly, we highlight telephone surveys as a survey mode that could be integrated to measure higher frequency outcomes such as labor, agricultural sales or food stocks. Telephone surveys have been heavily integrated into field operations in response to the global COVID-19 pandemic which resulted in limitations on in-person interviews. Literature on telephone surveys in developing countries existed before the pandemic (see for example, Dabalen et al. 2016), but the global pandemic has resulted in a renewed interest in telephone survey methodology. Existing literature suggests that response rates vary substantially by mode. Some response rates are 250 percent higher in computer-assisted telephone interviews (CATI) than interactive voice response surveys or SMS surveys based on metadata from 41 studies in 20 countries, but are significantly improved with small incentives (IPA, 2020). Drawbacks related to response rates and potential selection biases, particularly in the context of agricultural surveys, need to be further explored though improved telephone survey methodologies offer the potential for high frequency agricultural data collection with significant cost savings. Unit of analysis and proxy response are key open concerns for telephone surveys that use random digit dialing.

With integration of supplemental data sources, however, comes increased concern of respondent anonymity. While *open data* is the mantra of development institutions at large, as well as NSOs, respondent confidentiality must be maintained. This entails removing all information that could lead to the identification of respondents, including names, phone numbers, ID numbers, etc. However, integration with additional data sources complicates the anonymization practice, as it becomes easier to re-identify individuals through triangulation, especially those with uncommon characteristics (Heffetz and Ligett, 2014). Maximizing the value of survey data through public dissemination and integration with other sources, while maintaining the privacy of respondents, will encourage sustainable and effective survey operations going forward.

## 7.2 Measuring Theoretical Concepts More Precisely

The Ghosh and Glewwe volumes highlighted innovations in economic theory and how these ideas were translated to survey questions, generating the data for empirical analysis. As we have highlighted earlier, innovations in relating

consumption measures to welfare (Deaton and Zaidi, 2002), labor supply, returns to education, and the agricultural production function, among others, have resulted in improved survey design. There is still more work to do as an evolving scientific field to improve survey design to facilitate empirical analysis. We highlight a few areas of promise below but underscore that these examples are not an exhaustive research agenda.

Longstanding difficulties in measuring agricultural time use on plots or with livestock, not only create data that may be biased by non-classical measurement error but have implicitly restricted the set of potential econometric applications. The profit function is an important example of how restrictions in data quality prevent estimation of an underlying theoretical concept. Agricultural survey design has often adopted a production function approach as input quantities and outputs are more easily recalled by respondents. The limitation of such an approach, particularly for the LSMS-ISA surveys founded to measure welfare and its determinants, is that farmers do not maximize production and production does not necessarily relate to higher welfare as demonstrated in the food security and technology adoption literatures. The policy implications of not measuring profitability are potentially large for agricultural development.

As highlighted in Chapter 3, markets are often difficult to capture in household surveys as households only represent a partial segment of the demand and supply present in an economy. Recording the relevant prices in input, output, farm-gate, and market prices are vague boundaries in a continuum of economic activity and overlapping markets. The difficulty of accurately measuring prices and wages often limits potential structural estimation analysis. The concept ‘market access’ is inherently flawed when markets exist but may be seasonal or thin. For transactions that are infrequent such as in land or housing markets, the most valuable household assets may not be well captured in household survey samples but be the most consequential transactions from a welfare perspective. High frequency transactions are also consequential to household welfare but may be difficult to measure. Examples of this include food consumed outside of the household which may be difficult to record in household surveys by a main respondent who does not observe food consumption outside of the household (Oseni et al., 2017). In the case of small-scale food producers, frequent small transactions may be difficult to recall for respondents, particularly as recall periods increase. Food markets are fundamental to the agricultural supply chain, but their measurement remains challenging as represented in the above examples.

As a last example, we cite the measurement of preferences and agency as two theoretical concepts that have been widespread importance in empirical analysis. While the literature on risk preferences is extensive in the laboratory setting, preferences, including risk and time inconsistency which are foundational to understanding poverty and investment dynamics, remain omitted in most nationally representative survey data. The insights and econometric applications from behavioral economics have not been fully integrated into multi-topic

household questionnaires. A more widely emphasized, but not fully adapted literature on women's empowerment (Alkire et al., 2013; Glennester et al., 2018) is an important example of the importance of improved measures of agency to complement sex-disaggregated data that is largely available in nationally representative surveys.

### 7.3 Open Measurement Questions for Validation Research

Our review of the agricultural survey design literature leaves us impressed by the substantial progress the research, policy and national statistical office stakeholders have achieved since the Ghosh and Glewwe (2000) volume on measuring agricultural information. The range of empirical applications that innovations in plot and sex-disaggregated data in agricultural survey design have opened is impressive. We highlight that agricultural survey design is in no way a closed topic. A large and important research agenda of open measurement questions remains with important research and policy implications for farmers throughout the world.

In different sections of this volume, we have highlighted open measurement questions. Agricultural survey design is predicated on land measurement. Though a large literature has focused on the biases of self-reports, open questions on remotely sensed land measures, including the effect of slope and the feasibility of measurement at the plot level in smallholder farming systems, are important to further improve this literature. From a production perspective, open production measurement questions related to root, tuber and tree crops as well as appropriately attributing intercropped plots for yield estimation are important production systems with remaining measurement challenges. With respect to agricultural output, post-harvest loss remains an active area of measurement research, particularly as it relates not only to food waste, but improved profitability measures. In the livestock sector, livestock labor measurement is even more challenging due to unit of analysis and attribution to particular livestock in both sedentary and pastoralist systems of production. There continues to be much to learn with respect to alternative survey modes and the integration of survey modes, particularly when face-to-face interviews may be limited. As stakeholders have responded to public demands to make nationally representative data publicly available, a concurrent set of privacy and geo-referencing challenges are important to balance against the empirical possibilities facilitated by such data. We underscore that earlier measurement challenges, such as those that were identified after the Ghosh and Glewwe (2000) volumes have led to innovations in agricultural survey design. The continued interaction of not a sole set of stakeholders, but the collaboration between research, policy, national statistical office, and donors will facilitate future innovations in agricultural survey design.





