Not Your Average Job:
Irregular Schedules, Recall Bias, and Farm Labor Measurement in Tanzania[1]

Vellore Arthi
University of Oxford

Kathleen Beegle
World Bank

Joachim De Weerdt
IOB, University of Antwerp LICOS, KU Leuven

Amparo Palacios-López[2]
World Bank

ABSTRACT

A good understanding of the constraints to agricultural growth in Africa relies on the accurate measurement of small-holder labor. Yet, serious weaknesses in these statistics persist. The extent of bias in small-holder labor data is examined by conducting a randomized survey experiment amongst farming households in rural Tanzania. Agricultural labor estimates obtained from weekly surveys are compared to those reporting in a single end-of-season recall survey. The results show strong evidence of recall bias: people in traditional recall-style modules report working up to 4 times as many hours per person-plot compared to those reporting labor on a weekly basis. When hours are aggregated to the household level, however, this discrepancy disappears, a factor driven by the under-reporting by recall households of people and plots active in agricultural work. Evidence suggests that these competing forms of recall bias are driven not only by failures in memory, but also by the mental burdens of reporting on highly variable agricultural work patterns to provide a "typical" estimate. These results contribute to debates on the agricultural productivity gap and imply that current assessments of agricultural labor productivity may be vastly underestimated low due to recall bias in the measurement of farm labor or assumptions about the work being full-time.

[1] This study is an output from a project funded by the UK Department for International Development (DFID), the Institute for the Study of Labor (IZA), and the World Bank. The authors gratefully acknowledge the comments of participants at the Centre for the Study of African Economies (CSAE) Research Workshop, European Survey Research Association (ESRA) Conference, Growth and Labour Markets in Low Income Countries Programme (GLM | LIC) Conferences, Northeast Universities Development Consortium (NEUDC), Structural Transformation of African Agriculture and Rural Spaces (STAARS) Conference, and seminar series at the University of Antwerp, the University of Namur, the University of Washington and World Bank. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of DFID, IZA or the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, nor those of the Executive Directors of the World Bank or the governments they represent. The data was collected on surveybe and fieldwork expertly implemented by Economic Development Initiatives (EDI).

[2] Corresponding author: apalacioslopez@worldbank.org

INTRODUCTION

Of the 1.4 billion people living in extreme poverty, the majority reside in rural areas and rely on agriculture as a source of income and livelihood (Olinto et al., 2013). In Sub-Saharan Africa, nearly 75 percent of the extreme poor reside in rural areas, and over 90 percent participate in agriculture. Small-holder agriculture is the predominant form of farm organization, with 33 million small farms holding less than two hectares and representing 80 percent of all farms in Africa (FAO, 2009). In these farms, agricultural practices are typically labor intensive, with the majority of the labor being provided by household members.[3]

Accordingly, the labor of household members in agriculture is a key asset for poor households, and its accurate measurement is essential to developing sound policy. Despite the importance of the agricultural sector in reducing poverty and food insecurity (Irz et al., 2001; Chen and Ravallion, 2007; Ligon and Sadoulet, 2007), serious weaknesses in agricultural statistics persist (FAO, 2008). In this study, we examine one aspect, measures of family farm labor.

To assess the degree of recall bias in household farm labor, we conduct a survey experiment over the long rainy season, January-June 2014, in the Mara region of Tanzania. Small-holder farming households are randomly assigned to one of four survey designs: (1) households reporting agricultural labor in weekly face-to-face visits; (2) households reporting agricultural labor in weekly phone surveys; (3) households reporting agricultural labor in a single post-harvest recall survey, per the Tanzania National Panel Survey (NPS); and (4) households reporting agricultural labor in a shorter version of the Tanzania NPS post-harvest recall survey. Household labor information collected in weekly visits, our resource-intensive 'golden standard,' is then compared to data reported after the harvest. After establishing the magnitude of recall bias, we investigate the mechanisms by which it arises.

We find strong evidence of recall bias in the reporting of family farm labor, but due to competing forms of recall bias in reporting of hours and the number of plots and active household members, the degree of distortion in reporting depends on the level of data aggregation. Labor data collected on a weekly basis, whether in person or by phone, are similar, albeit sometimes not statistically identical. Likewise, the labor data reported by recall in our two recall designs are also quite alike. However, there are striking and economically meaningful differences between the weekly and recall data. Respondents in recall-style modules report working up to nearly 4 times as many hours per person per plot, compared to labor reported on a weekly basis. When hours are aggregated to the household level, however, this discrepancy disappears. This is driven by two factors: under-reporting by recall households of both working household members and plots under cultivation. Evidence suggests that these competing forms of recall bias are driven not only by failures in memory, but also by the mental burdens of computing a "typical" figures when agricultural work patterns are highly variable during the season.

Our results have important implications for development policy, and fill key gaps in the literature concerning survey methods and the quality of agricultural labor data. This is one of the few studies to test the accuracy of agricultural labor data in developing-country settings. While

[3] ILO's LABORSTA database suggest that in Tanzania in 2006 only 4% of agricultural workers were employees.

labor data have been an essential ingredient in a broad range of important studies on small-holder agriculture in developing countries,[4] scant attention has been paid thus far to the quality and robustness of the underlying data on family farm labor. By showing evidence that agricultural labor inputs may be substantially overestimated due to recall bias, we challenge the reliability of the traditional end-of-season labor estimates commonly used in development economics.

These findings regarding the overestimation of agricultural labor inputs contribute to academic and policy debates concerning the agricultural productivity gap and the degree to which rural labor may be misallocated in developing economies. Our study complements Gollin, Lagakos, and Waugh (2014) and McCullough (2015), which question the accuracy of current labor measures. These studies reconsider the agricultural productivity gap after adjusting for labor data quality. By conducting comparisons at the per-hour level (McCullough, 2015) and by adjusting for sectoral differences in hours worked as well as for levels of human capital (Gollin, Lagakos, and Waugh, 2014), both studies find that the difference in the productivity between agricultural and non-farming sectors is narrower than typically thought. Our study suggests that upward recall bias in reported hours of farm labor may further reduce these improved estimates of farm labor inputs and so further narrow the measured productivity gap.[5] If the results of our survey hold in other parts of Africa, labor productivity in African agriculture may be considerably underestimated.

Although our results call into question the accuracy of current farm labor data, they also suggest specific ways in which the accuracy of labor measurement may be improved. The consistency of the labor reported between face-to-face and phone surveys suggest that season-long phone surveys are an option to reduce error in the measurement of rural agricultural labor.

The rest of the paper proceeds as follows. In Section 2, we offer background on labor measurement. In Section 3, we provide an overview of the empirical approach, including details of the survey experiment. In Section 4, we present the results of our experiment on the reporting of labor inputs in family farming, and outline the sources of bias in recall data. Section 6 concludes.

1. MEASURING LABOR

2.1. Current practice

---

[4] Examples include farm household models (Rosenzweig and Wolpin, 1985; Barnum and Squire, 1979; Singh et al., 1986; Benjamin, 1992), shadow wages in family farm labor (Jacoby, 1993), tradeoffs between hired and household labor (Johnston and Le Roux, 2007; Chowdhury, 2010; Deolalikar and Vijverberg, 1987), how household allocate labor to farm and off-farm work (Matshe and Young, 2004; Lanjouw and Lanjouw, 2001; Shapiro, 1990) and intra-household labor allocation choices (Udry, 1996).

[5] For instance, intersecting with Gollin, Lagakos, and Waugh's (2015) interest in the contribution of cross-sectoral differences in human capital to labor productivity, we find that more highly educated respondents produce less recall bias in reported family farm labor than do their less well-educated counterparts (not reported). If those with greater levels of human capital are less likely to overestimate their labor in recall—perhaps because they are better able to cope with the cognitive burdens of remembering and inferring irregular labor—then their higher labor productivity may not be entirely attributable to true differences in productivity driven by skill and education, but rather, to differences in the quality of labor reporting by level of education.

The wealth of evidence on the quality and reliability of labor statistics in household surveys comes largely from the United States (see Bound, Brown, and Mathiowetz, 2001, for a thorough review). In the context of developing and agriculturally-driven countries, for contrast, little is known about the extent to which the design of surveys influences labor statistics. Clearly, it is difficult to extrapolate from the studies conducted in the United States to the African context. Moreover, the existing literature on data quality and survey methods in low-income settings rarely pertains to farm labor (see Bardasi et al., 2011). It has been noted that International Labor Organization (ILO) recommendations for measuring labor are likely to be inadequate in settings like rural Tanzania, where the majority of labor is found in the informal self-employed and farm sectors (World Bank 2014).

A review of existing surveys that collect labor data in Africa shows that, in practice, the capture of labor market statistics in household surveys varies widely. The recall period, the sequencing of questions, the use of screening questions, the seasonal timing of the survey, the granularity of reporting requested, the unit over which labor is reported, and the choice of respondent can vary across surveys, both within and across countries. Inconsistencies in data collection methods hampers comparisons over time and space when survey methods differ, as has been shown in the context of welfare, poverty and hunger measurement (Backiny-Yetna et al. 2014; Beegle et al. 2011; Beegle et al. 2016; and De Weerdt et al. 2016) and labor measurement (Bardasi et al., 2011).

A national integrated or multi-topic household surveys in Africa generally collect data on agricultural labor in two ways.[6] In one approach, general labor information, including agricultural labor, is collected in a labor module. In another, specific agricultural labor data is collected in an agriculture module (as done in the Living Standards Measurement Surveys – Integrated Surveys on Agriculture, LSMS-ISA). In the former case, information on labor for each household member above some specified age is collected in reference to the last 7 days or maybe last 12 months (Anderson Schaffner, 2000). The person's labor input is not differentiated by plot, by crop, or by farm activity (such as weeding, harvesting, etc.). Instead, in the agricultural module outlined by Reardon and Glewwe (2000), the total days of labor at the household level over the last completed season are collected for each plot and by specific activities. An expanded agricultural module would have the same questions asked for each household member (as in the LSMS-ISA).[7] A common feature in these surveys is that labor is collected from a single interview.

While considered to be an improvement over the more general labor force questions, the expanded LSMS-ISA agricultural module has several potential drawbacks. First, it is time-consuming to collect this detailed information. Second, the burden on respondents is substantial: respondents are asked to provide information that they may never have considered (e.g. labor by activity for each plot). Third, there is potential for problems in recall and memory.

2.2. What complicates the measurement of small-holder farm labor?

[6] Apart from multi-topic household surveys, small-holder information can be collected from specialized farm surveys. These often entail visiting the household at multiple times, particularly those surveys utilizing resident enumerators (e.g. agricultural extension agents or other Ministry of Agriculture staff). However, these surveys typically do not collect details on household farm labor.

[7] The LSMS-ISA program has conducted in Burkina Faso, Ethiopia, Malawi, Mali, Niger, Nigeria, Tanzania, and Uganda; for more detail, see www.worldbank.org/lsms.

