

Measuring Farm Labor: Survey Experimental Evidence from Ghana

Isis Gaddis, Gbemisola Oseni, Amparo Palacios-Lopez, and Janneke Pieters

Abstract

This study examines recall bias in farm labor through a randomized survey experiment in Ghana, comparing farm labor estimates from an end-of-season recall survey with data collected weekly throughout the agricultural season. Recall households report 10 percent more farm labor per person-plot, which can be explained by recall households' underreporting of "marginal" plots and household workers. This "selective" omission by recall households, denoted as listing bias, alters the composition of plots and workers across treatment arms and inflates average farm labor hours per person-plot in the recall group. Since listing bias, in this setting, dominates other forms of recall bias at higher levels of aggregation (i.e., when farm labor per person-plot is summed at the plot, person, or household level), farm labor productivity is overestimated for recall households. Consistent with the notion that recall bias is linked to the cognitive burden of reporting on past events, there is no recall bias among more educated households.

JEL classification: C83, J22, O12, Q12

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

Isis Gaddis (corresponding author) is a Senior Economist with the Gender Group at the World Bank and a Research Fellow at the Institute of Labor Economics (IZA); her email address is igaddis@worldbank.org. Gbemisola Oseni is a Senior Economist with the Development Data Group of the World Bank; her email address is goseni@worldbank.org. Amparo Palacios-Lopez is also a Senior Economist with the Development Data Group of the World Bank; her email address is apalacioslopez@worldbank.org. Janneke Pieters is an Assistant Professor with the Development Economics Group at Wageningen University and a Research Fellow at the Institute of Labor Economics (IZA); her email address is janneke.pieters@wur.nl. This work was supported by 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 (through the Trust Fund number 012853) and by 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 (grant number 2015-2070). The authors gratefully acknowledge the comments from participants at the Annual Conference of the German Economic Association Research Group on Development Economics, the GLMILIC 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, Wageningen University, and the World Bank. The authors would further like to thank two anonymous referees for constructive comments. 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

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 (ILO 2016; World Bank 2017). 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. Agricultural growth has been found to be more effective than nonagricultural growth in reducing extreme poverty (Christiaensen, Demery, and Kuhl 2011).

One of the most important agriculture inputs 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 or days 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 Santaaulalia-Llopis 2017; Gollin and Udry 2019), agricultural household models (e.g., LaFave and Thomas 2016), rural labor markets (e.g., Dillon, Brummund, and Mwabu 2019), and gender differences in agriculture (e.g., O’Sullivan et al. 2014; Damon and McCarthy 2019). A range of recent studies, for example, highlight lack of access to family members’ labor, particularly male labor, as one of the main factors constraining women farmers and a key determinant of the gender gap in agricultural productivity (O’Sullivan et al. 2014; Kilic, Palacios-Lopez, and Goldstein 2015).

Measuring farm labor, however, is 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*).¹ While this extended reference period minimizes the impact of seasonality, it can lead to difficulties for respondents when recalling 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 low labor activity after planting and prior to harvest. In addition, since irrigation is limited and rain-fed agriculture is the norm across much of Africa, the reliance on weather leads to added unpredictability. 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 2017), 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.

This paper builds upon the results from Tanzania to estimate recall bias in farm labor in the Ashanti and Brong Ahafo regions of Ghana. Two approaches to collecting farm labor data are compared. One group of farmers was interviewed weekly about farm labor for each of the seven days preceding the interviews, which is considered 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 season, hence using the traditional end-of-season recall method. Comparing these two groups provides an estimate of bias in the end-of-season recall of farm labor hours, denoted here as *recall bias*, that is, recall bias at the intensive margin. While this research design is broadly similar to the one used in Tanzania, this study made one key methodological refinement by visiting recall households at the beginning of the agricultural season

governments they represent. A supplementary online appendix for this article can be found at *The World Bank Economic Review* website.

1 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.

to collect an initial listing of plots and household workers. This allows for further analysis on the extent to which the recall method can lead to biased reporting on the number of cultivated plots and household farm workers, denoted here as *listing bias*, that is, recall bias at the extensive margin. This alteration in the research design builds upon the results reported in [Arthi et al. \(2018\)](#), which showed a significant underreporting of plots and households. However, the Tanzania experiment visited recall households only once, at the end of the season, and this limited the types of analyses that could be performed. In addition, while both research settings are characterized by smallholder agriculture, especially maize farming, there are important differences between the two study contexts. For example, smallholder farmers in Ghana's Ashanti and Brong Ahafo regions cultivate fewer but larger plots than farmers in the Mara region of Tanzania. Labor markets in the study area in Ghana are also more integrated, as evidenced by the greater prevalence of agricultural wage labor. Hence, one objective of the present study is to explore whether the Tanzania results hold in a somewhat different setting, and to assess the relationship between recall bias and contextual factors.

This paper makes several contributions. First, it shows that recall bias in farm labor per person per plot (i.e., the unit at which labor is reported by the survey respondents) is much smaller in the Ghanaian than in the Tanzanian setting, which highlights that estimates of recall bias are highly context-dependent and cautions against extrapolating recall bias across populations. Second, the results demonstrate that, in the research setting of Ghana, recall bias in farm labor at the person-plot level can be accounted for by bias in listings of plots and household farm workers. This complements the findings of [Arthi et al. \(2018\)](#), who document that listing bias runs counter to recall bias in farm labor at the person-plot level. The results reveal that recall households report more hours of farm labor per person per plot *because* they fail to list "marginal" plots and farm workers. This "selective" omission of marginal plots and household workers by recall households alters the composition of plots and workers across treatment arms and inflates average farm labor hours per person per plot in the recall group. Listing bias, hence, not only counteracts but *explains* recall bias in farm labor at the person-plot level, at least in this context. Surprisingly, once the analysis controls for listing bias, farmers in this study setting report fairly accurately on the number of hours of farm labor provided per person per plot.

Third, recall bias is strongly associated with educational attainment. There is no significant recall bias in hours worked per person per plot for more educated households, who tend to be better at recalling all their plots. This is consistent with the notion that recall bias is linked to the cognitive burden of reporting on events that occurred in the past. This educational gradient may also explain, at least in part, why recall bias in this study for Ghana is much lower than [Arthi et al. \(2018\)](#) found for Tanzania, given that education levels among sampled farmers are higher in Ghana than in Tanzania.

Fourth, there are important, yet complex, gender differences in recall bias in this study setting. Female farm workers are less likely to be listed than male farm workers, but this gender difference in listing bias is mitigated by the fact that women also work fewer hours per plot than men.

And finally, this paper illustrates the implications of its findings for empirical research, especially the analysis of agricultural productivity. The results in this study provide nuance to the conclusions reached by [Arthi et al. \(2018\)](#), who argue that recall-style modules tend to understate agricultural productivity. In settings such as the one in this study, where listing bias is the dominant source of reporting bias, the recall method underestimates farm labor at higher levels of aggregation, such as the plot or household levels, and overestimates the labor productivity of household-operated farms.

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 the role of listing bias. Section 5 analyzes heterogeneity by education and gender. Section 6 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 7 documents recall bias at different levels of aggregation and implications for the analysis of agricultural productivity. Section 8 concludes.

2. Measuring Farm Labor

Measuring labor in smallholder farming is inherently complex and prone to measurement error.² Like informal household enterprises in other sectors, smallholder farmers rarely keep any records; thus respondents 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 farm labor for the entire season.

The 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), differ widely across surveys. Household 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.³ However, these modules differ with respect to the unit of data collection, recall period, sequencing of questions, selection of respondent(s), and so on. 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 of the analysis in this paper.

Unit of Data Collection

The majority of household surveys with a focus on agriculture use the plot as the main unit of data collection (FAO 2017). This reflects a 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” agricultural sectors, which includes most smallholder farming in Sub-Saharan Africa, data on farm inputs and practices should be 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 when they describe agricultural production activities. In such a situation, plot-level data collection reflects the natural flow of conversation between the enumerator 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 intrahousehold differences (e.g., productivity gaps between male- and female-managed plots).

For data on farm labor, the 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 (for men, women, and children separately) about the number of days, average hours per day, and average number of workers who provided family labor on each plot during the last major/minor seasons.⁴ 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, including the 2015/16 Ghana agricultural labor survey used in this study.⁵

2 This section draws on Arthi et al. (2018).

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

4 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.