## Chapter 8

## References

- Abay, K.A., Abate, G.T., Barrett, C.B., and Bernard, T. 2019. “Correlated Non-Classical Measurement Errors, ‘Second Best’ Policy Inference and the Inverse Size-Productivity Relationship in Agriculture.” *Journal of Development Economics* 139(C), 171–184.
- Akogun, Oladele, Andrew Dillon, Jed Friedman, Ashesh Prasann, and Pieter Serneels. 2021. “Productivity and Health: Physical Activity as a Measure of Effort,” *World Bank Economic Review*, forthcoming.
- Alkire, S., Meinzen-Dick, R., Peterman, A., Quisumbing, A., Seymour, G., and Vaz, A. 2013. “The Women’s Empowerment in Agriculture Index.” *World Development* 52, 71–91.
- Ambler, K., Doss, C., Kieran, C., and Passarelli, S. 2017. “He says, she says: Exploring Patterns of Spousal Agreement in Bangladesh.” International Food Policy Research Institute.
- Anagol, S., Etang, A., and Karlan, D. 2017. “Continued Existence of Cows Disproves Central Tenets of Capitalism?” *Economic Development and Cultural Change* 65(4), 583-618.
- Antman, F., and McKenzie, D. J. 2007. “Earnings Mobility and Measurement Error: A Pseudo-Panel Approach.” *Economic Development and Cultural Change* 56, 125-161. <https://doi.org/10.1086/520561>
- Arthi, V., Beegle, K., De Weerd, J., and Palacios-López, A. 2018. “Not Your Average Job: Measuring Farm Labor in Tanzania.” *Journal of Development Economics* 130, 160-172.
- Asian Development Bank. 2019. “The CAPI Effect: Boosting Survey Data through Mobile Technology. A Special Supplement of Key Indicators for Asia and the Pacific 2019.” Manila: ADB.

Ashour, M., Gilligan, D.O., Hoel, J.B., and Karachiwalla, N.I. 2019. “Do Beliefs About Herbicide Quality Correspond with Actual Quality in Local Markets? Evidence from Uganda.” *Journal of Development Studies* 55, 1285–1306. <https://doi.org/10.1080/00220388.2018.1464143>

Bardasi, E., Beegle, K., Dillon, A., and Serneels, P. 2011. “Do Labor Statistics Depend on How and to Whom the Questions are Asked? Results from a Survey Experiment in Tanzania.” *World Bank Economic Review* 25, 418–447.

Bautista, R. 2012. “An Overlooked Approach in Survey Research: Total Survey Error.” In *Handbook of survey methodology for the social sciences* (p. 37–49). Springer.

Beaman, L. and Dillon, A. 2012. “Do Household Definitions Matter in Survey Design? Results from a Randomized Survey Experiment in Mali.” *Journal of Development Economics* 98, 124–135. <https://doi.org/10.1016/j.jdeveco.2011.06.005>

Beegle, K., Carletto, C., and Himelein, K. 2012. “Reliability of Recall in Agricultural Data.” *Journal of Development Economics* 98, 34–41.

Benedetti, L., De Baets, B., Nopens, I., and Vanrolleghem, P. A. 2010. “Multi-criteria analysis of wastewater treatment plant design and control scenarios under uncertainty.” *Environmental Modelling and Software* 25, 616–621. <https://doi:10.1016/j.envsoft.2009.06.003>

Benes, Elisa M., and Kieran Walsh. 2018a. “ILO LFS Pilot Studies in Follow Up to the 19th ICLS: Background, Objectives, and Methodology.” Statistical Methodology Series 1 (May). Geneva; International Labour Office. [https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms\\_627873.pdf](https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms_627873.pdf).

Benes, Elisa M., and Kieran Walsh. 2018b. “Measuring Employment in Labour Force Surveys: Main Findings from the ILO LFS Pilot Studies.” Statistical Methodology Series 4 (July). Geneva; International Labour Office. [https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms\\_635732.pdf](https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms_635732.pdf).

Berazneva, J., McBride, L., Sheahan, M., and Güereña, D. 2018. “Empirical Assessment of Subjective and Objective Soil Fertility Metrics in East Africa: Implications for Researchers and Policy Makers.” *World Development* 105, 367–382.

Bevis, L.EM., and Barrett, C.B., 2020. “Close to the Edge: High Productivity at Plot Peripheries and the Inverse Size-Productivity Relationship.” *Journal of Development Economics* 143, 102377.

Bowley, A. L., and Burnett-Hurst, A. R. (1926). *Livelihood and Poverty* (London, 1915). *AL Bowley and Margaret Hogg, Has Poverty Diminished*, 15.