Features of small-holder farming

The estimation of labor inputs on small-holder farms is complex and vulnerable to mis-reporting.[8] Small-holder farms typically mostly employ family labor and so there is no wage income in which to anchor recall. Written records are rarely kept and the respondent must rely on recall to report on past events. To arrive at the total amount of labor allocated by a household to farming, the household must accurately report the plots under cultivation, the specific household members that worked on each plot, the activities performed, and their timing and duration. Farming is a seasonal activity and work patterns are irregular during the season. Reporting "typical" or "average" time farming after the completion of the season requires remembering distant events and making complicated mental calculations. Alternatively, reporting hours worked in the last 7 days at any single point during the agricultural season will not necessarily be indicative of total labor during the season if labor inputs vary a lot during the season.

Insights from cognitive psychology

The design of the survey instrument itself may also influence the quality of data on family farm labor. Considering common survey practices and features of small-holder farm labor alongside insights from the social and cognitive psychology literature, there is particular call for caution when interpreting the farm labor data taken from household surveys.

Perhaps the most important aspect in our context is the implications of the recall period. Here, the effects can operate through faults in memory. Forgetting an event is considered more likely as time passes. Alternatively, telescoping, by which an event is remembered to have occurred more recently than it actually did, can result in memory-driven distortions, particularly for longer recall periods (Sudman and Bradburn, 1973). An example of this would be a respondent who worked on the farm 35 days ago, but who reports having worked on the farm in the past 30 days. Beegle, Carletto, and Himelein (2012) find little evidence that longer recall periods lead to less reliable reporting of hired farm labor in Malawi, Kenya and Rwanda.

The length of recall periods on survey responses may matter beyond the implications of memory process. It can affect how a respondent interprets the question. Schwarz (2007) provides evidence that in longer recall periods only salient events are reported. For example, in a survey in which respondents are asked how many times they have been angry over a period of time, Schwarz finds that if the recall period is 1 day, the respondent assumes that minor irritations should be counted. Extending the recall period to 1 year leads the respondent to believe that only very serious anger incidences should be reported. The study concludes that the shift in inferred pragmatic meaning

[8] Measurement problems are not restricted to labor. For instance, intercropping, continuous planting, extended harvest periods, and multiple plots of small sizes and irregular shapes can make reporting on most inputs and outputs difficult. Although several strategies are proposed in the literature to account for mixed-stand crops, no method has yet gained wide acceptance (Fermont and Benson, 2011). The introduction of GPS devices has improved measures of land holdings, but methods for collecting production and input data are not much different now than in the last several decades (Deininger et.al, 2011).

makes it difficult to disentangle effects of question interpretation and forgetting. Das et al. (2012) find a similar pattern in the self-reporting of past health, whereby smaller illness events are ignored and forgotten as the recall period increases. They also find heterogeneity in these effects by income, driven by the normalization by the poor of what would otherwise (i.e. for richer people) be salient illness events worthy of medical treatment. In our context, when asked to report on labor in the last week, a farmer may interpret the question differently compared to one who is asked to report on several months' worth of labor at the end of the season. Our results suggest that even seemingly straightforward questions, such as how many plots the farmer has cultivated, and who has worked on them, are affected by the recall period.

Beyond the length of the recall period, there are aspects of the cognitive and communicative processes that affect survey responses. Menon (1993) shows that for infrequent and salient events, respondents are likely to "recall and count" individual events since they are stored episodically and remain in memory for a longer time. In the absence of episodic event information that is easily retrieved, respondents will rely on other strategies. For regular events, such as "I visit my grandmother every Saturday," respondents are not likely to use the recall-and-count strategy, relying instead on the information they have stored about the event's periodicity. Such rate-based estimations may be adjusted by memories of non-occurrence ("except when I'm on holiday") or more frequent occurrence ("also on her birthday if that doesn't fall on a Saturday"). Menon (1993) notes that counting the occurrence of events that are neither salient nor regular requires much more cognitive effort on the part of the respondent. Thus, where work is neither salient nor regular—as may be the case for small-holder farmers labor over the agricultural season—respondents are unable to use rate-based or recall-and-count strategies, and are so likely to yield erroneous reports of labor.

In the absence of episodic or rate-based information, respondents may revert to their general assumptions about the state of the world in their search for an answer. These assumptions then form a benchmark that is used to infer previous behavior. Indeed, as will be discussed below, the spuriously high recall-surveyed labor we find in this study can stem from this sort of inference. Schwarz and Oyserman (2001) cite evidence that retrospective estimates of income and of tobacco, marijuana, and alcohol consumption are unduly influenced by people's income and consumption habits at the time of the interview. That is, they infer their previous behavior from their current or recent behavior. Similarly, de Nicola and Giné (2014) show that survey responses on income from small-scale boat owners in coastal India rely more on inference and less on true recollection as the recall period increases. The authors show that while this bias has little influence on the mean (because in their case fishermen base their inferences on average earnings), it does lead to an underestimation of income variability as the recall period increases. The information and assumptions held by respondents are also important when people report on the behavior of others, a common practice when collecting labor data in household surveys (Bardasi et al., 2011; de Nicola and Giné, 2014).

Respondents may also be suggestible and base their inferences on what they believe *should* have occurred. Ross and Conway (1986) allowed students to participate in a skills-training program that did not, in fact, influence their skills. After participating in this study, the students quantified their pre-training skills as having been lower than what they had originally assessed them to be prior to receiving the skills training. The authors argue that the students reconstructed their past, guided by their subjective theories over what the skills training ought to have done. If African farmers hold implicit theories about the link between, say, labor inputs and production, then the report on the one may influence the report on the other. For example, in an end-of-season recall survey, labor may be retrospectively overstated during good harvests and understated during bad harvests.

2. EXPERIMENTAL DESIGN AND CONTEXT

The goal of this study is to examine biases, of the like described above, in agricultural labor data collected from household surveys. We focus on potential bias introduced by the length of the recall period and the frequency of reporting. We conduct a large randomized survey experiment amongst small-holder farming households in rural Tanzania, through which we compare agricultural labor information collected in weekly surveys (our benchmark for the true labor estimates) with that collected in a single, end-of-season survey. Understanding farm productivity at the lowest level entails studying inputs and yields at the plot level. From a broader perspective, questions of productivity in small-holder farming may require analysis of aggregated measures. We focus on both by examining plot-person-level labor reporting as well as aggregate household measures of family labor.

3.1. Experimental design

We conduct a survey experiment among 854 farming households in 18 enumeration areas in the Mara region of rural northern Tanzania. Labor input was measured for the main rainy (locally called *masika*) season of 2014, running roughly from January to June 2014. Households were randomly assigned to one of four survey designs within each of the 18 enumeration areas. The differences in the four survey arms are the manner and frequency with which they are contacted.[9]

Two survey designs entailed weekly interviews throughout the entire Masika season either in person or by phone.[10] Face-to-face baseline survey was conducted in January 2014 and a face-to-face endline survey was fielded in July-September 2014.

The other two survey designs entailed one recall survey fielded at the end of the agricultural season in July-September 2014. This survey differs from the endline surveys received by the weekly households in that collected information on labor from January to June.

The four alternative survey designs are as follows:

- ***Weekly Visit (benchmark):*** *Weekly face-to-face surveys for the duration of the Masika season*
  For Weekly Visit households, a baseline survey is conducted in January 2014, followed by weekly face-to-face surveys conducted by enumerators through the end of June 2014, and

[9] The data were collected with Computer Assisted Personal Interviewing (CAPI) using the software program *surveybe*.
[10] All weekly households received a mobile phone, but recall households did not. With mobile phone ownership already very widespread at 72% of households in our sample this is unlikely to influence results.

an endline survey (July-September 2014) to collect farm production information. For each plot, household members that worked on the plot in the past week are identified, and the hours for each day that they worked on the plot during the past week are reported.[11]

- ***Weekly Phone:*** *Weekly phone surveys for the duration of the Masika season*[12]
  For Weekly Phone households, an face-to-face baseline survey is conducted in January 2014 (during which time households were provided with a mobile phone by which to respond to subsequent surveys), followed by weekly phone surveys through the end of June 2014, and an in face-to-face endline survey (July-September 2014) to collect farm production information. For each plot, household members that worked on the plot in the past week are identified, and the hours for each day that they worked on the plot during the past week are reported.

- ***Recall NPS:*** *Face-to-face survey at the end of the Masika season, standard NPS module*
  For Recall NPS households, a face-to-face endline survey is conducted after the harvest (July-September 2014), during which both labor and farm production information is collected. The agricultural labor module is identical to that in the Tanzania National Panel Survey, Waves 3 (2012/2013) and 4 (2014/2015). For each plot, the household members that worked on that plot at any point during the season are identified, and the following information is reported: (i) total days spent on the plot over the season in each of four activities (land preparation and planting; weeding; ridging, fertilizer application, and other non-harvest activities; harvesting) and (ii) typical hours per day worked in each of these four activities.

- ***Recall ALT:*** *Face-to-face survey at the end of the Masika season, alternate survey module*
  For Recall ALT households, a face-to-face endline survey is conducted after the harvest (July-September 2014), during which both labor and farm production information is collected. For each plot, the household members that worked on that plot at any point during the season are identified, and the following information is reported: (i) total weeks worked on the plot over the season (irrespective of activity), (ii) approximate number of days per week worked, and (iii) approximate number of hours per day worked.

Throughout this paper, we establish the magnitude of bias by comparing to the Weekly Visit design. This is based on the premise that the figures reported in the Weekly Visit design are likely to be the closest to the "truth". For Weekly Visit and Weekly Phone respondents were probed for day-by-day responses which were specific to the plot and to the person. Accordingly, these

[11] In addition, after reporting the hours per person per day over the last week, the range of activities done during that time is recorded (land preparation and planting; weeding; ridging, fertilizer application and other non-harvest activities; harvesting), but hours are not specified for each activity.

[12] The Weekly Phone interview design draws on lessons summarized by Dillon (2012), who uses a phone survey to collect information on purchased-input application among cotton farmers in Tanzania. Similar recent work has used phone surveys to collect high-frequency data on economic activity (see Garlick, Orkin, and Quinn (2015) for a review the literature on phone-based strategies for collecting household and enterprise data).

interviews minimized the need for respondents to make any complicated calculations or inferences regarding season-wide labor. While we cannot exclude the possibility that forgetting or telescoping still lead to some bias, we assume that the short 1-week period and specificity reduces the influence of forgetting. Anchoring the reporting to the previous interview reduced the possibility of telescoping.

Table 1 presents descriptive statistics of household characteristics across the four survey designs, drawing on the endline survey (post-harvest). For these set of traits, households are well-balanced across the different survey designs.

In order to be sure that the differences across the four arms of the experiment are not plagued by confounding factors, we consider some identification concerns. First households are randomized within villages to account for micro agro-ecological patterns affecting household labor (which we may not capture through data sources). This raises the possibility of intra-cluster contamination, where one person's response is influenced by another's design status. We opt for within-village randomization because we believe, *a priori*, that such contamination is unlikely because villages are relatively large and diffuse.

Secondly, the weekly visits themselves could have caused differential labor (akin to Hawthorne effects). We cannot rule this out but note two points. There is general increase in hours as the season progresses (which might occur if a respondent's started work harder or over-report work as a result of being interviewed frequently); nor do we find a decline in hours over the season in the weekly interviews (due to respondent fatigue). There is also no little difference between the face-to-face and phone interviews, whereas one would expect Hawthorne effects to be stronger for in-person visits.