5 Other examples are multiple rounds of nationally representative LSMS-ISA surveys in Burkina Faso, Ethiopia, Malawi, Mali, Niger, Nigeria, Tanzania, and Uganda (World Bank 2015b) or specialized surveys with a focus on agriculture (e.g., Walker 2009 – Ghana; Mwangi et al. 2017 – Tanzania, Uganda). Another example would be the Agricultural

A major advantage of collecting data on farm labor by person and by plot (denoted here as *person-plot level*) is that, due to the high level of granularity, it can accommodate different analytic approaches at the analysis stage. The main unit at which farm labor is reported and/or embedded in the empirical analysis of several recent studies, is shown in table A1.1 (appendix A1). Studies concerned with gender differences in agriculture tend to aggregate at the plot level (often with normalization for plot size). This allows for the comparison of productivity levels between plots managed by male and female farmers and for assessing the role of labor shortages as one of the underlying constraints (Fisher and Kandiwa 2014; O’Sullivan et al. 2014; Aguilar et al. 2015; Andrews, Golan, and Lay 2015; Kilic, Palacios-Lopez, and Goldstein 2015; Oseni et al. 2015; Palacios-Lopez and Lopez 2015; Ali et al. 2016; Damon and McCarthy 2019). Other analyses of agricultural productivity aggregate either at the plot level (Gollin and Udry 2019) or at the household level (Restuccia and Santaaulalia-Llopis 2017; André, Delesalle, and Dumas 2019; Dillon, Brummund, and Mwabu 2019). McCullough (2017) aggregates farm labor at the person level to compare labor productivity between farm and nonfarm 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 – that is, *plot*, *person*, and *household levels* – open for analysis. This, of course, assumes that greater granularity in data collection does not undermine data quality, so that there is no analytical cost to greater disaggregation. The results from this paper caution that this may not necessarily be the case, as different levels of data collection may be more or less prone to different forms of recall bias.

Despite the advantages highlighted above, 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, they may forget to mention marginal plots, that is, plots that are not farmed intensively during the season. Hence, enumerators need to probe carefully to ensure that all relevant plots are listed as per the standardized definition used in the survey.⁶ In addition, it may be difficult for the respondent to identify which members of the household worked on each plot during the reference period, particularly if the responses are provided by a proxy respondent on behalf of other household members. Finally, respondents may not find the concept of farm labor per person per plot intuitively meaningful, and therefore face difficulties in reporting on the number of days and hours that individual household members worked on each plot. Arthi et al. (2018) highlight that participants of focus group discussions in the Mara region of Tanzania requested to report on farm labor with reference to units of one or two acres, which is somewhat larger than the average plot size in the region. These challenges can be compounded by the length of the recall period, as discussed in the next section.

In this paper, and consistent with Arthi et al. (2018), the principal unit of data collection and analysis is farm labor per person per plot. However, higher levels of aggregation, that is, farm labor per plot, per person and per household, are also reported. The latter are, as shown in table A1.1 (appendix A1), more widely used at the analysis stage. This helps in assessing the potential direction of bias for analyses that use higher levels of aggregation. There are two main advantages to using the person-plot level as a starting point. First, to provide guidance to survey designers, it appears useful to assess recall bias directly at the unit at which the data are typically collected in household surveys. Second, this approach helps further

Integrated Survey (AGRIS) designed by the Food and Agriculture Organization of the United Nations (FAO). They ask, for 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.

6 The survey definition of a plot 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).

disentangle the proximate causes of bias at higher levels of aggregation – that is, incomplete listings of plots and household members (recall bias in listings or, simply, *listing bias* – that is, bias at the extensive margin) and recall bias in terms of the number of days and hours of farm labor at the person-plot level, conditional on being listed (i.e., bias at the intensive margin).

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 earlier, which asked respondents to recall each household member's labor input during the “long rainy season 2014.” The main advantages of using an extended reference period are the reduction in cost and logistical burden of data collection and ensuring consistency between measures of farm labor and measures of seasonal crop production. This consistency is crucial to an analysis of labor or total factor productivity. The flip side of using an extended reference period is that it increases the burden on the respondents, 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 the recall period (see [Schwarz and Oyserman 2001](#); [Arthi et al. 2018](#) for more detailed discussions). The most obvious explanation is that respondents forget to report events that took place long ago. Conversely, forward telescoping, where respondents report events that occurred outside the reference period, may be more prevalent with shorter reference periods ([Kjellsson, Clarke, and Gerdtham 2014](#)). In addition, there is evidence that the length of the recall period may have a bearing on how respondents interpret survey questions. [Winkielman, Knäuper, and Schwarz \(1998\)](#), for example, show that the longer the recall period, the more likely respondents are to interpret questions about frequencies as referring to rare, major events only (see also [Igou, Bless, and Schwarz 2002](#)). This potential change in question interpretation makes it difficult to attribute a negative correlation between the frequencies of reported events and the length of the reference period to forgetting.

The salience and periodicity of the event to be reported also have important implications for the respondents' strategy to respond to survey questions ([Menon 1993](#); [Arthi et al. 2018](#)). Respondents tend to use rate-based estimations to report on regular events with a low level of salience (e.g., weekly hours at the gym), but recall and count strategies for irregular and more salient events (e.g., number of marathons run in a year). With respect to events that are neither regular nor salient, such as farm labor provided over a long reference period, both strategies break down, and respondents may resort to impression-based estimates, based on generic assumptions or value judgments (e.g., the importance of the event in one's daily life), which are particularly prone to bias ([Conrad, Brown, and Cashman 1998](#)).

Several studies have found a positive association between the level of education – or direct measures of cognitive skills – and recall ability. [Peters \(1988\)](#), in 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 United States). While evidence from the United States may not necessarily extend to developing countries, given potentially large differences in education systems, survey methods, and so on, a few studies have found similar associations in Asia and Africa. [Beckett 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. This literature suggests that the accuracy of reports on farm labor may be positively related to educational attainment, as a proxy for cognitive skills.

3. Study Design and Context

This paper draws on the 2015 Ghana agricultural labor survey ([World Bank 2015a](#)), a specialized dataset collected by the Institute of Statistical, Social and Economic Research (ISSER) of the University of Ghana and the World Bank during the 2015 main rainy season (roughly March to September 2015) in the Mampong Municipal, Ejura Sekyedumasi, Nkoranza South, and Pru districts, which are located in the Ashanti and Brong Ahafo regions 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 – a recall arm, a weekly visit arm, and a weekly phone arm.⁷ Since the phone arm was intended primarily as a proof of concept (i.e., to document implementation challenges of soliciting high-frequency labor data using phone surveys), the results from this arm are not reported in this paper.⁸ The survey was implemented using Computer-Assisted Personal Interview (CAPI). All households were administered a baseline survey at the beginning of the 2015 long rainy season, which collected a listing of plots and household members, and identified which household members had already worked on each plot since the beginning of land preparation for the long rainy season.⁹ In addition, the baseline survey captured core 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 questionnaires, that is, a multitopic household survey with an agricultural production module.¹⁰ Farm labor hours were captured as follows in the recall and weekly visit arms (see the supplementary online appendix S1, available at *The World Bank Economic Review* website, for further details):

- A. **Recall:** The households in this group were interviewed only twice, during the baseline in March 2015 and the endline in October/November 2015. In the endline survey, the agricultural labor module for this group asked one household respondent (the most knowledgeable person regarding the activities on the plot) to identify all household members who worked on each of the household's plots during the long rainy season 2015, and for those members, data were collected on (i) total days per person per plot spent across five activities and (ii) typical hours per day worked per person per plot on each activity.¹¹ This group mimics the way LSMS-ISA and other agricultural surveys collect data on farm labor (see section 2). It is important to note that even though the CAPI application at endline was pre-populated with the baseline rosters of household members and plots, enumerators were instructed (at endline) to ask for additional household members and plots, that may not have been listed at baseline, and to add those to the respective rosters. This step occurred prior to the collection of farm labor data, so that respondents could immediately report on farm labor for newly listed plots and household members.
- B. **Weekly visit:** In addition to the baseline survey in March 2015 and the endline survey in October/November 2015, the households in this group were visited weekly from April to September 2015, that is, over the course of the 2015 rainy season. Farm labor for these households was captured as follows. At baseline (visit 0), respondents were asked about the total number of days and hours each
- 7 Although the sample of 720 agricultural households was not designed to be representative of Ghana, neither the whole country nor meaningful parts of it, the districts were selected to capture some variation in agroecological characteristics.
 - 8 The design of the phone arm was essentially the same as for the weekly visit group. However, instead of being visited in person, households were called once a week.
 - 9 As per the experience of the study team, smallholder farmers in developing countries typically understand what is meant by the main agricultural season (which in the case of Ghana is the long rainy season), and when it begins. For the reference period (i.e., the long rainy season 2015), the survey instrument was designed to capture work that had been performed on the households' plots for any type of crop – seasonal or nonseasonal. This was also highlighted in the enumerator instructions, which explicitly directed them to include work on plots with tree and permanent crops.
 - 10 The baseline and endline surveys were timed as per the production cycle of maize – the most important crop in this area.
 - 11 Five activities are distinguished : (1) land preparation and planting; (2) weeding; (3) ridging, fertilizing, and other non-harvest activities; (4) harvesting; and (4) supervision.

household member had worked on each plot since the beginning of land preparation for the main rainy season. This was to ensure that the survey did not miss out on any labor input for the current season that may have occurred before the beginning of field work. At the first weekly visit (visit 1), respondents were asked to report on the number of hours each household member had worked on each plot since the baseline visit, and the range of activities performed (but not hours per activity). During subsequent weekly visits (visits 2–23) respondents were asked to report on the number of hours each household member had worked on each plot during each day over the past week, and the range of activities performed. At endline (visit 24), respondents were asked to report on the total number of days and the average number of hours per day each household member had worked on each 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. Similar to the recall group, the CAPI application for both the weekly visit surveys and the endline was pre-populated with plots and household members that had been listed during earlier interviews. However, enumerators were instructed to inquire, at each weekly visit and at endline, about additional household members and plots that had not been previously listed. These newly listed household members and plots were immediately added to the rosters, so that respondents could report on farm labor for newly listed plots and household members.

