RECALL LENGTH AND MEASUREMENT ERROR IN AGRICULTURAL SURVEYS

Philip Wollburg[a]
Marco Tiberti[a]
Alberto Zezza[a]

50x2030 WORKING PAPER SERIES
FEBRUARY 2021

Abstract

This paper assesses the relationship between the length of recall and nonrandom error in agricultural survey data. Using data from the World Bank's Living Standards Measurement Study–Integrated Surveys on Agriculture in Malawi and Tanzania, the paper shows that key input and output variables are systematically related to the length of the recall period, indicating the presence of nonrandom measurement error. With longer recall periods, farmers report greater quantities of harvest, labor, and fertilizer inputs. Farmers list fewer plots as the recall period increases. The paper argues that it is plausible that farmers overestimate plot-level outcomes, or they forget some of their more marginal plots due to longer recall periods. The analysis also finds evidence of measurement error related to the length of recall in common measures of agricultural productivity. The size of the recall effect typically varies between 2 and 5 percent per additional month of recall length, which is economically significant. With data reliability affecting policy effectiveness, improving agricultural survey data quality remains an important concern. Mainstreaming objective measures where possible and reducing the risk of recall error through shorter recall periods appear to be promising avenues to improve the quality of key variables in agricultural surveys.

JEL classifications: Q12, Q18, O12, C81, C83

Keywords: agriculture, measurement error, recall, survey design

---

[a] Development Data Group, The World Bank. The authors would like to thank for their comments Chris Barrett (Cornell University), Isis Gaddis (World Bank and IZA), Lakshman Nagraj Rao (Asian Development Bank), Gero Carletto (World Bank), Aparajita Goyal (World Bank), Sydney Gourlay (World Bank), Talip Kilic (World Bank), Heather Moylan (World Bank), Ilaria Lanzoni (World Bank), and seminar participants at the 2019 AIEAA Conference and the Joint UNU-MERIT/School of Governance Seminar in November 2019. This analysis was supported by the 50x2030 Trust Fund.

This is a publication of the 50x2030 Initiative to Close the Agricultural Data Gap, a multi-partner program that seeks to bridge the global agricultural data gap by transforming data systems in 50 countries in Africa, Asia, the Middle East and Latin America by 2030. For more information on the Initiative, visit 50x2030.org.

1. Introduction

For many developing countries, boosting agricultural productivity and incomes of small food producers are key policy objectives in the effort to alleviate poverty and improve food security. The international community has also embraced those goals, making them prominent targets in the Sustainable Development Goals (SDG) monitoring agenda under SDG 2, Zero Hunger.[1] High-quality, reliable farm output and input data are indispensable both for measuring progress towards the SDGs and for effective policy and program design and analysis.

Household and farm surveys are the most important source of agricultural data, rivalled in some countries by administrative data. In all measurement, survey-based or otherwise, there is some amount of measurement error. In surveys, the necessary reliance on imperfect respondent recall is a common source of error. Measurement in agricultural surveys in low-income countries is particularly difficult because agricultural operations are complex and seasonally variable, respondents are often illiterate, familiarity with standard units of measures is the exception rather than the norm, and a host of other factors. When data are of poor quality, and particularly when they are affected by systematic biases, policies based on that information may turn out to be less effective, if not misguided.

Methodological research shows that survey design matters and that conscious efforts to adopt improved design and survey administration choices can contribute significantly to minimizing measurement error, including recall error,[2] and improve the quality of agricultural survey data in developing countries (Carletto et al., 2015). This paper contributes to these efforts by focusing on the length of the recall period and how it affects the reliability of agricultural survey data. The length of the recall period is thought to be associated with recall decay, in which survey respondents forget details of events or forget events entirely, leading to reporting inaccuracies (Sudman and Bradburn, 1973). In agricultural surveys, recall length also often differs across the various activities recorded because farming activities, such as planting, plot maintenance, or harvest, take place at different times during the agricultural year. Depending on the timing of survey administration, respondents may be asked to recall the details of events sometimes many months in the past and the recall period may vary substantially across households or farms in the same sample if the fieldwork is completed over several months. In addition, some countries have two or more cropping seasons in one agricultural year and agricultural activities can be very different in both extent and nature depending on the season, which can make it even harder to recall events correctly.

Survey designers and planners have some control over the length of the recall period by choosing the number and timing of field visits to farms. This choice, however, has clear cost implications. Visiting households or farms multiple times in order to shorten the length of the recall period is more expensive and logistically demanding than collecting information only once during the agricultural year. In practice, surveys differ widely with respect to the number and timing of visits.[3] Survey designers may also choose to rely on more

[1] Especially relevant are indicators 2.3.1 *Productivity of small-scale food producers* and 2.3.2 *Income of small-scale food producers.*

[2] See, for example, Kasprzyk (2005), Zezza et al. (2016), Gaddis et al. (2019) and Kilic et al. (2018).

[3] For example, the Tanzania National Panel Survey relies on one visit per household per year to collect information about Tanzania's two cropping seasons. Data collection begins after the main season's harvest and is rolled out over the course of 12 months. The AGRIS model, developed as part of the Global Strategy for Agricultural and Rural Statistics and published at the end of 2017 (Global Strategy to improve Agricultural and Rural Statistics, 2017), proposes a core module administered once per year

objective 'gold standard' measures of key variables, such as GPS land area measurement or crop cutting, if only for a subsample of farms. That choice, too, has cost implications. To evaluate the trade-off between quality and cost, survey designers require an empirically-informed understanding of the extent to which data quality varies with differences in the recall period, among other data quality considerations. Policy makers and other data users benefit from an appreciation of the limitations and biases in existing data related to recall length.

Previous work by Beegle et al. (2012a) found no consistent evidence of significant recall length effects in agricultural surveys, using data from the early 2000s. In contrast, in this paper, we find a significant effect of the recall length on key agricultural statistics. We take advantage of the more recent availability of high-quality and very detailed agricultural survey data to overcome some of the limitations of those data and assess and quantify the impact of the recall length on the quality of agricultural data. Using data from nationally representative surveys conducted by the National Statistical offices of Tanzania and Malawi with support from the World Bank's Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) program, we find that longer recall periods are associated with overreporting of plot-level outcomes – production quantity, labor input, and fertilizer input – consistently across the three data sets in the analysis. At the same time, respondents appear to forget listing plots as the recall period increases. Our results point to the need to devise and promote the adoption of survey design choices that minimize this source of measurement error.

The remainder of the paper is structured as follows: Section 2 reviews the literature on error in the measurement of agricultural outcomes, especially that related to recall length. Section 3 describes the data used, the variables generated, and the empirical strategy this analysis employs. Section 4 presents the main results. Section 5 discusses conclusions, policy implications, and offers recommendations.

2. Background and related literature

This paper contributes to a large body of literature exploring the effect of respondent recall on measurement accuracy in survey data. Systematic reviews of this literature are provided in Biemer et al. (2011), Bound et al. (2001), Groves and Lyberg (2010), and Meyer et al. (2015), among others. The role of recall error in survey data has been explored in a range of topics, including consumption expenditure and food intake (e.g. Beegle et al., 2012b; Troubat and Grünberger, 2017; Backiny-Yetna et al., 2017; Brzozowski et al., 2017; Engle-Stone et al., 2017; D'Alessio, 2017; Schündeln, 2018; Zezza et al., 2017), household enterprises (De Mel et al., 2009; Liedholm, 1991) and income measurement (Moore et al., 2000).

One strand of the literature focuses on the cognitive processes underlying survey response and has built an understanding of how these can lead to measurement error. Sudman and Bradburn (1973) distinguish between *recall decay* and *telescoping*. Telescoping refers to inaccurately remembering *timing* of events, leading to an event of interest incorrectly being moved into the reference period, which, in turn, may lead

at the end of the main agricultural season, complemented by an Economy module implemented in four visits during the agricultural year. Similarly, the Uganda Annual Agricultural Survey 2019 (AAS), is planned with a post-planting and a post-harvest visit in both agricultural seasons, for a total of four visits. In contrast, many LSMS-ISA surveys are now implementing two visits, one at the end of the main planting season and one at the end of the main harvest. Finally, some specialized farm surveys visit households/farms multiple times over the course of the agricultural year.

to over-reporting. In recall decay, respondents forget details of events or forget events entirely. The length of the recall period is arguably the most important driver of recall decay. As the recall length increases, recall decay occurs, and respondents no longer carry the requested information in memory, they turn to *reconstruction* strategies (Moore et al., 2000). Common reconstruction strategies include basing the response on typical behavior or average circumstances, which may lead to reporting error; whether it is nonrandom depends on the type of reconstruction strategy and may be related to respondent characteristics, such as cognitive ability and education. Reconstruction is also associated with *rounding* of quantities, which is common in all domains of subjective reporting (Roberts and Brewer, 2001) and for agriculture has been documented for instance with respect to land area measurement by Carletto et al. (2011). Finally, the *salience* of events is thought to counteract recall decay as respondents are able to recall salient events more easily, though some research has found salience being associated with over-estimation (see discussion in Bound et al., 2001).

There is some evidence of the effect of recall error on each of the outcomes of interest of this study – land input, crop production, labor and agricultural input use in low-income settings. Beegle et al. (2012a) explore the effect of recall length on the reliability of both agricultural input and output data, while others have focused on error affecting particular aspects of measurement of the agricultural production process (Arthi et al., 2018; Gaddis et al., 2019; Kilic et al., 2018; Zezza et al., 2016).

Regarding the measurement of **agricultural land**, a recent literature explores the biases in farmer-reported land area estimates (e.g. Carletto et al., 2011; De Groote and Traoré, 2005; Dillon et al., 2019; Goldstein and Udry, 1999; Keita and Carfagna, 2009; Kilic et al., 2017; Schøning, 2005), though these studies do not explore the role of recall length in affecting the error in farmer-reported land area. Two recent studies, Arthi et al. (2018) and Gaddis et al. (2019), find that respondents tend to undercount the number of plots under cultivation when using end-of-season recall to elicit the information.

With respect to **crop production** measurement, Beegle et al. (2012a) find no consistent evidence that the length of the recall period affects farmers' harvest estimates in data from Kenya, Malawi, and Rwanda from the early 2000s. In contrast, Deininger et al. (2012) use a different setup to assess recall error: they compare production estimates from farmer recall at the end of the season to production estimates based on a continuously administered harvest diary for a wide range of crops in Uganda. They find that harvest quantities based on end-of-season recall diverge significantly from those recorded in harvest diaries, which are deemed more reliable. In most cases, end-of-season recall is associated with under-reporting, especially for extended-harvest crops such as cassava or banana. Cash crop production, in contrast, is significantly over-reported. Similarly, Kilic et al. (2018) compare weekly harvest diaries to a 12-month recall period and a 6-month recall period each. All methods are benchmarked against crop-cutting estimates, considered the 'gold standard' for estimating crop production. The authors show significant under-reporting in the recall data, especially with the 12-month reference period, relative to the diary and crop cutting methods. A different strand of the literature found that reported harvest quantity is inversely related to land area: respondents tend to over-report maize harvest on small plots and under-report it on large plots (Desiere and Jolliffe, 2018; Gourlay et al., 2017).