Thirdly, we note that self-reporting rates are similar across survey designs. In the Weekly Visit group, interviewers are instructed where possible to collect information directly from respondents in order to avoid proxy reporting. Meanwhile, for the Weekly Phone interviews, one household member typically reports on his/herself as well as on other household members, although the possibility exists that people may self-report by turn. In both recall survey designs, and consistent with current common practice, interviewers are instructed to ask the most knowledgeable person in the household to report on family farm labor. Despite differences in the instructions given to enumerators and in the feasibility of self-reporting by survey type, the degree of self-reporting achieved is in fact very similar across the four survey designs of the study. The response rates for self-reporting respondents were as follows: Weekly Visit (35%), Weekly Phone (33%), Recall NPS (27%), and Recall ALT (28%).

Finally, attrition was minimal. Weekly-surveyed households which dropped within the first 5 weeks following the baseline interview were replaced at random from the list of unassigned households. In the Weekly Visit group, 17 (7%) households surveyed in the baseline later dropped out of the study; these were replaced by 14 households, for a total of 212 Weekly Visit households reporting data for the main season. In the Weekly Phone group, 14 (6.2%) households dropped out and 12 were added as replacements, for a total of 212 households reporting agricultural labor throughout the season. Replacements were made in this manner up to the 6th week of the weekly interviews. None of the recall-survey households declines to participate.

3.2. Farming Practices in Mara

Although its location on the edge of Lake Victoria enables a small fishing industry, the Mara region is primarily agricultural. Here, the bulk of farming activity takes place over the main rainy season (locally known as Masika), which runs roughly from January to June. The two main crops cultivated in the villages in our study are maize and cassava. Maize has a fixed seasonal cycle of land preparation, planting, weeding, and harvesting, a cycle which is governed by the starting of the rains.[13] Cassava, by contrast, has no specific cultivation cycle and is grown throughout the year. Harvesting of cassava occurs throughout the year, depending on food needs in the household. Households frequently diversify their cultivation, intercropping these two staples with beans, sweet potatoes, and sorghum.

Before comparing labor reporting by survey design, we first provide basic household descriptive statistics and outline general time use patterns in the data. The farm households we study are large, with an average household size of 6.3 members. Households are typically composed of about one-third children under 10 and membership is 50/50 by gender. Household average farm landholdings is 1.6 hectares across 4.8 plots.[14] These plots tend not to be located adjacent to the household's dwelling, nor are they typically adjacent to each other. On average, households report plots to be a 26-minute walk from the primary residence.[15]

Most people aged 10 and over engaged in household farm labor. Table 2 provides an overview of the activities of these household members in our sample from the weekly visit data. Consistent with the agricultural character of the region, the most common activity is work on the household farm, with 87% of people spending at least one day in this activity over the season. Paid work, whether agricultural or otherwise, is rare: only 16% of people engage in any paid agricultural work for others and 11% perform paid non-agricultural work. A large share of people spend at least some time collecting firewood and water. About a quarter spend at least one day in school and just under half have are sick for at least one day over the season.

While important, family farm labor is perhaps less frequent than might be expected: people spend on average 1.9 days per week working on their household farms, conditional on having reported any work that week. We will show below that this does not necessarily imply a regular weekly work pattern. There is considerable irregularity and cyclicality in agricultural work. As suggested in a number of studies of farm labor in Sub-Saharan Africa (see discussion in Arthi and Fenske, 2016), we find that the agricultural workday typically lasts 4-5 hours.[16] It is much shorter than hours spent in non-agricultural and market activities (such as paid non-agricultural work, non-

---

[13] Our experiment starts at the beginning of the maize cycle, in January 2014, and follows respondents to the completion of the harvest in August/September.

[14] Note that these figures draw on the weekly data; as will be discussed below, recall households report fewer plots and, thus, smaller total landholdings.

[15] The time to commute to and from plots is not included in any of the working time reported in this study. Households were explicitly instructed to exclude commuting time when reporting time worked in farming activities.

[16] Although some of the discrepancy between the "true" and assumed agricultural workday is a function of recall bias, the work reported by recall-surveyed individuals suggest that even they have shorter agricultural workdays, and fewer agricultural workdays per week, than are typically assumed.

agricultural household business, fishing, livestock keeping, and schooling), conditional on such work.

The largest portion of each workday is devoted to household agriculture. Figure 1 gives an overview of the hours per day across activities. Figure 1a averages across all people for all days and Figure 1b excludes weekends and days the person was ill. Roughly a third of the total 3.6 to 4.2 working hours , respectively, are devoted to agricultural activities. These figures obscure important distributional differences, which we will return to in later sections. Finally, Figure 1c shows the allocation of time on days when at least some household farm activity is reported. On average, 5.8 hours are spent across all activities, of which 78% is spent in household farming. The remainder of the time is made up largely of collecting water, tending to livestock, and attending school.

3. RESULTS

4.1. Main Results

Over-Reporting of Hours on the Farm

To examine the implications of survey design on the reporting of household farm labor, we start from the lowest unit: the number of hours each household member spent on each household plot over the course of the entire season (henceforth, person-plot hours).[17]

Throughout the analysis presented here, and unless otherwise specified, "plots" refers to plots on which any household member was reported to have worked at any point during the season. This measure of plots depends on the actual incidence of labor (rather than on the plot's stated use) and so does not include plots held fallow, rented out, etc., for which no household labor was reported. The analysis is restricted to household members aged 10 or older who report working on any household plot during the season (a "person").[18] By this definition, then, any specific person-plot hours can be zero.

Panel A of Table 3 reports the results at the person-plot level. Hours per day, exaggerated by roughly 11%, are more accurately reported in recall than are other aspects of time use such as

[17] The Recall NPS households were asked to report the number of days spent performing each of 4 agricultural activities. They do not provide the specific days on which these activities occurred, so we do not know if reporting one day in weeding and one day in planting was in fact two separate days of work, or a single day in which both of these two activities were performed. To compute total time in hours for the Recall NPS group, we choose to compute an upper bound for the number of days by assuming that each activity-day reported was sequential or mutually exclusive—that is, that people did not perform more than one activity on the same day. This choice is supported by the similarity between this measure and the days figure reported by Recall ALT households. It is also supported by the activity patterns of the weekly-surveyed households, where we find evidence that agricultural workers overwhelmingly tend to pursue one agricultural activity in a given work day. The typical length of an agricultural workday (roughly 4-5 hours) as reported across the other three arms of the study is similar to the mean hours per activity worked in the Recall NPS survey, further supporting this interpretation.

[18] Of the 3,707 individuals ages 10 and older in the 854 households in our study, 821 reported no agricultural work and are excluded from the analysis.

hours, days, or weeks.[19] By contrast, total weeks in recall are higher than in Weekly Visits by 128%, and total days are higher by 179-223%. The cumulative effect of the exaggerated days and weeks in the recall modules results in a striking divergence in the time spent by people working on a given plot: while in the Weekly Visit group, the person-plot average of total season hours is 39.5, this number jumps to 121.3 and 146.3 in Recall NPS and Recall ALT, respectively. Total hours worked per person-plot are 3 and 3.7 times higher in the recall surveys than in our preferred benchmark, the Weekly Visit estimates.[20] These results show considerable recall bias in season-wide person-plot hours, driven primarily by error in the least granular time unit reported (days in the case of Recall NPS and weeks in the case of Recall ALT).

Aggregation of Hours and Competing Sources of Recall Bias

Does the stark difference in reported hours we find by person-plot persist at all levels? Aggregating from the labor at the person-plot level to that at the household level entails introducing data on both the number of household members working in agriculture and the number of household plots on which farming takes place. If both of these new components are recorded with the same accuracy in the recall data as in the weekly data, then the degree of over-reporting in hours per person-plot (presented in Panel A of Table 3) will be roughly the same as that aggregated by household. Panel B-D in Table 3 presents this aggregation, presenting person-level (i.e. all labor performed by a given person on any plot; panel B), plot-level (i.e. all labor performed on a given plot by any person; panel C), and finally household-level (panel D) statistics. The large difference between the weekly and recall surveys virtually disappears in aggregation.

This is due to considerable under-reporting by recall households of the number of plots cultivated and the number people farming. Table 7 shows that 1.5 (or roughly 33%) fewer household members report having worked in farming in Recall NPS and Recall ALT than in the Weekly Visit group. The number of plots in recall is underestimated by roughly 35%, or 1.6 plots. As was the case in hours, the number of people and plots reported as active in agricultural labor is essentially identical between the two weekly survey designs and between the two recall designs. In Section 4.2, we investigate which people and plots may be systematically under-reported in recall surveys and why.

[19] As we will show in Section 4.2, this is consistent with the fact that hours worked per day are more regular and less variable, relative to weeks and days per week worked.

[20] Additional comparisons can be made with our survey experiment data and the data from the three waves of the Tanzanian NPS. This is a national panel survey where sampled households are interviewed once during each survey wave (randomly across 12 months). We can compare the NPS with our weekly data since in each NPS interview, members are asked if they worked in agriculture, livestock or fisheries in the last 7 days. This is a broader set of activities than the set here, which is restricted to time spent on the plot. For the NPS subsample of rural households in or near the Mara region, both participation and hours are significantly higher in the NPS than in our weekly data (results not reported). Hours conditional on working are closer: approximately 26 to 20 hours for the NPS and our weekly data, respectively. This suggests that the respondents in the NPS are interpreting the question not literally about hours in the last 7 days, but perhaps are reporting a "typical" when working number of hours.

When comparing NPS estimates to those obtained in our end-of-season recall modules, we find that both the total days worked on plots and the average hours per working day on plots are roughly the same as in the NPS (26 days and 4.9 hours per working day in the NPS as compared to 29 days and 4.6 and 4.8 hours found in the two recall designs when analysis is conducted conditional on realized person-plot combinations (not reported)).

Returning to Table 3, we can see that the significant difference in the number of active workers between recall- and weekly-surveyed households plays out in an important way when aggregating labor at each stage. When hours are aggregated to either the person or the plot level, the total season hours remain higher in recall than in the weekly surveys, but the gap seen for person-plot hours shrinks. Recall ALT hours are 3.7 times higher as those in Weekly Visit at the person-plot level, but they are only 1.9 and 2.5 times higher at the person and plot levels, respectively.

When looking at total time from the household perspective (in hours, days, weeks), there are few statistically significant or economically meaningful differences between the four survey designs. It appears that three wrongs make a right: the competing manifestations of recall bias (i.e. the mis-reporting by recall households of days/weeks, plots, and workers) offset each other when looking at average household labor. While the reliability of labor data generated by recall surveys may be sufficient for household labor supply measures, they may be problematic for other applications, such as plot-level productivity analysis or calculations of labor force participation in agriculture.