- Brown, J. F., and Pervez, M. S. 2014. "Merging Remote Sensing Data and National Agricultural Statistics to Model Change in Irrigated Agriculture." *Agricultural Systems* 127, 28-40.
- Caeyers, B., Chalmers, N., and De Weerd, J. 2012. "Improving Consumption Measurement and Other Survey Data through CAPI: Evidence from a Randomized Experiment." *Journal of Development Economics* 98, 19-33.
- Carfagna, E., and Gallego, F. J. (2005). "Using Remote Sensing for Agricultural Statistics." *International statistical review* 73(3), 389-404.
- Carletto, C., Aynekulu, E., Gourlay, S., and Shepherd, K. 2017a. Collecting the Dirt on Soils: Advancements in Plot-level Soil Testing and Implications for Agricultural Statistics. World Bank Policy Research Working Paper Series No. 8057. World Bank, Washington, DC.
- Carletto, C., Gourlay, S., Murray, S., and Zezza, A. 2017. "Cheaper, Faster, and More than Good Enough: Is GPS the new gold standard in land area measurement?" *Survey Research Methods* (v11) No. 3: 235-265.
- Carletto, C., Gourlay, S., and Winters, P. 2015. "From Guesstimates to GP-Stimates: Land Area Measurement and Implications for Agricultural Analysis." *Journal of African Economics* 24, 593-628.
- Carletto, G., Himelein, K., Kilic, T., Murray, S., Oseni, M., Scott, K. and Steele, D. 2010. "Improving the Availability, Quality and Policy-Relevance of Agricultural Data: The Living Standards Measurement Study-Integrated Surveys on Agriculture." In *Third Wye City Group Global Conference on Agricultural and Rural Household Statistics*. Washington, DC.
- Carletto, G., Savastano, S., and 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>
- Carletto, G., Gourlay, S., Murray, S., and Zezza, A. 2016. "Land Area Measurement in Household Surveys: Empirical Evidence and Practical Guidance for Effective Data Collection (English)." LSMS Guidebook Washington, D.C.; World Bank Group. <http://documents.worldbank.org/curated/en/606691587036985925/Land-Area-Measurement-in-Household-Surveys-Empirical-Evidence-and-Practical-Guidance-for-Effective-Data-Collection>
- Casaburi, L., and Macchiavello, R. 2019. "Demand and Supply of Infrequent Payments as a Commitment Device: Evidence from Kenya." *American Economic Review* 109, 523-55. <https://doi.org/10.1257/aer.20180281>
- Chambers, R. G. 1988. *Applied Production Economics: a Dual Approach*. Cambridge University Press.
- Chambers, R. G., and Quiggin, J. 2000. *Uncertainty, Production, Choice, and Agency: the State-Contingent Approach*. Cambridge University Press.

- Coates, J.C., Webb, P., Houser, R.F., Rogers, B.L., and Wilde, P. 2010. “He Said, She Said: Who Should Speak for Households about Experiences of Food Insecurity in Bangladesh?” *Food Security* 2, 81–95.
- Converse, J. M., and Presser, S. 1986. Quantitative applications in the social sciences. *Survey Questions: Handcrafting the Standardized Questionnaire*. Newbury Park, CA: Sage Publications, 10.
- Dang, H. L., Li, E., Nuberg, I., and Bruwer, J. 2014. Understanding Farmers’ Adaptation Intention to Climate Change: A Structural Equation Modelling Study in the Mekong Delta, Vietnam. *Environmental Science and Policy* 41, 11–22. <https://doi.org/10.1016/j.envsci.2014.04.002>
- De Groote, H. and Traoré, O. 2005. “The Cost of Accuracy in Crop Area Estimation.” *Agricultural Systems*. 84, 21–38.
- De Heer, W., de Leeuw, E.D., and van der Zouwen, J. 1999. Methodological Issues in Survey Research: a Historical Review.” *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 64, 25–48.
- De Nicola, F., and Giné, X. 2014. “How Accurate are Recall Rata? Evidence from Coastal India.” *Journal of Development Economics* 106, 52–65.
- De Weerdt, J., Beegle, K., and Gibson, J. 2019. “What Can We Learn from Experimenting with Survey Methods?” (No. 41819), LICOS Discussion Papers. LICOS - Centre for Institutions and Economic Performance, KU Leuven.
- Deaton, A. 1985. “Panel Data from Time Series of Cross-Sections.” *Journal of Econometrics* 30, 109–126. [https://doi.org/10.1016/0304-4076\(85\)90134-4](https://doi.org/10.1016/0304-4076(85)90134-4)
- Deaton, A. and Zaidi, S. 2002. *Guidelines for constructing consumption aggregates for welfare analysis*. Washington, DC: World Bank.
- Deininger, K., Carletto, C., Savastano, S., and Muwonge, J. 2012. “Can diaries help in improving agricultural production statistics? Evidence from Uganda.” *Journal of Development Economics* 98, 42–50.
- Desiere, S. 2015. “Area Measurement in Agricultural Surveys: GPS or Compass and Rope?” In S. Desiere (Ed.), “From Raw Numbers to Robust Evidence. Finding Fact, Avoiding Fiction” (Chapter, Area Measurement in Agricultural Surveys: GPS or Compass and Rope?). Doctoral dissertation, Ghent University.
- Desiere, S., Jolliffe, D. 2018. “Land Productivity and Plot Size: Is Measurement Error Driving the Inverse Relationship?” *Journal of Development Economics* 130, 84–98.
- Di Maio, M., and Fiala, N. 2019. “Be Wary of Those Who Ask: A Randomized Experiment on the Size and Determinants of the Enumerator Effect.” *The World Bank Economic Review*, lhy024, <https://doi.org/10.1093/wber/lhy024>
- Dillon, A., Bardasi, E., Beegle, K., and Serneels, P. 2012. “Explaining Variation in Child Labor Statistics.” *Journal of Development Economics*, 98, 136–147. <https://doi.org/10.1016/j.jdeveco.2011.06.002>