Beegle et al. (2012a) assess the effect of recall length on farmer reported **labor inputs**, finding no consistent evidence of recall decay affecting reported quantities. Arthi et al. (2018) and Gaddis et al. (2019), on the other hand, find significant measurement error in recall-based farm labor estimates. The two studies

compare end-of-season recall with weekly work diaries, which are considered more reliable, among farmers in Tanzania and Ghana. They find that end-of-season recall leads to over-reporting of labor use at the individual-by-plot level ('recall bias') relative to diary-keeping, but at the same time to undercounting of plots cultivated and of individuals who worked on them ('listing bias'). In contrast, Gollin (2019) argues that diary methods may lead to under-counting of farm labor given the eclectic set of tasks that 'farm labor' comprises. Seymour et al. (2017) find significant differences in the recording of time spent working between recall and diary methods in Uganda and Bangladesh.

Finally, empirical evidence on recall error (and measurement error in general) in **agricultural inputs** data (fertilizer, agro-chemicals, seeds, and others) is scant. Beegle et al. (2012a) find little systematic evidence for recall bias in farmer self-reported fertilizer usage in Malawi and Kenya, though results vary somewhat depending on respondent characteristics. Gollin (2019) argues that farmers likely recall quantities (and prices) of purchased inputs mostly accurately but also points to evidence (e.g. in Ashour et al., 2017; Bold et al., 2017) suggesting that fertilizer and agro-chemical counterfeiting and adulteration may leave farmers unsure about the quality of product they apply to their land.

3. Data and empirical strategy

Data

The study uses three data sets from nationally representative household surveys in Malawi and Tanzania: the Tanzania National Panel Survey (TNPS) 2012/13, the Fourth Malawi Integrated Household Survey 2016/17 (IHS4) and the Malawi Integrated Household Panel Survey 2016/17 (IHPS).

Tanzania NPS 2012/13 is the third wave of the Tanzania National Panel Survey and was collected between October 2012 and November 2013. The NPS sample used a multi-stage clustered sampling design covering a total is 5,015 urban and rural households. With households selected to be interviewed over the course of 14 months to account for seasonality in consumption, the data have variation in the recall period, which can be exploited to assess the impact of recall decay on data quality. The reference agricultural season is the long rainy season of 2012, which all farms in the sample reported on.

The IHS4 2016/17 includes both a cross-sectional and a panel component. The cross-sectional sample includes 12,480 households surveyed in 780 enumeration areas. Households were visited once throughout the 12 months of fieldwork between April 2016 and April 2017. The IHS4 data were collected in the same way as the NPS 2012/13 and used for analysis following the same rationale. For our analysis, we use the 2015/16 rainy season as the reference agricultural season. Data collection began in April 2016, at which point not all households had finished the 2015/16 rainy season harvest. These households instead reported on the harvest of the previous, 2014/15 rainy season. For comparability, we dropped these households from the analysis. In addition, some households stated having harvested in the 2015/16 season, but harvest dates revealed respondents were referring to the previous season, and vice versa. We dropped these households as well.

The IHPS 2016/17 sample includes 1,989 households which were interviewed twice between April 2016 and April 2017, in one post-planting and one post-harvest visit relative to the 2015/16 rainy season. Given

this two-visit setup, the recall length is shorter, especially for planting-related activities. Including this data set thus allows us to assess if recall effects vanish with somewhat shorter recall periods.

All three surveys are part of the Living Standards Measurement Study – Integrated Survey on Agriculture (LSMS-ISA) and, as such, the data sets contain an integrated household and agricultural component. The household survey component collects detailed socioeconomic information, including household-level data on consumption, income, assets and housing, and individual-level data on demographics, education, and health. The agricultural component collects detailed information, among other items, on agricultural inputs used and outputs produced, as well as output disposition, at the plot-level. The integrated survey further allows breaking down which household members are involved in agricultural production, who owns the means of production, and who is responsible for managing the household farm's plots of land.

As ancillary data to validate our findings on farmer-reported crop production, we use the second round of the Methodological Experiment on Measuring Maize Productivity, Soil Fertility and Variety (MAPS) survey. The MAPS data, which Gourlay et al. (2017) discuss in detail, were collected in two rounds in 2015 and 2016 in Uganda, containing 900 maize-growing households in round 1 and 489 in round 2. The MAPS survey employed objective and subjective survey methods to collect information on maize production (crop cutting), area, soil fertility, and maize variety identification.

Outcomes of interest

The analysis focuses on some of the agricultural outcomes of major policy interest, namely land input, crop production, labor and input use. The main dependent variable is the approximate length of the recall period between the interview and the activity relating to the outcome of interest (Table 1).

Land input reporting is assessed in two ways. First, as the total number of agricultural plots (or parcels) cultivated by the household. This variable is simply a count of the households' agricultural plots (and parcels) which respondents list at the beginning of the agricultural questionnaire of the survey. This information sheds light on whether a longer recall period leads households to forget listing some plots (Gaddis et al., 2019). The median (mean) number of plots listed per household is 2 in Tanzania NPS 2012/13 (mean: 2.2 plots) and Malawi IHPS 2016/17 (mean: 2.1 plots) and 1 in Malawi IHS4 (mean: 1.6 plots; Table 2).

Second, as the difference between GPS-measured plot area and farmer self-reported plot area. Given the prevalence of measurement error in land area reporting, LSMS-ISA surveys, including the data used in this analysis, implement an objective, GPS-based area measurement, while retaining farmer self-reported plot area alongside. Comparing these two measures allows assessing whether recall decay exacerbates the error in self-reported land area. This variable is constructed by subtracting the plot area variable based on GPS measurement from the self-reported plot area variable, so that area under-reporting has a negative sign while area over-reporting has a positive sign.

Plot sizes, measured by GPS, differ substantially across the three data sets. In Tanzania NPS 2012/13, the average plot size of 1.23 hectares (median: 0.49 hectares) is much larger than in Malawi at 0.37 hectares in the IHS4 (median: 0.30 hectares) and 0.34 hectares in the IHPS data set (median: 0.26 hectares; Table 2).

The inverse relationship between measured plot size and error in self-reported plot size discussed in the literature is also common to all three data sets used in this analysis (Table A.2).

Crop production is assessed, first, through the quantity of maize harvested per plot by the farm-household during the reference agricultural season. We choose maize since it is the most important staple crop in Malawi and Tanzania, though our findings on maize do not necessarily translate to other crops (see results in Deininger et al., 2012). In addition, we analyze the amount of maize harvest stored and the amount sold at the time of the interview.

Harvest quantities are recorded separately for each crop and plot, and all surveys allowed for reporting in local non-standard units. To construct a standardized maize harvest quantity, non-standard units are converted into kilograms using correspondence tables specific to each survey. Stored and sold harvest is reported at the household level. The crop production variables are summarized for each dataset separately in Table 2. The median maize harvest quantity per plot is 200 kg in Tanzania NPS 2012/13 (mean: 303.7 kg), 213.4 in Malawi IHS4 (mean: 320.5 kg), and 180 kg in Malawi IHPS 2016/17 (mean: 286.2 kg). While harvest per plot is similar, maize yield (harvest per area) is higher in Malawi IHS4 and IHPS 2016/17 than in Tanzania NPS 2012/13, where plots are larger on average (Figure 2, row 1).

The three main surveys used in this analysis rely on farmer self-reporting of the quantity of maize harvested. This means that there is no approximately objective measure of maize harvest to benchmark the results. We therefore make use of ancillary data from the second round of the Methodological Experiment on Measuring Maize Productivity, Soil Fertility and Variety (MAPS). These data contain full crop-cuts (considered benchmark estimates) for maize of 211 entire plots. The data set has two main limitations for our analysis: the data collection was so swift that the recall length is short with little variation between 0 and 4 months, and, at 211 plots, the sample size is small. The full crop cuts are used to quantify the error in farmer-reported maize harvest and relate it to the length of the recall period. On those plots, the mean farmer-reported harvest is 159 kg (median: 70 kg), while the mean crop cut estimate is 113.5 kg (median: 53.45 kg; Table 5).

For farm labor inputs, the headline outcome is total person-days of labor used per plot during the reference agricultural season. The variable is constructed by summing person-days worked by all household members and hired workers in all activities related to crop production, which are recorded separately for each plot. Family labor makes up the lion's share of total labor input in all three data sets, accounting for between 91 percent in Tanzania NPS 2012/13, 92 percent in Malawi IHPS 2016/17, and 96 percent in IHS4 (cross section). Total labor input is somewhat higher in Tanzania at 70 person-days than in Malawi with 64 and 62 person-days per plot in the cross section and the panel, respectively. However, given larger plot size in Tanzania, labor intensity (labor per area) is lower there than in the two Malawi samples (Figure 2, row 2). Twenty-six percent of plots in Tanzania NPS 2012/13 use hired labor, compared to 13 percent in Malawi IHS4 and 29 percent in IHPS 2016/17 (Table 2). Once again, the data exhibit an inverse relationship between labor intensity and plot size (Figure 2, row 2).

For an analysis of recall length, the total person-days variable has some drawbacks. As it is constructed by summing various work activities (planting, fertilizing, harvest, etc.), which all take place at different times during the agricultural season, there is no one single recall period for total labor person-days. To account for this, we assess whether the effect of recall length differs by type of activity (planting, fertilizing, harvest,

etc.) using appropriately defined recall length variables for each activity (Table 1, Table 3). Of the activities, most person-days are used in planting and maintenance activities, with fewer person-days used in harvest (Table 2). We also assess whether the recall effect differs by type of worker (household, hired). Labor inputs are summed to the plot-level, rather than assessed separately for each worker on each plot, because one respondent reports all labor inputs for a given plot, so that this measure likely best captures the recall effect.

