

# Measuring Farm Labor

## Survey Experimental Evidence from Ghana

*Isis Gaddis*

*Gbemisola Oseni*

*Amparo Palacios-Lopez*

*Janneke Pieters*



**WORLD BANK GROUP**

Development Economics  
Development Data Group  
January 2019

## Abstract

This study examines recall bias in farm labor by conducting a randomized survey experiment in Ghana. Hours of farm labor obtained from a recall survey conducted at the end of the season are compared with data collected weekly throughout the season. The study finds that the recall method overestimates farm labor per person per plot by about 10 percent, controlling for observable differences at baseline. Recall bias in farm labor per person per plot is accounted for by the fact that households in the recall group

report fewer marginal plots and farm workers, denoted here as listing bias. This listing bias also creates a countervailing effect on hours of farm labor at higher levels of aggregation, so that the recall method underestimates farm labor per plot and per household and overestimates the labor productivity of household-operated farms. Consistent with the notion that recall bias is linked to the cognitive burden of reporting on past events, the study finds that recall bias in farm labor has a strong educational gradient.

---

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

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

# Measuring Farm Labor: Survey Experimental Evidence from Ghana<sup>+</sup>

Isis Gaddis  
World Bank and IZA

Gbemisola Oseni  
World Bank

Amparo Palacios-Lopez  
World Bank

Janneke Pieters  
Wageningen University and IZA

**JEL:** C8, J22, O12, Q12

**Keywords:** Recall bias, measurement error, farm labor, agricultural productivity, gender, Ghana

<sup>+</sup> This study received funding from the U.K. Department for International Development (DFID) under the “Minding the (Data) Gap: Improving Measurements of Agricultural Productivity through Methodological Validation and Research” project and from the William and Flora Hewlett Foundation under the “Improving the Measurement of Subsistence Agriculture in the Framework of ICLS 19 and Its Implications for Gender Statistics” project. The authors gratefully acknowledge the comments from participants at the GLM|LIC Research Network Conference, the Conference of the Centre for the Study of African Economies, the International Conference on Globalization and Development, the Royal Economic Society Annual Conference, the World Statistics Congress and seminars at the ETH Zurich and the World Bank. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. The authors may be contacted at [igaddis@worldbank.org](mailto:igaddis@worldbank.org) or [apalacioslopez@worldbank.org](mailto:apalacioslopez@worldbank.org).

## 1. Introduction

Agriculture plays a key role in the economies of Sub-Saharan Africa. Across the continent, the sector contributes approximately 18 percent to GDP and accounts for 56 percent of the employed population (World Bank 2017, ILO 2016). Agriculture serves as a key livelihood strategy for many poor families in rural areas, where labor-intensive, smallholder farming is the predominant source of income. At the macro-level, agricultural growth has been found to be more effective than non-agricultural growth in reducing extreme poverty (Christiaensen, Demery and Kuhl 2011).

Designing policies to improve living conditions of smallholder farmers requires data on outputs and inputs. One of the most important inputs into agriculture in developing countries is the labor provided by family members on their own farms, denoted in this paper as *farm labor* and commonly measured in hours per person or hours per person per plot over an extended reference period, such as the last season or the last year. Data on farm labor are important for a range of literature strands in development economics – such as analyses of agricultural productivity (e.g. Restuccia and Santaaulia-Llopis 2017; Gollin and Udry 2017), agricultural household models (e.g. LaFave and Thomas 2016), rural labor markets (Dillon et al 2017) and gender differences in agriculture (O’Sullivan et al 2014; McCarthy et al 2016).

Measuring farm labor is, however, fraught with empirical difficulties. The most common approach is to ask survey respondents to recall the amount of time each member of the household spent on farm activities during the previous agricultural season (*end-of-season recall*).<sup>1</sup> While this extended reference period minimizes the impact of seasonality, it can lead to difficulties for respondents to remember how much time various members of the household worked on the farm over the entire season. Recall is further hindered by the informal nature of smallholder agriculture, where working hours are highly irregular, with periods of lower labor activity after planting and prior to harvest. In addition, since irrigation is limited and rain-fed agriculture the norm across much of Africa, there is added unpredictability with the reliance on weather. These features render measurement of farm labor in developing countries extraordinarily challenging.

While several recent papers highlight shortcomings in agricultural statistics in developing countries (Fermont and Benson 2011; Gollin, Lagakos and Waugh 2014a,b; de Janvry, Sadoulet and Suri 2016), there are very few empirical studies that evaluate the reliability of survey-based measures of agricultural labor. An exception is Arthi et al (2018), who show that end-of-season recall overestimates farm labor per person per plot by more than 200 percent in the Mara region of Tanzania. Their study also documents competing forms of recall bias, as end-of-season recall simultaneously leads to underreporting of cultivated plots and household farm workers. Because of these two counteracting effects, recall bias disappears if hours are aggregated to the household level.

Using a similar study design, we conducted a survey experiment in the Ashanti and Brang Ahafo regions of Ghana over the 2015-16 rainy season. One group of farmers was interviewed weekly about farm labor for each of the preceding seven days, what we consider the resource-intensive benchmark. Another group was interviewed at the end of the agricultural season (i.e. after harvest) about farm labor for the entire

---

<sup>1</sup> Variants of this approach are used, for example, by the surveys conducted under the umbrella of the World Bank’s Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) initiative.

season, hence using the traditional end-of-season recall method. By comparing these two groups we obtain an estimate of bias in the end-of-season recall of farm labor, what we denote as *recall bias*.

This paper makes the following contributions. First, we document that recall bias in farm labor is much lower in the Ghanaian than Tanzanian study context, which suggests important regional heterogeneity and exercising caution in extrapolating estimates of measurement bias across studies and country contexts. Second, we show that recall bias in farm labor per person per plot, which is the unit at which labor is reported by the survey respondents, is accounted for by recall bias in listings of plots and farm workers, what we denote as *listing bias*. This is an important refinement to the conclusions reached by Arthi et al (2018), who document that bias in listings of plots and household workers runs counter to recall bias in farm labor at the person-plot level. Our results show that recall households report more hours of farm labor per person per plot *because* they fail to list ‘marginal’ plots and farm workers. Listing bias hence not only counteracts, but *explains* recall bias in farm labor at the person-plot level. Third, our findings show that recall bias declines with the household’s level of education. This finding is consistent with the notion that recall bias is linked to the cognitive burden of reporting on past events. This educational gradient – together with variations in the design of the experiment – may also explain (at least in part) the difference in results between Arthi et al (2018) and this paper, given that education levels among farmers are higher in the Ghanaian districts covered by this study than in the Mara region of Tanzania.

The findings in this paper have several implications. In general, the presence of recall bias suggests that academics and policy makers ought to tread carefully when analyzing the state of agriculture, as biased data can lead to misguided policies. This paper, however, documents that the magnitude of recall bias depends on the level of aggregation and the characteristics of the population under investigation, which has implications for within- and cross-country comparisons. For example, the analysis in section 5 of this paper suggests that end-of-season recall significantly (both economically and statistically) overestimates measures of agricultural labor productivity. Finally, this study highlights the importance of investing in quality data to support evidence-based policy making and of periodically examining the reliability of such data.

The remainder of the paper is structured as follows. Section 2 provides an overview of the measurement of farm labor. Section 3 describes the experimental design of the study. Section 4 estimates recall bias in farm labor and explores proxy determinants, different levels of aggregation and heterogeneity across sub-populations. It also compares the results in Ghana to those obtained by Arthi et al (2018) for Tanzania and puts forth some tentative explanations for why they differ. Section 5 documents implications of recall bias in farm labor for the analysis of agricultural productivity. Section 6 concludes.

## **2. Measuring farm labor**

### **2.1. Overview**

Measuring labor in smallholder farming is inherently complex and prone to measurement error.<sup>2</sup> Like informal household enterprises in other sectors, smallholder farmers rarely keep any records, so that respondents need to rely on recall strategies to report on labor and other inputs. The seasonality and irregularity of smallholder agriculture pose additional challenges, especially if the objective is to obtain measures of labor for the entire season.

Current practices on how to collect data on farm labor, that is the labor provided by family members on their own farm (excluding hired labor), differs widely across surveys. Agricultural surveys, such as those conducted under the World Bank's Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) initiative, typically collect data on farm labor in a dedicated agricultural module.<sup>3</sup> However, these modules differ with respect to the unit of data collection, recall period, sequencing of questions, selection of respondent(s), etc. The following paragraphs elaborate on the first two of these issues, that is the unit of data collection and the recall period, which are important features for the analysis in this paper.

### **2.2. Unit of data collection**

Most (though not all) agricultural surveys use the plot as the main unit of data collection (FAO 2017). This reflects long-standing guidance for the design and implementation of household surveys in developing countries. Reardon and Glewwe (2000) recommend that in countries with 'hard-to-survey' agriculture, which includes most smallholder farming in Sub-Saharan Africa, data on farm inputs and practices is preferably collected at the plot-level (as opposed to the farm- or household-level). This is because a single farm may consist of multiple plots – cultivated under different production systems – and farmers tend to refer to each plot as they describe agricultural production activities. In such a situation, plot-level data collection reflects the natural flow of conversation between the interviewer and the farmer. In addition, collecting data at the plot level yields more observations than collecting data at the household-level, enables the analyst to control for land quality, and facilitates the analysis of agricultural productivity (as the plot can be used to link crops to inputs) and intra-household differences (e.g. productivity gaps between male- and female-managed plots).

