Measuring farm labor: survey experimental evidence from Ghana[+]

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

[*] World Bank [**] Wageningen University

October 2017

Preliminary, please do not quote

Abstract:

This study examines recall bias in farm labor by conducting a randomized survey experiment in Ghana. Estimates of farm labor obtained from a recall survey conducted at the end of the season are compared against data collected weekly throughout the season. We find that the recall method overestimates farm labor per person per plot by 15 percent, controlling for observable differences at baseline, which is much lower than in a similar study conducted in Tanzania (Arthi et al 2016). Recall bias in farm labor is accounted for by the fact that households in the recall group report fewer marginal plots and farm workers, what we call listing bias. This (negative) listing bias runs counter to (positive) recall bias in farm labor at the person-plot level and dominates at higher levels of aggregation, such as at the plot- and household-levels. Hence, the recall method underestimates farm labor per plot and per household and overestimates labor productivity for household-operated farms. Consistent with the notion that recall bias in farm labor is linked to the cognitive burden of reporting on past events, we find that recall bias has a strong educational gradient.

JEL: C8, J22, O12, Q12

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

+ The authors gratefully acknowledge funding from the U.K. Department for International Development (DFID) under the "Minding the (Data) Gap: Improving Measurements of Agricultural Productivity through Methodological Validation and Research" project and from the William and Flora Hewlett Foundation under the "Improving the Measurement of Subsistence Agriculture in the Framework of ICLS 19 and Its Implications for Gender Statistics" project. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

1. Introduction

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

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

Measuring farm labor is, however, fraught with empirical difficulties. The most common approach is to ask survey respondents to recall the 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 to remember how much time various members of the household worked on the farm over the entire season. Recall is further hindered by the informal nature of smallholder agriculture, where working hours are highly irregular, with periods of lower labor activity after planting and prior to harvest. In addition, since irrigation is limited and rain-fed agriculture the norm across much of Africa, there is added unpredictability with the reliance on weather. These features render measurement of farm labor in developing countries extraordinarily challenging.

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

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

[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.

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

This paper makes the following contributions. First, we document that recall bias in farm labor is much lower in the Ghanaian than Tanzanian study context, which suggests important regional heterogeneity and exercising caution in extrapolating estimates of measurement bias across studies and country contexts. Second, we show that recall bias in farm labor per person per plot, which is the unit at which labor is reported by the survey respondents, is accounted for by recall bias in listings of plots and farm workers, what we denote as *listing bias*. This is an important refinement to the conclusions reached by Arthi et al (2016), who document that bias in listings of plots and household workers runs counter to recall bias in farm labor at the person-plot level. Our results show that recall households report more hours of farm labor per person per plot *because* they fail to list 'marginal' plots and farm workers. Listing bias hence not only counteracts, but *explains* recall bias in farm labor at the person-plot level. Third, our findings show that recall bias declines with the household's level of education. This finding is consistent with the notion that recall bias is linked to the cognitive burden of reporting on past events. This educational gradient – together with variations in the design of the experiment – may also explain (at least in parts) the difference in results between Arthi et al (2016) and this paper, given that education levels among farmers are higher in the Ghanean districts covered by this study than in the Mara region of Tanzania.

The findings in this paper have several implications. In general, the presence of recall bias suggests that academics and policymakers ought to tread carefully when analyzing the state of agriculture, as biased data can lead to misguided policies. This paper, however, documents that the magnitude of recall bias depends on the level of aggregation and the characteristics of the population under investigation, which has implications for within- and cross-country comparisons. For example, the analysis in section 4 of this paper suggests that end-of-season recall significantly overestimates measures of agricultural productivity. Finally, this study highlights the importance of investing in quality data to support evidenced-based policy making and of periodically examining the reliability of such data.

The remainder of the paper is structured as follows. Section 2 describes the experimental design of the study. Section 3 estimates recall bias in farm labor and explores proxy determinants, different levels of aggregation and heterogeneity across sub-populations. It also compares the results in Ghana to those obtained by Arthi et al (2016) for Tanzania and puts forth some tentative explanations for why they differ. Section 4 documents implications of recall bias in farm labor for the analysis of agricultural productivity. Section 5 concludes.

2. Study design and context

The survey experiment was conducted during the main rainy season (roughly March to September 2015) in the Mampong Municipal, Ejura Sekyedumasi, Nkoranza South, and Pru districts of Ghana. A random sample of 720 agricultural households was selected from 20 enumeration areas (henceforth denoted as villages) and then randomly assigned to one of three alternative survey designs. All households were administered a baseline survey at the beginning of the season, which collected a roster of plots and household members working in agriculture, and basic demographic and plot-specific information. After the main harvest, households were administered an endline survey, which was modeled after the design of the LSMS-ISA survey series (i.e. a multi-topic household survey with an agricultural production module).[2] In terms of capturing farm labor, the three survey designs differed as follows:

A. **Weekly visit:** Following the baseline survey in March 2015, these households were visited weekly from April to September 2015, which in turn was followed by the endline survey in October/November 2015. At baseline (visit 0), households reported on the number of days worked and total number of hours worked, per person per plot, since starting preparing the plot for the season. During the season households were visited every week (visits 1-23), and asked to report on the number of hours worked on each day of the past week, per person per plot, and the range of activities (but not hours per activity).[3] At endline (visit 24), households report the total number of days and the average hours per day, per person per plot, since the last weekly visit. The design for this group minimizes recall periods and (mostly) avoids the need for complex aggregations –farm labor estimates for this group are considered as the benchmark.