Finally, input use is assessed through incidence of organic fertilizer application per plot and incidence and quantity of inorganic fertilizer application per plot. This analysis is possible because the three surveys record fertilizer use at the plot-level. Fertilizer use is considerably more common in Malawi than in Tanzania, for which Malawi's inorganic fertilizer subsidy program is likely responsible. Organic fertilizer is applied on 19 and 20 percent of plots in Malawi IHS4 and panel, respectively, and on 10 percent of plots in Tanzania NPS 2012/13. The difference is larger for inorganic fertilizer, which 53 and 61 percent of plots receive in Malawi IHS4 and IHPS 2016/17, respectively, compared to 9 percent in Tanzania. Among those plots to which inorganic fertilizer is applied, the applied quantities are similar across all three data sets at 66 kg per plot in Tanzania NPS 2012/13, 62 kg per plot in Malawi IHS4 cross-section, and 58 kg per plot in IHPS 2016/17 (Table 2). Application intensity (inorganic fertilizer applied per hectare) is again lower in Tanzania, owing to larger plot sizes, while fertilizer intensity is generally decreasing in plot size across all three data sets (Figure 2, row 3).

We apply the double Median Absolute Deviation method for outlier detection and correction to all constructed continuous variables in this analysis, that is, harvest quantity, storage and sales quantity, total, family, and hired labor person-days, organic and inorganic fertilizer quantity applied (Leys et al., 2013).

Determining recall length

The length of the recall period cannot be determined exactly in the survey data for two reasons. On the hand, we do not observe directly when farmers engage in the activities of interest (planting, fertilizing, working, harvesting, etc.). Instead, some information about the timing of the activities can be retrieved from farmer-reported dates and the cropping calendar. On the other hand, farmers engage in these activities over extended time periods, so that strictly no single recall length exists. We therefore proxy the length of the recall period for each outcome of interest.

During the interview, farmers are prompted to list all plots used for crop cultivation since the beginning of the agricultural season, so that this proxy for recall length seems an appropriate choice. The beginning of the planting season is determined by each country's cropping calendar and is identical for all households in each data set.

First, for harvest (and storage and sales) quantity, we use as recall length the distance in months between the interview date and the end of the harvest period. In all three surveys, farmers are asked to report harvest dates and this information is used to construct this second recall length variable. Harvest dates are likely recorded with error, being subject to memory decay over time in the same way as the outcome variables. To address this concern, we use the mode of harvest end dates at the cluster (enumeration area) level, rather than each household's individual maize harvest date, provided that there are at least 10 observations per cluster. Where there are fewer than 10 observations per cluster, the variable is instead based on the next

higher level of geographical aggregation (ward, district).[4] The interview date varies at the household level, so that the recall length variable for harvest input for plot *i* of household *j* is

$$T_i = Recall\ length_i^{Harvest} = \left(Date_j^{Interview} - mode_{cluster}(Date_i^{Harvest\ end})\right) \quad (1)$$

Table A.1 in the appendix presents an overview of the differences between the raw and corrected recall length variables for the three data sets. To assess whether this correction of the recall variable drives our results, we estimate one specification including the measure of difference between the raw and corrected recall variable.

Second, in the analysis of land inputs, the recall length to the beginning of the planting season is an appropriate choice. The beginning of the planting season likely varies regionally. In Malawi IHS4 and IHPS 2016/17, farmers report the month in which they began planting each of their plots. This information is used to capture the variation in season start. The recall variable is constructed following the same procedure[5] as for recall to harvest end, such that

$$T_i = Recall\ length_i^{Planting} = \left(Date_j^{Interview} - mode_{cluster}(Date_i^{Planting\ plot})\right) \quad (2)$$

In Malawi IHPS, the land information (and most input information) is collected during the first visit, and so the date of the first visit is used. The Tanzania NPS 2012/13 does not record individual planting dates. Instead, we rely on regionally disaggregated maize sowing time information from Arce and Caballero (2015) and the FAO Crop Calendar. The recall length to planting start variable for household *j* in region *k* is therefore:

$$T_j = Recall\ length_j^{PlantingNPS} = \left(Date_j^{Interview} - Date_k^{Sowing}\right) \quad (3)$$

Third, proxying the length of the recall period for labor inputs is challenging. Labor in crop production is spread out across the entire cropping cycle and comprises a set of disparate activities, from plot preparation, to planting, to fertilizing and weeding, to harvest and post-harvest work. The activities again differ in duration and salience and are carried out in part by household members and in part by hired workers – which is why we assess not only total labor input but analyze each activity in detail. The additional challenge in the task of constructing the recall length variable is that the timing of these activities is not always explicitly recorded. In Malawi IHS4 and IHPS 2016/17, farmers report the date of planting for each plot, how much later fertilizer was first applied, and the harvest period. In contrast, in Tanzania NPS 2012/13, only the timing of the harvest is recorded. Recall length variables therefore differ by labor activity and survey: for total labor input, we use recall length to planting from equations (2) and (3), respectively. Since total labor is the sum of the various activities during the reference season, this recall variable is a rough approximation; for harvest and post-harvest labor, recall length to end of harvest, as per equation (1), is used; for planting and plot preparation, and weeding and fertilizing, we again use recall length to planting (equations (2) and (3)).

[4] This is the case in 11 percent of observations in Tanzania NPS 2012/13, 10 percent in Malawi IHS4, and 3.5 percent in Malawi IHPS.

[5] Mode of beginning of planting on plot at the cluster, as long as there are at least 10 observations per cluster. If fewer than 10 observations per cluster, the variable is based on the next higher level of geographical aggregation. This is the case in 5.2 percent of observations in Malawi IHS4 and 1.6 percent in IHPS 2016/17.

Finally, constructing a sensible proxy for the recall length of fertilizer input has some of the same challenges as for labor. However, the application of fertilizer input covers shorter periods than the more continuous task of working on the plot. In Malawi IHS4 and IHPS 2016/17, we use the recall length to planting (equation (2)) when analyzing input use in the full sample (that is, both plots on which fertilizer was used and plots on which it was not used). When assessing fertilizer quantity reporting among plots receiving fertilizer, we can instead use the recall length to the first application of fertilizer on the plot. Fertilizer application is reported in weeks after planting the plot but transformed to months for the purpose of the recall variable. Since fertilizer application happens after planting, the recall length for fertilizer is shorter than for planting:

$$T_i = Recall\ length_i^{fertilizer} = Recall\ length_i^{Planting} - \ months\ to\ application_i \qquad (4)$$

In Tanzania NPS 2012/13, we once again use recall length to planting (equation (3)), lacking information on fertilizer application timing.

All recall variables are coded in increments of one month as this is the unit in which harvest, planting, and other dates are recorded in the data. The various recall variables vary in length between and within surveys. Recall length to end of harvest ranges from 2 to 18 months in Tanzania NPS 2012/13 (mean: 9.2), 0 to 13 months in Malawi IHS4 (mean: 8.9),[6] and 3 to 12 months in Malawi IHPS 2016/17 (mean: 5.7; Table 3; Figure 1).

Recall length to planting varies between 7 and 25 months in Tanzania NPS 2012/13 (mean: 14.7) and between 4 and 18 months in Malawi IHS4 2016/17 (mean: 13.4), though there are very few observations with a recall length shorter than 9 (Figure 1). In Malawi IHPS 2016/17, the range of recall length varies between 4 and 9 (mean: 5.7; Table 3). Most of the outcomes of interest which we analyze using recall length to beginning of season were collected during the first visit. The two-visit structure thus cuts the average recall length to beginning of planting to 5.7 months, relative to 13.4 (Malawi IHS4 2016/17 cross section) and 14.7 months (Tanzania NPS 2012/13). Recall length to first fertilizer application is minimally shorter than recall length to planting, suggesting most farmers start fertilizing quite soon after they finish planting (Figure 1; Table 3).

Since this analysis exploits variation in the timing of Tanzania NPS 2012/13 and Malawi IHS4 fieldwork, rather than explicitly randomizing by recall length to agricultural activities of interest, the data are not evenly spread across time and space (Figure 1). This is the case because, while the interviews are spread across 12 months, not all interviewed households are agricultural households and agricultural activities vary at the household or cluster level. Households also differ on observable characteristics over time, many of which are likely correlated with the outcomes of interest. To address this issue, we introduce a set of control variables. These fall into the following categories: regional and enumerator dummies, respondent characteristics (e.g. gender, age, education, whether respondent is plot manager), household characteristics

[6] There are three cases of a recall length to harvest end of zero months, which implies that in these cases farmers have finished their harvest within the month of being interviewed. This scenario can realistically occur in Malawi IHS4 because farmers are asked to identify the last rainy season completed, and the agricultural questionnaire is administered in reference to that season. The agricultural season spans a period of November to April and the three cases at hand are from households interviewed in April of 2016, so that it is very plausible that these farmers finished harvesting just before the interviews took place.

(e.g. consumption aggregate or wealth index, shocks, household head gender, education, age), plot characteristics (e.g. crop loss, land area, irrigation).

There are some marked differences in household characteristics and agricultural outcomes between the Malawi IHS4 and IHPS 2016/17 samples. This may be the result of several factors. First, the IHPS is a panel whose sample households have been interviewed several times before since the first round of the IHPS in 2010. Respondents are likely familiar with the survey process and the survey instrument and may have become better at responding to questions. At the same time, those respondents may act strategically and take shortcuts in answering in order to reduce the time of the interview, as they are familiar with the questionnaire structure. It is also likely that the sample of agricultural households in the panel survey that we select for this analysis have been active in farming for an extended period of time, having been followed since 2010. Second, data on agricultural outcomes were collected in two rather than in one visit. This means, on the one hand, that the recall length for many agricultural outcomes of interest is cut in half. It may further lead to differences in recorded agricultural outcomes beyond what the recall length can explain. Unfortunately, we cannot distinguish between the recall effect, the effect of forming part of a long-running panel, and the effect of a post-planting visit beyond recall length with the IHS4 and IHPS 2016/17 data sets, though these issues would make for interesting future research.

Empirical strategy

We use the following specification to assess the effect of the recall length on outcomes of interest:

$$Y_i = \alpha + \beta_1 T_i + \delta \mathbb{X}_i + \gamma + \mu + \varepsilon \quad (5)$$

with Y_i the outcome variable of household i (or plot i depending on the level of analysis), T_i explanatory variable of interest (planting or harvest recall length in months, depending on the outcome of interest), $\mathbb{X}_i$ a vector of control variables, γ an indicator for region, μ an enumerator fixed effect, and ε the error clustered at the EA level. Provided $\mathbb{X}_i$ captures all relevant confounding factors, the length of the recall period, T_i, should not be correlated with the outcome variables if farmers accurately recall them. Therefore, if β_1, the coefficient on T_i, is significantly different from zero, this is evidence of systematic measurement error.