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 households in the recall group, which allowed collecting 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, as described above, 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 facilitates the exploration of bias in listings of plots and household workers, but it may also explain some of the differences in results between this study and that of [Arthi et al. \(2018\)](#). This issue is further discussed in section 6, which compares the results obtained for Ghana and Tanzania.

The field work design randomized households within villages. The within-village randomization is aimed at balancing micro-agroecological characteristics that may affect the allocation of farm labor, and are difficult to capture in the available survey and satellite data. The fraction of households included in the study sample (across all treatment arms) at the village level varies between 25 and 63 percent. A potential concern is that the weekly visits affected reporting of recall households. Since such intra-village spillovers would be more likely to occur in villages where a larger proportion of the total household population was sampled, a potential test is to correlate mean person-plot hours of farm labor reported by recall households against the share of households sampled at the village level. The results show a small and statistically insignificant correlation of -0.14 , which provides confidence that intra-cluster contamination is minimal. Similarly, there is no significant correlation between the share of households sampled and village-specific regression estimates of recall bias in person-plot hours (results available on request).

The recall group followed the common practice for agricultural labor modules of asking the manager of the plot (primary decision maker of the day-to-day operations on the plot) to report on labor inputs for all household members. This practice implies that many responses are provided by proxy respondents in standard recall modules. In addition, agricultural modules are typically fielded during a single visit, which makes it operationally difficult to interview multiple household members (who may not be at home during the time of the visit, or may not have time to participate in the survey). Interviewers of the weekly visit

group were instructed to interview each member of the household if possible; however, if they were not at home during the interview, they too had to rely on proxy respondents. Moreover, in the weekly visit group, the respondent for questions on farm labor could vary across the 25 visits to the households. Self-reporting rates are slightly higher among weekly visit households (with 43 percent of person-plot farm labor hours being self-reported on most visits) than among recall households (with 33 percent being self-reported at endline). Though the difference in self-reporting across treatment arms is not very large, its endogeneity to treatment assignment and to the household member's profile of economic activity implies that proxy respondent effects cannot be convincingly analyzed with the setup used in this study, though such effects could, of course, play a role.

There was some attrition at different stages of field work. In the weekly visit group, 17 households from the original sample of 243 dropped out during the course of the season, resulting in a final sample of 226 households in this group. In the recall group, 6 out of 239 households dropped out, leaving a final sample of 233 households. Though attrition is related to treatment, this is unlikely to affect the estimates as the number of households lost is small. In table A1.2 (appendix A1), attritor households are similar to households in the final sample, with the exception of household size and distance between the household's residence and plots. As discussed below, these characteristics are unbalanced across treatment groups in the final sample, so that it is important to control for them in the analyses.

Household, individual, and plot characteristics for the final sample across the weekly visit and recall arms, drawing on the baseline survey, are shown in table 1. 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 of age – and cultivates 2.6 plots. The average plot is located within a 54-minute walk from the household's residence. Of all individuals aged 10 years 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, the 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 groundnut.

4. Recall Bias in Farm Labor

In the surveys administered to the recall and weekly visit households, farm labor is reported per person for each plot. As discussed in section 2, the main measure of farm labor is *season-wide hours per person per plot* (though other levels of aggregation are reported in section 7). In this measure, only those plots for which farm labor was reported at least once during the season are included. The analysis considers all individuals aged 10 years and older reporting at least 1 hour of labor on any household plot at any point in time during the agricultural season.

Main Results

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 up to arrive at season-wide hours of farm labor per person per plot. Recall households are only asked to provide a listing of plots and household members working in agriculture at baseline, while all information on farm labor is provided in the endline survey. Data on the number of days worked throughout the season and the typical number of hours worked per day are combined to calculate season-wide hours of farm labor per person per plot. As shown in table 2, recall households report 19 more hours

Table 1. Sample Characteristics in Baseline Survey by Treatment Arm

	Weekly visit	Recall
Household size	5.814 (0.228)	5.086** (0.168)
Male household head	0.752 (0.029)	0.807 (0.026)
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 listed	2.571 (0.085)	2.348* (0.088)
Number of plots intended for cultivation	2.509 (0.086)	2.236** (0.083)
<i>N (households)</i>	226	233
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)
Has no schooling (age 10+)	0.307 (0.015)	0.329 (0.016)
<i>N (individuals, all ages)</i>	1,314	1,185
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>	581	547

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

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

Table 2. Recall Bias in Farm Labor per Person per Plot, Descriptive Estimate

	Weekly visit	Recall	Difference
Season-wide hours per person per plot	106.40 (2.81)	125.72 (4.75)	19.32*** (5.16)
Number of person-plots	2,787	1,675	

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

Note: Reported hours worked have been winsorized at the top 1 percent of the distribution. Standard errors in parentheses. T-test on difference in means with *** denoting the difference is significant at the 1 percent level.

Table 3. Recall Bias in Farm Labor per Person per Plot, Regression Estimate

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

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

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 percent level.

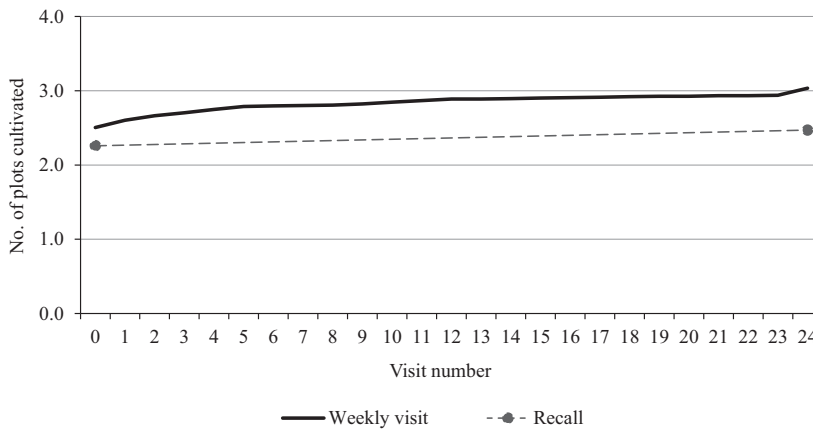
(18 percent) of farm labor per person per plot over the season. This difference is statistically significant at the 1 percent level.

As discussed in the previous section, households in the weekly visit group were slightly larger and cultivated slightly more plots at baseline. In addition, their plots were at a shorter distance from their residence than households in the recall group. It is important to adjust for these differences when making comparisons. In [table 3](#), recall and weekly visit households are pooled to regress season-wide hours per person per plot on an indicator variable, 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.¹² This delivers an estimate of recall bias of about 10 hours per person per plot, controlling for differences in observable characteristics at baseline. Though lower than the descriptive estimate above (which is due to the inclusion of control variables, which account for nearly half of the observed mean difference reported in [table 2](#)), 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.

A potential concern of the identification strategy used in this paper is that the weekly visits to farm households may have had behavioral impacts, which would confound the 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 the possibility that the survey methodology affected actual farm labor and not just its measurement, differences in maize yields between weekly visit and recall households for plots listed at baseline are analyzed.¹³ The focus is on maize because it is the most common crop in the study area, cultivated on just under half of all baseline plots (see [table 1](#)). Since the survey instrument and the timing of the interviews for capturing harvest and plot size were the same for weekly visit and recall households, they should not have an impact on measured yields among plots listed at baseline. But because the harvest is measured at endline, any differences in behavior and labor input across the two groups during the season would be expected to affect yields. The results in [table A1.3](#) (appendix A1) show reassuringly that there is no significant difference in maize yields between weekly visit and recall households on baseline plots. This provides further confidence that the survey method did not affect farm labor, at least not to the extent that farm labor affected maize yields.

- 12 The results are very similar using inverse probability weighting to account for the lack of balance in some of the baseline characteristics.
- 13 As shown later in this paper, 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 therefore important to compare baseline plots.