B. **Weekly phone:** The design for this group is essentially the same as for the weekly visit group, what differs is the method of data collection. While households in the weekly visit group received weekly face-to-face visits during the season, households in the weekly phone group received weekly phone calls. The main purpose of this treatment arm is to explore the potential of soliciting high-frequency labor data using phone surveys, an aspect of the experiment that will be explored in a companion paper.

C. **Recall:** For this group, there were no visits between the baseline in March 2015 and the endline in October/November 2015. In the endline survey, the agricultural labor module for this group identified the household members that worked on each of the household's plots during the season, and for those members, data was collected on (i) total days spent across five activities (land preparation and planting; weeding; ridging, fertilizer application and other non-harvest activities; harvesting; supervision) and (ii) typical hours per day worked by an individual on these five activities.

For the sake of parsimony, the focus of this paper is the comparison of the weekly visit group and the recall group – in other words, we do *not* include any analysis of the data from the weekly phone group.

While the overall design of this study is very similar to Arthi et al (2016), there is one significant difference: This study fielded a baseline survey to *all* households, including recall households, which allowed us to collect a baseline listing of plots and household workers. This listing of plots and household workers plays a crucial role, because all data on farm labor is collected at the person-plot level. However, respondents could add plots and household workers that were not listed at baseline during subsequent visits – that is during visits 1 to 24 (weekly visits, endline) in the weekly visit group, and during visit 24 (endline) in the recall group. This design variation follows from the results in Arthi et al (2016), who document that underreporting of plots and household workers in the recall group counteracts overestimation of farm labor per person per plot. Having access to a baseline survey for all households allows us to explore in greater detail bias in listings of plots and household workers, but it may also explain some of the difference in

[2] The baseline and endline surveys were timed as per the production calendar of maize, the most important crop in this area.
[3] We distinguish between five activities - land preparation and planting; weeding; ridging, fertilizing and other non-harvest activities; harvesting; and supervision.

results between this study and Arthi et al (2016). We will revert to this issue in section 3.5, where we compare the results obtained for Ghana and Tanzania.

It is important to emphasize that the approach adopted in this study of using the plot as the main unit of data collection is common practice in many (though not all) agricultural surveys and censuses (FAO 2017). As discussed in Reardon and Glewwe (2000), in countries characterized by 'hard-to-survey' agriculture (which includes most of Africa), data on farm inputs and practices is preferably collected at the plot-level (as opposed to the farm-level). This is because a single farm may consist of multiple plots – cultivated under different production systems – and farmers tend to refer to each plot as they describe agricultural production activities. Using the plot as the unit of analysis hence reflects the natural flow of conversation between the interviewer and the farmer. In addition, plot-level data collection yields more observations and facilitates the analysis of agricultural productivity (as the plot can be used to link crops to inputs) and intrahousehold allocations (e.g. differences between plots managed by male and female farmers). The person-plot level then is the natural unit of reporting for data on farm labor provided by individual members of the household.

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

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

Table 1 summarizes household and plot characteristics across the weekly visit and recall arms, drawing on the baseline survey. For most of the traits presented in this table, households are well balanced across survey arms. In the weekly visit group, the average household consists of 5.8 members, of whom 1.6 are children younger than 10 years old, and cultivates 2.6 plots. The average plot is located within 53 minutes' walk from the households' residence. Of all individuals aged 10 and older, 55 percent worked on one or more household plots between the start of the season and the baseline survey. However, households in the recall group are significantly smaller and have fewer cultivated plots compared to households in the weekly visit group. Because of these baseline differences, our empirical analysis controls for the number of plots and household workers at baseline. Households in the recall group also have a greater distance to their plots, which are located at almost 10 minutes more walking time than in the weekly visit group. Cropping patterns are similar across arms, with maize being the most important crop across all arms, followed by yam and groundnuts.

Table 1: Sample characteristics in baseline survey

	Weekly visit	Recall
Age	22.974	22.744
	(0.514)	(0.584)
Male	0.493	0.500
	(0.014)	(0.015)
Worked on household plot since start of season (age 10+)	0.550	0.582
	(0.016)	(0.017)
Enrolled in school (age 4-15)	0.848	0.894**
	(0.017)	(0.015)
Has no schooling (age 10+)	0.307	0.329
	(0.015)	(0.016)
N (individuals, all ages)	1314	1185
Household size	5.814	5.086**
	(0.228)	(0.168)
Male household head	0.752	0.807
	(0.029)	(0.026)
Household head single/divorced/widowed	0.212	0.262
	(0.027)	(0.029)
Number of children (age <10)	1.646	1.575
	(0.100)	(0.096)
Number of persons who worked on household plot (age 10+)	2.226	2.043
	(0.093)	(0.097)
Number of plots per household	2.571	2.348*
	(0.085)	(0.088)
N (households)	226	233
Distance plot to residence (min. walking)	53.601	62.706***
	(1.958)	(2.234)
Proportion of plots cultivated during main rainy season	0.976	0.952**
	(0.006)	(0.009)
Proportion of plots cultivating beans/peas	0.119	0.122
	(0.013)	(0.014)
Proportion of plots cultivating cassava	0.100	0.122
	(0.012)	(0.014)
Proportion of plots cultivating groundnuts	0.203	0.219
	(0.017)	(0.018)
Proportion of plots cultivating maize	0.487	0.439
	(0.021)	(0.021)
Proportion of plots cultivating yam	0.222	0.258
	(0.017)	(0.019)
N (plots)	581	547

Note: Standard errors in parentheses. T-test for difference between the two groups were done: *** denotes the difference is significant at the 1% level; ** at the 5% level; * at the10% level.