4.2. Irregularity in Working Patterns

The dramatic difference between the recall and weekly surveys in hours worked—as well as in people and plots active in agriculture— is clear and worrisome evidence that recall bias exists in both the intensive and extensive margins of end-of-season labor reporting. What drives these large and systematic gaps? In our context, the competing forms of recall bias, hours per person-plot being over-reported and people and plots active in agriculture being under-reported, nearly cancel each other out in the aggregate. These manifestations of recall bias can be explained by the same phenomenon: the cognitive burdens of irregular working patterns. We discuss the mechanisms by which irregularity drives each of these results in turn.

Reporting of Time Worked

If forgetting were the chief mechanism by which recall bias manifests in the person-plot labor data, one might expect weekly interviews to yield higher season-total estimates than end-of-season interviews. As the direction of the bias runs counter to this explanation, forgetting does not drive our results. Instead, we propose that the hours discrepancy arises from reliance on flawed inference in the recall interviews. Weekly data show that people on average do some agricultural work in 11 out of the season's 26 weeks and on 46 out of the season's roughly 182 days. Over such a long period, Recall NPS and Recall ALT respondents are unlikely to use recall-and-count strategies when reporting total days and total weeks, respectively. That is, rather than attempting to actually count the days and weeks worked, respondents may revert to short-cuts, such as inferring season-wide work from, say, their average workload or their most recent, intense, or otherwise psychologically salient period of work. Inference may drive the weeks (Recall ALT) and days (Recall NPS) estimates. For more granular time units, it does so by the very design of the standard survey: for days (Recall ALT) and hours (both recall modules), a memory-based recall-and-count strategy is precluded, since the question asks respondents to provide an “approximate” or “typical” number.

Whether inference results in an accurate approximation of true hours worked may depend on the degree to which work takes place in regular, predictable, or uniform patterns. For the small-holders in our study, work schedules are both variable (that is, they are different from week to another) and irregular (that is, there is no systematic or predictable pattern to the variability in work across weeks). Table 3 shows that agricultural labor does not take place every day, nor does it even necessarily take place every week. However, these facts alone need not undermine the success of inference, if there exists some regular cycle or pattern to the work, which respondents can use as a relatively accurate rule of thumb.

To uncover what, if any, labor patterns exist during the season, we examine the weekly data.[21] First, we calculate the modal days spent farming for those weeks with any farm work. In Table 4, for each mode of days worked, we show the distribution of days worked. The distribution of workdays is essentially bimodal, with many people generally working in agriculture once a week (24%) and another group working 6 times a week (29%). Even though farming is the predominant activity in the region, the farming workweek is short and the majority of people farm very little each week.[22]

There is also substantial deviation from the modal working pattern. For example, out of those with a modal farm workweek of 6 days, less than half (42%) of their weeks entailed 6 days of work. For these people, 15% of their working weeks consist of 5 working days and 9% entail only 1 working day. The proportion of all workweeks conforming to the people’s modal workday (represented by the bolded diagonal in the table), usually represent under half of the weeks, except for mode 1 persons, who work one day a week in 55% of their working weeks. The proportion of weeks not conforming to the modal working pattern are relatively evenly spread from 1 to 7 working days. From this data, it is clear that even a person’s typical workweek is not that typical and that their work patterns in an atypical week vary widely.

The case with respect to hours, however, is somewhat different. In Table 5 we present modal number of hours worked per day in farming. Two patterns emerge. First, in contrast to the bimodal days-per-week patterns above, nearly half of farming workdays consist of 4 hours of work. Second, a larger share of a person’s days is spent working the same number of hours as their modal hours. That is, a larger share of days worked are on the bolded diagonal here compared to in Table 4. Hours worked per day show less variation than days worked per week. Inferences based on a “typical” workday are likelier to be accurate than those based on a “typical” workweek.

Together, these results suggest that there may be no real “typical” working pattern to which farmers can reliably refer in constructing their survey response. Furthermore, we find that the spacing of workdays or working weeks is not consistent over the season, and that the variation in days per week or hours per day observed in Tables 4 and 5 is not driven by seasonality (not

[21] For simplicity’s sake, these statistics are calculated at the person level (i.e. summed across all plots on which that person worked), rather than the person-plot level; if anything, this will understate the degree of irregularity in person-plot working patterns, since there is considerable irregularity in work on a specific plot (see below in this subsection).

[22] This reality has implications for the traditional calculations of agricultural labor and agricultural labor productivity, which tend to assume full-time engagement in farming. Even the recall-based weeks worked and hours worked per day, which we posit are over-estimates, are not high enough to support these standard assumptions.

reported), meaning that any mental short-cuts or rules of thumb used in inference (e.g. “I may not work every day, but I usually work every three days” or “I typically work 4 days a week”) may produce inaccurate estimates of season-long labor.

Having postulated that irregularity in work schedules is likely to contribute to inaccuracy in labor measurement, with error increasing as schedules become more irregular, and having shown evidence of few systematic patterns in small-holder working schedules, we turn to the question of how people arrive at the labor figures they report.

Several possibilities are raised in the cognitive psychology literature. First, people might infer their labor by extrapolating from salient episodes of work, such as the busiest workweek. Second, they might base their inference on the most recent workweek. Third, as de Nicola and Giné (2014) find in the case of earnings, people might attempt to calculate a total from their knowledge of averages. In all cases, the season-wide total is built on the basis of some subset of the season.

To examine which subset of the season may be used as the reference period for their season-wide inference, we use the weekly data. Here, we compare season-long extrapolations based on the person-level average, peak, recent, and harvest work periods in the Weekly Visit and Weekly Phone data to the recall data. These estimates are presented in Table 6. Unlike de Nicola and Giné (2014), it appears that farmers in our data do not base season-long estimates on the average workweek. This finding is consistent with the idea that it may be difficult for people to mentally calculate an average in the absence of a clear and regular work pattern. There is little evidence that respondents are inferring their season-long labor based on their average weekly labor at the harvest, the scaled total of which is much higher than the reported figures for Recall NPS and Recall ALT. Nor does it appear that respondents are making inferences based on their busiest week, which are 2-3 times higher than those reported in the recall survey designs. Instead, although none of these reference periods provides a close approximation of the recall-reported person-season hours totals, the totals inferred from the most recent work experiences appear the closest to those obtained by recall, a finding consistent with those in Schwarz and Oyserman (2001).[23]

Alternately, it might be that people look to the units, or levels of aggregation, that are intuitive or meaningful to them to form estimates of labor. If farmers tend to think about labor at the person rather than the plot level, then they may erroneously substitute for person-plot level labor with their person-level estimates. Simulations using weekly data show that such a scenario may indeed be the case: we arrive at a total of 187 hours worked over the season in the Weekly Visits and 212 hours in the Weekly Phone data. These calculations are close to those reported at the person-plot level for Recall NPS and Recall ALT (121 hours and 146 hours, respectively) than those these groups report at the person level. Similarly, if, as Weekly Visit data at the person-plot level shows, farmers only perform 2.5 weeks of work—over the entire season—on a given plot, the possibility that reporting as if they work on every plot up to as often as they work on any plot

[23] It is worth noting that since in our context, the most recent period coincides with both the peak work period and a particularly culturally and economically salient one (namely, the harvest), it is difficult to disentangle the effects of recency from those of work intensity or salience. For instance, if Recall NPS and Recall ALT household members reported labor based on the work they performed during the last weeks of the season, this could be because the most recent work performed is the easiest to remember, because this is a peak and time-bound work period, or because the work period coincides with the harvest, where work is most salient in terms of income gains.

(e.g. 12.8 plot-weeks at the person-level in Weekly Visit interviews) introduces error in plot-person labor calculations is large. Accordingly, it appears as though farmers may report work that occurs across all plots as if it occurred on every plot, inflating person-plot labor estimates.

Reporting of People and Plots

Unlike in the case of working time, the direction of bias in the reporting of people and plots active in agriculture makes forgetting a plausible explanation for the observed weekly survey-recall survey gap. Indeed, it may be a more plausible explanation in this case than mis-measurement due to poor inference, since there is less of a gray area in reporting on the extensive margin of agricultural labor. That is, it is unlikely that an person would have to infer their participation (or a plot's engagement) in agricultural work. What, then, drives certain plots and people to be forgotten? Here, as before, inconsistency plays a role, albeit perhaps a weaker one: where the work associated with them is rare or infrequent, people and plots are unlikely to be recalled as active in farming.

Reporting of Plots

First, we examine the possibility that in recall surveys, households may have forgotten (or otherwise failed to report) plots which may have fallen out of the study over the course of the season for legitimate reasons. Put another way, we ask whether for recall-surveyed households the people and plots active in agriculture are actually net end-of-season figures, while in the weekly-surveyed household, they may amount to cumulative or gross figures.

Panel B of Table 7 shows a clear gap in the reporting of plots in the recalls, where roughly half the active agricultural plots are reportd as in the weekly surveys. To track the addition and subtraction of plots over time, we turn to the weekly data. In the recall survey, plots are reported in the post-harvest endline survey (July-September 2014), whereas weekly-surveyed households first list plots in the baseline survey (January 2014). For these weekly-surveyed households, the plot roster is then updated each week for the duration of the season. This allows for changes to the original plot listing—for instance, because it is necessary to add plots that were mistakenly forgotten at baseline, to add plots that are brought into cultivation after the past weekly survey, or to drop plots which households planned to and subsequently decided not to cultivate. Figure 2 shows the increase over the course of the season in the cumulative plots reported by each group. Over the course of the weekly data collection, the number of plots per household in the Weekly Visit group grows from an average of 3.4 to 5.1 in the first 20 weeks of interviews.[24] This stands in contrast to the 2.8 plots per household reported by Recall NPS and Recall ALT households.

[24] Part of the change in plots reported over the season could be phrasing/context of the survey. In baseline households were asked "Please list all plots anyone in this household currently owns or cultivates". Over the subsequent weeks the survey was clearly (by default) referencing the 2014 long rainy season. In the endline survey, recall households were asked "please list all plots anyone in your household owned or cultivated during the 2014 long rainy season (Masika season)." Nevertheless, since our standard for the inclusion of a plot in the data depends on whether labor actually was reported on the plot during the Masika season, mis-reporting based on a misunderstanding of the baseline question's intent is unlikely to drive the weekly-recall gap in plots.

This is only slightly larger than the number of plots reported in the three waves of the Tanzania National Panel Survey (reporting plots in the 2008, 2010 and 2012 long rainy seasons) for rural farm households in the Mara or bordering regions, where the mean plots per household is 2.2. However, it is much smaller than even the smallest-ever number of plots reported by households in the weekly surveys.

The main reason for households to make changes to the plot list obtained in the baseline is that some plots are stated to have been erroneously forgotten (Table 8). Far fewer plots (91) in the Weekly Visit group are dropped over the course of the season; the chief reason given is that the plots were originally listed in error. The number of plots dropped over the course of the season is not high enough to account for the discrepancy between the plots reported by recall and by weekly interviews. Furthermore, households are more likely to add plots over the season than drop. Even if restricting the number of plots per household to those plots which faced no change over the season (i.e. plots that were never dropped early nor added late), on the premise that perhaps recall households only tended to remember to report those plots which were consistently present to be reported on, the weekly-recall plot gap remains (not reported).