For data on farm labor, recommended practice is to collect such data either in terms of aggregate person-days per plot, or to ask, specifically, about each household member's labor input on each plot (Reardon and Glewwe 2000). The former approach is used in the 2009 Ghana Socio-Economic Panel Survey, which asks, separately for men, women and children, about the number of days, average hours per day and average

---

<sup>2</sup> This section draws on Arthi et al (2018).

<sup>3</sup> Multi-topic household surveys typically also administer a general labor module, which gathers information on different types of employment and work (e.g. farm work, non-farm household enterprise work, wage work, etc.) for each household member. These person-level measures of farm labor typically refer to the past 7 days and cannot easily be extrapolated to the entire agricultural season.

number of workers who provided family labor on each plot during the last major/minor season.<sup>4</sup> An example of the latter approach is the 2014/15 Tanzania National Panel Survey (NPS), which asks “during the long rainy season 2014, how many days did [NAME] spend on the following activities on this [PLOT]?” This is followed by a question about the “typical number of hours per day worked by an individual on this [PLOT]”. Many other agricultural surveys follow variants of this latter approach,<sup>5</sup> including the 2015/16 Ghana agricultural labor survey used in this study.

A major advantage of collecting data on farm labor by person and by plot (denoted here as *person-plot level*) is that it accommodate different analytic approaches at the analysis stage. Table A1 (appendix) shows for several recent studies the main unit at which farm labor is reported and/or embedded in the empirical analysis. Studies concerned with gender differences in agriculture tend to aggregate at the plot-level (sometimes with normalization for plot size), because this allows to compare levels of productivity between plots managed by male and female farmers and assess the role of labor shortages as one of the underlying constraints (Aguilar et al. 2016; Ali et al. 2016; Fisher and Kandiwa 2014; Kilic et al. 2015; McCarthy et al. 2016; Oseni et al. 2016; O’Sullivan et al. 2014; Palacios-Lopez and Lopez 2015). Other analyses of agricultural productivity aggregate either at the plot-level (Gollin and Udry 2017) or at the household-level (Andre et al. 2017; Dillon et al. 2017; Restuccia and Santaaulalia-Llopis 2017). McCullough (2017) aggregates farm labor at the person-level to compare labor productivity between farm and non-farm household enterprises. This illustrates that different levels of aggregation serve different purposes. Collecting data on farm labor at the person-plot level keeps all higher levels of aggregation – i.e. *plot-, person- and household-levels* – open for analysis.

Despite these advantages, data collection at the person-plot level can be challenging in practice. In a first step, enumerators need to compile a comprehensive list (roster) of household members and agricultural plots. Plot rosters, in particular, are notoriously difficult to compile. While most farmers have an intuitive understanding of what constitutes a plot of land, their understanding does not always align with the survey-definition of a plot<sup>6</sup> and they may forget to mention marginal plots. Enumerators hence need to probe carefully to ensure that all relevant plots are listed as per the standardized definition used in the survey. Another difficulty is that respondents may not find the concept of farm labor per person per plot intuitively meaningful, and hence face difficulties in reporting on the number of days and hours that individual household members worked on each plot. These challenges can be further compounded by the length of the recall period, as discussed in the next section.

In this paper, the principal unit of data collection and analysis is farm labor per person per plot. However, we also report on higher levels of aggregation, i.e. farm labor per plot, per person and per household, which are, as shown in Table A1, more widely used at the analysis stage. This allows us to

---

<sup>4</sup> The Ghana Living Standards Survey (GLSS) 5 uses an even more aggregated approach, asking about the number of male, female and children’s hours of family labor on each plot during the last major/minor seasons.

<sup>5</sup> Other examples are multiple rounds of nationally representative LSMS-ISA surveys in Burkina Faso, Ethiopia, Malawi, Mali, Niger, Nigeria, Tanzania and Uganda (World Bank 2015) or specialized surveys with a focus on agriculture (e.g. Walker 2009 - Ghana, Winowiecki et al. 2016 – Tanzania, Uganda). Another example would be the Agricultural Integrated Survey (AGRIS) designed by the Food and Agriculture Organization of the United Nations (FAO). They ask the holding about the total number of months worked within the last N months, followed by questions on the average number of days per month and average number of hours per day. These questions are at the agricultural holding level and asked to every household member.

<sup>6</sup> This definition may be, for example, a “continuous piece of land on which a unique crop or a mixture of crops is grown, under a uniform, consistent crop management system” (World Bank 2016).

gauge the potential direction of bias in different literature strands. Using the person-plot level as a starting point has two main advantages. First, to provide guidance to survey designers, it appears useful to assess recall bias directly at the unit at which the data are reported by the respondents. Second, this approach enables us to further disentangle the proximate causes of bias at higher levels of aggregation – i.e. incomplete listings of plots and household members (recall bias in listings or, simply, *listing bias* – i.e. bias at the extensive margin) and recall bias in terms of days and hours of farm labor at the person-plot level, conditional on being listed (i.e. bias at the intensive margin).

### 2.3 Recall bias

Most agricultural surveys ask respondents to report on farm labor over an extended reference period, such as the previous agricultural season. An example of this approach is the Tanzania NPS introduced in the previous section, which asked respondents to recall each household member’s labor input during the “long rainy season 2014”. The main advantage of using an extended reference period is that it minimizes the impact of seasonality and assures consistency between measures of farm labor and measures of seasonal crop production. This consistency is important when data on agricultural output are linked to data on family labor, for example to analyze agricultural labor or total factor productivity. The flip side of using an extended reference period is that it increases the burden on the respondent, by requiring them to recall events that occurred several months in the past and to perform complex mental calculations to arrive at averages.

Different cognitive processes suggest that recall bias may be linked to the length of recall period (see Arthi et al 2018 for further discussion). The most obvious explanation is a simple failure of memory, where respondents forget to report events that took place long ago. However, there is also evidence from cognitive research that the length of the recall period may have a bearing on how respondents interpret survey questions. Schwarz (2007), for example, shows that the longer the recall period, the more likely respondents are to interpret a given question as referring to major events only. This potential change in inferred meaning makes it difficult to distinguish between the effects of question interpretation and forgetting.

Several studies have found a positive association between the level of education – or direct measures of cognitive skills – and recall ability. Peters (1988), comparing lifecycle data from a retrospective marital history with panel reports for the same individuals in the United States, shows that more educated respondents report more consistently on marital events. McAuliffe, DiFranceisco and Reed (2010) find that arithmetic skills predict the accuracy of retrospective self-reports of sexual activity (also in the US). While evidence from the United States may not necessarily extend to developing countries, given potentially large differences in the quality of education systems, survey methods, etc., a few studies have found similar associations in Asia and Africa. Becket et al (1999, 2001) show that the reliability of reports in the Malaysian Family Life Surveys is higher for more educated respondents. Abebe (2013) finds that recall errors in sales revenues and output among Ethiopian shoemakers are negatively correlated with respondents’ years of schooling. Given this literature, we hypothesize that the accuracy of reports on farm labor is positively related to educational attainment, as a proxy for cognitive skills.

### 3. Study design and context

The data used in this paper were collected by the Institute of Statistical, Social and Economic Research (ISSER) of the University of Ghana and the World Bank during the main rainy season (roughly March to September 2015) in the Mampong Municipal, Ejura Sekyedumasi, Nkoranza South, and Pru districts of Ghana. A random sample of 720 agricultural households was selected from 20 enumeration areas (henceforth denoted as villages) and then randomly assigned to one of three alternative survey designs. All households were administered a baseline survey at the beginning of the season, which collected a roster of plots and household members working in agriculture, and basic demographic and plot-specific information. After the main harvest, households were administered an endline survey, which was modeled after the design of the LSMS-ISA survey series (i.e. a multi-topic household survey with an agricultural production module).<sup>7</sup> In terms of capturing farm labor, the three survey designs differed as follows:

- A. **Weekly visit:** Following the baseline survey in March 2015, these households were visited weekly from April to September 2015, which in turn was followed by the endline survey in October/November 2015. At baseline (visit 0), households reported on the number of days worked and total number of hours worked, per person per plot, since starting preparing the plot for the season. During the season households were visited every week (visits 1-23), and asked to report on the number of hours worked on each day of the past week, per person per plot, and the range of activities (but not hours per activity).<sup>8</sup> At endline (visit 24), households report the total number of days and the average hours per day, per person per plot, since the last weekly visit. The design for this group minimizes recall periods and (mostly) avoids the need for complex aggregations and calculation of averages –farm labor estimates for this group are considered as the benchmark.
- B. **Weekly phone:** The design for this group is essentially the same as for the weekly visit group, what differs is the method of data collection. While households in the weekly visit group received weekly face-to-face visits during the season, households in the weekly phone group received weekly phone calls. The main purpose of this treatment arm was to explore the potential of soliciting high-frequency labor data using phone surveys, as a proof of concept, an aspect of the experiment that will be explored in a companion paper.
- C. **Recall:** For this group, there were no visits between the baseline in March 2015 and the endline in October/November 2015. In the endline survey, the agricultural labor module for this group identified the household members that worked on each of the household's plots during the season, and for those members, data were collected on (i) total days per person per plot spent across the five activities and (ii) typical hours per day worked per person per plot on each activity. This group hence mimics the way LSMS-ISA and other agricultural surveys collect data on farm labor (see also section 2.2).