Figure 1. Cumulative Number of Cultivated Plots, by Visit Number

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

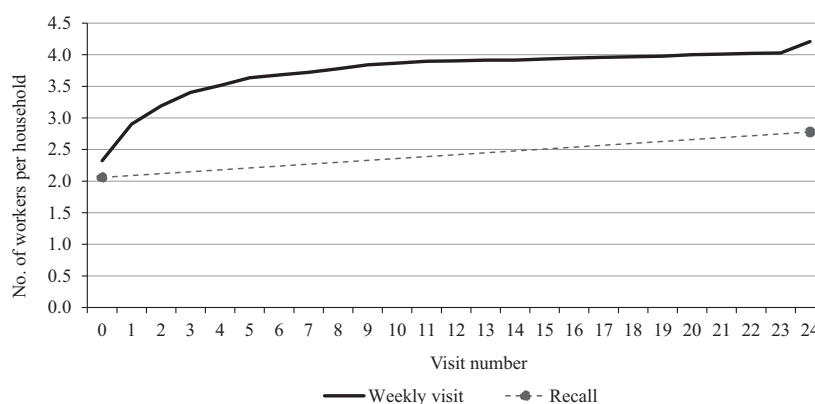
Note: Recall households were only interviewed at baseline (visit 0) and endline (visit 24). Cultivated plots are plots the household reported as used for cultivation.

Listings of Plots and Household Workers

Providing accurate information on farm labor at the person-plot level requires respondents to be able to recall several components of farm-related activities, including an accurate listing of cultivated plots, and an accurate listing of household members working on those plots. Arthi et al. (2018) show that in Tanzania both the number of plots and the number of farm workers were underreported in the recall survey, and that the least-salient persons and plots were omitted. This section explores the link between recall bias in listings of plots and workers on the one hand, and recall bias in person-plot hours on the other hand. The results show that listing bias can in fact account for recall bias in person-plot hours, by changing the composition of person-plots across treatment arms.

It is important to 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 they currently owned and/or had cultivated since the last visit but that 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.

This hypothesis is explored in fig. 1, which 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 across treatment arms 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.

Figure 2. Cumulative Number of Household Workers, by Visit Number

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

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.

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*** (0.059)	1.885*** (0.128)
Endline * Recall	-0.320*** (0.088)	-1.116*** (0.167)
N	908	908
R2	0.156	0.393

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

Note: Regressions control for household fixed effects; standard errors clustered at household level are in parentheses; *** denotes significance at the 1 percent level.

Similar considerations apply to the listing of household members working in agriculture. Households in the recall group reported at baseline and at endline which household members had been working on each plot since the beginning of the season. In contrast, households in the weekly visit group reported on a weekly basis.

This is visually assessed in [fig. 2](#), which shows the number of farm workers per household, which is 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.¹⁴ This listing bias mostly reflects that households in the weekly visit group report on more household workers engaged in farm labor compared with the recall group, though reporting of additional household members also plays a role (results available on request).

A double difference estimator is used to quantify the effect of the survey method on the number of plots and household workers reported ([table 4](#)). 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, respectively, in the recall group.¹⁵ Both double difference

14 Some increase in the number of farm workers between the baseline and endline is to be expected, due to seasonal differences in labor demand, in- and out-migration of household members, etc.

15 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 per Person per Plot

	Season-wide hours per person per plot			
	(1)	(2)	(3)	(4)
Recall	10.317** (5.203)	4.347 (12.914)	4.843 (8.222)	5.016 (13.540)
Plot listed at baseline		76.250*** (8.150)		69.612*** (7.913)
Recall * Plot listed at baseline		1.703 (13.989)		-4.933 (13.627)
Person listed as farm worker at baseline			91.235*** (6.198)	87.082*** (6.142)
Recall * Person listed as farm worker at baseline			-2.579 (10.324)	-2.087 (10.263)
N	4,462	4,462	4,462	4,462
R2	0.024	0.052	0.090	0.112

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

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 percent level; ** at the 5 percent level.

estimators of listing bias due to recall (-0.3 for plots and -1.1 for household workers) are statistically significant.

The question whether bias in listings of plots and household workers accounts for recall bias in farm labor at the person-plot level is explored in table 5. Column (1) repeats the benchmark estimate of recall bias in farm labor per person per plot from table 3. Column (2) further includes an indicator variable, which equals unity if the plot was listed at baseline as well as its interaction with the main recall variable. This specification allows for the testing of 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 as farm workers 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, 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 renders 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 at the person-plot level. Season-wide hours of farm labor per person per plot are higher for recall households *because* they fail to list all plots and household workers. These omitted plots and persons have, on average, fewer hours of farm labor than the plots and persons that were listed at baseline. In other words, recall households report fewer “marginal” plots and farm workers than weekly visit households, which inflates their average hours per person per plot.

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

Table 6. Individual Characteristics and Bias in Listings of Farm Workers

	Farm worker was added after baseline
Farmer by main occupation	−0.273 ^{***} (0.036)
Recall* Farmer by main occupation	0.138 ^{**} (0.062)
Male	0.060 ^{**} (0.030)
Recall* Male	0.111 ^{**} (0.051)
Age 15–50	−0.142 ^{***} (0.031)
Recall* Age 15–50	0.036 (0.053)
Household head	−0.384 ^{***} (0.035)
Recall* Household head	−0.064 (0.057)
N	1,521
R2	0.278

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

Note: Regressions control for household fixed effects. *** denotes significance at the 1 percent level; ** at the 5 percent level.

with 200 hours of labor. In such a scenario, one 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 the analysis controlled for plot type. In other words, it is the listing bias in plots that drives compositional differences across the two groups and overall recall bias in season-wide person-plot hours.

Besides fewer hours of farm labor, do other person or plot characteristics predict under-reporting? As discussed by [Beegle, Carletto, and Himelein \(2012\)](#), salience may be an important factor affecting this type of “extensive margin” recall bias. [Arthi et al. \(2018\)](#) show that the persons and plots reported by recall households are indeed relatively more salient (plots are more likely to be owned by the household and to be close to the residence, while persons are more likely to be male adults whose main occupation is farming). Since the study in Ghana collected baseline data for recall and weekly visit households, it is possible to analyze this question using a double difference estimation. An indicator variable of whether the person or plot was added *after* the baseline survey is regressed on various indicators of person or plot salience, interacted with the recall variable, and controlling for household fixed effects. The expectation is that more salient persons and plots are less likely to be added after baseline (as they would have been listed at baseline), and that this effect is weaker among recall households. In other words, persons and plots added after baseline by recall households are hypothesized to be more salient compared to persons and plots that weekly visit households added after baseline.

For persons, the empirical evidence supports this hypothesis. As shown in [table 6](#), recall households are more likely to add male workers and workers who state that their main occupation is farming. In terms of age and household headship, the results show that household heads and household members of prime working age (15–50 years) are less likely to be added after the baseline, but this effect holds for recall and weekly visit households alike.

For plots ([table 7](#)), there is no evidence that salience predicts why recall households are more prone to under-reporting of plots. Large plots are less likely to be added after baseline by recall and weekly

Table 7. Plot Characteristics and Bias in Listings of Plots

	Plot was added after baseline
Large plot (\geq 75 th percentile)	-0.101** (0.041)
Recall*Large plot	-0.035 (0.064)
Nearby (distance 30 min or less)	0.062 (0.049)
Recall*Nearby	-0.052 (0.075)
Plot owned	-0.090 (0.089)
Recall*Plot owned	0.059 (0.142)
Male manager	0.027 (0.051)
Recall*Male manager	-0.036 (0.084)
Any maize	-0.053 (0.035)
Recall*Any maize	0.028 (0.054)
N	1,226
R2	0.028

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

Note: Regressions control for household fixed effects. ** denotes significance at the 5 percent level.

visit households alike. Plot distance, ownership, gender of the manager, and maize cultivation are not associated with being listed.

Taken together, this section presents evidence that recall households omit plots and farm workers to a greater extent than weekly visit households, and that this listing bias can account for recall bias in person-plot farm labor hours through a compositional effect. As recall households omit persons and plots with fewer farm labor hours, their average person-plot hours are biased upwards. Gender and main occupation are significant predictors of omission of persons, while for plots there is no evidence that salience matters (other than the plot's average season-wide hours of farm labor per person). Conditional on being listed at baseline, there is no recall bias in person-plot hours. This suggests, perhaps surprisingly, that respondents in this setting can accurately recall farm labor hours conditional on listing persons and plots.

5. Heterogeneity: The Role of Education and Gender

This section explores whether recall bias varies by education and gender, starting with the relationship between education and recall bias. As discussed in section 2, several studies have found a positive association between the level of education – or direct measures of cognitive skills – and recall ability. Based on the literature, the accuracy of reports on farm labor is expected to be positively linked to educational attainment, as a proxy for cognitive skills, so that recall bias is lower among more educated households.