3. Recall bias in farm labor

3.1 Main results

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

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

Table 2: Recall bias in farm labor, descriptive estimate

	Weekly visit	Recall	Difference
Season-wide hours per person per plot	106.40	125.72	19.33***
	(2.81)	(4.75)	(5.16)
Number of person-plots	2787	1675	

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

As discussed in the previous section, households in the weekly visit group were slightly larger and cultivated slightly more plots at baseline than households in the recall group and it is important to adjust for these differences. **Table 3** pools recall and weekly visit households to regress season-wide hours per person per plot on an indicator variable (*Recall*), which equals to unity for households in the recall group and controls for the household number of plots and workers at baseline. This delivers an estimate of conditional recall bias of 16 hours (15 percent) per person per plot. Though slightly lower than the unconditional estimate, it is still significant at the 1 percent level.

Table 3: Recall bias in farm labor, regression estimate

	Season-wide hours per person per plot
Recall	16.182***
	(5.165)
Hh number of workers at baseline	1.329
	(1.615)
Hh number of plots at baseline	-10.160***
	(1.603)
Constant	130.361***
	(6.109)
N	4462
R2	0.012

*** denotes significance at the 1% level; ** at the 5% level; * at the10% level.

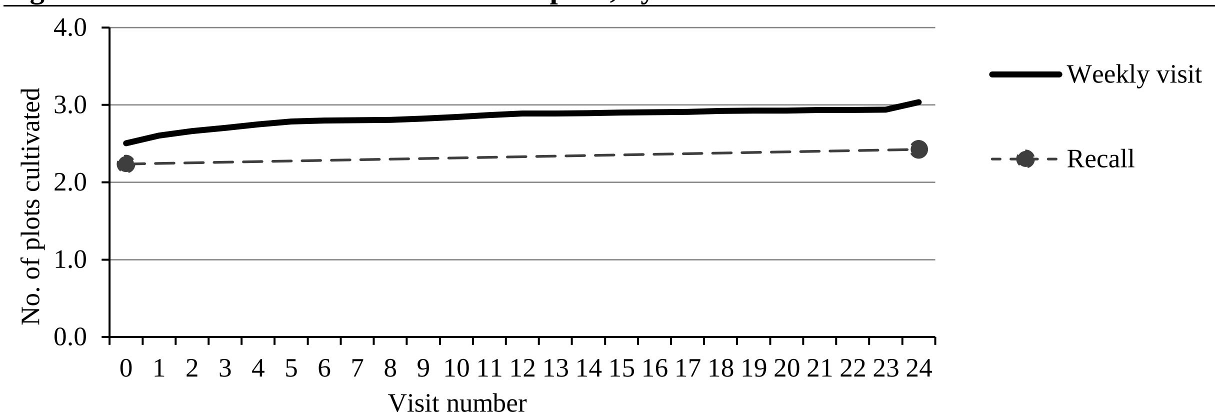
Providing accurate information of farm labor requires respondents to recall several components, including an accurate listing of cultivated plots, an accurate listing of household members working on those plots. The next section explores an important proxy determinant of recall bias in farm labor – bias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

3.2 Proxy determinants – listings of plots and household workers

Let's first recap how respondents in the recall and weekly visit groups provide a listing of plots cultivated by the household. Recall households were interviewed twice – at the beginning of the season (baseline = visit 0) and after the main harvest (endline = visit 24). 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. However, given the lag between the actual farm labor input and the timing of the endline, recall households may not report precisely on all the plots cultivated during the season. Conversely, households in the weekly visit group were visited 23 times between the baseline and the endline, and during each these visits they could add plots that were currently owned and/or had been cultivated since the last visit but had not yet been listed. These weekly visits could make it easier to recall and list all the plots cultivated during the season.

Figure 1 shows the cumulative number of cultivated plots listed per household, that is, the number of cultivated plots reported up to a given week (visit number) in the season. At baseline, households in the weekly visit group reported on average 2.5 plots for cultivation, compared with 2.2 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 2.4 plots for recall households. Both at baseline and at endline, the difference is statistically significant, but the fact that the gap doubles over time suggests that the survey method accounts for a significant part of the difference.