---

<sup>7</sup> The baseline and endline surveys were timed as per the production calendar of maize, the most important crop in this area.

<sup>8</sup> We distinguish between five activities - land preparation and planting; weeding; ridging, fertilizing and other non-harvest activities; harvesting; and supervision.

Since the weekly phone arm was primarily intended as a proof of concept, the focus of this paper is the comparison of the weekly visit group and the recall group – in other words, we do *not* include any analysis of the data from the weekly phone group.

While the overall design of this study is very similar to Arthi et al (2018), there is one significant difference: This study fielded a baseline survey to *all* households, including recall households, which allowed us to collect a baseline listing of plots and household workers. This listing of plots and household workers plays a crucial role, because all data on farm labor were collected at the person-plot level. However, respondents could add plots and household workers that were not listed at baseline during subsequent visits – that is during visits 1 to 24 (weekly visits, endline) in the weekly visit group, and during the endline in the recall group. This design variation follows from the results in Arthi et al (2018), who document that underreporting of plots and household workers in the recall group counteracts overestimation of farm labor per person per plot. Having access to a baseline survey for all households allows us to explore in greater detail bias in listings of plots and household workers, but it may also explain some of the difference in results between this study and Arthi et al (2018). We will revert to this issue in section 4.5, where we compare the results obtained for Ghana and Tanzania.

Our field work design randomized households within villages. We are confident that intra-cluster contamination among households is minimal given that villages tend to be large and the sampled households are rather dispersed. The within-village randomization is aimed at balancing micro-agroecological characteristics that may affect the allocation of household labor and are difficult to be captured in the available survey and satellite data.

Although the initial sample was 240 households per experiment arm, we had some attrition at different stages. In the weekly visit group, 20 households dropped out of which 9 were replaced with a final sample of 229 households. In the recall group, 7 households dropped out leaving a final sample of 233 households. In all subsequent analysis, households that were not interviewed at the endline are excluded, as well as households in the weekly visit arm that dropped out before week 16 (even if they re-appeared at endline).

**Table 1** summarizes household and plot characteristics across the weekly visit and recall arms, drawing on the baseline survey. For most of the traits presented in this table, households are well balanced across survey arms, but there are some exceptions. In the weekly visit group, the average household consists of 5.8 members, of whom 1.6 are children younger than 10 years old, and cultivates 2.6 plots. The average plot is located within 54 minutes' walk from the households' residence. Of all individuals aged 10 and older, 55 percent worked on one or more household plots between the start of land preparations and the baseline survey. However, households in the recall group are significantly smaller, have fewer cultivated plots and are further away from their plots, compared to households in the weekly visit group. Because of these baseline differences, our empirical analysis controls for household size, number of plots, and the average distance between the plots and the residence at baseline. Cropping patterns are very similar across arms, with maize being the most important crop in both arms, followed by yam and groundnuts.

**Table 1: Sample characteristics in baseline survey**

	Weekly visit	Recall
Age	22.974 (0.514)	22.744 (0.584)
Male	0.493 (0.014)	0.500 (0.015)
Worked on household plot since start of season (age 10+)	0.550 (0.016)	0.582 (0.017)
Enrolled in school	0.405 (0.014)	0.421 (0.014)
Has no schooling (age 10+)	0.307 (0.015)	0.329 (0.016)
<i>N (individuals, all ages)</i>	<i>1,314</i>	<i>1,185</i>
Household size	5.814 (0.228)	5.086** (0.168)
Male household head	0.752 (0.029)	0.807 (0.026)
Household head single/divorced/widowed	0.212 (0.027)	0.262 (0.029)
Number of children (age <10)	1.646 (0.100)	1.575 (0.096)
Number of persons who worked on household plot (age 10+)	2.226 (0.093)	2.043 (0.097)
Number of plots per household	2.571 (0.085)	2.348* (0.088)
<i>N (households)</i>	<i>226</i>	<i>233</i>
Distance plot to residence (min. walking)	53.601 (1.958)	62.706*** (2.234)
Proportion of plots cultivating beans/peas	0.119 (0.013)	0.122 (0.014)
Proportion of plots cultivating cassava	0.100 (0.012)	0.122 (0.014)
Proportion of plots cultivating groundnuts	0.203 (0.017)	0.219 (0.018)
Proportion of plots cultivating maize	0.487 (0.021)	0.439 (0.021)
Proportion of plots cultivating yam	0.222 (0.017)	0.258 (0.019)
<i>N (plots)</i>	<i>581</i>	<i>547</i>

Note: Standard errors in parentheses. T-test for difference between the two groups were done: \*\*\* denotes the difference is significant at the 1% level; \*\* at the 5% level; \* at the 10% level.

## 4. Recall bias in farm labor

### 4.1. Main results

In the surveys administered to the recall and weekly visit households, farm labor is reported per person and per plot. As discussed in section 2.2, our main measure of farm labor is *season-wide hours per person per plot*. In this measure, we include only those plots for which farm labor was reported at least once during the season. We consider all individuals aged 10 years and older reporting at least one hour of labor on any household plot at any point in time during the agricultural season.

Households in the weekly visit group report farm labor at 25 points in time, starting with the baseline survey (visit 0) and ending with the endline survey (visit 24). Baseline, weekly visit, and endline hours are then summed to arrive at season-wide hours of farm labor per person per plot. Recall households, though asked at baseline to provide a listing of plots and households members working in agriculture, provide all information on farm labor in the endline survey. We combine data on the number of days worked during the season and the typical number of hours worked per day to calculate season-wide hours of farm labor per person per plot. As shown in **Table 2**, recall households report 19 more hours (18 percent) of farm labor per person per plot over the season. This difference is statistically significant at the 1 percent level.

**Table 2: Recall bias in farm labor, descriptive estimate**

	Weekly visit	Recall	Difference
Season-wide hours per person per plot	106.40 (2.81)	125.72 (4.75)	19.33*** (5.16)
Number of person-plots	2,787	1,675	

Note: Reported days and hours worked have been winsorized at the top 1 percent of the distribution. Standard errors are shown in parentheses. T-test on difference in means with \*\*\* indicating the difference is significant at the 1% level.

As discussed in the previous section, households in the weekly visit group were slightly larger, cultivated slightly more plots at baseline, and their plots were at a shorter distance than households in the recall group - it is important to adjust for these differences. **Table 3** pools recall and weekly visit households to regress season-wide hours per person per plot on an indicator variable (*Recall*), which equals unity for households in the recall group and controls for the household number of plots and workers at baseline, and the average distance to their plots. This delivers an estimate of conditional recall bias of 10.3 hours per person per plot. Though lower than the unconditional estimate, it is still about 10 percent of the average in the weekly visit group, and the difference is statistically significant at the 5 percent level.

**Table 3: Recall bias in farm labor, regression estimate**

	Season-wide hours per person per plot
Recall	10.317** (5.203)
N	4,462
R2	0.021 0.024

Note: Regressions control for household size and number of plots, and the average distance between the household's residence and the plots (all at baseline). \*\* denotes significance at the 5% level.

A potential concern of our identification strategy is that weekly visits to farm households may have behavioral impacts, which would confound our estimate of recall bias. The analysis relies on the identifying assumption that the survey methodology affects the quality of the data only, not behavior. Though it seems unlikely, there is a possibility that weekly visits and repeated interviews about farm labor activities may have induced farmers in the weekly visit group to organize their farm work more carefully or to spend more time farming than they would otherwise do (akin to Hawthorne effects).

As a robustness check, and to rule out that the survey methodology affected farm labor, and not just its measurement, we analyze differences in maize yields between weekly visit and recall households for

plots listed at baseline.<sup>9</sup> We chose maize because it is the most common crop in our study area, cultivated on just under half of all baseline plots (see **Table 1**). Since the survey method for measuring harvest and plot size was the same for weekly visit and recall households, the survey method should not have a direct impact on measured yields among plots listed at baseline. On the other hand, harvest is measured at endline, so any differences in behavior and labor input across the two groups during the season would be expected to affect yields. The results in **Table A2 (appendix)** show reassuringly that there is no significant difference in maize yields between weekly visit and recall households on baseline plots. This provides us with further confidence that the survey method did not affect farm labor, at least not to the extent that farm labor affected maize yields.

#### **4.2. Proxy determinants - listings of plots and household workers**

Providing accurate information on farm labor requires respondents to recall several components, including an accurate listing of cultivated plots, and an accurate listing of household members working on those plots. This section explores proxy determinants of recall bias in farm labor – bias in listings of plots and household workers.