- Dillon, A., Gourlay, S., McGee, K., and Oseni, G. 2019. "Land measurement bias and its empirical implications: evidence from a validation exercise." *Economic Development and Culture Change* 67, 595–624.
- Dillon, A., Karlan, D., Udry, C., and Zinman, J. 2020. "Good Identification, Meet Good Data." *World Development* 127, 104796. <https://doi.org/10.1016/j.worlddev.2019.104796>
- Dillon, A. and Edouard Mensah. 2021. "Respondent Biases in Household Surveys" Global Poverty Lab Working paper.
- Dillon, A., and Rao, L.N. 2021. "Land Measurement Bias: Comparisons from Global Positioning System, Self-Reports, and Satellite Data." Global Poverty Lab Working Paper.
- Dixon-Mueller, R. B., and Anker, R. 1988. "Assessing Women's Economic Contributions to Development." World Employment Programme Background Paper 6.
- Donaldson, D., and Storeygard, A. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30, 171–98.
- Doss, C., and Kieran, C. 2014. "Standards for Collecting Sex-Disaggregated Data for Gender Analysis; a Guide for CGIAR Researchers." CGIAR Gender Research Network.
- Doss, C., Kieran, C., and Kilic, T. 2019. "Measuring Ownership, Control, and use of Assets." *Feminist Economics* 0, 1–25. <https://doi.org/10.1080/13545701.2019.1681591>
- Fabregas, R., Kremer, M., and Schilbach, F. 2019. "Realizing the Potential of Digital Development: The Case of Agricultural Advice." *Science* 366. <https://doi.org/10.1126/science.aay3038>
- Fafchamps, M., McKenzie, D., Quinn, S., and Woodruff, C. 2012. "Using PDA Consistency Checks to Increase the Precision of Profits and Sales Measurement in Panels." *Journal of Development Economics* 98, 51–57.
- FAO. 1982. "Estimation of Crop Areas and Yields in Agricultural Statistics." FAO Economic and Social Development Paper, 22. Rome: FAO.
- FAO. 2011. *World Livestock 2011 – Livestock in food security*. Rome: FAO.
- FAO. 2013. *Children's work in the livestock sector: Herding and beyond*. Rome: FAO.
- FAO. 2015a. *World Census of Agriculture 2020. Volume 1: Programme, Concepts and Definitions*. Rome: FAO.
- FAO. 2015b. *Handbook on Master Sampling Frames for Agricultural Statistics*. Rome: FAO
- FAO; The World Bank; UN-Habitat. 2019. *Measuring Individuals' Rights to Land: An Integrated Approach to Data Collection for SDG Indicators 1.4.2*

- and 5.a.1. Washington, DC: World Bank. © FAO, The World Bank, and UN-Habitat. License: CC BY-NC-SA 3.0 IGO
- Fermont, A., and Benson, T. 2011. “Estimating yield of food crops grown by smallholder farmers”. IFPRI Discussion Paper 01097. International Food Policy Research Institute, Washington, DC.
- Fisher, R. 1926. “The Arrangement of Field Experiments.” *Journal of the Ministry of Agriculture* 33, 503-515.
- Fitzgerald, J., Gottschalk, P., and Moffitt, R. 1998. “An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics.” *The Journal of Human Resources* 33, 251–299.
- Fowler Jr, F. J. 1995. *Improving survey questions: Design and evaluation*. Sage.
- Fox, K. A. 1986. “Agricultural Economists as World Leaders in Applied Econometrics, 1917–33.” *American Journal of Agricultural Economics* 68, 381-386. <https://doi.org/10.2307/1241449>
- Friedman, J., Gaddis, I., Kilic, T., Martuscelli, A., Palacios-Lopez, A., and Zezza, A. Forthcoming. “The Distribution of Effort: Physical Activity, Gender Roles, and Bargaining Power in Malawi.”
- Gaddis, I., Siwatu, G. O., Palacios-Lopez, A., and Pieters, J. 2019. “Measuring farm labor: survey experimental evidence from Ghana.” World Bank Policy Research Working Paper 8717. World Bank, Washington, DC.
- Garlick, R., Orkin, K., Quinn, S., 2020. “Call Me Maybe: Experimental Evidence on Frequency and Medium Effects in Microenterprise Surveys. *World Bank Economic Review* 34, 418–443. <https://doi.org/10.1093/wber/lhz021>
- Gideon, L. 2012. *Handbook of Survey Methodology for the Social Sciences*.
- Glennerster, R., Walsh, C., and Diaz-Martin, L. 2018. “A Practical Guide to Measuring Women’s and Girls Empowerment in Impact Evaluation.” <https://www.povertyactionlab.org/practical-guide-measuring-womens-and-girls-empowerment-impact-evaluations>
- Global Strategy to improve Agricultural and Rural Statistics. 2017. *Handbook on the Agricultural Integrated Survey*. AGRIS.
- Goldstein, M. and Udry, C. 1999. *Agricultural innovation and risk management in Ghana*. Unpublished Final Report IFPRI.
- Gourlay, S., Aynekulu, E., Carletto, C., and Shepherd, K. 2017. *Spectral Soil Analysis and Household Surveys*. Washington, DC: World Bank.
- Gourlay, S., Kilic, T., and 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.