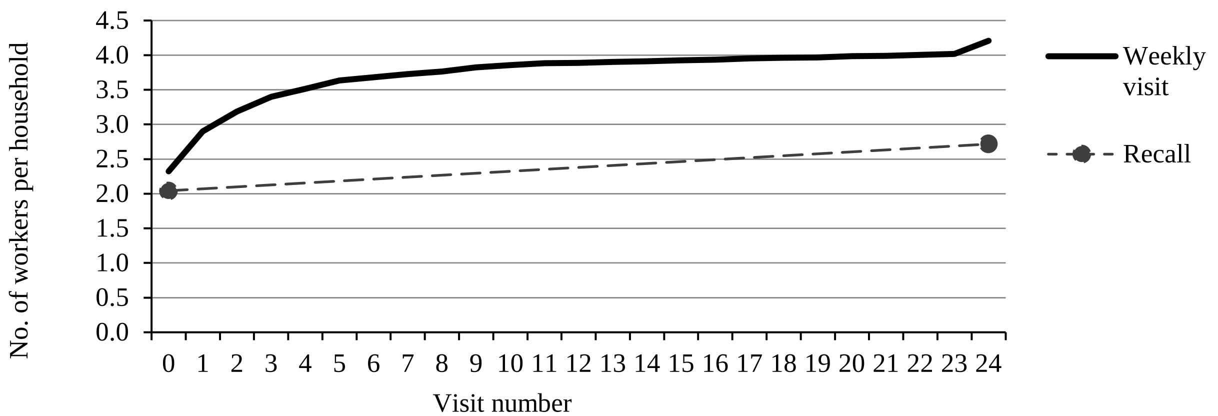
Figure 1: Cumulative number of cultivated plots, by visit number



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

We now turn to the listing of household members working in agriculture. Households in the recall group reported at baseline and at endline on which household members had been working on each plot. In contrast, households in the weekly visit group reported on a weekly basis. Figure 2 shows the number of farm workers per household, calculated as the cumulative total of persons aged 10 and older who have been listed as having worked on one of the household's plots. At endline, households in the recall arm report 2.7 farm workers on average, compared to a cumulative total of 4.2 workers in weekly visit households. The difference is much smaller at baseline (2.0 vs. 2.3 workers), again suggesting that the survey method matters for the number of household workers reported.

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



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

To quantify the effect of the survey method on the number of plots and household workers we use a double difference estimator (**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. Both double difference estimators of listing bias due to the recall method – -0.3 for plots and -1.2 for household workers – are statistically significant.

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

	No. of cultivated plots listed per household	No. of workers listed per household
Endline	0.531***	1.885***
	(0.063)	(0.118)
Endline * Recall	-0.342***	-1.211***
	(0.088)	(0.166)
Constant	2.368***	2.181***
	(0.031)	(0.059)
N	918	918
R2	0.150	0.386

*** denotes significance at the 1% level; ** at the 5% level; * at the10% level.

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

The results in columns (2) to (4) do not provide any evidence for the second hypothesis that recall bias differs between plots (household workers) listed at baseline and those that were added later, as the relevant interaction effects are insignificant in all three specifications. Conversely, we do see that reported season-wide hours per person per plot are significantly higher for plots and household workers listed at baseline than for plots and household workers added later (columns (2) to (4)) and controlling for this characteristic turns the recall variable insignificant. Together, these results provide significant support for the first hypothesis that compositional differences explain recall bias in farm labor. Season-wide hours of farm labor per person per plot are higher for recall households because the latter fail to list several plots and persons. These omitted plots and persons have, on average, less farm labor than the plots and persons that were listed at baseline. In other words, weekly visit households report more ‘marginal’ plots and farm workers than recall households, which reduces their total average hours per person per plot.

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

	Hours per person per plot			
	(1)	(2)	(3)	(4)
Recall	16.182***	10.289	2.612	-4.818
	(5.165)	(13.024)	(8.22)	(13.624)
Plot listed at baseline		73.393***		64.657***
		(8.194)		(7.928)
Recall * Plot listed at baseline		2.21		4.854
		(14.151)		(13.715)
Person listed as farm worker at baseline			97.359***	93.937***
			(6.278)	(6.234)
Recall * Person listed as farm worker at baseline			5.635	5.74
			(10.333)	(10.276)
Hh number of farm workers at baseline	1.329	0.551	-8.425***	-8.827***
	(1.615)	(1.605)	(1.638)	(1.629)
Hh number of plots at baseline	-10.160***	-13.382***	-11.388***	-14.223***
	(1.603)	(1.609)	(1.548)	(1.557)
Constant	130.361***	81.119***	105.126***	62.906***
	(6.109)	(8.643)	(6.599)	(8.676)
N	4462	4462	4462	4462
R2	0.012	0.038	0.087	0.107

*** denotes significance at the 1% level; ** at the 5% level; * at the10% level.

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

Different cognitive processes may explain why households in the recall group report fewer ‘marginal’ plots and fewer ‘marginal’ household workers. The most obvious explanation is a simple failure of memory, where respondents forget to report plots that were not farmed intensively and household workers who provided comparatively few hours of farm labor during the season. This failure of memory is plausibly linked to the length of the recall period and could hence why households in the recall group list fewer plots and household workers than households in the weekly visit group. However, there is also evidence from cognitive research that the length of the recall period may have a bearing on how respondents interpret survey questions. Schwarz (2007), for example, shows that the longer the recall period, the more likely respondents are to interpret a given question as referring to major events only (see Arthi et al 2016 for a discussion) and this could also explain why recall households report fewer marginal plots. This potential change in inferred meaning makes it difficult to distinguish between the effects of question interpretation and forgetting.