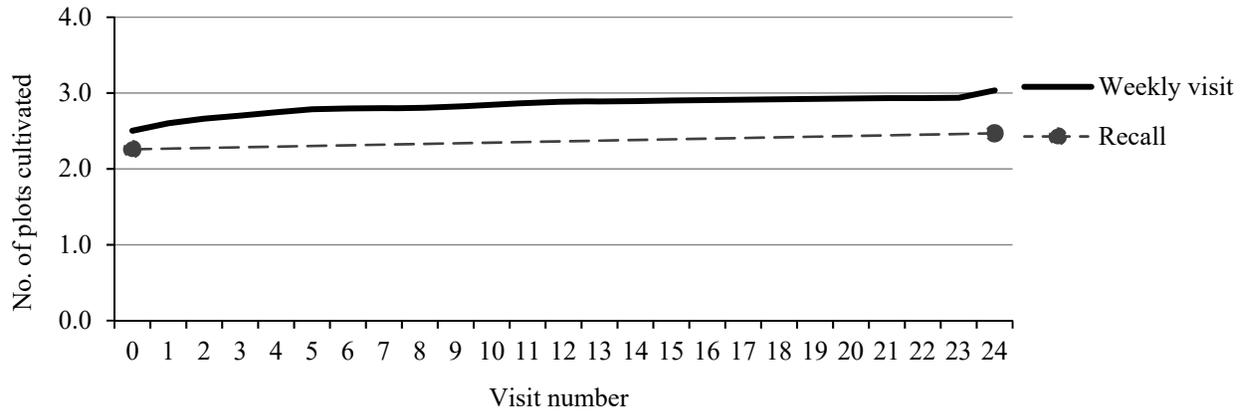
Let's first recap how respondents in the recall and weekly visit groups provide a listing of plots cultivated by the household. Recall households were interviewed twice – at the beginning of the season (baseline) and after the main harvest (endline). At baseline, respondents were asked to list all plots owned and/or expected to be cultivated during the 2015 long rainy season. At endline, respondents in the recall group were presented with the baseline list of plots and asked to add any additional plots that were owned and/or had been cultivated during the 2015 long rainy season but had not yet been listed. Conversely, households in the weekly visit group were visited 23 times between the baseline and the endline, and during each of these visits they could add plots that were currently owned and/or had been cultivated since the last visit but had not yet been listed. These weekly visits could make it easier to recall all plots cultivated during the season compared to the recall households, who face a lag of 23 weeks between the first visit and the endline and may have forgotten to list all additional plots that were added after baseline.

**Figure 1** shows the cumulative number of cultivated plots listed per household, that is, the number of cultivated plots reported up to a given week (visit number) in the season. At baseline, households in the weekly visit group reported on average 2.5 plots for cultivation, compared with 2.3 plots for recall households. By the time of the endline survey, households in the weekly visit group reported a cumulative total of 3.0 plots for cultivation, compared with only 2.5 plots for recall households. Both at baseline and at endline, the difference is statistically significant, but the fact that the gap doubles over time suggests that the survey method accounts for a significant part of the difference at endline.

---

<sup>9</sup> As we will show in the next section, recall households list fewer additional plots during the season than weekly visit households and plots that are listed during the season are structurally different from the ones listed at baseline. It is hence important to compare baseline plots.

**Figure 1: Cumulative number of cultivated plots, by visit number**

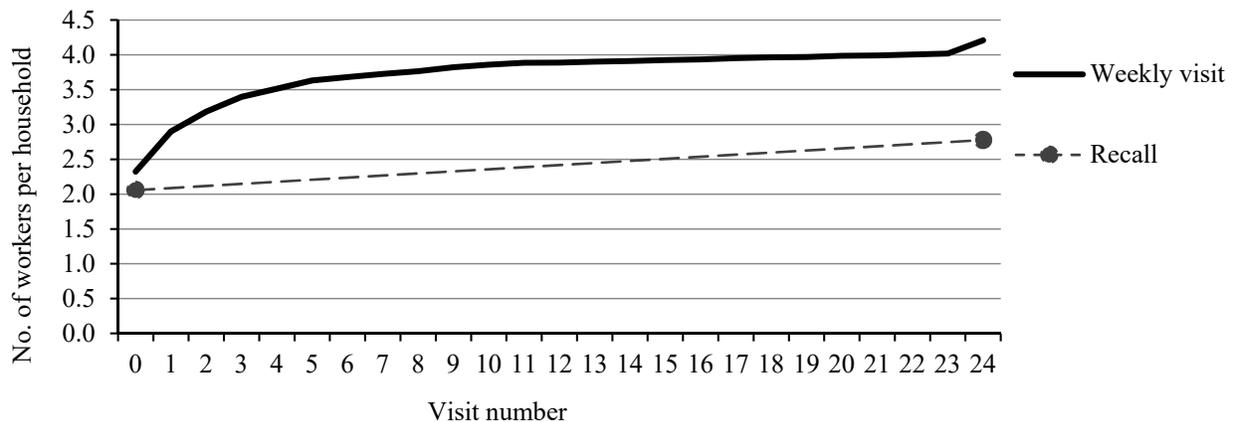


Note: Recall households were only interviewed at baseline (visit 0) and endline (visit 24). Cultivated plots are plots the household reported as used for cultivation.

We now turn to the listing of household members working in agriculture. Households in the recall group reported at baseline and at endline on which household members had been working on each plot. In contrast, households in the weekly visit group reported on a weekly basis.

Figure 2 shows the number of farm workers per household, calculated as the cumulative total of persons aged 10 and older who have been listed as having worked on one of the household's plots. At endline, households in the recall arm report 2.8 farm workers on average, compared to a cumulative total of 4.2 workers in weekly visit households. The difference is much smaller at baseline (2.1 vs. 2.3 workers), again suggesting that the survey method matters for the number of household workers reported. This listing bias mostly reflects that households in the weekly visit group report on more household workers engaged in agricultural labor compared with the recall group, though reporting of additional household members also plays a role.

**Figure 2: Cumulative number of household workers, by visit number**



Note: Recall households were only interviewed at baseline (visit 0) and endline (visit 24). Household workers denotes persons aged 10+ reported of doing any agricultural labor.

We use a double difference estimator (Table 4) to quantify the effect of the survey method on the number of plots and household workers reported. For households in the weekly visit group, the number of

plots and household workers increases by 0.5 and 1.9, respectively, between baseline and endline. This is much more than the increase of 0.2 and 0.7,<sup>10</sup> respectively, in the recall group. Both double difference estimators of listing bias due to recall – -0.3 for plots and -1.1 for household workers – are statistically significant.

**Table 4: Double difference estimator of bias in listings of plots and household workers**

	No. of cultivated plots listed per household	No. of workers listed per household
Endline	0.531*** (0.063)	1.885*** (0.118)
Endline * Recall	-0.320*** (0.088)	-1.116*** (0.166)
N	908	908
R2	0.156	0.393

Note: Regressions control for household fixed effects \*\*\* denotes significance at the 1% level.

**Table 5** turns to explore whether bias in listings of plots and household workers accounts for recall bias in farm labor at the person-plot level. Column (1) repeats the benchmark estimate of recall bias in farm labor from **Table 3**. Column (2) further includes an indicator variable, which equals unity if the plot was listed at baseline (*Baseline\_Plot*) as well as its interaction with the main recall variable (*Baseline\_Plot\*Recall*). This specification allows to test two hypotheses. First, season-wide hours per person per plot may differ between plots listed at baseline and plots that were added later, so that differences in the composition of plots between the recall and weekly visit groups may explain recall bias in the number of hours reported. Second, recall bias itself might differ between plots listed at baseline and plots added later. Column (3) uses an analogue specification but distinguishes between household members listed at baseline and those that were added later. Column (4) includes both sets of variables.

The results in columns (2) to (4) do not provide any evidence for the second hypothesis that recall bias differs between plots (household workers) listed at baseline and those that were added later, as the relevant interaction effects are insignificant in all three specifications. Conversely, we do see that reported season-wide hours per person per plot are significantly higher for plots and household workers listed at baseline than for plots and household workers added later (columns (2) to (4)) and controlling for this characteristic turns the main effect of the recall variable insignificant.

Together, these results provide significant support for the first hypothesis that compositional differences explain recall bias in farm labor. Season-wide hours of farm labor per person per plot are higher for recall households because the latter fail to list several plots and persons. These omitted plots and persons have, on average, less farm labor than the plots and persons that were listed at baseline. In other words, weekly visit households report more ‘marginal’ plots and farm workers than recall households, which reduces their total average hours per person per plot.

<sup>10</sup> This is the total effect of the Endline and the Endline\*Recall coefficients (i.e. 0.531-0.320=0.2).

**Table 5: Listings of plots and household workers and recall bias in farm labor**

	Hours per person per plot			
	(1)	(2)	(3)	(4)
Recall	10.317**	4.347	4.843	5.016
	(5.203)	(12.914)	(8.222)	(13.540)
Plot listed at baseline		76.250***		69.612***
		(8.150)		(7.913)
Recall * Plot listed at baseline		1.703		-4.933
		(13.989)		(13.627)
Person listed as farm worker at baseline			91.235***	87.082***
			(6.198)	(6.142)
Recall * Person listed as farm worker at baseline			-2.579	-2.087
			(10.324)	(10.263)
N	4462.000	4462.000	4462.000	4462.000
R2	0.024	0.052	0.090	0.112

Note: Regressions control for household size and number of plots, and the average distance between the household's residence and the plots (all at baseline). \*\*\* denotes significance at the 1% level; \*\* at the 5% level; \* at the 10% level.