- Grace, D., Mutua, F., Ochungo, P., Kruska, R., Jones, K., Brierley, L., and Ogutu, F. 2012. "Mapping of poverty and likely zoonoses hotspots." *Zoonoses Project 4*, 1-119.
- Grace, K., Brown, M., and McNally, A. 2014. "Examining the Link Between Food Prices and Food Insecurity: A Multi-level Analysis of Maize Price and Birthweight in Kenya." *Food Policy* 46, 56-65.
- Grosh, M. and Glewwe, P. 2000. *Designing household survey questionnaires for developing countries*. Washington DC: World Bank.
- Guarcello, L., Lyon, S., and Rosati, F. C. 2008. "Child Labor and Education for All: An Issue Paper." *The journal of the History of Childhood and Youth* 1, 254-266.
- Heffetz, O., and Ligett, K. 2014. "Privacy and Data-Based Research." *Journal of Economic Perspectives* 28(2), 75-98.
- Hengl, T., Heuvelink, G.B., Kempen, B., Leenaars, J.G., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Mendes de Jesus, J., Tamene, L. and Tondoh, J.E. 2015. "Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions." *PloS one*, 10(6), e0125814.
- Herberich, D. H., Levitt, S. D., and List, J. A. 2009. "Can Field Experiments Return Agricultural Economics to the Glory Days?" *American Journal of Agricultural Economics* 91, 1259-1265. <https://doi.org/10.1111/j.1467-8276.2009.01294.x>
- Hess, J., Moore, J., Pascale, J., Rothgeb, J., and Keeley, C. 2001. "The Effects of Person-Level versus Household-Level Questionnaire Design on Survey Estimates and Data Quality." *Public Opinion Quarterly* 65, 574-584. <http://doi:10.1086/323580>
- Hill, Z. 2004. "Reducing attrition in panel studies in developing countries." *International Journal of Epidemiology* 33, 493-498. <http://doi:10.1093/ije/dyh060>
- Himelein, K. 2016. "Interviewer Effects in Subjective Survey Questions: Evidence from Timor-Leste." *International Journal of Public Opinion Research* 28(4): 511-533
- Hoddinott, J., Headey, D., and Dereje, M. 2015. "Cows, Missing Milk Markets, and Nutrition in Rural Ethiopia." *The Journal of Development Studies* 51, 958-975.
- Holden, S., Ali, D., Deininger, K., and Hilhorst, T. 2016. A Land Tenure Module for LSMS. Centre for Land Tenure Studies Working Paper 1/16. Norwegian University of Life Sciences.
- Husak, G., and Grace, K. 2016. "In Search of a Global Model of Cultivation: Using Remote Sensing to Examine the Characteristics and Constraints of Agricultural Production in the Developing World." *Food Security* 8.

- Iarossi, G. 2006. *The Power of Survey Design : A User's Guide for Managing Surveys, Interpreting Results, and Influencing Respondents*. Washington, DC: The World Bank.
- Jacobs, K., Kes, A. 2015. "The Ambiguity of Joint Asset Ownership: Cautionary Tales from Uganda and South Africa." *The Feminist Economics* 21, 23–55.
- Jain, Meha. 2020. "The Benefits and Pitfalls of Using Satellite Data for Causal Inference". *Review of Environmental Economics and Policy* 14(1): 157-169.
- Janzen, S.A. 2018. "Child Labour Measurement: Whom Should We Ask?" *International Labour Review* 157, 169–191.
- Just, R. E., and Pope, R. D. 2001. "The agricultural producer: Theory and statistical measurement." Chapter 12 of *Agricultural Production* (Vol. 1, p. 629-741). Elsevier. [https://doi.org/10.1016/S1574-0072\(01\)10015-0](https://doi.org/10.1016/S1574-0072(01)10015-0)
- Keita, N. and Carfagna, E. 2009. "Use of Modern Geo-Positioning Devices in Agricultural Censuses and Surveys: Use of GPS for Crop Area Measurement." Bulletin of the International Statistical Institute, the 57th Session, 2009, Proceedings, Special Topics Contributed Paper Meetings (STCPM22), Durban.
- Kilic, T., Koolwal, G.B., and Moylan, H.G. 2020. "Are You Being Asked? Impacts of Respondent Selection on Measuring Employment." World Bank Policy Research Working Paper 9152. World Bank, Washington, DC.
- Kilic, T., Moylan, H., and Koolwal, G. 2020b. "Getting the (Gender-Disaggregated) Lay of the Land: Impact of Survey Respondent Selection on Measuring Land Ownership and Rights." World Bank Policy Research Working Paper 9151. World Bank, Washington, DC.
- Kilic, T., and Moylan, H. 2016. "Methodological Experiment on Measuring Asset Ownership from a Gender Perspective (MEXA)." World Bank, Washington, DC.
- Kilic, T., Moylan, H.G., Ilukor, J., Mtengula, C., and Pangapanga-Phiri, I. 2018. "Root for the Tubers: Extended-Harvest Crop Production and Productivity Measurement in Surveys." World Bank Policy Research Working Paper 8618. World Bank, Washington, DC.
- Kilic, T., Serajuddin, U., Uematsu, H., and Yoshida, N. 2017. "Costing household surveys for monitoring progress toward ending extreme poverty and boosting shared prosperity." World Bank Policy Research Working Paper 7951. World Bank, Washington, DC.
- Kilic, T., Van den Broeck, G., Koolwal, G., and Moylan, H. 2020a. "Are You Being Asked? Impacts of Respondent Selection on Measuring Employment." World Bank Policy Research Working Paper 9152. World Bank, Washington, DC.