3.3 Recall bias at different levels of aggregation

Listing bias has additional effects on estimates of farm labor if we consider alternative levels of aggregation. Sections 3.1 and 3.2 measured farm labor in season-wide hours per person per plot, which is the unit at which farm labor is reported in our survey. However, depending on the type of analysis, other levels of aggregation might be of interest. As we show in this section, recall bias in farm labor is strongly affected by the level of aggregation, both in direction and in magnitude.

Agricultural productivity analysis typically uses input and yields data per plot and hence requires total labor input per plot (e.g. season-wide hours *per plot*). Plot-level measures of farm labor are sensitive to household workers not being listed, because the labor input of these workers would not be accounted for. Conversely, labor market analysis is typically interested in labor per individual (e.g. season-wide hours *per person)*, as a measure of the intensive margin of labor supply. Such measures are sensitive to plots not being listed, given that an individual’s labor input on these non-reported plots would not be captured.

Table 6 summarizes farm labor at different levels of aggregation. Comparing recall households to weekly visit households, we find very similar farm labor *per person* (panel A). Hence the recall bias in season-wide hours per person per plot (in **Table 2**) is nullified by the lower number of plots listed. Farm labor *per plot* (panel B) is about 13 per cent lower for recall households and total farm labor *per household* (panel C) is about 30 percent lower compared to weekly visit households and both differences are statistically significant. At these levels of aggregation, listing bias hence dominates recall bias in farm labor per person per plot. Or in other words, because recall households report farm labor for too few plots and (especially) too few household workers, farm labor per plot and per household is *underestimated*, even though farm labor per person per plot is *overestimated*.

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

	Weekly visit	Recall	Difference
A. Per person (all household plots)			
Hours	333.95	332.68	-1.27
	(10.72)	(18.80)	(20.33)
Number of persons	888	633	
B. Per plot (all household persons)			
Hours	432.29	374.71	-57.57**
	(17.90)	(20.96)	(27.41)
Number of plots	686	562	
C. Per household (all persons and plots)			
Hours	1312.16	923.63	-388.53***
	(63.50)	(86.09)	(107.12)
Number of households	226	228	

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

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

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

	Hours per person	Hours per plot	Hours per household
Recall	10.232	-57.455**	-286.949***
	(20.258)	(26.863)	(101.389)
Hh number of farm workers at baseline	-4.015	79.637***	195.204***
	(6.304)	(9.803)	(37.015)
Hh number of plots at baseline	42.146***	-34.014***	152.437***
	(7.549)	(9.044)	(40.627)
Constant	250.138***	328.870***	532.920***
	(23.918)	(32.134)	(121.541)
N	1521	1248	454
R2	0.021	0.055	0.149

*** denotes significance at the 1% level; ** at the 5% level; * at the10% level.

3.4 Heterogeneity and mechanisms – the role of education and gender

To further disentangle the mechanisms behind recall bias, **Table 8** examines heterogeneity across subpopulations. For reference, column (1) repeats the main specification shown in **Table 3**, with a recall bias of 16 hours per person per plot.

We start with the relationship between education and recall bias. Several studies have found a positive association between the level of education – or direct measures of cognitive skills – and recall ability. Peters (1988), comparing lifecycle data from a retrospective marital history with panel reports for the same individuals in the United States, shows that more educated respondents report more consistently on marital events. McAuliffe, DiFranceisco and Reed (2010) find that arithmetic skills predict the accuracy of retrospective self-reports of sexual activity (also in the US). Similar evidence is available from developing countries, a distinction that is important considering differences in the quality of education systems. Becket et al (1999, 2001) show that the reliability of reports in the Malaysian Family Life Surveys is higher for more educated respondents. Abebe (2013) finds that recall errors in sales revenues and output among Ethiopian shoemakers are negatively correlated with respondent's years of schooling. Based on this literature, we expect the accuracy of reports on farm labor to be linked positively to educational attainment, as a proxy for cognitive skills, and recall bias should be lower among more educated households.

Table 8: Estimation of recall bias in farm labor – heterogeneity

	Labor hours per person per plot		
	(1)	(2)	(3)
Recall	16.182**	41.580***	22.737**
	(5.165)	(8.243)	(7.086)
Hh education above primary		-24.818***	
		(6.505)	
Recall * Above primary		-41.664***	
		(10.486)	
Female			-40.653***
			(6.246)
Recall * Female			-15.696
			(10.191)
Hh no. of farm workers at baseline	1.329	2.646	0.314
	(1.615)	(1.608)	(1.602)
Hh no. of plots at baseline	-10.160***	-8.729***	-10.117***
	(1.603)	(1.599)	(1.587)
Constant	130.361***	138.423***	153.130***
	(6.109)	(7.025)	(6.869)
N	4462	4462	4462
R-squared	0.012	0.029	0.032
Total recall bias for educated households or females		-0.0838	7.041
		(6.516)	(7.358)

Note: Hh education refers to the highest level of education attained by any household member. Standard errors in parentheses. *** denotes significance at the 1% level; ** at the 5% level; * at the10% level.