We estimate equation 3 with OLS when assessing the uncensored continuous outcome variables of interest, maize production, error in self-reporting, labor input in person-days. The number of plots listed and cultivated is a count variable, so we asses this outcome with both OLS and Poisson. For the binary outcome variables use of hired labor, use of organic and of inorganic fertilizer, we make use of Probit. Finally, the quantity of inorganic fertilizer applied per plot is truncated at zero. On the considerable share of plots to which no fertilizer is applied, the fertilizer quantity variable takes the value zero, while it is continuous and positive for the share of plots to which fertilizer was applied. We deal with this case by running both a Tobit model on the truncated variable as well two separate regressions, a Probit model on the binary choice whether inorganic fertilizer is used at all and OLS on the continuous quantity of application conditional on use.

To verify that the linear baseline specification in equation 3 accurately captures the relationship between recall length and the outcomes of interest, we run a set of regressions with added polynomial terms

(quadratic, cubic) of T_i (recall length). We compare the linear and non-linear models using a variant of the Bayesian Information Criterion (BIC') as a goodness-of-fit measure to guide which model is best suited.

Recall error may be exacerbated by certain household, individual, or farm characteristics. We explore interaction effects between the recall length and plot size, respondent characteristics (gender and education), and whether the respondent is also the plot manager. The latter is relevant because, to an extent, it is in the survey designers' purview to enforce that the respondent be the plot manager.

4. Results

We find that the recall length has a significant impact on reported outcomes in all areas of interest of this analysis. Farmers report significantly higher quantities of plot-level variables, that is maize harvest, labor and fertilizer input, as the recall length increases. In contrast, farmers report fewer cultivated plots of land in longer recall periods.

Land

Households list significantly fewer plots (and parcels) as the recall period increases. Results from the baseline linear OLS specification with a full set of controls, shown in Figure 3 and Figure 4 (and Table A.3 in the Appendix), are consistent across the three data sets. The results from a Poisson regression are in line with those from OLS. The magnitude of the effect varies by survey, however. It is small for Tanzania NPS 2012/13 at 0.03 fewer plots per month and Malawi IHS4 (cross section) at 0.05 plot per month, but larger in Malawi IHPS 2016/17 (panel) at 0.1 plot per month. This may be the result of several factors, which we cannot directly assess.

The results from a quadratic specification confirm the overall negative relationship between recall length and plots listed (Table A.4). The BIC' goodness-of-fit comparison suggests that the linear model is preferred in Tanzania NPS 2012/13 and Malawi IHS4 data and that the quadratic model provides a better fit in Malawi IHPS (Figure 4).

We have no way to observe the 'true' number of plots household-farms in our data operate, so we cannot ultimately determine the direction of the reporting bias. However, it seems plausible that farmers forget some of their less important plots as recall decays over time, as was observed previously for example in Gaddis et al. (2019) and Arthi et al. (2018).[7]

Turning now to farmer self-reported land area, we assess the effect of recall length on the magnitude of error in self-reported plot area, comparing it to the objective GPS-based measure of plot. We split the sample in plots whose area was over-stated and plots whose areas was under-stated (positive and negative

[7] The analysis in this paper relies on a different setup than Arthi et al. (2018) and Gaddis et al. (2019), who use a methodological experiment in which farmers are randomly assigned to report on labor inputs either through end-of-season recall modules or in weekly visits during the agricultural season. In contrast, we rely on variation in recall length stemming from variation in interview timing. The findings in Arthi et al. (2018) and Gaddis et al. (2019) are therefore a combination of a recall effect (end-of-season modules imply longer periods of time to recall) and a data collection effect (weekly visits in which respondents are prompted to report in detail on their plots). Our findings isolate and underscore the former.

error). We find no consistent evidence of the self-reporting error being associated with recall length: only in the Malawi IHS4 data set, the overstatement of plot area increases in recall length at a rate significantly different from zero. All other specifications show no significant correlation (Table A.5). Overall, this suggests that respondents' difficulties in accurately estimating land area are not associated with memory decay over time. It is plausible that farmers at the time of the interview still operate many of the same plots as at the beginning of the season and therefore have a belief or estimate of plot area based on relatively current information.[8]

Crop production

The maize harvest quantity as reported by farmers increases with the length of the recall period. The results are consistent across all three data sets, with point estimates ranging from 4.9 kg (2 percent) in Tanzania NPS 2012/13, to 11.97 kg (4 percent) in Malawi IHS4, to 15.85 kg (8 percent) for each additional month of recall length in the Malawi IHS4 panel (Figure 5, Table A.6, Table A.7). Based on BIC', a quadratic specification yields a better fit of the relationship in Malawi IHS4, suggesting that here the recall length effect is strongest in the first 3 to 8 months, leveling off after that (Table A.8, Figure 6).

The results support the hypothesis that there is non-random measurement error in the farmer-reported quantity of maize harvested related to the length of the recall period. However, this result is valid only if no relevant confounding factor is omitted in the regression. There is one potential confounding factor that merits discussion: the weight of maize produced can be reported in different states of the crop. For example, at the time of harvest, maize is still on the cob. Before being sold, stored, or consumed, maize grains may be removed from the maize cob and dried. In this process, the weight of the same harvest may vary depending on the state in which its weight is reported. Moreover, this may be related to the length of the recall period if, for example, farmers more often report the harvest weight of maize on the cob immediately after the harvest and the weight of dried grains more often as more time passes between the harvest and the interview. This, in turn, could mean that the observed recall effect is the result of changes in reported harvest state over time. We assess this possibility using information on the harvest state – shelled (grain) and unshelled (on the cob) – which is available for Malawi IHS4 and IHPS 2016/17 data in three steps. First, we assess whether the state in which harvest weight is reported changes with the length of the recall period. We find that the share of farmers reporting the weight of maize on the cob decreases with recall length at 0.8 percentage point per month in Malawi IHPS 2016/17 and 0.4 percentage point in Malawi IHS4. Next, we include the state in which harvest was reported as a control variable in the main regression. Finally, we convert the quantity of unshelled maize (on the cob) to grain-equivalent shelled weight and then repeat the analysis with this new variable. The unshelled to shelled conversion factor was obtained from the MAPS experiment in which maize was weighed twice, before and after shelling. The main results are robust to these two specifications. Further, we assess whether the correction of the harvest date and hence of the recall length variable (section 3) drives our results. We estimate the main specification including the difference (in months) between the raw and the corrected harvest date variable (results not shown). The inclusion of this correction term is not statistically significant in any of the data sets nor does it meaningfully change the coefficient on the recall length variable.

[8] This does not mean that the estimate is correct (Carletto et al., 2011; De Groote and Traoré, 2005; Dillon et al., 2019; Goldstein and Udry, 1999; Keita and Carfagna, 2009; Kilic et al., 2017; Schøning, 2005).

As an additional robustness check, we assess the relationship between recall length and the quantity of maize sold and maize in storage (results not shown). The rationale is as follows: Absent measurement error, the quantity of maize harvested should not be correlated with the timing of the interview. In contrast, as the time between harvest and interview increases, the expectation is that households deplete their storage of harvested maize, while selling more of it. Thus, there should be a negative correlation between the time passed since harvest and maize in storage and a positive correlation with maize sold. Finding that this is not the case may point to some other underlying dynamic and cast doubt on the results. We thus test this hypothesis in the data. The data indeed reflect a negative relationship between time elapsed since harvest – that is, recall length – and the quantity of maize in storage. We also find a positive relationship between time elapsed since harvest and the quantity of maize households report they have sold.

In light of the finding that reported harvest quantity increases with the recall period, the question of interest is whether this implies that harvest quantity is being increasingly over-reported the longer the recall period becomes. The logic of recall decay would suggest this is the case: farmers likely remember the amount harvested if the harvest has only just finished. But absent written records, they need to rely on inference or reconstruction more often if the harvest is many months in the past. If this logic holds, interviews conducted just after the harvest, with a short recall length, produce the most reliable estimates. Whether or not this logic holds is of course predicated on respondents having accurate knowledge of how much they harvested in the first place. Existing evidence shows that other factors, such as land area or respondent characteristics, are correlated with harvest mis-reporting, so even in cases of short recall periods self-reported harvest quantities may well be inaccurate.

Resolving this question systematically would require comparing farmer-reported harvest quantities to more objective benchmark estimates, for example based on crop cutting. Crop cutting is not available in the three main data sets used for this analysis. As an ancillary analysis, we use the MAPSII data set to benchmark the farmer self-reported harvest against the more objective full-plot crop cuts. We replicate the main specification for the effect of recall length on farmer-reported harvest in the sample of plots for which a full-plot crop cut is available. Then, we quantify the reporting error (farmer-reported harvest quantity minus full plot crop cut quantity) and assess the effect of the recall length on the reporting error. We find a significant effect of the recall length on self-reported harvest as well as on reporting error (Figure 7). This suggests that farmers are indeed increasingly over-reporting harvest with longer recall periods, at least in this sample.

Labor

Reported farm labor per plot increases significantly with recall length. The effect size is similar in Tanzania NPS 2012/13 and Malawi IHS4 at 1.6 (2.3 percent) and 1.2 person-days (2.8 percent) per plot per month of recall length, respectively (Figure 8, Figure 9, Table A.9, Table A.10, Table A.11).[9] A quadratic rather than linear specification for recall length suggests no mis-specification by the linear model, and the BIC' analysis shows the linear model provides a better fit in both data sets (results not shown).

[9] The comparison with the Malawi IHS4 panel is omitted because in that survey labor inputs are collected partially in the first visit (labor used for land preparation and planting, ridging and fertilizing) and partially in the second visit (harvest and post-harvest labor). Thus, there is no one unique recall period against which to evaluate total labor input, rendering the comparison meaningless.

The concern with total labor days is that it is made up from different work activities with different timings. We therefore assess the recall effect separately for each type of activity: Land preparation and planting; weeding, ridging, and fertilizing; and harvest and post-harvest activities. Generally, the positive association between recall length and labor input holds across activities in all three data sets, though not all point estimates are significant (Figure 10). Recall length has the clearly largest and most significant level effect on reporting of labor dedicated to 'weeding, ridging, and fertilizing' (Figure 10,
Table A.12, Table A.13, Table A.14). This category is arguably the least salient and captures labor applied on the plot between planting and harvest, a period of several months, which make these activities particularly hard to recall correctly.

We also test family and hired labor input per plot separately. The total labor input results are driven by family labor. On the one hand, the number of family labor days make up a large share of total labor days. On the other hand, recall length has a significant effect on family labor days in all three data sets, but not on hired labor (Figure 8, Table A.9, Table A.10). Further, there is no significant effect of recall length on the number of family members reported to have worked per plot and per household (results not shown).

The findings are consistent with plot-level over-reporting of labor inputs in long recall modules discussed in Arthi et al. (2018) and Gaddis et al. (2019).[10] While we cannot benchmark our results against a more objective measure of farm labor, it seems plausible that respondents here too are over-reporting as the recall period increases.