- Kilic, T., Zezza, A., Carletto, C., and Savastano, S. 2017). “Missing (ness) in Action: Selectivity Bias in GPS-based Land Area Measurements.” *World Development* 92, 143-157.
- Kosmowski, F., Aragaw, A., Kilian, A., Ambel, A., Ilukor, J., Yigezu, B. and Stevenson, J. 2018. “Varietal Identification in Household Surveys: Results from Three Household-based Methods against the Benchmark of DNA Fingerprinting in Southern Ethiopia”, *Experimental Agriculture* 55, pp. 371–385.
- Krosnick, J. A. 1999. “Survey research.” *Annual review of psychology*, 50, 537–567.
- Laajaj, R., and Macours, K. 2019. “Measuring Skills in Developing Countries.” *Journal of Human Resources*, 1018-9805R1.
- Lobell, D. B., Cassman, K. G., and Field, C. B. 2009. “Crop Yield Gaps: Their Importance, Magnitudes, and Causes.” *Annual Review of Environment and Resources*, 34, 179-204. <http://doi:10.1146/annurev.enviro.041008.093740>
- Lobell, D. B., Di Tommaso, S., You, C., Yacoubou Djima, I., Burke, M., and Kilic, T. 2020. “Sight for Sorghums: Comparisons of Satellite-and Ground-Based Sorghum Yield Estimates in Mali.” *Remote Sensing*, 12(1), 100.
- Lobell, D., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., and Murray, S. 2019. “Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis.” *American Journal of Agricultural Economics* aaz051, <https://doi.org/10.1093/ajae/aaz051>
- Manski, C. F., and Molinari, F. 2008. “Skip Sequencing: A Decision Problem in Questionnaire Design.” *The Annals of Applied Statistics* 2, 264–285.
- Michelson, H., Fairbairn, A., Ellison, B., Maertens, A., and Manyong, V. 2021. “Misperceived Quality: Fertilizer in Tanzania.” *Journal of Development Economics* 148, 102579. <https://doi.org/10.1016/j.jdeveco.2020.102579>
- Moore, J. C. 1988. “Self/proxy Response Status and Survey Response Quality.” *Journal of Official Statistics* 4, 155–172.
- O’Sullivan, M., Rao, A., Raka, B., Kajal, G., and Margaux, V. 2014. *Levelling the Field: Improving Opportunities for Women Farmers in Africa*. Washington, DC: World Bank.
- Oseni, G., Durazo, J., and McGee, K. 2017. The use of non-standard measurement units for the collection of food quantity: a guidebook for improving the measurement of food consumption and agricultural production in living standards surveys. LSMS Guidebook, Washington, DC: World Bank.
- Palacios-Lopez, A., Christiaensen, L., Kilic, T. 2017. “How much of the labor in African agriculture is provided by women?” *Food Policy* 67, 52–63. <https://doi.org/10.1016/j.foodpol.2016.09.017>
- Payne, S. L. 2014. *The Art of Asking Questions: Studies in Public Opinion*, 3 (Vol. 451). Princeton University Press.

- Pica-Ciamarra, U., Baker, D., Morgan, N., Zezza, A., Azzarri, C., Ly, C., Nsima, L., Nouala, S., Okello, and P., Sserugga, J. 2014. “Investing in the livestock sector: Why Good Numbers Matter, a Sourcebook for Decision Makers on How to Improve Livestock Data.” World Bank.
- Rozelle, S. 1991. *Rural household data collection in developing countries: Designing instruments and methods for collecting farm production data*. Technical report. World Bank: Washington, DC.
- Sagesaka, A., Palacios-Lopez, A., and Amankwah, A. 2020. *Measuring Work on Household-Farms using Household Surveys. A Practical Guidebook on Designing Household Surveys for Effective Data Collection of Work on Household Farms*. Technical report. World Bank: Washington, DC.
- Schöning, P. 2005. *Handheld GPS Equipment for Agricultural Statistics Surveys: Experiments on Area-Measurements Done during Fieldwork for the Uganda Pilot Census of Agriculture, 2003*. Statistisk sentralbyrå.
- Schündeln, M. 2018. “Multiple Visits and Data Quality in Household Surveys.” *Oxford Bulletin of Economics and Statistics* 80, 380–405.
- Serneels, P., Beegle, K., and Dillon, A. 2017. “Do returns to education depend on how and whom you ask?” *Economics of Education Review* 60, 5–19.
- Singh, I., Squire, L., and Strauss, J. 1986. *Agricultural household models: Extensions, applications, and policy*. Washington, DC: World Bank.
- Sudman, S., and Bradburn, N.M. 1973. “Effects of Time and Memory Factors on Response in Surveys.” *Journal Of American Statistics Association* 68, 805–815.
- Thomas, D., Witoelar, F., Frankenberg, E., Sikoki, B., Strauss, J., Sumantri, C., and Suriastini, W. 2012. “Cutting the Costs of Attrition: Results from the Indonesia Family Life Survey.” *Journal of Development Economics*, 98, 108–123. <https://doi.org/10.1016/j.jdeveco.2010.08>
- Udry, C. 1996. “Gender, Agricultural Production, and the Theory of the Household.” *Journal of Political Economy*, 104(5), 1010–1046.
- United Nations Statistics Division. 2019. *Guidelines for Producing Statistics on Asset Ownership from a Gender Perspective*. [https://unstats.un.org/edge/publications/docs/Guidelines\\_final.pdf](https://unstats.un.org/edge/publications/docs/Guidelines_final.pdf)
- Wineman, A., Anderson, C. L., Reynolds, T., and Biscaye, P. 2019. “Methods of Crop Yield Measurement on Multi-Cropped Plots: Examples from Tanzania,” *Food Security* 11, 1257–1273. <https://doi.org/10.1007/s12571-019-00980-5>
- Willis, K. J., Seddon, A. W., Long, P. R., Jeffers, E. S., Caithness, N., Thurston, M., and Macias-Fauria, M. 2015. “Remote Assessment of Locally Important Ecological Features across Landscapes: how Representative of Reality?” *Ecological Applications* 25, 1290–1302. <https://doi.org/10.1890/14-1431.1>