The results in column (2) of **Table 8** provide support for this hypothesis. To test for heterogeneity in recall bias by level of education, the regression includes an indicator variable that equals to unity if at least one member of the household has attained education above the primary level (*Hh education above primary*), and its interaction with the main recall variable (*Recall*Above primary*). Both variables are negative and significant, suggesting 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 suggest that there is basically no recall bias for households with above primary educational attainment (see bottom row in **Table 8**), while households with lower education on average overestimate season-wide hours by approximately 42 hours per person per plot. This is consistent with descriptive evidence in **Table A1** (appendix), which uses a more granular classification of educational attainment and also shows a strong educational gradient, though not conditioning on baseline characteristics.

We next turn to the gender differences – particularly the question whether recall bias differs between male and female workers. Time use data from Africa show that women often carry out different roles and activities simultaneously, rather than sequentially. This holds particularly for childcare, which is typically embedded within other economic activities (Blackden and Wodon 2006). This may make it more difficult to recall farm labor for female than for male household workers. Our empirical results portray a more nuanced picture. Column (3) includes an indicator variable that equals to unity if the worker is female (*Female*) and the interaction between this variable and the recall variable (*Recall* Female*). The results show that women work around 40 hours less than men on the farm over the season. The point estimates further suggest that season-wide hours are overestimated by 23 hours for males, compared with only 7 hours for females, but the difference is just below the 10 percent significance threshold. Counter to our expectations, there is no indication that recall bias is larger for women. If anything, our point estimates suggest the opposite.

Since the results in section 3.2 show that recall bias stems primarily from listing bias, we now turn to explore whether there is any gender difference to bias in the listing household workers. The first column of **Table 9** repeats the double-difference estimate of recall bias in the listing of household workers previously shown in **Table 4**. The second and third columns disaggregate between male and female household workers. The results show that female workers are more likely to be added after baseline than male workers (the coefficient on *Endline* is larger for females), and that recall bias in listings is somewhat larger for females than for males. Hence the difference in listing bias between women and men does not appear to account for the difference in recall bias in hours worked. However, a plausible explanation is that since women, on average, work fewer hours than men, listing bias has less of an impact on women's average hours per person per plot. In other words, on average, there is less of a difference between 'listed' and 'omitted' women than between 'listed' and 'omitted' men, which may explain why recall bias in farm labor is smaller for female household workers.

Table 9: Double difference estimator of bias in listings of male and female farm workers

	No. of farm workers listed per household		
	All	**Female**	**Male**
Endline	1.885***	1.175***	0.869***
	(0.118)	(0.083)	(0.067)
Endline * Recall	-1.211***	-0.844***	-0.526***
	(0.166)	(0.114)	(0.093)
Constant	2.181***	1.041***	1.208***
	(0.059)	(0.040)	(0.033)
N	918	882	897
R2	0.386	0.331	0.305

*** denotes significance at the 1% level; ** at the 5% level; * at the10% level.

3.5 Comparison between Ghana and Tanzania

As mentioned earlier, this study follows the design of a survey experiment conducted in the Mara region of Tanzania in 2014 (Arthi et al 2016). The main difference between both studies 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. The Tanzania approach allows for the comparison of the treatment group with what would be a 'typical' cross-sectional household survey that is conducted at the end of the agricultural season without any information from the start or during the season. Conversely, the Ghana approach resembles a panel survey and enables us to control for differences in baseline characteristics and disentangle the possible sources of recall bias by allowing to obtain more rigorous estimates of listing bias.

Table 10 summarizes the results of both studies. Columns 3 and 4 replicate **Table 2**, and columns 1 and 2 are the equivalent numbers for Tanzania. While the (unconditional) recall bias in Ghana is 18 percent, it is 207 percent in Tanzania. What would explain such large differences in the magnitude of recall bias?

Table 10: Farm labor per person per plot, Ghana and Tanzania

	Tanzania		**Ghana**	
	Weekly Visit	**Recall**	**Weekly Visit**	**Recall**
Hours	39.5	121.3***	106.4	125.7***
	(69.5)	(133.8)	(148)	(194.5)

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

Table 11 provides information on listing bias in Ghana and Tanzania, that helps to understand the difference in recall bias shown in the previous table. The simple difference estimator of listing bias presented in the top panel is equal to the difference in the average number of workers (plots) at endline between weekly visit and recall households and can be compared between Ghana and Tanzania. The bottom panel shows the double difference estimator of listing bias, which nets out baseline differences. Given that Tanzania did not collect baseline information for recall households, the double difference estimator can only be calculated for Ghana.

Table 11: Listing bias, Ghana and Tanzania

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

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

The simple difference estimator of bias in the listing of plots is significantly larger in Tanzania (2.1 fewer plots for recall households) than in Ghana (0.6 plots fewer plots in the recall group), while listing bias in household workers is similar (1.6 vs. 1.5 fewer workers in the recall group). In both experiments, the roster for weekly visit households is "cumulative," as shown in **Figures 1** and **2**, this means that it has been updated throughout the season to capture all plots and all household workers involved in farm work. Conversely, the recall households do not have a cumulative roster. The fact that listing bias in plots is smaller in Ghana than in Tanzania might partly reflect that recall households in Ghana were administered a baseline plot roster, which was updated once at endline and helped to facilitate recall. As illustrated in section 3.2, listing bias is one of the driving forces behind recall bias in farm labor – and this may hence explain the large differences between the two studies highlighted in **Table 10**. Unfortunately, we cannot compare the double difference estimator of listing bias across the two countries, which for Ghana is slightly smaller than the simple difference estimator.