Inputs

There are two main results regarding the effect of recall length on agricultural inputs (we analyze organic and inorganic fertilizer use in this study): first, respondents tend to report higher fertilizer quantities as the recall length increases. Second, respondents are slightly less likely to report incidence of fertilizer use, that is whether or not any fertilizer was applied to a given plot.

We use a linear OLS specification with a full set of controls to estimate the effect of recall length on quantity of inorganic fertilizer applied per plot, conditional on fertilizer use. The effect size is large in Malawi IHPS 2016/17 at 4.3 kg per plot per month (8.8 percent), and smaller in Malawi IHS4 at 0.9 kg (1.9 percent). In Tanzania NPS 2012/13, the coefficient has the same sign but is not statistically significant (Figure 12, Table A.15, Table A.16, Table A.17). Low fertilizer application rates lead to a small sample of households using inorganic fertilizer on their land at 675 of 7,322 plots, which may explain the recall effect being insignificant. The linear model is the preferred specification based on BIC' analysis.

Based on a Probit model, we find a small negative recall effect on the binary variable whether any inorganic fertilizer was applied in the case of Malawi IHS4 and IHPS 2016, and whether any organic fertilizer was

[10] Arthi et al. (2018) and Gaddis et al. (2019) find two countervailing types of reporting bias. First, end-of-season recall modules lead to an over-estimation of reported labor inputs per plot ('recall bias'). Second, end-of-season recall modules lead to an under-counting of plots listed and household members reported to have worked on them ('listing bias'). Our results echo those findings: we find longer recall periods are associated with fewer plots listed and higher labor inputs reported. A key difference is that we find no evidence of farmers reporting fewer members having worked on each plot or on the entire farm. Moreover, Arthi et al. (2018) and Gaddis et al. (2019) find a significant educational gradient in recall bias, while we find evidence of education interacting with the recall effect.

applied in the case of Tanzania NPS 2012/13 (Figure 11); that is, respondents are slightly less likely to report having used fertilizer as the recall length increases.

Recall error and other characteristics

While our focus is on the effect of the recall length on data quality, we always control for respondent characteristics, including via interactions with the length of the recall. Most of the respondent characteristics (age, education, gender, whether respondents are also plot managers) have no significant impact on the direction or magnitude of the effect of the recall length.[11] However, some are significantly correlated with our outcomes of interest. Female respondents tend to understate harvest and input quantities relative to male respondents. The same is true of respondents who are also plot managers (results not shown).[12] In contrast, respondent education is neither strongly correlated with levels of reported outcomes nor does it impact the recall length effect.

Unlike respondent characteristics, plot size consistently matters for the effect of recall length. In line with expectations, the recall effect on maize harvest quantity, labor and fertilizer inputs is attenuated, and sometimes even reversed, on larger plots in Tanzania NPS 2012/13 (Figure 14). Similarly, there is a significant interaction effect of recall length and plot size for harvest and fertilizer input in Malawi IHS4 and IHPS (Figure 15).

Recall error and productivity measurement

This section explores the implications of the recall effect for agricultural productivity measurement, in which context the input and output variables discussed in this analysis are frequently studied. Interest is in three of the most common measures of agricultural productivity, yield (output per unit of land), output per day of work, and output per worker (

Table 6).

The results discussed so far suggest that the impact of the recall length on the three productivity measures will vary. For yield (output per unit of land), the expected association is positive, since there is a positive effect of recall length on output (maize harvest), the numerator, while the plot area, measured with GPS, is not subject to recall error. For output per day of work, the results suggest an ambiguous effect. Both output, the numerator, and labor-days, the denominator, are positively associated with the recall length. The direction of the composite effect depends on which of the two dominates, and it is possible that there is no effect at all. Finally, for output per worker, the results suggest a positive recall effect. The recall effect on output, the numerator, is positive, and there is no effect on workers per plot, the denominator (

Table 6).

[11] Plots listed in the Malawi IHS4 cross section are the one case in which there is a significant interaction effect between respondent characteristics and recall length. Female respondents list fewer plots than male respondents (-0.52*** plots), but the effect of recall length is smaller among female respondents (-0.05** plots/month as opposed to -0.076*** plots/month).

[12] Plot managers are presumably best informed about harvest quantities and inputs used, suggesting that self-reporting from female respondents may be more reliable than that from male respondents.

Akin to the analysis discussed in the previous paragraphs, we use as a proxy for recall length the time in months between the interview day and the first day of fieldwork. This proxy has two advantages in the case at hand. First, productivity measures are composite indicators whose components – output, land area, labor inputs – each have different timing patterns, so that there is no obvious single recall length measure. Second, time between the interview and the beginning of fieldwork has a simple practical interpretation and can be determined by survey designers. We test the effect of the recall length both at the plot level and aggregated at the farm level in the Malawi IHS4 and Tanzania NPS 2012/13 data using an OLS specification as in equation 5.[13]

The results are in line with the prediction. Maize yield at the plot-level increases by 3.1 percent with each additional month of fieldwork in Tanzania NPS 2012/13 and by 5.1 percent in Malawi IHS4 (at the household level: 2.2 percent and 5.2 percent, respectively; Figure 16, Table A.19, Table A.20). Output (maize harvest) per day of labor input is also generally positively associated with fieldwork length, with the effect ranging from null to 3.1 percent per month (Tanzania NPS 2012/13: null effect at the plot-level, 1.9 percent at the household-level; Malawi IHS4: 2.7 percent at the plot-level, 3.1 percent at the household-level; Table A.19; Table A.20; Figure 16). Finally, output per worker is also positively associated with fieldwork length, with the effect size ranging between 2.4 and 5.1 percent per month (Tanzania NPS 2012/13: 2.4 at the plot-level, 2.5 percent at the household-level; Malawi IHS4: 5.1 percent at the plot-level, 3.0 percent at the household-level; Table A.19; Table A.20; Figure 16).

5. Conclusions, policy implications, and recommendations

In this paper, we set out to evaluate the impact of the recall length and related design choices on the quality of key agricultural input and output data – land, labor, fertilizer, and production. Our results demonstrate, consistently across the three data sets used in the analysis, that the survey-based estimates of these variables depend on the length of the recall period, which indicates the presence of nonrandom measurement error of economically significant size. With longer recall periods, farmers report higher quantities of harvest, labor and fertilizer inputs, all three of which are recorded at the plot-level in our data sets. At the same time, farmers list fewer plots as the recall period increases. We have argued that it is plausible that farmers over-estimate plot-level outcomes – harvest, labor and fertilizer inputs – while it is also plausible that they forget some of their more marginal plots, as their memory decays due to longer recall periods. We also show that the recall length has a meaningful impact on agricultural productivity measurement.

The reliability of agricultural data matters for policy effectiveness. Policy makers and the international community recognize agricultural input, output, and productivity outcomes as integral to agricultural growth and food security, and a large body of empirical evidence supports this view. These outcomes have therefore been made priorities within several development frameworks. Under the Sustainable Development Goals (SDGs), target 2.3 is to double the agricultural productivity and incomes of small-scale food producers by 2030. The target is monitored through indicator 2.3.1, volume of production per labor unit – that is, one of the productivity measures we found were subject to measurement error related to the

[13] The Malawi IHPS 2016/17 data are omitted because the two-visit structure of that survey makes time between interview and beginning of fieldwork difficult to interpret.

recall length. Similarly, African Union's (AU) Comprehensive Africa Agriculture Development Programme (CAADP) was conceived in 2003 to achieve economic growth, eliminate hunger, and reduce poverty through concerted investments in African agriculture. CAADP adopted the Malabo Declaration in 2014 in which AU member states resolved, among other things, to double agricultural productivity and pledged to allocate at least 10 percent of public expenditure towards agriculture. The CAADP results framework monitors progress towards its goals through a set of priority indicators and guides member countries' strategic investment decisions. CAADP target 2.1 – Increased agriculture production and productivity[14] is comprised of five priority indicators, focusing on growth in agricultural production volume and value added as well as land and labor productivity. For instance, priority indicator 2.1.5 measures yield for the five AU priority commodities, of which maize is one; indicator 2.1.2 tracks the level of agricultural production.

The results of our analysis show that recall error can lead to unreliable measurement of all these indicators at an economically significant level. This is especially true when it comes to monitoring their evolution over time. For illustration, maize yields in Africa (reflected in CAADP 2.1.5) have grown at an average annual rate of 0.8 percent between 2000 and 2017, according to FAOSTAT. Maize production (reflected partially in CAADP 2.1.2) has grown by an average of 3.9 percent per year during the same period. In comparison, our analysis indicates that an additional month in recall length can lead to a change of between 2 and 5 percent in reported maize yields and 2 and 7 percent in the reported maize harvest quantity. Further, an annual growth rate of around 7 percent would be required to achieve SDG target 3.2 of doubling labor productivity within about 10 years. We find an additional month of recall length is associated with a 2 to 3 percent increase in reported output per unit of labor. What this shows is that the recall effect is likely to introduce an amount of variability that makes tracking progress towards achieving SDG targets and CAADP goals challenging and sensitive to the survey design choices that affect the length of recall. Policy and investment priorities based on these indicators risk being ineffective or misguided. Improving data quality is critical to minimize this risk.

The results lend support to ongoing efforts to mainstream more objective 'gold standard' measures of key variables into the design and implementation of surveys. The use of GPS devices to measure the area of agricultural plots is already considered best practice and incorporated into many surveys concerned with agricultural productivity and land use. In the field of production measurement, the practice of crop cutting is used in many national production estimations and other surveys, and our findings offer strong reasons for doing so systematically. Finally, when it comes to accurately measuring labor and fertilizer input, as well as counting all plots used in agricultural production, the use of diaries appears worth exploring.

One strategy to improve agricultural data reliability by shortening recall periods and recall length variation in household and farm surveys is to field two visits, one in the post-planting and one in the post-harvest period, rather than just one. While this strategy has cost implications, as field teams need to visit farms twice rather just once, it has the additional advantage of providing countries with more frequent and timely data on planting in the ongoing agricultural season. Another avenue to explore is the use of higher-

[14] CAADP Indicator 2.1. *Increased agriculture production and productivity* contains the following sub-indicators: 2.1.1 *Agriculture value added*; 2.1.2 *Agriculture production index*; 2.1.3 *Agriculture value added per agricultural worker*; 2.1.4 *Agriculture value added per hectare of arable land*; 2.1.5 *Yields for the five AU priority commodities* (cassava, yams, maize, meat, and cow milk).

frequency phone surveys to collect data on input and even output variables, though the data quality implications of such a survey design ought to be rigorously tested before mainstreaming.