Panel A in table A1.4 (appendix A1) shows that recall households report, on average, 44.2 hours per person per plot more (about 25 percent) than weekly visit households among less educated households, whereas the difference is only 2.3 hours and not statistically significant among more educated households. The regression estimates of recall bias, controlling for baseline characteristics, are shown in table 8. To test for heterogeneity in recall bias by level of education, the regression includes an

Table 8. Recall Bias in Farm Labor per Person per Plot and Education, Regression Estimate

	Season-wide hours per person per plot
Recall	36.794*** (8.251)
Hh education above primary	-22.421*** (6.471)
Recall * Above primary	-42.694*** (10.444)
N	4,462
R2	0.040
Total recall bias for educated households	-5.900 (6.535)

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

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 percent level.

indicator variable that equals unity if at least one member of the household has attained education above the primary level, and its interaction with the main recall variable.¹⁶ The results provide support for the education hypothesis. Both variables are negative and significant, confirming the descriptive evidence that more educated households report 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), while households with lower education on average overestimate season-wide hours by approximately 37 hours per person per plot.

Since recall bias in person-plot hours is mainly driven by listing bias, the latter should also show an educational gradient. This is explored in [table 9](#), by analyzing heterogeneity in the listing of plots and household workers. The first two columns show the double difference estimate of listing bias in *plots* by household education, while the third and fourth columns estimate listing bias in *persons* by household education. 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 still show listing bias in workers (column 4). Hence more educated households do a better job of recalling all their plots, but still fail to list all farm workers (but note that listing bias in plots alone can account for recall bias in person-plot hours, as was shown in column (2) of [table 5](#)). This may be explained by the mobility of household members within the period of study, making it particularly difficult to recall all household members who worked on each plot but are not physically present at the time of the endline interview. As reported in [table A1.5](#) (appendix A1), furthermore, a comparison of less- and more-educated households within the weekly visit group shows that the more-educated households have more marginal workers, and in particular a higher number of workers whose main occupation is not farming (as shown

16 The highest level of educational attainment in the household is used to split the sample into less-educated (primary education or below) and more-educated households. This is based on the notion that the household's highest education level is a more suitable indicator of recall ability than the individual level of education of the person whose farm labor is reported, given the large share of proxy responses in the sample. An alternative approach, using the education level of the respondent, shows a regression estimate of recall bias of 11.6 hours for persons whose farm labor was reported by a respondent with primary education or less (significant at the 5 percent level), while recall bias is negative (-4.7) and statistically insignificant for persons whose farm labor was reported by a respondent with above primary education. Results are available from the authors.

Table 9. Double Difference Estimator of Bias in Listings of Plots and Household Workers by Education

	No. of plots listed per household		No. of workers listed per household	
	Low education	High education	Low education	High education
	(1)	(2)	(3)	(4)
Endline	0.600*** (0.086)	0.476*** (0.080)	1.630*** (0.159)	2.087*** (0.191)
Endline * Recall	-0.568*** (0.115)	-0.045 (0.136)	-1.162*** (0.202)	-1.058*** (0.261)
N	452	456	452	456
R2	0.179	0.175	0.373	0.422

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

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

in table 6, these workers are less likely to be listed by recall households). There is no difference in total plot-level hours and number of plots between less- and more-educated households. Hence, more-educated households may be better at listing farm workers as well as plots, *ceteris paribus*, but for farm workers this is counteracted by the fact that they have more marginal farm workers.

Recall bias may also differ 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. Moreover, the previous section showed that recall households are more likely to omit female workers than male workers, which also suggests that recall bias in person-plot hours should be larger for women.¹⁷

The empirical results portray a more nuanced picture. Panel B in table A1.4 (appendix A1) 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 expectation the recall bias appears to be greater for males than for females. Column (1) of table 10 presents the regression estimate, using an indicator variable that equals unity if the worker is female and the interaction between this variable and the recall variable. 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. Contrary to expectations, there is no indication that recall bias is larger for women. If anything, the point estimates suggest the opposite.

Additional regressions in columns (2) and (3) of table 10 show that for both men and women, listing bias can account for recall bias in person-plot hours. However, the difference in total hours between workers listed at baseline and those added later is much larger for men (men listed at baseline work 112 hours more than men added after the baseline, while women listed at baseline work 64 hours more than women added after the baseline). This suggests that for female farm workers, the impact of listing bias is mitigated as the baseline workers differ less from the “omitted” workers.

In all, the analysis in this section suggests that recall ability is positively related to educational attainment, since more-educated households are better able to accurately list all their plots at endline, at least in this study setting. Hence, there is no significant recall bias in hours worked per person per plot for more-educated households. Somewhat surprisingly, there is no significant recall bias for female farm workers

17 Double difference estimators by sex show that listing bias is statistically significant for both male and female workers and confirm that it is stronger for female workers. Results are available from the authors.

Table 10. Recall Bias in Farm Labor per Person per Plot and Gender, Regression Estimate

	Season-wide hours per person per plot		
	All (1)	Men (2)	Women (3)
Recall	17.328** (7.085)	8.854 (13.595)	2.207 (9.221)
Female	-39.695*** (6.204)		
Recall * Female	-16.379 (10.136)		
Person listed as farm worker at baseline		111.819*** (10.059)	64.071*** (7.128)
Recall * Person listed as farm worker at baseline		-0.594 (16.672)	-6.207 (11.905)
N	4,462	2,290	2,172
R2	0.043	0.108	0.064
Total recall bias for females	0.948 (7.373)		

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

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). Standard errors in parentheses. *** denotes significance at the 1 percent level; ** at the 5 percent level.

(as opposed to male farm workers) in this study. The estimation results indicate that this can be explained by the fact that omitted female workers are more similar to the listed female workers than is the case for male workers.

6. Comparison between Ghana and Tanzania

As mentioned earlier, the design of this study evolved from an earlier survey experiment, which was 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, which draws on an earlier roster of household members and plots. The latter facilitates controlling for differences in baseline characteristics and disentangling the possible sources of recall bias by allowing more rigorous estimates of listing bias.

The results of both studies are summarized in table 11. 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 (0.3 ha for weekly visit households) than in Ghana (1.2 ha). The large difference in recall bias (not controlling for baseline differences) is also striking – 18 percent in Ghana, versus 207 percent in Tanzania. What would explain such a large difference 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
Season-wide hours per person per plot	39.5 (1.03)	121.3*** (3.44)	106.4 (2.81)	125.7*** (4.75)

Source: Authors' analysis based on data from the 2014 Tanzania agricultural labor survey (Arthi et al. 2018) and the 2015 Ghana agricultural labor survey.

Note: Standard errors in parentheses. *** Denotes significantly different from weekly visit estimate at the 1 percent level.

Table 12. Bias in Listings of Plots and Household Workers, Ghana and Tanzania

		Tanzania	Ghana
Listing bias (simple difference)	Workers	-1.5***	-1.4***
	Plots	-2.3***	-0.6***
Listing bias (double difference)	Workers	N/A	-1.1***
	Plots	N/A	-0.3***

Source: Authors' analysis based on data from the 2014 Tanzania agricultural labor survey (Arthi et al. 2018) and the 2015 Ghana agricultural labor survey.

Note: *** Denotes the estimator of listing bias due to recall is significant at the 1 percent level.

The information on listing bias in Ghana and Tanzania, which helps to understand the difference in recall bias shown in table 11, are provided in table 12. 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.

The simple difference estimator of bias in the listing of plots is significantly larger in Tanzania (2.3 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 (1.5 vs. 1.4 fewer workers in the recall group). In both experiments, the roster for weekly visit households is “cumulative,” as shown in figs. 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. The fact that listing bias in plots is smaller in Ghana than in Tanzania might partly reflect the fact that recall households in Ghana were administered a baseline plot roster, which was updated once at endline and thereby helped to facilitate recall. As illustrated in section 4, listing bias is the driving force behind recall bias in farm labor per person per plot in Ghana – and this may partly explain the large difference between the two studies highlighted in table 11.¹⁸

Differences in agricultural production systems and rural labor markets could further contribute to the difference in listing bias of plots, though this hypothesis cannot be tested formally with the data at hand. Households in Ghana cultivate, on average, 2.6 plots of 1.2 ha, compared with 5.7 plots of 0.3 ha in Tanzania. The fact that households in Ghana cultivate significantly fewer, yet larger, plots than households in Tanzania could explain why they are able to provide a more accurate listing of plots already at baseline.¹⁹ Labor markets in the study area in Ghana are also more integrated, as evidenced by the

18 Unfortunately, it is not possible to compare the double difference estimator of listing bias across the two countries, which for Ghana is slightly smaller than the simple difference estimator.

19 A comparison of the distribution of person-plot hours between Ghana and Tanzania (not reported, but available from the authors on request) shows that weekly visit households in Tanzania report substantially more person-plots with less than 20 hours of farm labor compared to weekly visit households in Ghana, which could reflect the higher number of plots and smaller plot size in Tanzania. If such very low-frequency farm work is most likely to go unreported in

Table 13. Recall Bias in Aggregated Farm Labor, Descriptive Estimates

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

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

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 in parentheses. T-test on difference in means with *** denoting the difference is significant at the 1 percent level; ** at the 5 percent level.

greater prevalence of agricultural wage labor, and this may lead to a better mental accounting of plot-level labor inputs among the study population.²⁰

Finally, differences in educational attainment could also play a role. In the 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 it has been shown in section 5 that recall bias, and particularly listing bias for plots, declines with the level of education, the lower levels of education in Tanzania compared to Ghana are consistent with the observed differences in recall and listing bias.