By means of illustration, assume that there are two types of plots. Type A plots use 400 hours of labor per person and type B plots use 200 hours of labor per person. Households in the recall and weekly visit groups both cultivate on average 2 type A plots and 2 type B plots. Both groups correctly list type A plots and report 400 hours of farm labor per person for these plots. However, recall households on average only list one type B plot with 200 hours of labor, whereas weekly visit households list both type B plots with 200 hours of labor. In such a scenario, we would estimate total season hours per person per plot at 333 for recall households, compared with 300 for weekly visit households – a positive recall bias of 33 hours per person per plot, but this recall bias would disappear if we controlled for plot type. In other words, it is listing bias in plots that drives compositional differences across the two groups and overall recall bias in farm labor measured in season-wide hours per person per plot.

### 4.3. Recall bias at higher levels of aggregation

Listing bias has additional effects on estimates of farm labor if we consider alternative levels of aggregation. Sections 4.1 and 4.2 measured farm labor in season-wide hours per person per plot, which is the unit at which farm labor is reported in our survey. However, as discussed in section 2.2, other levels of aggregation might be of interest. As we show in this section, recall bias in farm labor is strongly affected by the level of aggregation, both in direction and in magnitude.

Agricultural productivity analysis, particularly if it has a focus on gender or other intra-household differences, typically uses input and yields data per plot and hence requires total labor input per plot. Plot-level measures of farm labor are sensitive to household workers not being listed, because the labor input of these workers would not be accounted for. Conversely, person-level analysis is more sensitive to plots not being listed, given that an individual's labor input on these non-reported plots would not be captured.

**Table 6** summarizes farm labor at different levels of aggregation. Comparing recall households to weekly visit households, we find no significant difference in farm labor *per person* (panel A). Hence the recall bias in season-wide hours per person per plot (in **Table 2**) is nullified by the lower number of plots

listed. Farm labor *per plot* (panel B) is about 13 percent lower for recall households and total farm labor *per household* (panel C) is about 30 percent lower compared to weekly visit households and both differences are statistically significant. At these levels of aggregation, listing bias hence dominates recall bias in farm labor per person per plot. Or in other words, because recall households report farm labor for too few plots and (especially) too few household workers, farm labor per plot and per household is *underestimated*, even though farm labor per person per plot is *overestimated*.

**Table 6: Recall bias in aggregated farm labor, descriptive estimates**

	Weekly visit	Recall	Difference
<b>A.</b> Hours per person (all household plots)	333.95 (10.72)	332.68 (18.80)	-1.27 (20.33)
Number of persons	888	633	
<b>B.</b> Hours per plot (all household persons)	432.29 (17.90)	374.71 (20.96)	-57.57** (27.41)
Number of plots	686	562	
<b>C.</b> Hours per household (all persons and plots)	1,312.16 (63.50)	923.63 (86.09)	-388.53*** (107.12)
Number of households	226	228	

Note: Includes individuals aged 10+ reported as having performed agricultural labor at any point in time during the season, and plots for which any agricultural labor was reported at any point in time during the season. Reported hours worked have been winsorized at the top 1 percent of the distribution. Standard errors are shown in parentheses.

**Table 7** provides regression estimates that confirm significant negative recall bias in hours per plot and hours per household, conditional on the baseline number of plots, household size, and the average distance to the plots. The conditional bias in hours per household is smaller than the unconditional bias reported in **Table 6**, but still large and statistically significant.

**Table 7: Recall bias in aggregated farm labor, regression estimates**

	Hours per person	Hours per plot	Hours per household
Recall	-3.886 (20.421)	-53.686** (26.992)	-276.062*** (99.237)
N	1521	1248	454
R2	0.036	0.067	0.204

Note: Regressions control for household size, number of plots, and the average distance between the household's residence and the plots (all at baseline). \*\* denotes significance at the 5% level, \*\*\* at the 1% level.

#### 4.4. Heterogeneity and mechanisms – the role of education and gender

To further disentangle the mechanisms behind recall bias, this section examines heterogeneity across subpopulations. We start with the relationship between education and recall bias. As discussed in section 2.3, several studies have found a positive association between the level of education – or direct measures of cognitive skills – and recall ability. Based on this literature, we expect the accuracy of reports on farm

labor to be linked positively to educational attainment, as a proxy for cognitive skills, so that recall bias is lower among more educated households.<sup>11</sup>

Second, we consider gender differences – particularly the question whether recall bias differs between male and female workers. Time use data from Africa show that women often carry out different roles and activities simultaneously, rather than sequentially. This holds particularly for childcare, which is typically embedded within other economic activities (Blackden and Wodon 2006). This may make it more difficult to recall farm labor for female than for male household workers.

In **Table 8**, we compare mean hours per person per plot between weekly visit and recall households for less educated households, more educated households, female workers, and male workers. Panel A shows that in less educated households, the average hours of farm labor are about 25 percent higher than in more educated households, and importantly there is no evidence of recall bias in the latter group. Recall households report, on average, 44.2 hours per person per plot more than weekly visit households among less educated households, whereas the difference is only 2.3 hours (and not statistically significant) among more educated households. Panel B summarizes hours per person per plot for male and female workers. Male farmers work almost 50 percent more hours per plot than female farmers, but contrary to our expectation the recall bias appears to be greater for males than for females.

**Table 8: Heterogeneity in recall bias, descriptive estimate**

<b>A: By education</b>	Weekly visit	Recall	Difference	Weekly visit	Recall	Difference
	Hh education primary or below			Hh education above primary		
Season-wide hours per person per plot	123.8 (4.87)	168.0 (10.25)	44.20*** (10.20)	96.1 (3.39)	98.3 (3.93)	2.27 (5.36)
Number of person-plots	1,037	658		1,750	1,017	
<b>B: By gender</b>	Weekly visit	Recall	Difference	Weekly visit	Recall	Difference
	Male workers			Female workers		
Season-wide hours per person per plot	126.49 (4.56)	151.94 (7.34)	25.45*** (8.18)	85.84 (3.14)	96.71 (5.68)	10.87* (5.97)
Number of person-plots	1,410	880		1,337	795	

Note: Reported hours worked have been winsorized at the top 1 percent of the distribution. Standard errors are shown in parentheses. T-test on difference in means with \*\*\* indicating the difference is significant at the 1% level and \* at the 10% level.

**Table 9** shows regression estimates of recall bias, controlling for baseline characteristics. For reference, column (1) repeats the main specification shown in **Table 3**, with a recall bias of 10.3 hours per person per plot. To test for heterogeneity in recall bias by level of education, in column (2), the regression includes an indicator variable that equals unity if at least one member of the household has attained education above the primary level (*Hh education above primary*), and its interaction with the main recall variable (*Recall\*Above primary*). The results provide support for the education hypothesis. Both variables are negative and significant, confirming the descriptive evidence that more educated households report

<sup>11</sup> We use the highest level of educational attainment in the household to split the sample into less educated (primary education or below) and more educated households.

fewer hours per person per plot and that recall bias for them is significantly lower than for less educated households. The estimates confirm that there is no significant recall bias for households with above primary educational attainment (see bottom row in **Table 9**), while households with lower education on average overestimate season-wide hours by approximately 37 hours per person per plot.

In terms of gender differences, our empirical results portray a more nuanced picture than we expected. Column (3) of **Table 9** includes an indicator variable that equals unity if the worker is female (*Female*) and the interaction between this variable and the recall variable (*Recall\*Female*). The results show that women work around 40 hours per plot less than men over the season. The point estimates further suggest that season-wide hours per plot are overestimated by 17 hours for males, compared with only 1 hour for females (see bottom row), but the difference as indicated by the interaction term is just below the 10 percent significance threshold. Counter to our expectations, there is no indication that recall bias is larger for women. If anything, our point estimates suggest the opposite.

**Table 9: Estimation of recall bias in farm labor – heterogeneity**

	Labor hours per person per plot		
	No heterogeneity (1)	By education (2)	By gender (3)
Recall	10.317** (5.203)	36.794*** (8.251)	17.328** (7.085)
Hh education above primary		-22.421*** (6.471)	
Recall * Above primary		-42.694*** (10.444)	
Female			-39.695*** (6.204)
Recall * Female			-16.379 (10.136)
N	4,462	4,462	4,462
R2	0.024	0.040	0.043
<b>Total recall bias for educated households (2) or females (3)</b>		-5.900 (6.535)	0.948 (7.373)

Note: Regressions control for the household numbers of workers and plots, and the average distance between the household's residence and the plots (all at baseline). Hh education refers to the highest level of education attained by any household member. Standard errors in parentheses. \*\*\* denotes significance at the 1% level; \*\* at the 5% level; \* at the 10% level.