Besides variations in the design of the experiments, differences in educational attainment could also play an important role. In our study area in Ghana, 50 percent of households have at least one member with above primary education, compared with only 34 percent of households in the Mara region of Tanzania.

Since recall bias declines with level of education, these figures suggest that recall bias in farm labor ought to be larger in Tanzania than in Ghana.

4. Implications of recall bias for agricultural productivity measures

Measures of farm labor are commonly used to estimate labor productivity in agriculture and the latter will be affected by recall bias in farm labor. Because labor appears in the denominator, positive recall bias (where the recall method *overestimates* farm labor) leads to an *underestimation* of agricultural productivity, and vice versa.

As discussed in section 2, the endline fielded a standard agricultural production module to all households, which asked to report the quantity and value of crop harvested during the preceding agricultural season (for each plot and crop). We aggregate across all plots and crops to arrive at total production (in Cedi) at the household level. To estimate labor productivity, total production is normalized on the number of farm workers and/or farm labor hours at the household-level.[4]

Table 12 presents raw estimates and adjusted estimates. The former use the household's own-reported value of crop harvest, whereas the latter combine the reported quantity of the harvest with the median value per kg reported for that crop. The left panel uses all households, while the right panel only uses households with completed crop data (i.e. completed harvest). All estimates are winsorized at 1 percent.

Table 12: Labor productivity in agriculture, recall and weekly visit groups

	All households			**Households with complete crop data**		
	Recall	Weekly Visit	**Dif p-value**	Recall	Weekly Visit	**Dif p-value**
Value per hour	8.029	4.378	0.000	7.591	4.920	0.013
	(0.800)	(0.574)		(0.790)	(0.706)	
Adjusted value per hour*	7.020	3.976	0.000	7.007	4.483	0.009
	(0.682)	(0.507)		(0.729)	(0.623)	
Value per worker	1502.183	983.471	0.004	1447.906	1033.565	0.035
	(146.612)	(104.344)		(152.819)	(119.519)	
Adjusted value per worker*	1294.981	879.581	0.004	1292.141	915.382	0.017
	(113.851)	(85.379)		(123.599)	(95.503)	
N	219	218		185	174	

Note: Households with zero harvest are excluded from the analysis. All variables are winsorized (1%). Values are in Cedi. The last 3 columns exclude households for which harvest data is incomplete due to missing conversion factors (units to kgs).
*Adjusted value per hour or per day combines household's own reported quantity harvested with the median value per kg reported for that crop.

Irrespectively of the estimation method used, labor productivity in agriculture is significantly overestimated among recall households. This, of course, reflects the results in **Table 6**, which show that the recall method underestimates farm labor per plot and per household. Listing bias in household workers unambiguously leads to an overestimation of labor productivity, as some of the labor input that went into the production of output is not captured. Conversely, listing bias in plots has more ambiguous effects, as

[4] To keep matters simple, we ignore hired labor.

the omitted plots affect both the enumerator (in terms of non-measured output) and the enumerator (in terms of non-measured labor input).

5. Conclusion

We conducted a randomized survey experiment in Ghana. Estimates of farm labor obtained from a recall survey conducted at the end of the season, which asks survey respondents to recall the time each member of the household spent on farm activities during the previous agricultural season, are compared against data collected weekly throughout the season.

The results indicate that the recall method overestimates farm labor per person per plot by 15 percent, conditional on difference in baseline characteristics. This recall bias is significantly lower than in a similar study conducted in Tanzania. This recall bias in farm labor is accounted for by listing bias, as households in the recall group report fewer marginal plots and farm workers. Moreover, listing bias runs counter to recall bias at the person-plot level and dominates at higher levels of aggregation. The end-of-season recall method thus *underestimates* farm labor per plot and per household, even though it overestimates farm labor per person per plot. As result, household-level estimates of labor productivity in agriculture are significantly overestimated by the recall method.

We also moved beyond proxy determinants to understand the deeper forces behind recall bias. Consistent with the notion that recall bias is linked to the cognitive burden of reporting on past events, we find that better educated households recall farm labor with greater accuracy. This educational gradient – together with variations in the design of the experiment – most likely explains why this study finds much lower recall bias in farm labor in the Ghanaian context than the previous study found for the Mara region in Tanzania.

References

Abebe, Girum. 2013. "Recall Bias in Retrospective Surveys: Evidence from Enterprise-Level Data." *Asian-African Journal of Economics and Econometrics* 13(1): 17-33.

Arthi, V., K. Beegle, J. de Weerdt, A. Palacios-Lopez. 2016. "Not Your Average Job. Measuring Farm Labor in Tanzania." Policy Research Working Paper 7773, Washington DC, World Bank.