7. Implications of Recall Bias for Empirical Research

Sections 4 and 5 analyzed farm labor in season-wide hours per person per plot, which is the unit at which farm labor is reported in this survey. However, as discussed in section 2, other levels of aggregation (particularly farm labor per plot and per household) are of greater interest for most data users. As will be discussed in this section, listing bias – that is, the disproportionate omission of plots and household workers by recall households – mechanically depresses total farm labor hours (aggregated at the plot or household level) for this group. Due to this counteracting effect of listing bias on recall bias in farm labor at the person-plot level, “net” recall bias in farm labor is strongly affected by the aggregation level at which it is reported, both in direction and in magnitude.

Recall Bias at Different Levels of Aggregation

Farm labor hours at different levels of aggregation are summarized in table 13. Comparing recall households to weekly visit households, there is 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.²¹ Farm labor *per plot* (panel B) is about 13 percent lower for recall households, and

end-of-season recall surveys, this may explain why recall bias is larger in Tanzania. In addition, the comparison indicates that recall households in Tanzania are *less* likely to report person-plots with 0–60 hours of farm labor compared to recall households in Ghana, which suggests that Tanzanian recall households had more difficulty reporting marginal person-plot hours. There is no evidence that farm labor is more irregular in Ghana than in Tanzania (results available on request).

20 In Ghana, 90 percent of weekly visit households used hired labor on at least one plot during the season, compared to 28 percent of households in Tanzania.

21 One may expect season-wide hours per person to be underestimated by recall households, due to their omission of plots. However, recall households omit only 0.3 plots (see table 4), which limits the effect this would have on hours per

Table 14. 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	1,521	1,248	454
R2	0.036	0.067	0.204

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

Note: 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 percent level; ** at the 5 percent level.

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 other forms of recall bias in farm labor per person per plot.²² 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*.

In table 14, regression estimates are provided that confirm a significant and negative recall bias in hours per plot and hours per household, controlling for the baseline number of plots, household size, and the average distance to the plots. This is smaller than the descriptive bias reported in table 13, but still large and statistically significant.

Implications 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. As shown in table A1.6 (appendix A1), common measures of agricultural productivity – such as labor productivity per-worker, labor productivity per-hour, and land productivity per-acre – can be 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 nonlisted persons is not captured in the total hours that enter through the denominator. Listing bias in plots has ambiguous effects: since output is measured at the crop-plot level, any omitted plot reduces both the numerator (in terms of nonmeasured output) and the denominator (in terms of nonmeasured hours). The net 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, recall bias in hours per person-plot conditional on being listed needs to be considered. If this bias is positive, as was the case in Tanzania, this would reduce measured productivity by increasing the denominator. However, in the context of this study in Ghana, there is no such recall bias conditional on being listed. Because of all these counteracting effects, the net effect of listing bias in plots, persons, and conditional recall bias is unclear a priori.

person. Furthermore, recall households report slightly more hours per person-plot conditional on the plot being listed at baseline (table 5), which, in this sample, offsets the mechanical reduction in person-hours due to the omission of plots.

- 22 In this setting, listing bias dominates other forms of bias at the plot level and household level, even though recall households omit relatively marginal plots and household workers. This shows that even though these marginal plots and household workers account for fewer farm labor hours than other (i.e., nonmarginal) plots and workers, they are still economically meaningful. In general, the extent to which listing bias outweighs other forms of recall bias in farm labor hours per person per plot at higher aggregation levels depends on the magnitude of each bias.

Table 15. Recall Bias in Land and Labor Productivity (Maize)

	Land productivity (i.e., yields)	Labor productivity (per hour)	Labor productivity (per worker)
	Kg/acre (1)	Kg/hour (2)	Kg/worker (3)
Recall	56.216 (51.136)	6.231* (3.345)	436.454*** (127.417)
N	486	486	486
R2	0.013	0.046	0.101
Control mean	727.35	15.87	902.56

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

Note: All dependent variables are winsorized (1 percent). 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 percent level; * at the 10 percent level. Control mean shows the mean value of the dependent variable for weekly visit households.

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 nonlisted 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 there is no potentially counteracting effect of conditional recall bias. 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 numerator) but may not necessarily reduce the number of workers (the denominator), unless some workers worked exclusively on nonlisted plots.

To assess the impact on different measures of productivity empirically, data on crop production collected at endline are used. As discussed in section 3, the endline survey fielded a standard agricultural production module to all households, which asked households to report the quantity and value of crop harvested during the preceding (i.e., just completed) 2015 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, plot-level measures of land and labor productivity for maize (the most important crop in the sample), in kg per acre, per labor hour, and per worker, are computed.²³ All measures are winsorized at 1 percent, and the analysis is conducted at the plot level. The regression estimates of recall bias are presented in table 15.

The analysis shows no evidence of significant recall bias in maize yields (column 1), which broadly confirms the expectation that measures of land productivity should be fairly robust to the types of measurement bias that are the focus of this study. Maize yields 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, measures of labor productivity in agriculture (columns 2 and 3) are significantly overestimated among recall households in this setting. In line with expectations, the recall effect is more significant (both economically and statistically) for labor productivity per worker (column 3) than for labor productivity per hour (column 2): this is because listing bias has a proportionately larger effect on the number of workers than on the number of hours per listed plot, as the omitted workers are the ones working relatively few hours.

23 Hired labor is not included to maintain simplicity in the comparisons.

8. Conclusion

Data on agricultural labor are a key ingredient in understanding agricultural productivity in developing countries and informing policy making in rural areas. However, due to the seasonality and irregularity of farmers' work schedules such data are difficult to collect, and research on the reliability of agricultural labor data and the best survey methods to collect them is scarce. This paper contributes to this literature by analyzing the data of a randomized survey experiment conducted in Ghana. The study was designed to allow for the comparison of farm labor estimates obtained from a recall survey conducted at the end of the season with data collected weekly throughout the season.

The results show that, in this particular study context, the recall method overestimates farm labor per person per plot by about 10 percent, controlling for differences in baseline characteristics. This recall bias in farm labor at the person-plot level is accounted for by listing bias, as households in the recall group omit marginal plots and farm workers (those with relatively few hours of farm labor over the season). This selective omission of marginal plots and household workers by recall households changes the composition of plots and workers across treatment arms and inflates average farm labor hours per person per plot in the recall group. Besides working few hours, the omitted workers are more likely to be females and workers for whom farming is not their main occupation. Surprisingly, once the analysis controls for listing bias, farmers in this setting report fairly accurately on the number of hours of farm labor provided per person per plot. The results are consistent with the notion that recall households, due to the long reference period, forget to list marginal plots and household members. However, it could also be the case that the length of the reference period subtly changes the way farmers interpret questions about farm labor to exclude marginal plots and farm workers from their response. The study design does not allow disentangling these two mechanisms.

The paper further explored heterogeneity by education and gender. More-educated households are better able to accurately list all their plots, even though they still underreport the number of farm workers. As a result, there is no significant recall bias in hours worked per person per plot for more-educated households (as opposed to households with lower levels of education). This educational gradient suggests that recall and listing bias may be explained by the cognitive burden of reporting on past events. Even though recall households are more likely to omit female workers than male workers, gender difference in recall bias are not significant. This is explained by the fact that females work fewer hours per plot compared to males, so that the omitted marginal female workers are more similar to the listed female workers than is the case for male workers.

While it is common practice in household surveys to collect farm labor at the person-plot level, higher levels of aggregation (especially the plot level) are more widely used at the analysis stage. Because recall households report too few plots and too few farm workers, the end-of-season recall method *underestimates* farm labor per plot (and also per household). In settings such as this in Ghana, where listing bias is the dominant source of reporting bias, plot-level estimates of labor productivity in agriculture are therefore significantly *overestimated* by the recall method.