Plot traits might reveal which plots are likelier to be forgotten. We do this by comparing the proportion of plots by characteristic across the four arms of the study. In Table 9, however, we show that except perhaps for an over-representation of nearby plots in the recall surveys, the plots characteristics are very similar, suggesting that recall farmers do not systematically forget certain types of plots.[25]

Instead, it could be possible that the plots that are forgotten are the prime responsibility or exclusive domain of people who are themselves likely to have gone unreported. If this were the case, then the omission of a person (say, one who worked very little, or very infrequently) from the household's reporting may also result in the omission of her/his plot. To test this, we look for significant differences in the average number of plots worked on per person and the average number of people working per plot. These figures are presented in Table 7 and suggest both that a given plot will have many different people working on it (3.2 on average in Weekly Visit households) and that a given person will work on many different plots (3.6 on average). Thus, it is unlikely that the omission of a single household member would necessarily result in the omission in "their" plot(s). However, there are significant differences by the length of the recall period in these figures, which suggests that some core set of workers and plots is being reported in recall-surveyed household, while some other workers and plots are not.

Reporting of People

We now perform a similar analysis of which household workers may be likelier to be forgotten in recall-based reporting and why.

Firstly, as above, we examine the entry and exit of people from the household using weekly data. In the baseline survey, households report which household member work on each of the household's plots. Figure 3 shows that in this earliest survey, an average of 2.1 people per Weekly

[25] Here it should be noted that the large and significant differences by ownership status are driven by technical difficulties in the collection of ownership information for some weekly-surveyed households.

Visit household are reported as engaged in agricultural work. This number grows quickly in the first several weeks, eventually ending up at roughly double the initial figure. For contrast, the recall-surveyed households are asked at the end of the season to report which household members worked on which of the household's plots; they report an average of 2.8 workers per household. Figure 3 indicates that there are differences in the rates of increase by traits, with women having a steeper curve, indicating that they tended to be not reported at baseline but subsequently working on the plot during the season. Nevertheless, Table 10 shows that women are not systematically forgotten, except perhaps to a small degree in Recall ALT.

Although gender does not make someone more or less likely to be reported by survey design, there is evidence that children, who may be less important to household farm production, and who we might expect to be marginalized within the household, are more likely to be forgotten. There is little difference across arms of the study in those who self-identify as farmers, whereas we might expect that people who identify farming as their livelihood would be less likely to be forgotten in end-of-season surveys. Indeed, in the Recall ALT group alone, proportion of farmers is statistically significantly larger than in Weekly Visit data. However, given the fact that most household members in the sample—whether identifying as farmers or not—are active in agricultural work, this occupational designation may have little practical meaning.

Finally, we look to actual working patterns to uncover the types of members who may be under-represented in the recall surveys. While the average number of household members working for an above-mean number of weeks is 1.7 in Weekly Visit households and 1.8 in Weekly Phone households, these figures are much closer to (but still somewhat lower than) the 2.8 workers per household reported by recall. Table 10 shows this idea more starkly. Those who work infrequently during the season are dramatically underrepresented, whether at the person-plot-days, or person-days levels: for instance, Recall NPS reports a household workforce composed of 13% working fewer than 10 days per person-plot, while Weekly Visit reports that 56% of their household workers fall within this category. The total season hours reported by Weekly Visit and Weekly Phone interviews which above-mean weeks are very similar to those reported by recall (not reported), further supporting the idea that those who work infrequently are likelier to be forgotten in the recall survey. With the paucity of person-plot weeks and days outlined in Table 3, and with the wide spacing of work events they suggest, the work schedules of those working infrequently are almost certainly highly irregular and difficult both to remember and to make inferences about. Accordingly, it may be the case that the bulk of members whom recall-surveyed households remember to report as active in agriculture are those who make large and/or consistent labor contributions during the season; meanwhile, there appears to be a subset of workers—perhaps those working little, infrequently, or irregularly—who stand a chance of being forgotten.

Here (by forgetting), as in the case of mis-measurement of working time (by inference), it appears that the lack of a "typical" pattern of work serves to make end-of-season surveys more cognitively burdensome than their weekly alternatives and to exacerbate recall bias in the reporting of family farm labor.

4. CONCLUSION

How accurate are data on household farm labor? Our survey experiment finds that recall data collected in the post-harvest period leads to overestimates of the time members spent on specific plots over the course of the season, in some cases by a factor of 3.7. Yet, this over-reporting is counterbalanced by considerable under-reporting—by up to 50% each—in the number of plots household members engaged in family farming. Accordingly, at the household levels of aggregation, the total season hours vary little with the recall period. Recall bias appears to result both from forgetting and from the extrapolation of season-wide labor from erroneous inferences about past labor. Both of these distortions have their roots in the irregular nature of farm working schedules and practices in our study region. In the absence of a "typical" schedule for work, or a "typical" and consistent level of engagement with workers and plots, traditional end-of-season recall surveys force respondents into cognitively taxing calculations which result in labor inferences that appear to be based on recent rather than representative experiences, the omission of members only intermittently engaged in family farm labor, and the exclusion of plots further from the house and thus less salient in the memory.

This paper makes two contributions to the literature. The first is to the literature on measurement. If our results hold in other settings, then agriculture-based low-income countries, asking about farm activities 6-12 months after they have ended will tend to exaggerate estimates of the total days and hours members spend working on their plots and farms. These findings may even translate outside the context of agriculture, for instance to settings in which some but not other components of the labor calculation face considerable variability (e.g. Dupas et al., 2015).

Clearly survey designers should tread carefully when asking questions about the frequency of non-salient and irregular events. But what is the alternative? The benchmark Weekly Visit approach used here is an expensive one that is unlikely to be a realistic prospect at the larger scale necessary for national labor surveys. A result that comes out strongly in this study is the strong performance of the phone surveys, which show little difference from the benchmark Weekly Visit. Crucially, given the significantly lower transportation costs involved, phone surveys are also by design likely to be less expensive to implement than face-to-face high-frequency alternatives—but how much cheaper?

We use the cost data available from our survey experiment to mimic a scenario wherein an existing household baseline survey adds either short face-to-face surveys or short phone surveys. The results of this costing exercise are presented in Table 11. We assume that all fixed costs related to training and preparation have been subsumed by baseline interview, and focus instead on the increase in variable costs of conducting 1, 10, 20, 25 or 30 visits or phone calls. The table shows that while phone calls are nearly 2.5 times as inexpensive revisits, the costs are non-negligible in relation to the cost of the baseline round of the survey. The cost of a single round of phone surveys is 6% of the value of the baseline survey. Notably, this estimate is very close to the 7% figure reported by Dillon (2012). This implies that contacting all respondents 10 times by phone would increase the cost of the survey by 54%, while calling all respondents 30 times would increase costs by 162%. Our particular experiment required 24 calls to cover the complete agricultural season, but this is highly context-specific, and other surveys may be able to achieve gains in accuracy with fewer points of contact.

These numbers suggest that in practice the use of high-frequency phone surveys to collect more reliable labor data may remain quite expensive. That said, they may be a viable option in surveys that already call respondents for other purposes, such as to ensure the continued participation of respondents, to keep track of respondents who relocate, or to collect data requiring a high frequency or a quick turnaround (Dillon, 2012; Garlick, Orkin, and Quinn. 2015).

Given the importance of cognitive burdens in driving mis-measurement in labor data obtained by recall, another approach is to design surveys in ways that minimize these burdens. For instance, where the analytical demands on the data make this possible, questions could be posed in ways that are more intuitive to, and better aligned with, the ways farmers remember and make inferences about their work.[26] Similarly, data collectors can attempt to shorten the recall period so that labor reporting is likelier to be based on memory than on inference.

Another approach involves managing and correcting for known shortcomings in recall survey data. For instance, by assessing the degree of irregularity in farming practices in the survey context, data collectors will be better able to anticipate whether, and to what extent, the resulting labor data will be reliable. They may also use high-frequency surveys like the ones used in this study, which dramatically shorten the traditional season-long recall period—whether as an approach to large-scale data collection unto itself, or whether as a means to create a consistent adjustment factor which can be applied to past and future recall surveys in the traditional vein. Of course, whether the latter is a reasonable approach to correcting systematic bias in reporting will depend on the specifics of the research context and the degree of variability in these specifics within a given survey group, e.g. the region, crops, degree of irregularity in farming, degree of individual responsibility over plots, prevalence of other types of economic activity, and the uses to which the resulting data will be put.

The second contribution of this study is to the debate on the agricultural productivity gap (Gollin, Lagakos, and Waugh 2014). Systematically overestimated measures of how much people work on small-holder farms leads to underestimates of labor productivity in agricultural.

This study may have a rather narrow concept of farm labor, when in fact there is more to farming than going to the field. For instance, it may fail to capture the farmer's day in sufficient detail, accounting for time spent fixing tools, planning for contingencies, negotiating land and labor agreements, and all the other economic and social interactions that are crucial to farm life. Whether the issue is as lofty as fostering structural transformation or as modest as improving data quality, it is clear that a better understanding of the farming context, including the patterns—or the lack of them—in time use, is key.

[26] Indeed, survey experiments which test the level (e.g. person-plot, person, etc.) at which individuals provide the most accurate labor histories would be a promising area for future research.

REFERENCES

Anderson Schaffner, Julie. 2000. "Employment" in Paul Glewwe and Margaret Grosh (eds). Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Development Study. Oxford University Press.

Arthi, Vellore and James Fenske. 2016. "Intra-household Labor Allocation in Colonial Nigeria." *Explorations in Economic History* 60:69-92.

Prospere Backiny-Yetna, Diane Steele, and Ismael Yacoubou Djima. 2014. "The Impact of Household Food Consumption Data Collection Methods on Poverty and Inequality Measures in Niger." World Bank Policy Research Working Paper 7090.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaimany. 2015. "The Misallocation of Labor in Village Economies." Mimeo, London School of Economics.

Bardasi, Elena, Kathleen Beegle, Andrew Dillon, 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.

Barnum, Howard and Lynn Squire. 1979. "An Econometric Application of the Theory of the Farm-Household." *Journal of Development Economics* 6:79-102.

Beegle, Kathleen, Calogero Carletto, and Kristen Himelein. 2012. "Reliability of Recall in Agricultural Data." *Journal of Development Economics* 98:34-41.

Beegle, Kathleen, Luc Christiaensen, Andrew Dabalen, and Isis Gaddis. 2016. "Poverty in a Rising Africa." Washington, DC: World Bank.

Beegle, Kathleen, Joachim De Weerdt, Jed Friedman and John Gibson. 2012. "Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania." *Journal of Development Economics* 98:3-18.

Benjamin, Dwayne. 1992. "Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models." Econometrica 60(2):287-322.

Bound, John, C. Brown and N. Mathiowetz. 2001. "Measurement Error in Survey Data." in Handbook of Econometrics Vol. 5. ed. J. Heckman and E. Leamer. Amsterdam: North-Holland, Elsevier Science.

Chen, Shaohua, and Martin Ravallion. 2007. "Absolute Poverty Measures for the Developing World, 1981–2004." Proceedings of the National Academy of Sciences 104(16):757–762.

Chowdhury, Nasima T. 2010. "The Relative Efficiency of Hired and Family Labour in Bangladesh Agriculture." South Asian Network for Development and Environmental Economics Working Paper No. 49-10.