Wineman, A., Njagi, T., Anderson, C.L., Reynolds, T.W., Alia, D.Y., Wainaina, P., Njue, E., Biscaye, P. and Ayieko, M.W. 2020. "A Case of Mistaken Identity? Measuring Rates of Improved Seed Adoption in Tanzania using DNA Fingerprinting." *Journal of Agricultural Economics*, 71(3), 719-741.

Witoelar, F. 2011. *Tracking in Longitudinal Household Surveys*. Washington, DC: The World Bank, Development Research Group, LSMS-ISA Integrated Surveys on Agriculture.

Wollburg, P., Tiberti, M., and Zezza, A. 2020. "Recall Length and Measurement Error in Agricultural Surveys." *Food Policy* 102003. <https://doi.org/10.1016/j.foodpol.2020.102003>

Working, H. 1925. "The Statistical Determination of Demand Curves." *The Quarterly Journal of Economics* 39, 503-543. <https://doi.org/10.2307/1883264>

Zeza, A., Federighi, G., Kalilou, A. A., and Hiernaux, P. 2016a. "Milking the Data: Measuring Milk Off-Take in Extensive Livestock Systems. Experimental Evidence from Niger." *Food policy* 59, 174-186.

Zeza, A, Pica-Ciamarra, U, Mugera, HK, Mwisomba, T, and Okello, P. 2016b. *Measuring the Role of Livestock in the Household Economy: A Guidebook for Designing Household Survey Questionnaires (English)*. LSMS Guidebook Washington, D.C. World Bank. <http://documents.worldbank.org/curated/en/108351587037911099/Measuring-the-Role-of-Livestock-in-the-Household-Economy-A-Guidebook-for-Designing-Household-Survey-Questionnaires>

Zwane, A.P., Zinman, J., Dusen, E.V., Pariente, W., Null, C., Miguel, E., Kremer, M., Karlan, D.S., Hornbeck, R., Giné, X., Duflo, E., Devoto, F., Crepon, B., and Banerjee, A. 2011. Being surveyed can change later behavior and related parameter estimates. *Proceedings of the National Academy of Sciences* 108, 1821–1826. <https://doi.org/10.1073/pnas.1000776108>



# Appendix I: Agricultural Reference Questionnaire

Refer to attachment.

[\[#{#\\_Toc70494919 .anchor}\]](#)



# Appendix II: Livestock Reference Questionnaire

Refer to attachment which is extracted from Zezza et al. (2016b).





## Appendix III: Glossary

Term	Description
Administrative data	Data generated as a by-product of registration and other administrative and service delivery functions of government entities or other organizations. Examples include registries of large farms or information collected through agricultural extension services.
Agricultural season	Agricultural season refers to the portion of a year in which seasonal crops are typically grown in a given locale. This is dictated by local climatic conditions. In agricultural survey instruments, the 'reference agricultural season' is used to demarcate and refer to the production cycle of interest.
Agro-forestry	Land use system in which trees or shrubs are deliberately grown on a plot of land together with crops and animals.

Term	Description
Area sampling	Sampling method used when no complete sampling frame is available, based on dividing the land area under study into smaller areas and sampling from the list of those smaller areas.
Attrition	In longitudinal surveys, attrition refers to the loss of survey participants (households, holdings, individuals) over time, with each subsequent round of the survey.
Bias	In statistics, an estimator is biased when its expected value differs from the true value of the underlying parameter.
Census	The full enumeration or count of an entire population of interest, such as farms or agricultural holdings (agricultural census) or households (population and housing census).
Compass and rope	A method to measure the area of a unit of land reliably using poles, ropes, compasses.
Computer Assisted Personal Interviewing (CAPI)	A survey mode in which the survey questionnaire is pre-programmed on a tablet or computer and administered to the respondent in person by an interviewer.

Term	Description
Consumption aggregate	Total value of items consumed by the household over a given reference period, common welfare metric used widely to determine whether a household is considered poor.
Crop cutting	Crop cutting is a more objective measurement method for crop production than farmer recall, considered the gold standard for measuring crop production and yields. Randomly located subplots within a given plots are harvested and weighted before and after drying.
Crop rotation	Farming method whereby crops on a plot are grown one after the other with the aim of maintaining fertile soil on the plot.
Crop Yield	Quantity of crop obtained per unit of land area used.
Data collection mode	Data collection mode or survey mode is the method used to collect the data. Examples include face-to-face, telephone, mail, and web-based surveys.
Enumerator	Interviewer in surveys.
External validity	Extent to which results from a study or experiment can be generalized beyond its original context.
Geospatial data	Data related by geographic location, such as maps, satellite or remote sensing data.