The findings of this paper offer lessons for researchers and survey practitioners. First, the difference in the magnitude of recall bias between the Ghana and Tanzania studies (Arthi et al. 2018) highlights the need to understand the context in which a specific household survey will be implemented. A potential explanation for this difference in results is that the higher levels of education among smallholder farmers in the Ghanaian setting, combined with socioeconomic and agroecological characteristics of the study area (i.e., fewer but larger plots in Ghana and more integrated rural labor markets) attenuate recall bias in farm labor hours at the person-plot level compared to the Tanzanian setting. Second, the effect of the recall method on bias in farm labor hours depends, both in direction and in magnitude, on the level at which the data are aggregated at the analysis stage. This calls for exercising caution in extrapolating estimates of recall bias from one study to another. Third, the results highlight the need to pay careful

attention to the listings of household members and plots, and the identification of cultivated plots and household members that are engaged in farm work, during survey implementation. Accurate capture of cultivated plots and household farm workers is not only important in and by itself, but it is also an important strategy to minimize recall bias in farm labor hours at all aggregation levels. Future methodological survey research should explore alternative methods to mitigate listing bias; for example, the use of specific terms (e.g., asking explicitly for small plots, kitchen gardens, etc., using local terminology whenever possible), clarifying statements (e.g., “Please identify all household members that worked on each plot, *even if they only did so for a few hours.*”), or increased use of self-respondents (as opposed to proxy respondents). In addition, it would be very useful to have a better understanding of the time profile of listing bias – such as if the respondents’ adding of plots and workers levels off at a specific point during the agricultural season or after a certain number of interactions with the interviewer – as this would provide further insights on what modifications to the way surveys are operationalized might help to reduce listing bias.

Appendix A1: Additional Tables

Table A1.1. Measures of Farm Labor and Aggregation Approaches Used by Empirical Studies

Author(s)	Literature strand	Aggregation of labor data for analysis
Aguilar et al. (2015)	Gender differences	Labor per plot (normalized for plot size)
Ali et al. (2016)	Gender differences	Labor per plot (normalized for plot size)
André, Delesalle, and Dumas (2019)	Child labor	Labor per household
Andrews, Golan, and Lay (2015)	Gender differences	Labor per plot (normalized for plot size)
Damon and McCarthy (2019)	Gender differences	Labor per plot (normalized for plot size)
de la O Campos, Covarrubias, and Patron (2016)	Gender differences	Labor per plot (normalized for plot size)
Dillon, Brummund, and Mwabu (2019)	Rural labor markets	Labor per household
Fisher and Kandiwa (2014)	Gender differences	Labor per plot (normalized for plot size)
Gollin and Udry (2019)	Agricultural productivity	Labor per plot (normalized for plot size)
Kilic, Palacios-Lopez, and Goldstein (2015)	Gender differences	Labor per plot (normalized for plot size)
McCullough (2017)	Productivity gaps across sectors	Labor per person
Oseni et al. (2015)	Gender differences	Labor per plot (normalized for plot size)
O’Sullivan et al. (2014)	Gender differences	Labor per plot (normalized for plot size)
Palacios-Lopez and Lopez (2015)	Gender differences	Labor per plot (normalized for plot size)
Restuccia and Santaaulalia-Llopis (2017)	Agricultural productivity	Labor per household

Source: Authors’ compilation.

Table A1.2. Sample Characteristics in Baseline Survey by Attrition Status

	Attritor	Non-attritor
Household size	3.739 (0.427)	5.444*** (0.142)
Male household head	0.739 (0.094)	0.780 (0.019)
Number of children (age < 10)	1.261 (0.283)	1.610 (0.069)
Number of persons who worked on household plot (age 10+)	1.304 (0.277)	2.133*** (0.067)
Number of plots	2.043 (0.213)	2.458 (0.061)
<i>N (households)</i>	23	459
Age	23.767 (2.160)	22.865 (0.387)
Male	0.465 (0.054)	0.497 (0.010)
Worked on household plot since start of season (age 10+)	0.566 (0.069)	0.565 (0.012)
Has no schooling (age 10+)	0.439 (0.066)	0.317* (0.011)
<i>N (individuals, all ages)</i>	86	2,499
Distance plot to residence (min. walking)	35.213 (4.159)	58.016*** (1.486)
Proportion of plots cultivating beans/peas	0.106 (0.045)	0.121 (0.010)
Proportion of plots cultivating cassava	0.106 (0.045)	0.111 (0.009)
Proportion of plots cultivating groundnuts	0.191 (0.058)	0.211 (0.012)
Proportion of plots cultivating maize	0.532 (0.074)	0.464 (0.015)
Proportion of plots cultivating yam	0.149 (0.052)	0.239 (0.013)
<i>N (plots)</i>	47	1,128

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

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

Table A1.3. Effect of the Recall Method on Maize Yields of Baseline Plots

	Kg/acre (1)
Recall	15.713 (55.525)
N	413
R ²	0.007

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

Note: Dependent variables are winsorized (1 percent). 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 A1.4. Heterogeneity in Recall Bias in Farm Labor per Person per Plot, Descriptive Estimates

A: By education	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: By gender	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	

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

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 *** denotes the difference is significant at the 1 percent level; * at the 10 percent level.

Table A1.5. Summary Statistics Weekly Visit Sample, by Household Education

	Low education	High education
Persons		
Minimum person-hours	210.050 (26.892)	119.603*** (13.731)
Mean person-hours	409.816 (27.525)	331.118** (18.102)
Maximum person-hours	655.930 (34.603)	621.579 (28.340)
No. of workers	3.430 (0.178)	4.325*** (0.197)
No. of workers, main occupation farming	2.390 (0.160)	2.714 (0.173)
No. of workers, other main occupation	1.040 (0.119)	1.611*** (0.153)
No. of male workers	1.750 (0.126)	2.151** (0.119)
No. of female workers	1.680 (0.106)	2.175*** (0.129)
Plots		
Minimum plot-hours	306.800 (45.804)	288.310 (32.429)
Mean plot-hours	496.546 (45.569)	514.667 (40.671)
Maximum plot-hours	739.920 (52.936)	776.532 (59.828)
No. of plots	2.990 (0.127)	3.071 (0.139)
N	100	126

Source: Authors' analysis based on data from the 2015 Ghana agricultural labor survey.

Note: The rows minimum (maximum) person-hours show the average, across households, of hours worked by the household member who worked the smallest (highest) number of hours; similar for plot-hours. Standard errors in parentheses. T-test for difference between the two groups were done: *** denotes the difference is significant at the 1 percent level; ** at the 5 percent level.

Table A1.6. Effect of the Recall Method on Measures of Agricultural Productivity, Mechanisms

Productivity concept	Plot		Unit of analysis		Household Direction of bias (recall method)
	Measure	Direction of bias (recall method)	Measure	Measure	
Labor productivity	per-worker Output (plot)/workers (plot)	<ul style="list-style-type: none"> * Listing bias, persons: Pr↑ * Listing bias, plots: Pr↓ (nonlisted plots more productive than average) or Pr↑ (nonlisted plots less productive than average), ambiguous * Conditional recall bias in hours per person-plot: no effect → <i>net effect ambiguous, expectation Pr↑</i>	Output (household)/workers (household)	<ul style="list-style-type: none"> * Listing bias, persons: Pr↑ * Listing bias, plots: probably Pr↓ (unless some workers worked exclusively on nonlisted plot) * Conditional recall bias in hours per person-plot: no effect → <i>net effect ambiguous</i>	
	per-hour Output (plot)/hours (plot)	<ul style="list-style-type: none"> * Listing bias, persons: Pr↑ * Listing bias, plots: Pr↓ (nonlisted plots more productive than average) or Pr↑ (nonlisted plots less productive than average) * Conditional recall bias in hours per person-plot: Pr↓ (if bias is positive) or Pr↑ (if bias is negative) → <i>net effect ambiguous</i>	Output (household)/hours (household)	<ul style="list-style-type: none"> * Listing bias, persons: Pr↑ * Listing bias, plots: Pr↓ (nonlisted plots more productive than average) or Pr↑ (nonlisted plots less productive than average) * Conditional recall bias in hours per person-plot: Pr↓ (if bias is positive) or Pr↑ (if bias is negative) → <i>net effect ambiguous</i>	
Land productivity	per-acre Output (plot)/acres (plot)	<ul style="list-style-type: none"> * Listing bias, persons: no effect * Listing bias, plots: Pr↓ (nonlisted plots more productive than average) or Pr↑ (nonlisted plots less productive than average) * Conditional recall bias in hours per person-plot: no effect → <i>net effect ambiguous</i>	Output (household)/acres (household)	<ul style="list-style-type: none"> * Listing bias, persons: no effect * Listing bias, plots: Pr↓ (nonlisted plots more productive than average) or Pr↑ (nonlisted plots less productive than average) * Conditional recall bias in hours per person-plot: no effect → <i>net effect ambiguous, expect effect to be relatively small</i>	
	per-acre Output (plot)/acres (plot)	<ul style="list-style-type: none"> * Listing bias, persons: no effect * Listing bias, plots: Pr↓ (if bias is positive) or Pr↑ (if bias is negative) → <i>net effect ambiguous</i>	Output (household)/acres (household)	<ul style="list-style-type: none"> * Listing bias, persons: no effect * Listing bias, plots: Pr↓ (nonlisted plots more productive than average) or Pr↑ (nonlisted plots less productive than average) * Conditional recall bias in hours per person-plot: no effect → <i>net effect ambiguous, expect effect to be relatively small</i>	

Source: Authors' compilation.
Note: Pr denotes productivity.