Our descriptive and econometric results above suggest that perhaps recall bias is lower for groups of farm workers with fewer hours of farm labor: both for the more educated households as well as for female workers we find that average hours per person per plot are significantly lower than for less educated households and male workers, respectively. The results in section 4.2 showed that recall bias stems primarily from listing bias, hence if particular sub-groups work fewer hours on average, perhaps the impact of listing bias is mitigated as the baseline workers and plots differ less from the 'omitted' workers and plots. On the other hand, and particularly with respect to education heterogeneity, if recall bias is lower due to higher cognitive ability, we would expect to see that listing bias itself is lower among more educated households. We now turn to explore this, by analyzing education and gender heterogeneity in the listing of plots and household workers. The first two columns of **Table 10** show the double-difference estimate of listing bias in *plots* by household education. The third and fourth columns estimate listing bias in *persons*

by household education, and finally, the fifth and sixth columns disaggregate listing of *persons* between male and female household workers.

The results show that less educated households have significant listing bias in plots as well as persons. Yet the more educated households have no listing bias in plots (column 2), although they do show significant listing bias in persons (column 4). Hence more educated households do a better job of recalling all their plots, which is in line with our expectations and the cognitive skills interpretation, though they still fail to list all farm workers. This may be explained by the mobility of household members within the period of study, making it difficult to recall all household members who worked on each plot but are not physically present at the time of the endline interview.

Turning to gender (columns 5 and 6), we find that female workers are more likely to be added after baseline than male workers (the coefficient on *Endline* is larger for females), and that listing bias is somewhat larger for females than for males. It therefore seems likely that the reason why we see no significant recall bias in hours per person per plot for female workers is that women work fewer hours on average than men: listing bias has less of an impact if the difference between listed and omitted workers is smaller.<sup>12</sup>

**Table 10: Double difference estimator of bias in listings by education and gender**

	No. of plots listed per household		No. of workers listed per household		No. of workers listed per household	
	Low education (1)	High education (2)	Low education (3)	High education (4)	Males (5)	Females (6)
Endline	0.600*** (0.086)	0.476*** (0.089)	1.630*** (0.148)	2.087*** (0.178)	0.869*** (0.067)	1.175*** (0.083)
Endline * Recall	-0.568*** (0.115)	-0.045 (0.133)	-1.162*** (0.199)	-1.058*** (0.266)	-0.501*** (0.093)	-0.824*** (0.115)
N	452	456	452	456	887	872
R2	0.179	0.175	0.373	0.422	0.313	0.335

Note: All estimations include household fixed effects; education refers to the highest level of education attained by any household member. Standard errors in parentheses. \*\*\* denotes significance at the 1% level.

In all, our analysis in this section suggests that recall ability is positively related to educational attainment, since we find that more educated households are better able to accurately list all their plots at endline. They do still underreport the number of farm workers, but we find no significant recall bias in hours worked per person per plot in more educated households. Second, against our expectations, there is no significant recall bias for female farm workers (as opposed to male farm workers). Our estimation results indicate further that this can be explained by the fact that females work substantially fewer hours per plot compared to males. In general, these results suggest that recall bias will be larger for sub-groups of workers among which there is greater variation in the intensity of farm labor.

<sup>12</sup> Additional regressions confirm that the difference in hours worked between workers listed at baseline and those added later is larger for men than for women. Hence, even though listing bias is larger for female than for male workers, this listing bias has less of an effect, because the ‘omitted’ female workers are more similar to the ‘listed’ female workers than it is the case for male workers.

#### 4.5. Comparison between Ghana and Tanzania

As mentioned earlier, this study follows the design of a survey experiment conducted in the Mara region of Tanzania in 2014 (Arthi et al 2018). The main difference in design of the two experiments is related to the way information was captured for the recall group. In the Tanzania case, the recall households were interviewed only once at the end of the season. In Ghana, the recall households were visited at the beginning of the agricultural season for the baseline survey, which collected an initial listing of plots and household workers, and again at the end of the season. The Tanzania approach allows for the comparison of the treatment group with what would be a ‘typical’ cross-sectional household survey that is conducted at the end of the agricultural season without any information from the start or during the season. Conversely, the Ghana approach resembles a panel survey and enables us to control for differences in baseline characteristics and disentangle the possible sources of recall bias by allowing more rigorous estimates of listing bias.

**Table 11** summarizes the results of both studies. Columns 3 and 4 replicate **Table 2**, and columns 1 and 2 are the equivalent numbers for Tanzania. Weekly visit households in Tanzania report 39.5 hours of farm labor per person per plot, which is significantly less than in Ghana (106.4 hours). However, this difference can be easily explained by differences in plot size, with plots being much smaller in Tanzania (0.33 ha) than in Ghana (1.19). What is much more striking is the large difference in (unconditional) recall bias – 18 percent in Ghana, versus 207 percent in Tanzania. What would explain such large differences in the magnitude of recall bias?

**Table 11: Farm labor per person per plot, Ghana and Tanzania**

	Tanzania		Ghana	
	Weekly Visit	Recall	Weekly Visit	Recall
Hours	39.5 (69.5)	121.3*** (133.8)	106.4 (2.81)	125.7*** (4.75)
Plot Size (ha)	0.33 (0.007)	0.31 (0.01)	1.19 (0.02)	1.32*** (0.03)

Note: \*\*\* Denotes significantly different from weekly visit estimate at 1%. Standard errors are reported in parenthesis below the coefficient estimates.

**Table 12** provides information on listing bias in Ghana and Tanzania, that helps to understand the difference in recall bias shown in the previous table. The simple difference estimator of listing bias presented in the top panel is equal to the difference in the average number of workers (plots) at endline between weekly visit and recall households and can be compared between Ghana and Tanzania. The bottom panel shows the double difference estimator of listing bias, which nets out baseline differences. Given that Tanzania did not collect baseline information for recall households, the double difference estimator can only be calculated for Ghana.

**Table 12: Listing bias, Ghana and Tanzania**

		Tanzania	Ghana
Listing Bias (Simple Difference)	Workers	-1.6***	-1.5***
	Plots	-2.1***	-0.6***
Listing Bias (Double Difference)	Workers	N/A	-1.2***
	Plots	N/A	-0.3***

Note: \*\*\* Denotes significantly different from weekly visit estimate at 1%.

The simple difference estimator of bias in the listing of plots is significantly larger in Tanzania (2.1 fewer plots for recall households) than in Ghana (0.6 plots fewer plots in the recall group), while listing bias in household workers is similar (1.6 vs. 1.5 fewer workers in the recall group). In both experiments, the roster for weekly visit households is ‘cumulative’, as shown in **Figures 1 and 2**, this means that it has been updated throughout the season to capture all plots and all household workers involved in farm work. Conversely, the recall households do not have a cumulative roster. The fact that listing bias in plots is smaller in Ghana than in Tanzania might partly reflect that recall households in Ghana were administered a baseline plot roster, which was updated once at endline and helped to facilitate recall. As illustrated in section 4.2, listing bias is one of the driving forces behind recall bias in farm labor – and this may hence explain the large differences between the two studies highlighted in **Table 10**. Unfortunately, we cannot compare the double difference estimator of listing bias across the two countries, which for Ghana is slightly smaller than the simple difference estimator.

Besides variations in the design of the experiments, differences in educational attainment could also play an important role. In our study area in Ghana, 50 percent of households have at least one member with above primary education, compared with only 34 percent of households in the Mara region of Tanzania. Since we have shown in section 4.4 that recall bias, and particularly listing bias for plots, declines with the level of education, the lower levels of education in Tanzania than in Ghana are consistent with the observed differences in recall and listing bias.

## **5. Implications of recall bias for agricultural productivity measures**

Measures of farm labor can be of interest in isolation but are most commonly used in the context of analyzing agricultural productivity. **Table A3 (appendix)** shows how three common measures of agricultural productivity – labor productivity per-worker, labor productivity per-hour and land productivity per-acre – are affected by listing bias in plots and persons and recall bias in the number of hours per person-plot conditional on being listed. The table further distinguishes according to whether the unit of analysis is the plot or the household. Let’s take the example of labor productivity per-hour measured at the household-level. Listing bias in persons increases the measured labor productivity per-hour, because the labor input (measured in hours) of the non-listed persons is not captured in the total hours that enter through the denominator. Listing bias in plots has ambiguous effects, as the omitted plots reduce both the numerator (in terms of non-measured output) and the denominator (in terms of non-measured hours). This effect depends on whether labor productivity is systematically lower or higher on omitted plots versus listed plots. Together, these two forms of listing bias in persons and plots are likely to increase household level per-hour labor productivity, but the net effect might be negative if omitted plots have substantially higher labor productivity, which seems unlikely. Additionally, we should consider the positive recall bias in hours per person-plot conditional on being listed, which reduces measured productivity by increasing the numerator. Because of all these counteracting effects, the net effect of listing bias in plots, persons, and recall bias on the number of hours per person-plot conditional on being listed is unclear a priori.