In sum, there is a need for survey practitioners and agencies charged with collecting agricultural sector data to carefully consider the data quality implications of survey design choices that implicitly affect the length of the recall period for the respondent, and to develop and promote the adoption of survey designs that can ensure data quality without excessive implications for the survey budgets and logistics.

Figures and Tables

Table 1. Outcome and main explanatory variables of interest

	Dependent variable (Y_i)		**Unit of measurement**	**Level of analysis**	**Main independent variable** (T_i)
Output/ Production	Maize quantity produced		kg	Plot-level	Recall length (months) to harvest end
	Harvest disposition	Maize quantity sold		Household/farm-level	
		Maize in storage			
Inputs – Land	Agricultural plots and parcels listed and cultivated by the household		number	Household/farm-level	Recall length (months) to planting of plot*
	Difference between self-reported and GPS-measured plot land area		hectares	Plot-level	
Inputs – Labor	Total labor inputs by type, household and external	Total labor	person-days	Plot-level	Recall length (months) to planting of plot*
		Household member labor			
		Use of hired labor	binary		
	Labor by activity	Plot preparation and planting labor	person-days	Plot-level	Recall length (months) to planting of plot*
		Person-days in plot maintenance			
		Person-days in harvesting and post-harvest activities.			Recall length (months) to harvest end
Inputs – other	Organic fertilizer use		binary	Plot-level	Recall length (months) to planting of plot*
	Inorganic fertilizer use				
	Inorganic fertilizer application quantity		kg		Recall length (months) to first fertilizer application*

* In the Tanzania NPS 2012/13 survey, these variables are not available at the plot-level. Regional sowing season timing information is used instead.

Table 2. Outcome variables

		Tanzania NPS 2012/13				Malawi IHS4 2016/17				Malawi IHPS4 2016/17			
	Outcome	**Mean**	**Median**	**SD**	**N**	**Mean**	**Median**	**SD**	**N**	**Mean**	**Median**	**SD**	**N**
Land	Parcels					1.426	1	0.667	6,799	1.780	2	0.999	1,803
	Plots	2.257	2	1.343	3,300	1.587	1	0.842	6,790	2.130	2	1.429	1,801
	Plots cultivated	1.874	2	1.253	3,300	1.532	1	0.818	6,790	2.014	2	1.338	1,801
	GPS measured area, ha	1.233	0.486	3.145	7,447	0.371	0.295	0.317	9,277	0.338	0.255	0.321	3,573
	Self-reported area, ha	1.164	0.405	3.743	7,447	0.362	0.405	0.261	10,379	0.360	0.304	0.328	3,805
	Error in self-reporting (SR-GPS, ha)	-0.052	0.0202	0.516	5,396	-0.003	0.0121	0.150	9,267	0.0164	0.0283	0.152	3,571
Harvest	Maize harvest quantity (kg)	303.7	200	349.0	3,079	320.5	213.4	336.5	7,110	286.2	180	341.1	2,535
	Sales of maize (kg)	91.11	0	336.2	3,300	39.74	0	139.6	5,828	35.13	0	125.7	1,674
	Storage of maize (kg)	38.44	0	112.3	2,067	30.03	0	278.4	5,810	105.2	0	220.2	1,675
Labor	Total labor (person-days)	70.13	45	83.39	7,447	63.76	56	39.94	10,403	61.95	51	45.06	3,812
	Family labor (person-days)	63.52	40	78.39	7,447	61.27	54	40.11	10,403	57.32	47	44.97	3,812
	Hired labor (Y/N)	0.263	0	0.440	7,447	0.135	0	0.341	10,403	0.294	0	0.455	3,812
	Hired labor (person-days)	24.81	13	34.62	1,742	13.11	9	13.95	1,400	3.805	0	9.014	3,812
	Land prep + planting (HH person-days)	24.34	14	33.49	7,447	27.60	24	21.40	10,403	28.14	22	26.33	3,812
	Weeding + fertilizing (HH person-days)	25.86	15	33.18	7,447	22.82	19	18.54	10,403	21.47	16	20.48	3,812
	Harvest + post-harvest (HH person-days)	19.93	8	32.61	7,447	8.148	6	8.379	10,403	7.710	5	9.561	3,812
	HH members working on plot	2.320	2	1.747	7,447	2.377	2	1.294	10,403	2.752	2	1.550	3,812
Inputs	Organic fertilizer use (Y/N)	0.0974	0	0.296	7,447	0.187	0	0.390	10,403	0.201	0	0.401	3,812
	Inorganic fertilizer use (Y/N)	0.0928	0	0.290	7,447	0.533	1	0.499	10,403	0.609	1	0.488	3,812
	Inorganic fertilizer (kg)	6.175	0	36.61	7,447	33.17	10	45.75	10,403	35.23	20	44.69	3,812
	Inorganic fertilizer use (kg|>0)	66.55	50	102.2	691	62.39	50	45.98	5,531	57.86	50	44.40	2,321
Productivity	Maize yield (kg/ha)	1,053	0	793.3	4,971	756.2	0	596.4	3,841	527.6	0	373.8	2,616
	Labor intensity (person-days/ha)	229.1	0	183.6	1,013	208.1	0	177.4	726.8	138.1	0	88.00	889.6
	Labor productivity (output/person-day)	6.102	0	3.980	43.16	4.408	0	2.881	32.49	5.139	0	3.333	30.25
	Output per worker	129.6	0	88.55	776.9	134.2	0	90	824.2	118.6	0	72	780
	Fertilizer intensity (kg/ha)	128.2	0	42.60	18,533	187.1	0	101.1	5,148	12.92	0	0	6,178
	Fertilizer intensity (kg/ha|>0)	241.1	1.602	164.7	18,533	292.2	1.602	205.9	5,148	140.8	0.414	79.71	6,178

Table 3. Recall length

Variable	Mean	Min	Median	Max	SD	N
			Tanzania NPS 2012/13			
Recall length, harvest	9.292	2	10	18	3.546	3,079
Recall length, planting	14.71	7	14	25	4.068	7,447
			Malawi IHS4 2016/17			
Recall length, harvest	8.843	0	9	13	2.475	7,110
Recall length, planting	13.41	4	14	18	2.336	10,360
Recall length, fertilizer	12.77	4	13	17	2.347	5,523
			Malawi IHPS 2016/17			
Recall length, harvest	5.676	3	5	12	1.649	2,549
Recall length, planting finish	5.662	4	6	9	1.087	3,592
Recall length, fertilizer	5.046	3	5	8	1.108	2,187

Table 4. Control variables for regression

Group of control variables Xi	Variables Malawi IHS4 and IHPS4	Variables Tanzania NPS 2012/13
Respondent characteristics	Respondent gender, age, education (years), literacy.	Respondent age, gender, education (years).
Household characteristics	Drought shock, flood shock, income loss shock; agricultural asset index, household wealth index; household head literacy, gender, age; dependency ratio, household size.	Household size, production shocks, household consumption expenditure, household head gender.
Plot characteristics	Plot owned, cash crop grown, erosion control terraces, swampland, fertilizer use, plot size, any hired labor, family labor input, pre-harvest losses, intercropped.	Plot area; Main crop, irrigation, respondent is plot owner, respondent, or user; use of fertilizer; any hired labor; family labor input.
Regional controls	District dummies.	Region dummies.
Enumerator controls	Enumerator dummies.	Enumerator dummies.

Table 5. Maize harvest, MAPSII, full crop-cut plots

Variable	Mean	Min	Median	Max	SD	N
Full crop cut (CC) maize harvest, kg	113.5	0	53.45	1,633	178.6	211
Self-reported (SR) maize harvest, kg	159.5	0	70	3,000	289.3	211
Error in self-reporting, SR-CC, kg	46.00	-259.9	6.306	2,943	219.5	211