References

- Abebe, G. 2013. "Recall Bias in Retrospective Surveys: Evidence from Enterprise-Level Data." *Asian-African Journal of Economics and Econometrics* 13 (1): 17–33.
- Aguilar, A., E. Carranza, M. Goldstein, T. Kilic, and G. Oseni. 2015. "Decomposition of Gender Differentials in Agricultural Productivity in Ethiopia." *Agricultural Economics* 46 (3): 311–34.
- Ali, D., D. Bowen, K. Deininger, and M. Duponchel. 2016. "Investigating the Gender Gap in Agricultural Productivity: Evidence from Uganda." *World Development* 87: 152–70.
- André, P., E. Delesalle, and C. Dumas. 2019. "Returns to Farm Child Labor in Tanzania." Discussion Paper No. 2019-5. Institut de Recherches Économiques et Sociales de l'Université catholique de Louvain, Louvain-la-Neuve, Belgium.
- Andrews, M. J., J. Golan, and J. Lay. 2015. "Inefficiency of Male and Female Labor Supply in Agricultural Households: Evidence from Uganda." *American Journal of Agricultural Economics* 97 (3): 998–1019.
- Arthi, V., K. Beegle, J. De Weerd, and A. Palacios-Lopez. 2018. "Not Your Average Job: Measuring Farm Labor in Tanzania." *Journal of Development Economics* 130: 160–72.
- Beckett, M., J. Da Vanzo, N. Sastry, C. Panis, and C. Peterson. 1999. "The Quality of Retrospective Reports in the Malaysian Family Life Survey." Working Paper Series 99-13, RAND, Santa Monica, CA, USA.
- . 2001. "The Quality of Retrospective Data: An Examination of Long-Term Recall in a Developing Country." *Journal of Human Resources* 36 (3): 593–625.
- Beegle, K., C. Carletto, and K. Himelein. 2012. "Reliability of Recall in Agricultural Data." *Journal of Development Economics* 98 (1): 34–41.
- Blackden, M. C., and Q. Wodon. 2006. "Gender, Time Use, and Poverty in Sub-Saharan Africa." Working Paper No. 73, World Bank, Washington, DC, USA.
- Christiaensen, L., L. Demery, and J. Kuhl. 2011. "The (Evolving) Role of Agriculture in Poverty Reduction – An Empirical Perspective." *Journal of Development Economics* 96 (2): 239–54.
- Conrad, F. G., N. R. Brown, and E. R. Cashman. 1998. "Strategies for Estimating Behavioural Frequency in Survey Interviews." *Memory* 6 (4): 339–66.
- Damon, A. L., and A. S. McCarthy. 2019. "Partnerships and Production: Agriculture and Polygyny in Tanzanian Households." *Agricultural Economics* 50 (5): 527–42.
- de Janvry, A., E. Sadoulet, and T. Suri. 2017. "Field Experiments in Developing Country Agriculture." In *Handbook of Field Experiments*. Vol. 2. Edited by E. Duflo and A. V. Banerjee, 427–66. Amsterdam: North-Holland.
- de la O Campos, A. P., K. A. Covarrubias, and A. P. Patron. 2016. "How Does the Choice of the Gender Indicator Affect the Analysis of Gender Differences in Agricultural Productivity? Evidence from Uganda." *World Development* 77: 17–33.
- Dillon, B., P. Brummund, and G. Mwabu. 2019. "Asymmetric Non-separation and Rural Labor Markets." *Journal of Development Economics* 139: 78–96.
- FAO (Food and Agriculture Organization of the United Nations). 2017. "Measuring Work in Agricultural Surveys and Censuses: A Review." Unpublished.
- Fermont, A., and T. Benson. 2011. "Estimating Yield of Food Crops Grown by Smallholder Farmers. A Review in the Ugandan Context." IFPRI Discussion Paper 01097. International Food Policy Research Institute, Washington, DC, USA.
- Fisher, M., and V. Kandiwa. 2014. "Can Agricultural Input Subsidies Reduce the Gender Gap in Modern Maize Adoption? Evidence from Malawi." *Food Policy* 45: 101–11.
- Gollin, D., and C. R. Udry. 2019. "Heterogeneity, Measurement Error and Misallocation: Evidence from African Agriculture." NBER Working Paper No. 25440, National Bureau of Economic Research, Cambridge, MA USA.
- Gollin, D., D. Lagakos, and M. E. Waugh. 2014a. "Agricultural Productivity Differences across Countries." *American Economic Review: Papers & Proceedings* 104 (5): 165–70.
- . 2014b. "The Agricultural Productivity Gap." *Quarterly Journal of Economics* 129 (2): 939–93.
- Igou, E., H. Bless, and N. Schwarz. 2002. "Making Sense of Standardized Survey Questions: The Influence of Reference Periods and Their Repetition." *Communication Monographs* 69 (2): 179–87.
- ILO (International Labour Organization). 2016. ILOSTAT – Employment by Sector (ILO modeled estimates, Nov. 2016). Accessed June 23, 2017. http://www.ilo.org/ilostat/faces/oracle/webcenter/portalapp/pagehierarchy/Page3.jsp?MBI_ID=33.

- Kilic, T., A. Palacios-Lopez, and M. Goldstein. 2015. "Caught in a Productivity Trap: A Distributional Perspective on Gender Differences in Malawian Agriculture." *World Development* 70: 416–63.
- Kjellsson, G., P. Clarke, and U.-G. Gerdtham. 2014. "Forgetting to Remember or Remembering to Forget: A Study of the Recall Period Length in Health Care Survey Questions." *Journal of Health Economics* 35: 34–46.
- LaFave, D., and D. Thomas. 2016. "Farms, Families and Markets: New Evidence on Completeness of Markets in Agricultural Settings." *Econometrica* 84 (5): 1917–60.
- McAuliffe, T. L., W. DiFranceisco, and B. R. Reed. 2010. "Low Numeracy Predicts Reduced Accuracy of Retrospective Reports of Frequency of Sexual Behavior." *AIDS Behav* 14 (6): 1320–29.
- McCullough, E. B. 2017. "Labor Productivity and Employment Gaps in Sub-Saharan Africa." *Food Policy* 67: 133–52.
- Menon, G. 1993. "The Effects of Accessibility of Information in Memory on Judgments of Behavioral Frequencies." *Journal of Consumer Research* 20 (3): 431–40.
- Mwungu, C. M., C. Mwongera, K. M. Shikuku, F. N. Nyakundi, J. Twyman, L. A. Winowiecki, E. L. Ampaire, M. Acosta, and P. Läderach, 2017. "Survey Data of Intra-household Decision Making and Smallholder Agricultural Production in Northern Uganda and Southern Tanzania." *Data in Brief* 14: 302–6.
- Oseni, G., P. Corral, M. Goldstein, and P. Winters. 2015. "Explaining Gender Differentials in Agricultural Production in Nigeria." *Agricultural Economics* 46 (3): 285–310.
- O'Sullivan, M., A. Rao, R. Banerjee, K. Gulati, and M. Vinez. 2014. "Levelling the Field: Improving Opportunities for Women Farmers in Africa." Documents & Reports. World Bank and ONE Campaign, Washington, DC, USA.
- Palacios-Lopez, A., and R. Lopez. 2015. "The Gender Gap in Agricultural Productivity: The Role of Market Imperfections." *Journal of Development Studies* 51 (9): 1175–92.
- Peters, H. E. 1988. "Retrospective Versus Panel Data in Analyzing Lifecycle Events." *Journal of Human Resources* 23 (4): 488–513.
- Reardon, T., and P. Glewwe. 2000. "Agriculture." In *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*. Vol. 2. Edited by M. Grosh and P. Glewwe, 139–81. Washington, DC: World Bank.
- Restuccia, D., and R. Santaaulalia-Llopis. 2017. "Land Misallocation and Productivity." NBER Working Paper No. 23128, National Bureau of Economic Research, Cambridge, MA, USA.
- Schwarz, N., and D. Oyserman. 2001. "Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction." *American Journal of Evaluation* 22 (2): 127–60.
- Walker, T. 2009. "Appendices: Technical Details on the 2009 Household Survey." <http://barrett.dyson.cornell.edu/files/research/databases/ghana/2009-Walker-Appendix-Summary.pdf>.
- Winkielman, P., B. Knäuper, and N. Schwarz. 1998. "Looking Back at Anger: Reference Periods Change the Interpretation of Emotion Frequency Questions." *Journal of Personality and Social Psychology* 75 (3): 719–28.
- World Bank. 2015a. Ghana Agricultural Labor Survey. <http://surveys.worldbank.org/lsms/our-work/survey-methods/work-and-employment>.
- . 2015b. LSMS – Integrated Surveys on Agriculture. <http://surveys.worldbank.org/lsms/programs/integrated-surveys-agriculture-ISA>.
- . 2016. "Malawi Labor Experiment. Agriculture Questionnaire. Interviewer Manual." Unpublished.
- . 2017. World Development Indicators. Accessed June 23, 2017. <http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators>.