Even though the net effects may be ambiguous, the table allows making the following predictions. First, measures of land productivity will in general be less affected by recall bias in farm labor than measures of labor productivity. However, land productivity can still be affected by listing bias in plots. Depending on whether the non-listed plots are more (less) productive than the listed plots, the recall method will

underestimate (overestimate) measured land productivity. Second, of all the productivity measures, labor productivity per worker measured at the plot-level is most likely to be overestimated by the recall method. This is because listing bias in persons reduces the denominator and (unlike with labor productivity per hour) is not counteracted by positive recall bias in hours per person-plot conditional on being listed. Third, even if labor productivity per worker is overestimated at the plot-level, it could still be underestimated at the household-level. This is because listing bias in plots reduces measured output (in the enumerator), but may not necessarily reduce the number of workers (the denominator), unless some workers worked exclusively on non-listed plots.

To assess the impact on different measures of productivity empirically, we use data on crop production collected at endline. As discussed in section 2, the endline survey fielded a standard agricultural production module to all households, which asked to report the quantity and value of crop harvested during the preceding agricultural season (for each plot and crop). Hence households in the weekly visit and recall groups reported on their plot-level harvest in the same manner. Using these harvest data, we measure plot-level yields and labor productivity for maize as the maize harvest in kg or in cedi per acre, per labor hour, and per worker.<sup>13</sup> All measures are winsorized at 1 percent and the analysis is conducted at the plot-level. **Table 13** presents regression estimates of recall bias.

**Table 13: Yields and labor productivity in maize production, recall and weekly visit groups**

	Land productivity		Labor productivity (per-hour)		Labor productivity (per-worker)	
	Cedi/acre (1)	Kg/acre (2)	Cedi/hour (3)	Kg/hour (4)	Cedi/worker (5)	Kg/worker (6)
Recall	21.543 (31.903)	28.358 (48.880)	2.633 (1.691)	4.656** (2.359)	264.322*** (88.471)	415.993*** (118.681)
N	493	491	487	485	487	485
R2	0.019	0.014	0.040	0.045	0.099	0.112
Control Mean	457.62	728.15	9.20	13.79	555.65	837.26

Note: All dependent variables are winsorized (1%). Regressions control for household size, number of plots, and the average distance between the household's residence and the plots (all at baseline). Standard errors in parentheses. \*\*\* denotes significance at the 1% level; \*\* at the 5% level. Control Mean shows the mean value of the dependent variable for Weekly Visit households.

We find no evidence of significant recall bias in maize yields, which broadly confirms our expectations that measures of land productivity should be fairly robust to the types of measurement bias that are the focus of this study. Maize yields (columns 1 and 2) are somewhat higher among recall households, but not statistically significantly so, suggesting that listing bias in plots does not affect yield measures (or, in other words, any productivity differences between omitted and listed plots are too small to affect average maize yields). However, for three of the four measures of labor productivity in agriculture (columns 4 to 6), we find it is significantly overestimated among recall households. In line with our expectations, the recall effect is more significant (both economically and statistically) for labor productivity per-worker than for labor productivity per hour – this is because listing bias has a proportionately larger effect on the number of

<sup>13</sup> Hired labor is not included to maintain simplicity in the comparisons.

workers than on the number of hours per listed plot, as the omitted workers are the ones working relatively few hours.

## 6. Conclusion

Measuring labor in agriculture is a key input to the analysis of agricultural labor productivity, which in turn is an important metric in the design of policies for developing countries. Nevertheless, research on the best methods to measure agricultural labor is scarce. We contribute to the literature on measuring agricultural labor by analyzing the data of a randomized survey methodological study conducted in Ghana. The study was designed to allow for the comparison of estimates of farm labor obtained from a recall survey conducted at the end of the season with data collected weekly throughout the season.

The results indicate that the recall method overestimates farm labor per person per plot by about 12 percent, conditional on differences in baseline characteristics. This recall bias in farm labor is accounted for by listing bias, as households in the recall group report significantly fewer marginal plots and farm workers. Moreover, even though farm labor per person per plot is overestimated, the recall bias is very different at higher levels of aggregation. Because recall households report too few plots and too few workers, the end-of-season recall method *underestimates* farm labor per plot and total farm labor per household. As a result, plot-level estimates of labor productivity in agriculture are significantly overestimated by the recall method.

We also moved beyond proxy determinants to understand the deeper forces behind recall bias. Consistent with the notion that recall bias is linked to the cognitive burden of reporting on past events, we find that better educated households recall farm labor with greater accuracy. This educational gradient – together with variations in the design of the experiment – most likely explains why this study finds much lower recall bias in farm labor in the Ghanaian context than the previous study found for the Mara region in Tanzania (Arthi et al. 2018). We find that more educated households are better able to accurately list all their plots at endline. Additionally, we find no significant recall bias in hours worked per person per plot in more educated households, even when they underreport the number of farm workers.

Contrary to our expectations, we do not find significant recall bias for female farm workers (as opposed to male farm workers). This is explained by the fact that females work substantially fewer hours per plot compared to males, thus the omitted female farm workers differ relatively less from the listed female farm workers. This finding illustrates that recall bias in farm labor per person per plot is likely to be greater for sub-groups of workers among which there is greater variation in the intensity of farm labor.

Our findings have implications for survey design. First, the differences in results between this paper and the experiment in Tanzania (Arthi et al. 2018) highlight the need to understand the context in which the survey will be implemented. Second, data collection agencies should increase their efforts on ensuring that listings of plots and household members are complete to minimize the effect of listing bias on the collection of labor and other agricultural inputs. Third, it is extremely important to have clarity on the objective of the survey, whether the survey will be used for analysis of employment status, land productivity and/or labor productivity; each objective may require different levels of granularity of the data collected and may be more prone to different approaches for data collection.

## References

- Abebe, Girum. 2013. "Recall Bias in Retrospective Surveys: Evidence from Enterprise-Level Data." *Asian-African Journal of Economics and Econometrics* 13(1): 17-33.
- Aguilar, Arturo, Eliana Carranza, Markus Goldstein, Talip Kilic and Gbemisola Oseni. 2015. "Decomposition of gender differentials in agricultural productivity in Ethiopia." *Agricultural Economics* 46(3): 311–34.
- Ali, Daniel, Derick Bowen, Klaus Deininger and Marguerite Duponchel. 2016. "Investigating the Gender Gap in Agricultural Productivity: Evidence from Uganda." *World Development* 87: 152-70.
- Andre, Pierre, Esther Delesalle and Christelle Dumas. 2017. "Returns to Farm Child Labor in Tanzania." Retrieved from: [https://economics.handels.gu.se/digitalAssets/1643/1643674\\_57.-delesalle\\_esther-returns-to-farm.pdf](https://economics.handels.gu.se/digitalAssets/1643/1643674_57.-delesalle_esther-returns-to-farm.pdf)
- Arthi, Vellore, Kathleen Beegle, Joachim de Weerd and Amparo Palacios-Lopez. 2018. "Not Your Average Job. Measuring Farm Labor in Tanzania." *Journal of Development Economics* 130: 160-72.
- Bardasi, Elena, Kathleen Beegle, Andrew Dillon and Pieter Serneels. 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(3): 418-47.
- Beckett, Megan, Julie DaVanzo, Narayan Sastry, Constantijn Panis and Christine Peterson. 1999. "The Quality of Retrospective Reports in the Malaysian Family Life Survey." Working Paper Series 99-13, RAND.
- Beckett, Megan, Julie DaVanzo, Narayan Sastry, Constantijn Panis and Christine Peterson. 2001. "The Quality of Retrospective Data: An Examination of Long-Term Recall in a Developing Country." *Journal of Human Resources* 36(3): 593-625.
- Blackden, Mark C. and Quentin Wodon. 2006. "Gender, Time Use, and Poverty in Sub-Saharan Africa." Working Paper No. 73. World Bank: Washington, DC.
- Christiaensen, Luc, Lionel Demery and Jesper Kuhl. 2011. "The (Evolving) Role of Agriculture in Poverty Reduction – An Empirical Perspective." *Journal of Development Economics* 96(2): 239-54.
- De Janvry, Alain, Elisabeth Sadoulet and Tavneet Suri. 2016. "Field Experiments in Developing Country Agriculture." In: Abhijit Banerjee and Esther Duflo (eds.). *A Handbook of Economic Field Experiments*. North-Holland: Amsterdam.
- Dillon, Brian, Peter Brummund and Germano Mwabu. 2017. "How Complete Are Labor Markets in East Africa? Evidence from Panel Data in Four Countries." Retrieved from: [http://faculty.washington.edu/bdillon2/CV\\_papers/DBM-labor-markets-170510.pdf](http://faculty.washington.edu/bdillon2/CV_papers/DBM-labor-markets-170510.pdf)
- FAO (Food and Agriculture Organization of the United Nations). 2017. *Measuring Work in Agricultural Surveys and Censuses: A Review*. Mimeo.
- Fermont, Anneke and Todd Benson. 2011. "Estimating Yield of Food Crops Grown by Smallholder Farmers. A Review in the Ugandan Context." IFPRI Discussion Paper 01097. Washington, DC: International Food Policy Research Institute (IFPRI).