Figure 1. Distribution of recall length variables

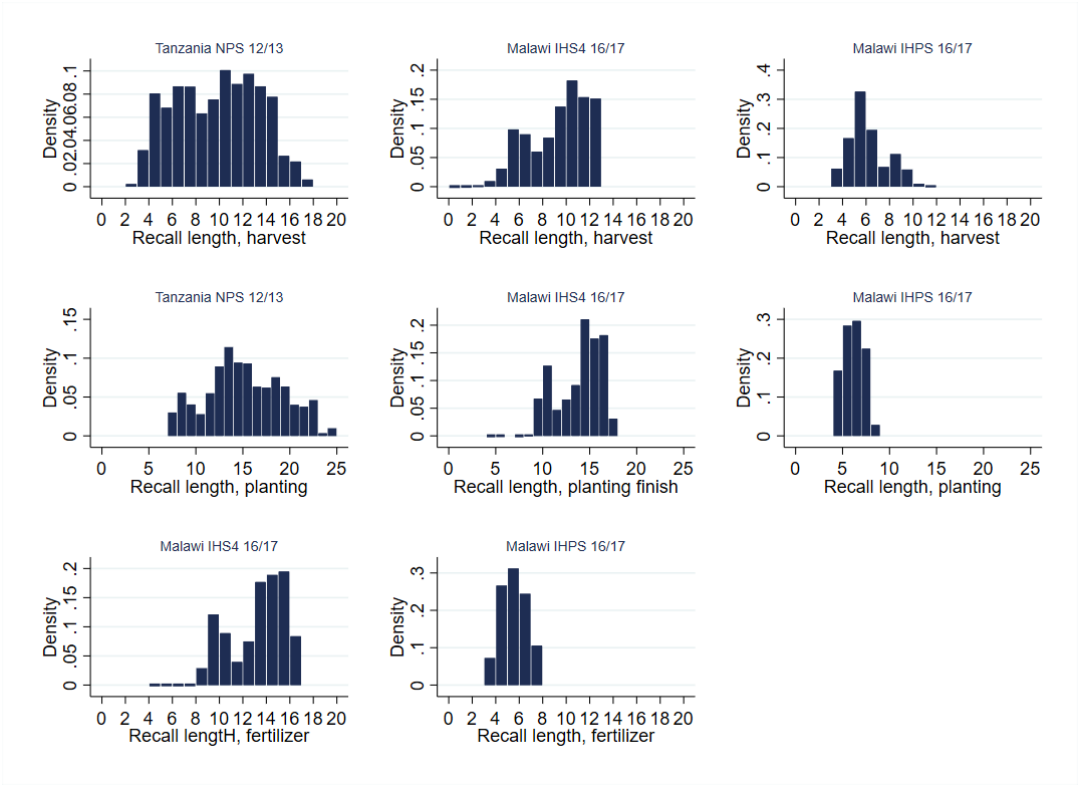

Figure 2. Local polynomial fit between input and output per area against area

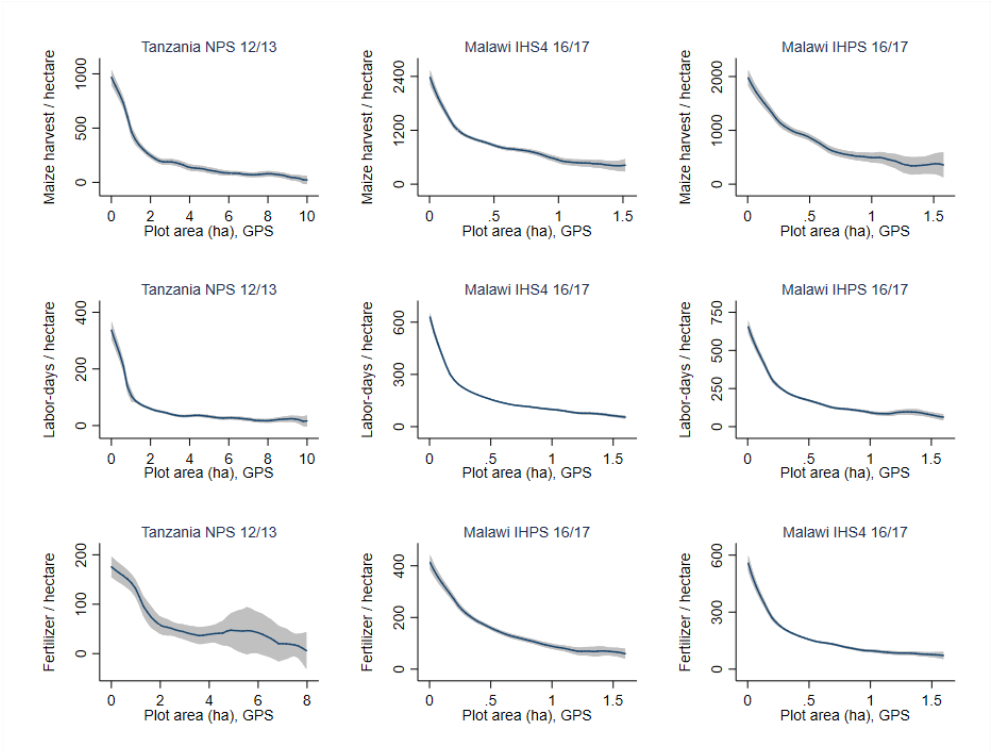

Figure 3. Effect of recall length on plots listed and cultivated, point estimates

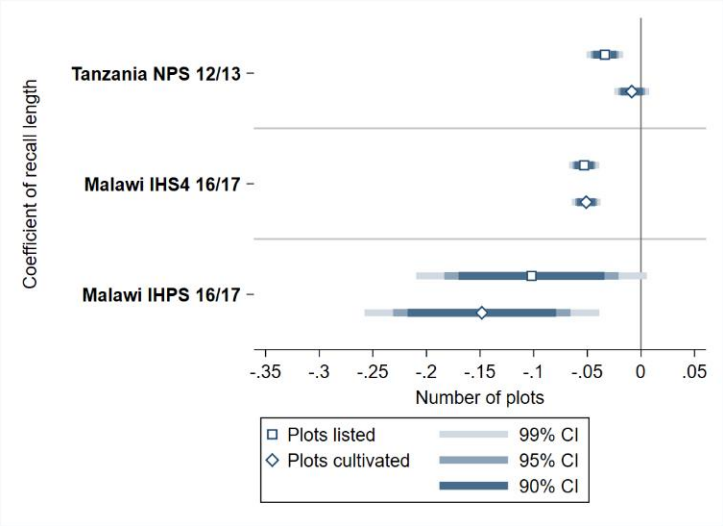

Figure 4. Plots listed against recall length

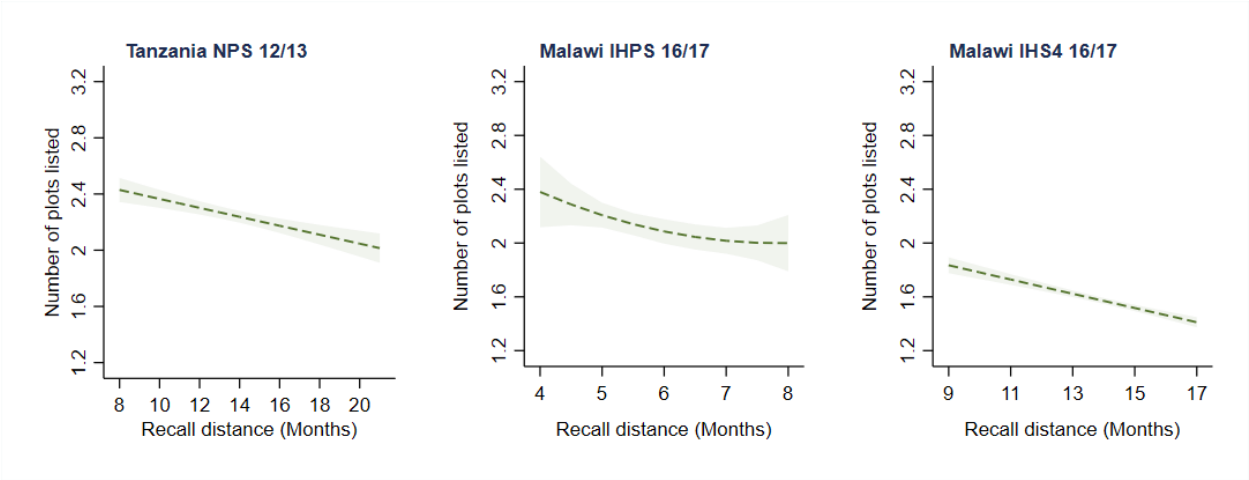

Figure 5. Effect of recall length on reported maize harvest, point estimates

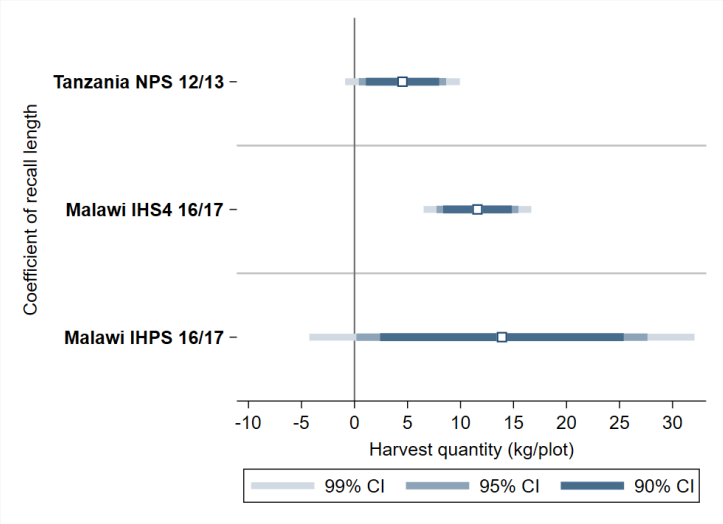

Figure 6. Maize harvest quantity against recall length

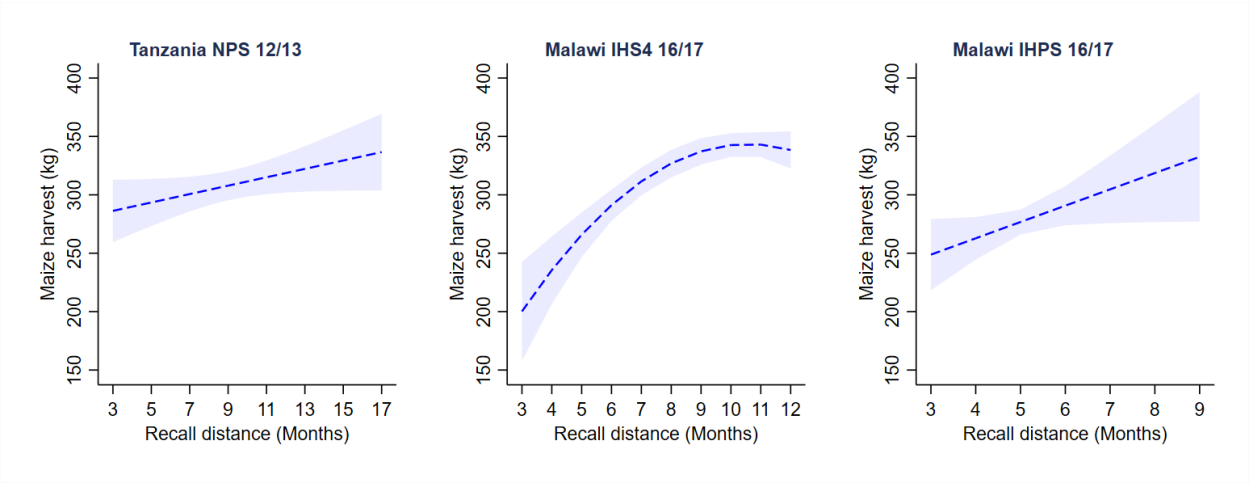

Figure 7. Self-reported harvest and error in self-reported harvest (self-report – crop cut), Uganda MAPSII

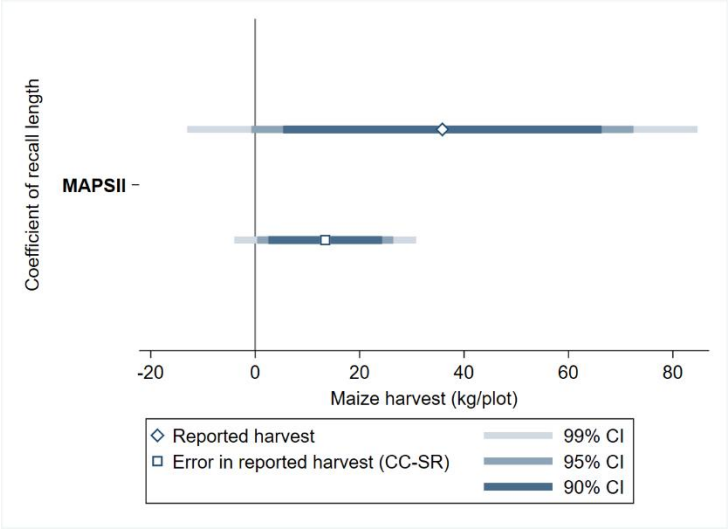

Figure 8. Effect of recall length on labor inputs, point estimates

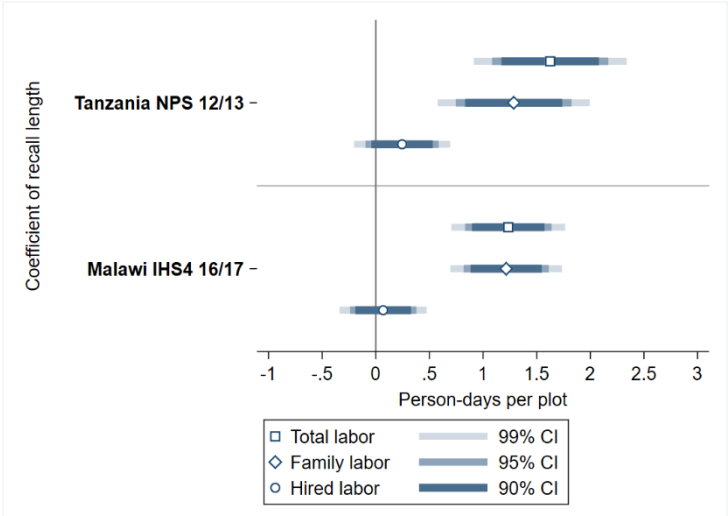

Figure 9. Total labor on plot against recall length.