Das, Jishnu, Jeffrey Hammer, and Carolina Sanchez-Paramo. 2012. "The Impact of Recall Periods on Reported Morbidity and Health Seeking Behavior." *Journal of Development Economics*. 98(1): 76-88.

de Nicola, Francesca and Xavier Giné. 2014. "How Accurate are Recall Data? Evidence from Coastal India." *Journal of Development Economics* 106:52-65.

De Weerdt, Joachim, Kathleen Beegle, Jed Friedman and John Gibson. 2016. "The Challenge of Measuring Hunger through Survey." *Economic Development and Cultural Change*, forthcoming.

Deininger, Klaus, Calogero Carletto, Sara Savastano, James Muwonge. 2011. "Can Diaries Help in Improving Agricultural Production Statistics? Evidence from Uganda." *Journal of Development Economics* 98:42-50.

Deolalikar Anil, and Wim Vijverberg. 1987. "A Test of Heterogeneity of Family and Hired Labour in Asian Agriculture." *Oxford Bulletin of Economics and Statistics* 49(3): 291-305.

Dillon, Brian. 2012. "Using Mobile Phones to Collect Panel Data in Developing Countries." *Journal of International Development* 24(4):518–27.

Dupas, Pascaline, Jonathan Robinson and Santiago Saavedra. 2015. The Daily Grind: Cash Needs, Labor Supply and Self-control. Mimeo.

Food and Agricultural Organization of the United Nations, 2008. "The Agricultural Bulletin Board on Data Collection, Dissemination and Quality of Statistics." Rome, Italy.

Food and Agricultural Organization of the United Nations. 2009. "How to Feed the World in 2050. High Level Expert Forum – The Special Challenge for Sub-Saharan Africa." Rome, Italy

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

Garlick, Rob, Kate Orkin, and Simon Quinn. 2015. "Call Me Maybe: Experimental Evidence on Using Mobile Phones to Survey African Microenterprises." mimeo, Duke University.

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

Irz, X., Lin, L., Thirtle, C., & Wiggins, S. 2001. "Agricultural Productivity Growth and Poverty Alleviation." *Development Policy Review* 19(4):449-466.

Jacoby, Hanan. 1993. "Shadow Wages and Peasant Family Labour Supply: An Econometric Application to the Peruvian Sierra." *The Review of Economic Studies* 60(4):903-921.

Johnston, Deborah, and Hester Le Roux. 2007. "Leaving the Household Out of Family Labour? The Implications for the Size-Efficiency Debate." *European Journal of Development Research* 19(3):355-371.

Lanjouw, O. Jean and Peter Lanjouw. 2001. "The Rural Non-Farm Sector: Issues and Evidence from Developing Countries." *Agricultural Economics* 26(1):1-23.

Ligon, Ethan, and Elisabeth Sadoulet. 2007. "Estimating the Effects of Aggregate Agricultural Growth on the Distribution of Expenditures." Background paper for the WDR 2008.

Matshe Innocent, and Trevor Young. 2004. “Off-farm Labour Allocation Decisions in Small-Scale Rural Households in Zimbabwe.” *Agricultural Economics* 30(3):175-186.

McCullough, Ellen B. 2015. “Labor Productivity and Employment Gaps in Sub-Saharan Africa.” World Bank Policy Research Working Paper 7234.

Menon, Geeta. 1993. “The Effects of Accessibility of Information in Memory on Judgments of Behavioral Frequencies.” *Journal of Consumer Research* 20:431-440.

Olinto, Pedro, Kathleen Beegle, Carlos Sobrado, and Hiroki Uematsu. 2013. “The State of the Poor: Where Are The Poor, Where Is Extreme Poverty Harder to End, and What Is the Current Profile of the World’s Poor?" World Bank Economic Premise October 2013.

Reardon, Tom and Paul Glewwe. 2000. “Agriculture.” in Paul Glewwe and Margaret Grosh (eds) Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Development Study. Oxford University Press.

Rosenzweig, Mark and Kenneth Wolpin. 1985. “Specific Experience, Household Structure, and Intergenerational Family Land and Labor Arrangements in Developing Countries.” *Quarterly Journal of Economics* 100(4):961-987.

Ross, Michael and Conway, Michael. 1986. Remembering one’s own past: the reconstruction of personal histories. In R. M. Sorrentino & E.T. Higgins (Eds.), Handbook of motivation and cognition. Pp. 122-144. New York: Guilford

Schwarz, Norbert. 2007. Cognitive Aspects of Survey Methodology. *Applied Cognitive Psychology* 21:277-287.

Schwarz, Norbert and Daphna Oyserman. 2001. “Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction.” American Journal of Evaluation 22(2): 127-160.

Shapiro, David. 1990. “Farm Size, Household Size and Composition, and Women’s Contribution to Agricultural Production: Evidence from Zaire.” *Journal of Development Studies 2*7(1):1-21.

Singh, Inderjit, Lynn Squire and John Strauss (eds) 1986. Agricultural Household Models: Extensions, Applications and Policy. Baltimore: Johns Hopkins University Press.

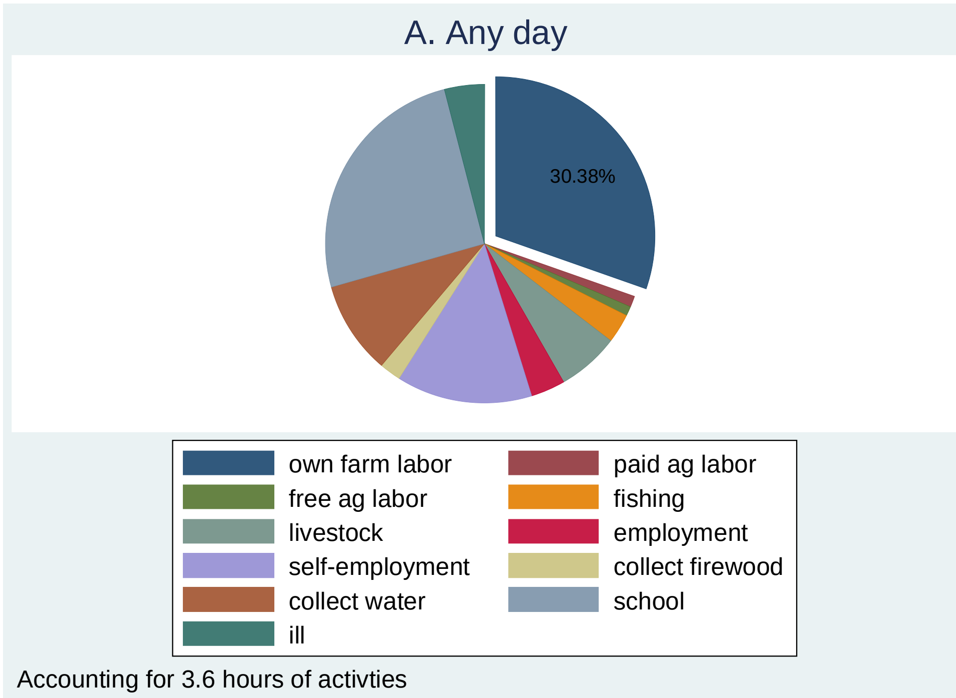
Sudman, Seymour and Norman Bradburn. 1973. “Effects of Time and Memory Factors on Response in Surveys.” *Journal of the American Statistical Association* 68(344):805-815.

Udry, Christopher. 1996. “Gender, Agricultural Production, and the Theory of the Household.” *Journal of Political Economy* 104(5):1010-1046.

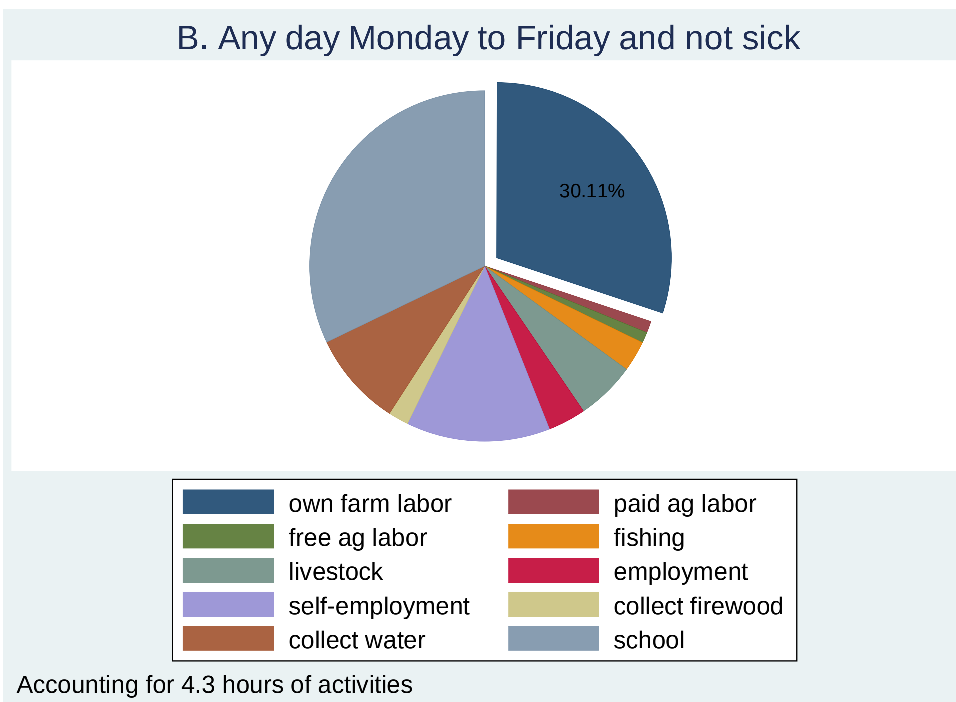
World Bank. 2014. “Tanzania - Productive Jobs Wanted (Vol. 2) : Final Report.” Washington, DC: World Bank Group.