- Fisher, Monica and Kadiwa Vongai. 2014. "Can Agricultural Input Subsidies Reduce the Gender Gap in Modern Maize Adoption? Evidence from Malawi." *Food Policy* 45: 101-11.
- Gollin, Douglas and Christopher Udry. 2017. "Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture." Retrieved from: [http://egcenter.economics.yale.edu/sites/default/files/files/Conference%202017%20Agri-Devo%20speakers/Gollin\\_Udry\\_Heterogeneity%2C%20Measurement%20Error%2C%20and%20Misallocation.pdf](http://egcenter.economics.yale.edu/sites/default/files/files/Conference%202017%20Agri-Devo%20speakers/Gollin_Udry_Heterogeneity%2C%20Measurement%20Error%2C%20and%20Misallocation.pdf)
- Gollin, Douglas, David Lagakos and Michael E. Waugh. 2014a. "Agricultural Productivity Differences Across Countries." *American Economic Review* 104(5): 165-70.
- Gollin, Douglas, David Lagakos and Michael E. Waugh. 2014b. "The Agricultural Productivity Gap." *Quarterly Journal of Economics* 129(2): 939-93.
- ILO (International Labor Organization). 2016. ILOSTAT – Employment by Sector. Retrieved from: [http://www.ilo.org/ilostat/faces/oracle/webcenter/portalapp/pagehierarchy/Page3.jsp?MBI\\_ID=33](http://www.ilo.org/ilostat/faces/oracle/webcenter/portalapp/pagehierarchy/Page3.jsp?MBI_ID=33)
- Kilic, Talip, Amparo Palacios-Lopez and Markus Goldstein. 2015. "Caught in a Productivity Trap: A Distributional Perspective on Gender Differences in Malawian Agriculture." *World Development* 70: 416-63.
- LaFave, Daniel and Duncan Thomas. 2016. "Farms, Families and Markets: New Evidence on Completeness of Markets in Agricultural Settings." *Econometrica* 84(5): 1917-60.
- McAuliffe, Timothy L., Wayne DiFranceisco and Barbara R. Reed. 2010. "Low Numeracy Predicts Reduced Accuracy of Retrospective Reports of Frequency of Sexual Behavior." *AIDS Behav* 14(6): 1320-29.
- McCarthy, Aine Seitz, Amy L. Damon and Vincent Seigerink. 2016. "Favoritism and Farming: Agricultural Productivity and Polygyny in Tanzania." Retrieved from: <http://pages.vassar.edu/lacdevconf2016/files/2016/08/Favoritism-and-Farming-Agricultural-Productivity-and-Polygyny-in-Tanzania.pdf>
- McCullough, Ellen B. 2017. "Labor Productivity and Employment Gaps in Sub-Saharan Africa." *Food Policy* 67: 133-52.
- Oseni, Gbemisola, Paul Corral, Markus Goldstein and Paul Winters. 2015. "Explaining gender Differentials in Agricultural Production in Nigeria." *Agricultural Economics* 46(3): 285–310.
- O’Sullivan, Michael, Arathi Rao, Raka Banerjee, Kajal Gulati and Margaux Vinez. 2014. *Levelling the Field: Improving Opportunities for Women Farmers in Africa*. World Bank and One Campaign, Washington, DC.
- Palacios-López, Amparo & Ramón López, 2015. "The Gender Gap in Agricultural Productivity: The Role of Market Imperfections." *The Journal of Development Studies* 51(9): 1,175-92.
- Peters, H. Elizabeth. 1988. "Retrospective Versus Panel Data in Analyzing Lifecycle Events." *Journal of Human Resources* 23(4): 488-513.

- Reardon, Thomas and Paul Glewwe. 2000. Agriculture. In: Margaret Grosh and Paul Glewwe (eds.) *Designing Household Survey Questionnaires for Developing Countries. Lessons from 15 Years of the Living Standards Measurement Study*. Volume 2. Washington, DC: World Bank.
- Restuccia, Diego and Raul Santaaulalia-Llopis. 2017. "Land Misallocation and Productivity." NBER Working Paper 23128, National Bureau of Economic Research, Cambridge, MA.
- Walker, Thomas. "Appendices. Technical Details on the 2009 Household Survey." Retrieved from: <http://barrett.dyson.cornell.edu/files/research/databases/ghana/2009-Walker-Appendix-Summary.pdf>
- Winowiecki, Leigh, Caroline Mwongera, Jennifer Twyman, Kelvin Shikuku, Edidah Ampaire, Chris Miyinzi, Mariola Acosta, Wendy Okolo, Peter Läderach. 2016. "Intra-household and farm production decision making survey in rural Tanzania and Uganda." Retrieved from: [doi:10.7910/DVN/0ZEXKC](https://doi.org/10.7910/DVN/0ZEXKC).
- World Bank. 2015. LSMS – Integrated Surveys on Agriculture. Retrieved from: <http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:23512006~pagePK:64168445~piPK:64168309~theSitePK:3358997,00.html>
- World Bank. 2016. Malawi Labor Experiment. Agriculture Questionnaire. Interviewer Manual. Unpublished.
- World Bank. 2017. World Development Indicators. Retrieved from: <http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators>

## Appendix

**Table A1: Measures of farm labor and aggregation approaches used by empirical studies**

<b>Author</b>	<b>Literature strand</b>	<b>Aggregation of labor data for analysis</b>
Andre, Delesalle and Dumas (2017)	Child labor	Labor per household
Aguilar et al. (2016)	Gender differences	Labor per plot (normalized for plot size)
Ali et al. (2016)	Gender differences	Labor per plot (normalized for plot size)
de la O Campos et al. (2016)	Gender differences	Labor per plot (normalized for plot size)
Dillon et al. (2017)	Rural labor markets	Labor per household
Fisher and Kandiwa (2014)	Gender differences	Labor per plot (normalized for plot size)
Gollin and Udry (2017)	Agricultural productivity	Labor per plot (normalized for plot size)
Kilic et al. (2015)	Gender differences	Labor per plot (normalized for plot size)
McCarthy, Damon and Seigerink (2016)	Gender differences	Labor per plot (normalized for plot size)
McCullough (2017)	Productivity gaps across sectors	Labor per person
Oseni et al. (2016)	Gender differences	Labor per plot (normalized for plot size)
O'Sullivan et al (2014)	Gender differences	Labor per plot
Palacios-Lopez & Lopez (2015)	Gender differences	Labor per plot (normalized for plot size)
Restuccia and Santaaulalia-Llopis (2017)	Agricultural productivity	Labor per household

**Table A2: Effect of the recall method on maize yields of baseline plots**

	Kg/acre (1)	Cedi/acre (2)
Recall	-15.762 (52.897)	-2.288 (34.471)
N	416	417
R2	0.009	0.013

Note: Dependent variables are winsorized (1%). Regressions control for household size, number of plots, and the average distance between the household's residence and the plots (all at baseline). Standard errors in parentheses.

**Table A3: Effect of the recall method on measures of agricultural productivity, mechanisms**

Productivity concept		Unit of analysis				
		Plot		Household		
	Measure	Direction of bias (recall method)		Measure	Direction of bias (recall method)	
<b>Labor productivity</b>	<b>per-worker</b>	Output (plot) / workers (plot)	* <b>Listing bias, persons:</b> Pr↑ * <b>Listing bias, plots:</b> Pr↓ (non-listed plots more productive than average) or Pr↑ (non-listed plots less productive than average), ambiguous * <b>Conditional recall bias in hours per person-plot:</b> no effect → <i>net effect ambiguous, expectation Pr↑</i>		Output (household) / workers (household)	* <b>Listing bias, persons:</b> Pr↑ * <b>Listing bias, plots:</b> probably Pr↓ (unless some workers worked exclusively on non-listed plot) * <b>Conditional recall bias in hours per person-plot:</b> no effect → <i>net effect ambiguous</i>
	<b>per-hour</b>	Output (plot) / hours (plot)	* <b>Listing bias, persons:</b> Pr↑ * <b>Listing bias, plots:</b> Pr↓ (non-listed plots more productive than average) or Pr↑ (non-listed plots less productive than average) * <b>Conditional recall bias in hours per person-plot:</b> Pr↓ → <i>net effect ambiguous</i>		Output (household) / hours (household)	* <b>Listing bias, persons:</b> Pr↑ * <b>Listing bias, plots:</b> Pr↓ (non-listed plots more productive than average) or Pr↑ (non-listed plots less productive than average) * <b>Conditional recall bias in hours per person-plot:</b> Pr↓ → <i>net effect ambiguous</i>
<b>Land productivity</b>	<b>per-acre</b>	Output (plot) / acres (plot)	* <b>Listing bias, persons:</b> no effect * <b>Listing bias, plots:</b> Pr↓ (non-listed plots more productive than average) or Pr↑ (non-listed plots less productive than average) * <b>Conditional recall bias in hours per person-plot:</b> no effect → <i>net effect ambiguous, expect effect to be relatively small</i>		Output (household) / acres (household)	* <b>Listing bias, persons:</b> no effect * <b>Listing bias, plots:</b> Pr↓ (non-listed plots more productive than average) or Pr↑ (non-listed plots less productive than average) * <b>Conditional recall bias in hours per person-plot:</b> no effect → <i>net effect ambiguous, expect effect to be relatively small</i>

Note: Pr denotes productivity.