Term	Description
Gold standard	In survey design, a method for collecting data on a given parameter that is regarded as yielding highly reliable data.
Head of the household	Individual household member who is (considered by other household members) the main decision maker in a given household.
Holding	Economic unit of agricultural production under single management, comprising all livestock kept and all land used wholly or partly for agricultural production purposes, without regard to title, legal form, or size (FAO, WCA 2020).
Household	A household usually refers to a group of individuals who eat meals together and live in the same dwelling. In addition, according to the production-based definition of the household, it consists of adult individuals co-mingle revenues from agricultural production and non-farm enterprise for consumption. The production-based definition is used less frequently.
Household member	Individual who is a member of a given household.
Intercropping	Cultivation of more than one temporary and/or permanent crop simultaneously on the same plot.
Internal validity	Extent to which a causal effect is soundly identified in a study or experiment.

Term	Description
Land tenure	Land tenure defines the property and use rights of an individual or group with respect to land.
Longitudinal survey	A panel or longitudinal survey is a survey in which the same survey units (households, holdings, individuals) are re-visited for at least one round after the baseline survey round.
Measurement error	Deviation of a measured value of a parameter from its true value.
Multi-topic household survey	Sample survey with the household as its main unit of analysis and that studies multiple issues affecting the household, such as education, employment, poverty, migration, among others.
National Statistical Office (NSO)	Government agency in charge of producing official statistics.
Nonresponse	In surveys, nonresponse refers to selected respondents altogether failing to participate in the survey (unit nonresponse) or failing to respond to certain questions of the survey (item nonresponse).
Non-sampling error	Survey error not due to sampling, such as response errors, when survey respondents misreport on survey questions for example because survey questions are poorly framed, or errors resulting from nonresponse.

Term	Description
Panel survey	A panel or longitudinal survey is a survey in which the same survey units (households, holdings, individuals) are re-visited for at least one round after the baseline survey round.
Paper Assisted Personal interviewing (PAPI)	Survey interview mode whereby the enumerator uses pen and paper to fill the questionnaire.
Parcel	A parcel is a piece of land of one land tenure type entirely surrounded by other land, water, road, forest or other features not forming part of the holding, or forming part of the holding under a different land tenure type. A parcel may comprise one or more plots.
Pastoralism	Livestock production system involving the grazing of livestock herds on rangelands and large pastures typically practiced by nomadic people who move with their herds.
Permanent crops	Crops with a growing cycle of more than one year (FAO, WCA 2020), and which do not need to be replanted after one growing cycle. Also referred to as perennial crops.
Plot	A plot is defined as a continuous piece of land on which a specific crop or a mixture of crops is grown, or which is fallow or waiting to be planted.

Term	Description
Plot manager	The manager of an agricultural plot is typically the main decision maker regarding crops planted, inputs applied, and output harvested.
Post-harvest questionnaire	Questionnaire administered during the post-harvest visit of the survey, that is, a survey visit taking place after the farm's main harvest of temporary/seasonal crops has been completed.
Post-planting questionnaire	Questionnaire administered during the post-planting visit of the survey, that is a survey visit taking place after the farm has completed the planting period of its main temporary/seasonal crops.
Productivity	In this volume, productivity refers to agricultural productivity, whether yield (output per area of land), labor productivity (output per unit of labor or per worker), or total factor productivity, taking into account all inputs and outputs.
Proxy response	In proxy response, an individual responds on behalf another individual, rather than all individuals responding directly for themselves.
Randomized controlled trial	Scientific experiment to test the causal effect of a treatment or an intervention while controlling for confounding factors through experimental design.
Recall	Recall refers to the process of respondents self-reporting of past events of interest.

Term	Description
Recall period	Length of the period between the interview and an event the respondent is requested to recall.
Reference period	The period of reference for a survey question, for example, the reference agricultural season or the last 7 days.
Representativeness	A sample is representative if it accurately reflects the study population of interest with respect to a set of predefined characteristics.
Research design	Methods, techniques, and strategies chosen to answer a given research question.
Root crops	Root crops or root vegetables are those crops whose roots are meant for consumption, such as cassava or onions.
Sample survey	Survey based on a sample of the entire population of interest, rather than full enumeration, usually designed to be representative of the population of interest.
Sampling	Selecting a predetermined number of units from a population of interest, such as farms or households.
Sampling error	Deviation of sample means from true population means leading to reduced representativeness of the sample.
Sampling frame	A list of units, such as farms, from which a survey sample is drawn.



Term	Description
Sex-disaggregated data	Data that allows distinguishing between women and men at the level of the individual.
Skip sequences/patterns	In survey questionnaires, a skip pattern routes the respondent to a specific question based on his or her answer to a previous question.
Smallholder	A smallholder or smallholder farmer is a person who owns and/or operates a small-scale agricultural holding, whether defined by land area, production volume, production technology, or other factors.
Social desirability bias	Social desirability bias refers to a response bias whereby survey respondents answer survey questions in a way they perceive as pleasing the interviewer.
Survey design	Methods and techniques for developing and implementing a survey.
Survey experiment	Scientific experiment to test the impact of survey design on measurement through surveys.
Survey methodology	The study of survey methods.
Survey respondent	Individual who responds to survey questions.
Temporary crops	Crops with at least one growing cycle (planting and harvesting) per agricultural year, often once per reference agricultural season.

Term	Description
Total survey error	The sum of all errors from design, implementation, data processing and analysis of surveys.
Tropical livestock units	A standardized unit for measuring the stock of live animals independent of their breed or size.
Unit of analysis	The unit of analysis or unit of observation is the unit of interest of a study or survey, such as farms, individuals, households, or enterprises. The unit of analysis may vary in one survey or study.
Welfare	Measure of material wellbeing.