Figure 1: Activities in an average day

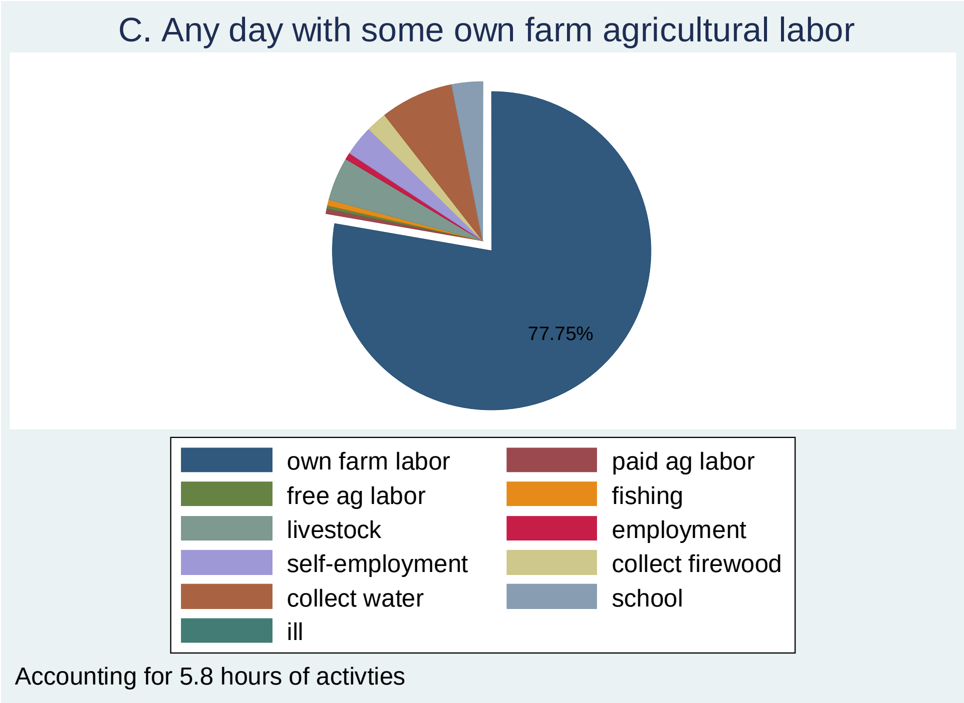
Panel A



Panel B

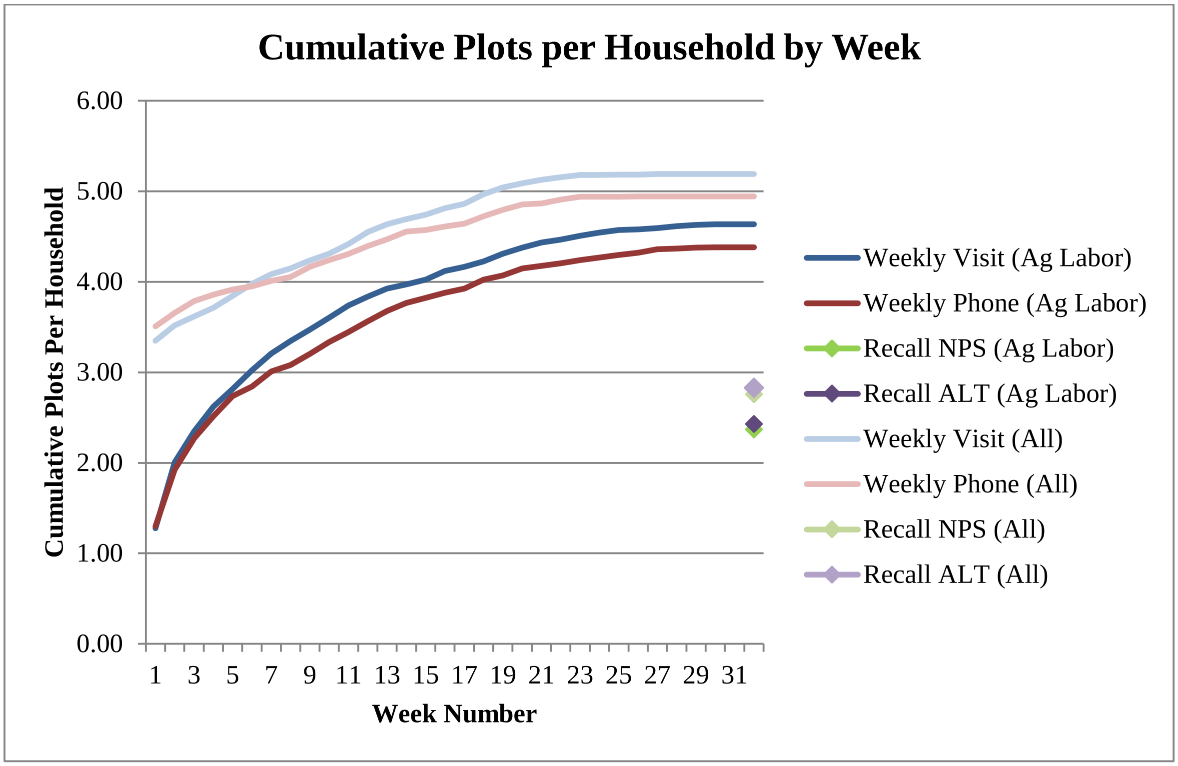


Panel C



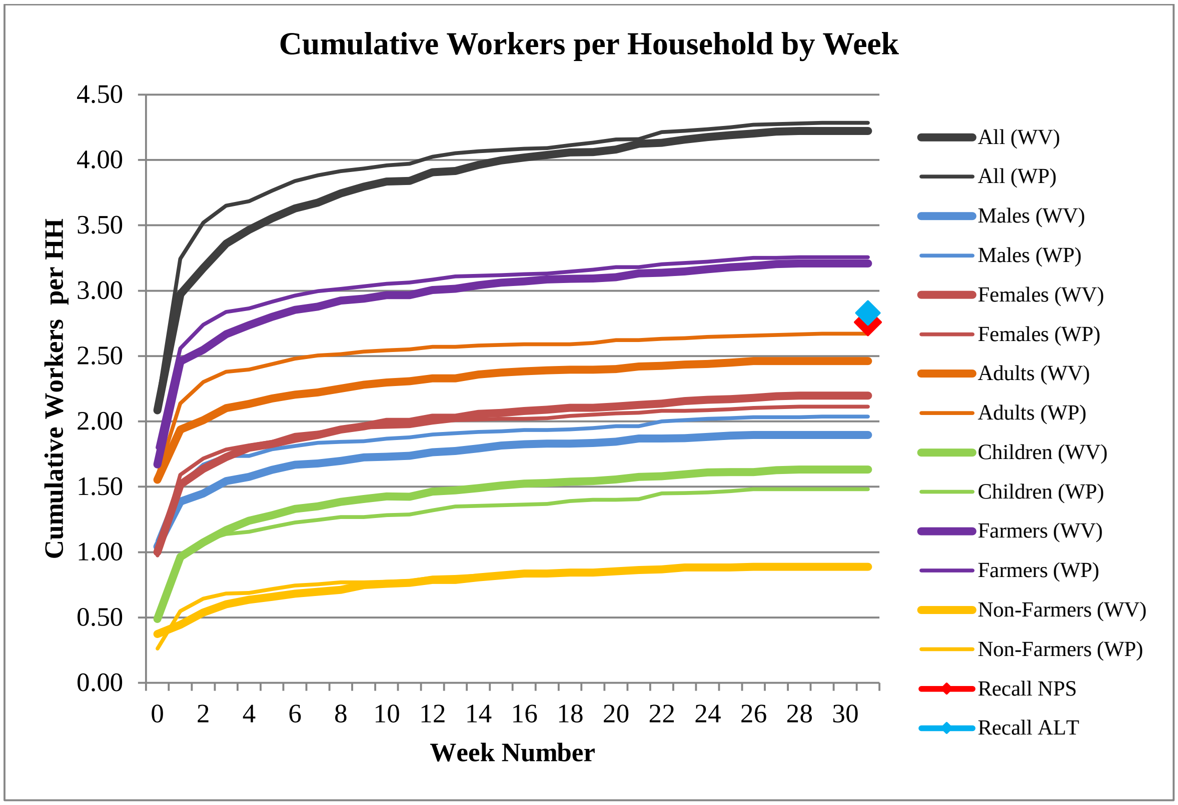
Note: All panels of Figure 1 are based on Weekly Visit data for household members aged 10 years and over. The data in Panels A and B pertains to all individuals, not just to those individuals reporting agricultural labor at any point in the season.

Figure 2: Plots Reported Over Duration of the Season



Note: "Ag Labor" refers to those plots on which own-household agricultural labor was reported at any point in the season, while "All" refers to any household plot reported, irrespective of its stated or actual use. Week number 0 refers to the baseline questionnaire, while week 31 refers to the endline.

Figure 3: Family Farm Workers Reported Over Duration of the Season



Note: Week number 0 refers to the baseline questionnaire, while week 31 refers to the endline. All individuals in this figure are ones who reported own-household agricultural labor at some point in the season. For ease in reading, only the average number of workers for Recall NPS and Recall ALT are provided above as a comparison to weekly data. Adults are defined as individuals 20 and over, and children as those aged 10-19, inclusive. Farmers are those who self-report their occupation as being in farming.

Table 1: Sample Characteristics

	Weekly Visit	Weekly Phone	Recall NPS	Recall ALT
Individuals (*N*=5,375)				
Age	20.98	22.47*	22.34*	21.60
	(20.12)	(20.47)	(20.70)	(19.71)
Proportion aged 10 years and over	0.63	0.67**	0.63	0.63
Proportion male	0.49	0.48	0.49	0.51
Proportion in school	0.28	0.32**	0.30	0.30
Proportion living with spouse	0.27	0.31*	0.28	0.27
Proportion literate	0.58	0.61	0.56	0.56
Proportion father deceased	0.28	0.26	0.29	0.28
Proportion mother deceased	0.16	0.17	0.17	0.16
Proportion visit health care provider past 4 weeks	0.16	0.15	0.15	0.14
Households (*N*=854)				
Household size	6.44	6.54	6.27	6.21
	(3.1)	(3.3)	(2.9)	(2.4)
Rooms in dwelling	2.93	3.08	2.86	2.98
	(1.2)	(1.3)	(1.1)	(1.2)
Minutes to water source	58.49	55.01	54.81	53.50
	(48.3)	(43.4)	(45.7)	(41.5)
Proportion with good walls	0.47	0.48	0.40	0.44
Proportion with good roof	0.74	0.78	0.76	0.78
Proportion with good floor	0.22	0.32**	0.24	0.31**
Number of households	212	212	212	218

Note: Table uses endline data. Mean values which are significantly different from the mean for the Weekly Visit group are denoted as follows: *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Overview of Activities during the Agricultural Season

Activity	**Share of Individuals reporting the activity at least once over season**	**Average days per week in activity, conditional on reporting the activity at least once over the season***	**Hours per day in activity, conditional on activity that day**
Household farm	0.88	1.88	4.49
Paid agricultural	0.16	0.34	4.65
Free agricultural, other hh	0.21	0.28	4.38
Fishing	0.10	1.24	6.38
Livestock	0.27	1.08	5.08
Paid non-agricultural	0.11	1.00	8.38
Non-agricultural business	0.31	1.43	7.59
Collecting firewood	0.56	0.49	2.01
Collecting water	0.73	2.72	1.23
Schooling	0.27	2.76	7.86
Sick	0.49	N/A	N/A

Note: The table is based on Weekly Visit data, and is restricted to individuals aged 10 years and over.