Bardasi, Elena, Kathleen Beegle, Andrew Dillon, and Pieter Serneels. 2011. "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania." *World Bank Economic Review* 25(3): 418-447.

Beckett, Megan, Julie DaVanzo, Narayan Sastry, Constantijn Panis and Christine Peterson. 1999. "The Quality of Retrospective Reports in the Malaysian Family Life Survey." Working Paper Series 99-13, RAND.

Beckett, Megan, Julie DaVanzo, Narayan Sastry, Constantijn Panis and Christine Peterson. 2001. "The Quality of Retrospective Data: An Examination of Long-Term Recall in a Developing Country." *Journal of Human Resources* 36(3): 593-625.

Blackden, Mark C. and Quentin Wodon. 2006. "Gender, Time Use, and Poverty in Sub-Saharan Africa." Working Paper No. 73. World Bank: Washington, DC.

Christiaensen, Luc, Lionel Demery and Jesper Kuhl. 2011. "The (Evolving) Role of Agriculture in Poverty Reduction – An Empirical Perspective." Journal of Development Economics 96(2): 239-54.

De Janvry, Alain, Elisabeth Sadoulet and Tavneet Suri. 2016. "Field Experiments in Developing Country Agriculture." In: Abhijit Banerjee and Esther Duflo (eds.). A Handbook of Economic Field Experiments. North-Holland: Amsterdam.

Dillon, Brian, Peter Brummund and Germano Mwabu. 2017. "How Complete Are Labor Markets in East Africa? Evidence from Panel Data in Four Countries." Retrieved from: http://faculty.washington.edu/bdillon2/CV_papers/DBM-labor-markets-170510.pdf

Doss, Cheryl R. 2017. "Women and Agricultural Productivity: Reframing the Issue." *Development Policy Review*: 1-16. Retrieved from: https://doi.org/10.1111/dpr.12243.

FAO (Food and Agriculture Organization of the United Nations). 2017. *Measuring Work in Agricultural Surveys and Censuses: A Review*. Mimeo.

Fermont, Anneke and Todd Benson. 2011. "Estimating Yield of Food Crops Grown by Smallholder Farmers. A Review in the Ugandan Context." IFPRI Discussion Paper 01097. Washington, DC: International Food Policy Research Institute (IFPRI).

Gollin, Douglas and Christopher Udry. 2017. "Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture." Retrieved from: http://egcenter.economics.yale.edu/sites/default/files/files/Conference%202017%20Agri-Devo%20speakers/Gollin_Udry_Heterogeneity%2C%20Measurement%20Error%2C%20and%20Misallocation.pdf

Gollin, Douglas, David Lagakos and Michael E. Waugh. 2014a. "Agricultural Productivity Differences Across Countries." *American Economic Review* 104(5): 165-70.

Gollin, Douglas, David Lagakos and Michael E. Waugh. 2014b. "The Agricultural Productivity Gap." *Quarterly Journal of Economics* 129(2): 939-93.

ILO (International Labor Organization). 2016. ILOSTAT – Employment by Sector. Retrieved from: http://www.ilo.org/ilostat/faces/oracle/webcenter/portalapp/pagehierarchy/Page3.jspx?MBI_ID=33

LaFave, Daniel and Duncan Thomas. 2016. "Farms, Families and Markets: New Evidence on Completeness of Markets in Agricultural Settings." *Econometrica* 84(5): 1917-60.

McAuliffe, Timothy L., Wayne DiFranceisco and Barbara R. Reed. 2010. "Low Numeracy Predicts Reduced Accuracy of Retrospective Reports of Frequency of Sexual Behavior." *AIDS Behav* 14(6): 1320-29.

McCullough, Ellen B. 2017. "Labor Productivity and Employment Gaps in Sub-Saharan Africa." *Food Policy* 67: 133-52.

O'Sullivan, Michael, Arathi Rao, Raka Banerjee, Kajal Gulati and Margaux Vinez. 2014. *Levelling the Field: Improving Opportunities for Women Farmers in Africa*. World Bank and One Campaign, Washington, DC.

Peters, H. Elizabeth. 1988. "Retrospective Versus Panel Data in Analyzing Lifecycle Events." *Journal of Human Resources* 23(4): 488-513.

Reardon, Thomas and Paul Glewwe. 2000. Agriculture. In: Margaret Grosh and Paul Glewwe (eds.) *Designing Household Survey Questionnaires for Developing Countries. Lessons from 15 Years of the Living Standards Measurement Study*. Volume 2. Washington, DC: World Bank.

Restuccia, Diego and Raul Santaeulalia-Llopis. 2017. "Land Misallocation and Productivity." NBER Working Paper 23128, National Bureau of Economic Research, Cambridge, MA.

World Bank. 2017. World Development Indicators. Retrieved from: http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators

Appendix

Table A1: Labor input per person per plot, by highest education level in the household

	Less than primary education		Primary education		Above primary education	
	Weekly visit	Recall	Weekly visit	Recall	Weekly visit	Recall
Hours	140.0	196.1	104.3	130.8	96.1	98.3
	(174.5)	(297.9)	(129.6)	(202.2)	(141.8)	(125.4)
N	568	375	469	283	1750	1017

Note: See notes Table 1. Hh education refers to the highest level of education attained by any household member.