Table 3: Total Hours and Days of Agricultural Labor Reported Over Season

	Weekly Visit	Weekly Phone	Recall NPS	Recall ALT
A. Per person-plot				
Hours	39.5	48.8***	121.3***	146.3***
	(69.5)	(85.2)	(133.8)	(159.3)
Days	9.2	10.7***	25.7***	29.8***
	(14.2)	(14.9)	(24.6)	(29.6)
Weeks	2.5	2.6	N/A	5.7***
	(3.2)	(3.1)		(5.2)
Hours per day worked	4.1	4.4***	4.6***	4.6***
	(1.3)	(1.5)	(1.2)	(1.1)
B. Per person (all household plots)				
Hours	201.0	228.3***	313.5***	389.56***
	(196.6)	(222.8)	(332.2)	(436.8)
Days	46.4	49.6*	66.5***	79.3***
	(40.9)	(39.4)	(62.0)	(80.5)
Weeks	12.8	12.0*	N/A	15.3***
	(9.0)	(8.3)		(13.8)
Hours per day worked	4.1	4.3***	4.6***	4.6***
	(1.1)	(1.2)	(1.1)	(1.1)
C. Per plot (all household persons)				
Hours	183.0	223.1***	363.9***	452.4***
	(232.3)	(298.5)	(457.59)	(522.7)
Days	42.2	48.5***	77.2***	92.1***
	(49.1)	(52.9)	(82.2)	(99.0)
Weeks	11.7	11.8	N/A	17.7***
	(11.2)	(11.3)		(17.7)
Hours per day worked	4.1	4.3***	4.6***	4.7**
	(1.1)	(1.4)	(1.1)	(1.2)
D. Per Household (all persons and all plots)				
Hours	848.6	977.6*	865.1	1104.1**
	(699.7)	(823.2)	(1151.3)	(1548.3)
Days	195.8	212.3	183.5	224.9
	(151.8)	(147.4)	(213.4)	(288.2)
Weeks	54.0	51.9	N/A	43.3**
	(35.1)	(32.8)		(50.2)
Hours per day worked	4.1	4.2	4.6***	4.7***
	(0.8)	(1.2)	(1.1)	(1.1)

Note: Mean values which are significantly different from the mean for the Weekly Visit group are denoted as follows: *** p<0.01, ** p<0.05, * p<0.1. All of the calculations are restricted to those aged 10 and older who reported having performed agricultural labor at any point in the season, and those plots reporting a positive number of hours of agricultural labor at any point in the season. The calculations are based on all plausible (but not necessarily realized) person-plot combinations per the preceding definition of individuals and plots. “N/A” indicates that the information is not collected in the survey design.

Table 4: Modal Days Farmed per Week Farmed

Modal days	Frequency (%)	Distribution of days farmed, for a given mode (%)						
		1	**2**	**3**	**4**	**5**	**6**	**7**
1	24.4	**55.7**	14.9	7.8	6.4	5.6	7.3	2.3
2	12.1	17.9	**41.0**	11.4	8.41	8.2	8.5	4.6
3	7.2	14.7	14.7	**33.8**	11.1	11.8	10.0	3.9
4	6.4	11.8	13.8	12.8	**34.7**	11.2	11.8	3.9
5	10.3	12.2	13.0	13.6	11.3	**34.5**	11.7	3.7
6	29.0	9.1	7.6	9.0	11.0	15.4	**41.8**	6.1
7	10.5	6.29	8.7	8.4	9.5	11.1	15.3	**40.9**

Note: This table is based on the data for Weekly Visit individuals aged 10 or over and considering weeks in which some own-household agricultural labor was reported. We do not consider work reported in the baseline, since working patterns cannot be discerned from the data therein. The table can be read as follows. 29.02% of considered individuals have a modal working week of 6 days (in weeks with any own-household agricultural work). 41.75% of their working weeks actually entailed working six days, while 9.13% of their weeks they worked one day.

Table 5: Modal Hours Farmed per Day Farmed

Modal hours	Frequency (%)	Distribution of hours farmed, for a given mode (%)				
		2	**3**	**4**	**5**	**6**
1-2	5.4	**49.0**	13.5	21.0	6.6	9.9
3	12.9	10.9	**53.9**	20.5	8.9	5.9
4	48.3	4.5	14.8	**56.9**	13.7	10.1
5	15.0	3.2	10.9	25.9	**46.5**	13.5
6+	18.3	3.5	8.9	18.3	16.0	**53.3**

Note: This table is based on the data for Weekly Visit individuals aged 10 or over. We do not consider work reported in the baseline, since working patterns cannot be discerned from the data therein. Less than 2 percent of all observations on hours per day were under 2 hours; 7 percent were more than 6 hours.

Table 6: Scaled Comparisons to Reported Total Person-Level Season Hours

	Weekly Visit	Weekly Phone	Recall NPS	Recall ALT
Actual reported hours	201.0	228.3	313.7	389.5
	(196.6)	(222.8)	(332.5)	(436.9)
Scaling based on time unit:				
Hours in busiest week (scaled up by 26 weeks)	939.4	1011.1		
	(642.2)	(694.7)		
Hours in most recent week (scaled up by 26 weeks)	392.9	498.2		
	(348.9)	(348.9)		
Hours in average harvest week (scaled up by 26 weeks)	432.9	629.2		
	(532.2)	(654.4)		
Hours in average week (scaled up by 26 weeks)	410.5	484.4		
	(229.1)	(244.1)		

Note: In this table, all figures are reported at the person level. Scaling is based on the variation in weekly data, and is compared to the actual reported figure amongst recall-surveyed individuals.

Table 7: People and Plots Active in Household Farming

	Weekly Visit	Weekly Phone	Recall NPS	Recall ALT
A. People per household				
All people	4.9	5.0	4.0***	3.9***
	(2.4)	(2.4)	(2.2)	(1.9)
People working on the farm	4.2	4.3	2.6***	2.7***
	(2.1)	(2.2)	(1.5)	(1.4)
Plots worked per person	3.6	3.5	2.3***	2.4***
	(1.9)	(1.9)	(1.3)	(1.3)
B. Plots per household				
All plots	5.2	5.0	2.8***	2.8***
	(2.4)	(2.2)	(1.6)	(1.6)
Plots cultivated	4.6	4.4	2.4***	2.4***
	(2.2)	(2.0)	(1.3)	(1.3)
Plots cultivated, exc. plots dropped	4.4	4.2	N/A	N/A
	(2.2)	(2.0)		
Plots cultivated, exc. plots added	3.1	3.2	N/A	N/A
	(1.4)	(1.3)		
Plots cultivated, exc. plots dropped and added	2.9	3.1	N/A	N/A
	(1.4)	(1.3)		
People working per plot cultivated	3.2	3.4*	2.5***	2.7***
	(1.8)	(1.9)	(1.3)	(1.4)

Note: Mean values which are significantly different from the mean for the Weekly Visit group are denoted as follows: *** p<0.01, ** p<0.05, * p<0.1. All of the calculations in Panel A are restricted to those aged 10 and older. “All plots” refers to all plots reported by the household, including plots which are fallow, rented out, and cultivated (including those owned and rented in). “Plots cultivated” refers to those plots on which agricultural labor was actually reported as taking place.

Table 8: Changes in Plot Listings Over the Season

	Weekly Visit	Weekly Phone
A. Plots added after baseline		
Forgot to list before	247 [216]	122 [107]
Started renting plot	95 [83]	91 [85]
Split off plot	13 [9]	38 [29]
Plot was given to household	23 [17]	18 [16]
Bought plot	8 [8]	16 [15]
Other reason	3 [2]	10 [6]
No reason given, added in week 1	2 [2]	1 [0]
No reason given, added after week 1	7 [0]	10 [2]
Total	398 [337]	360 [260]
B. Plots dropped before endline		
No longer renting plot	30 [21]	9 [8]
No longer cultivating plot	12 [7]	4 [4]
Sold plot	7 [6]	0 [0]
Gave away plot	8 [4]	1 [1]
Other	0 [0]	1 [1]
Total	91 [56]	49 [33]

Note: The figures presented without brackets are the number of plots in the designated category, and figures in brackets are the subset of these plots reporting any agricultural labor during the season. The list includes 25 plots on which agricultural labor was reported which were added late, only to be later dropped; and 18 plots on which agricultural labor was reported which were dropped, only to be added back later.

Table 9: Characteristics of Plots Reporting Agricultural Labor

	Weekly Visit	Weekly Phone	Recall NPS	Recall ALT
Mean plot size (ha)	0.39	0.41	0.36	0.36
	(0.38)	(0.41)	(0.32)	(0.35)
Plot size (ha), proportion:				
(0, 0.5]	0.67	0.64	0.71	0.68
(0.5, 1]	0.14	0.15	0.12	0.14
(1, 1.5]	0.04	0.05	0.05	0.03
(1.5, 2]	0.01	0.01	0.01	0.01
(2, 3.5]	0.01	0.01	0.00	0.00
Unknown	0.13	0.14	0.12	0.13
Mean distance from residence (min)	31.57	33.73	31.22	29.53
	(37.08)	(35.87)	(40.92)	(39.77)
Distance (min), proportion:				
(0, 30]	0.66	0.60***	0.74***	0.75***
(30, 60]	0.23	0.27**	0.15***	0.16***
(60, 90]	0.06	0.07	0.03**	0.03**
(90, 120]	0.02	0.03*	0.06***	0.04**
(120, 240]	0.03	0.03	0.03	0.02
Ownership status, proportion:				
Owned	0.68	0.69	0.83***	0.82***
Used free	0.05	0.04	0.05	0.07
Rented in	0.13	0.15	0.11	0.11
Unknown^	0.14	0.13	0.00***	0.00***
Proportion cultivating any maize	0.39	0.38	0.39	0.42
Proportion cultivating sweet potato	0.25	0.24	0.30*	0.30*
Proportion cultivating cassava	0.37	0.39	0.42*	0.40
Proportion cultivating no maize, sweet potato, or cassava	0.22	0.17**	0.20	0.17**

Note: ^ Ownership status was erroneously missing for some plots due to data processing error. For mean and proportion values which are significantly different from those for the Weekly Visit group are denoted as follows: *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Characteristics of People Reporting Agricultural Labor

	Weekly Visit	Weekly Phone	Recall NPS	Recall ALT
Proportion adults (ages 20 and up)	0.60	0.65**	0.74***	0.73***
Proportion children (ages 10-19)	0.40	0.35**	0.26***	0.27***
Proportion men	0.47	0.49	0.49	0.52**
Proportion women	0.53	0.51	0.51	0.48**
Education level, proportion:				
Below primary	0.69	0.67	0.72	0.75***
Primary	0.01	0.00*	0.01	0.01
Above primary	0.24	0.25	0.15***	0.09***
Proportion stated occupation farmer	0.78	0.78	0.82	0.83**
Proportion working <10 days (pp)	0.56	0.50***	0.13***	0.22***
Proportion working <20 days (pp)	0.78	0.76*	0.38***	0.44***
Proportion working <30 days (pp)	0.87	0.87	0.61***	0.57***
Proportion working <10 days (p)	0.19	0.16	0.06***	0.09***
Proportion working <20 days (p)	0.35	0.30**	0.16***	0.22***
Proportion working <30 days (p)	0.46	0.41**	0.29***	0.32***

Note: Values which are significantly different from those for the Weekly Visit group are denoted as follows: *** p<0.01, ** p<0.05, * p<0.1. The designation "pp" refers to per person-plot, while "p" refers to per person.

Table 11: Per-Household Interviewing Cost Increases

# Interviews	Weekly Visit	Weekly Phone
1	14%	6%
10	139%	54%
20	277%	108%
25	346%	135%
30	416%	162%

Note: The costs presented in this table are the cost increases in US Dollars, per household, relative to the cost of an LSMS-type (baseline) survey.