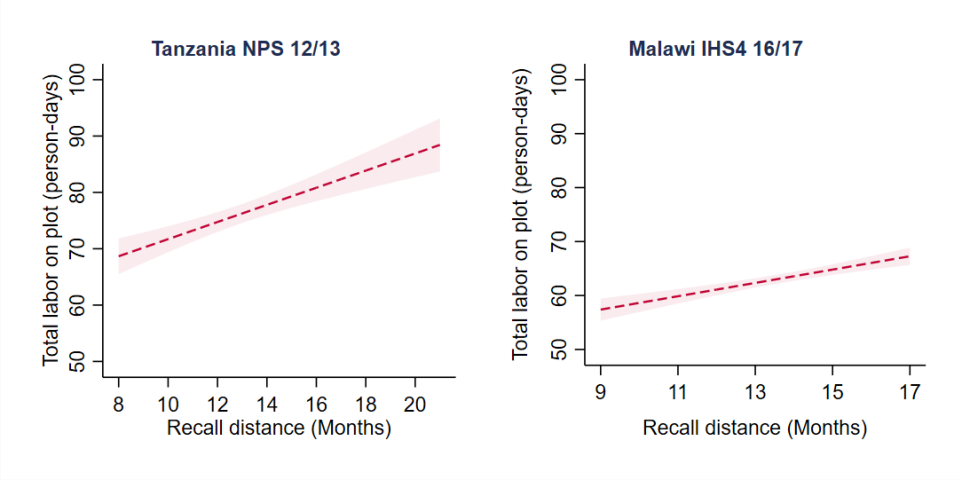

Figure 10. Effect of recall length on labor input by activity, point estimates

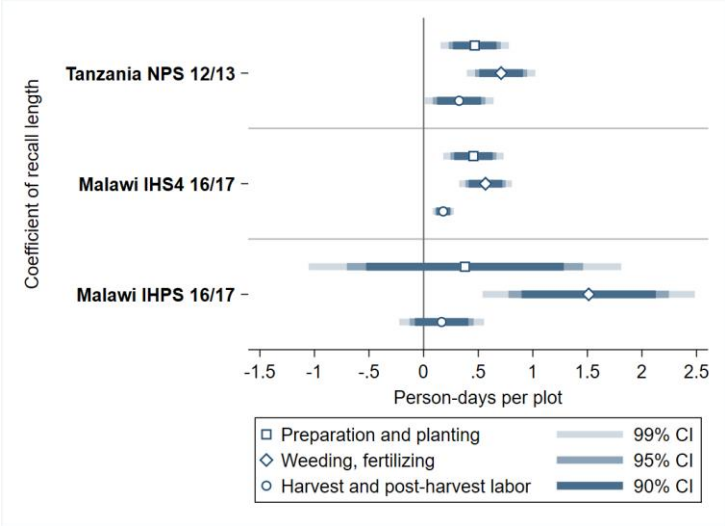

Figure 11. Effect of recall length on whether respondent reported using fertilizer on plot, point estimates

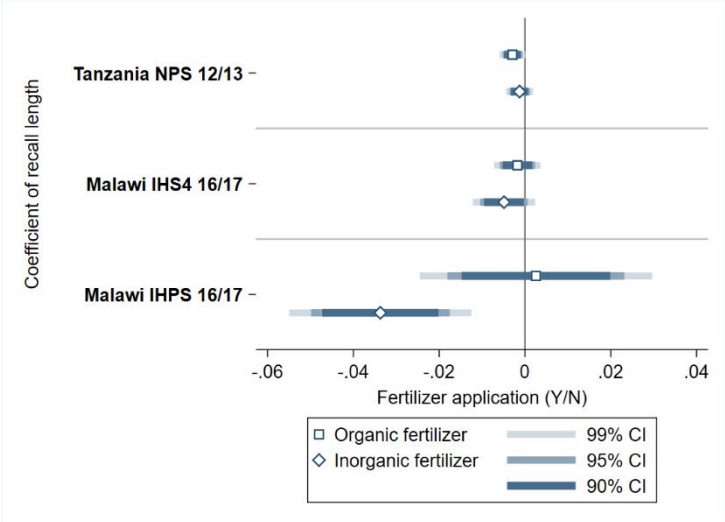

Figure 12. Effect of recall length on reported fertilizer quantity per plot, point estimates

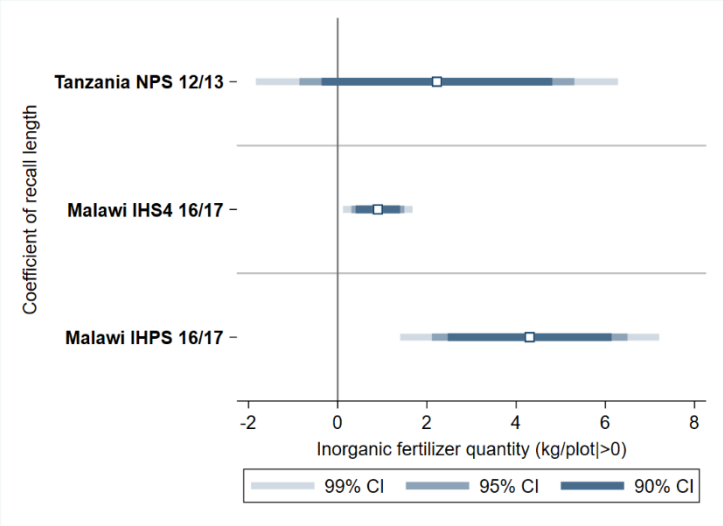

Figure 13. reported inorganic fertilizer amount (kg|>0) against recall length

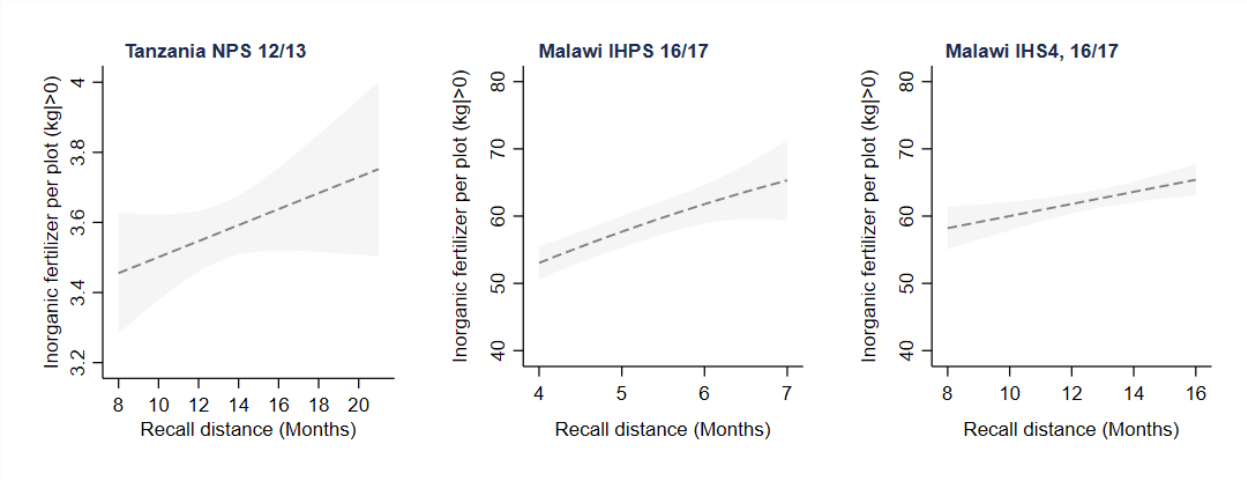

Figure 14. Selected outcomes against recall length at five quintiles of plot area, Tanzania NPS 2012/13

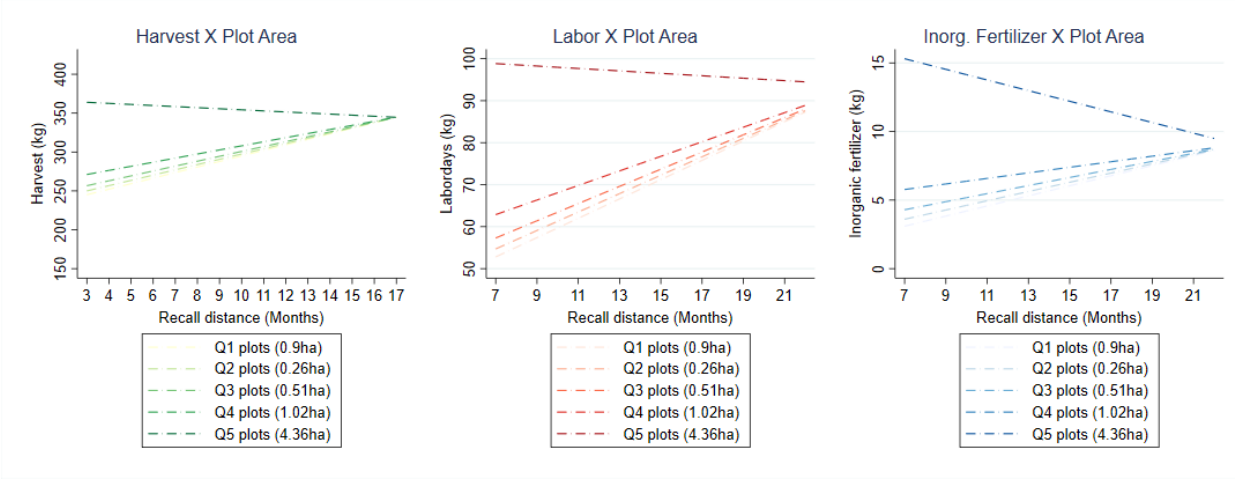

Figure 15. Selected outcomes against recall length at five quintiles of plot area, Malawi IHS4 and IHPS

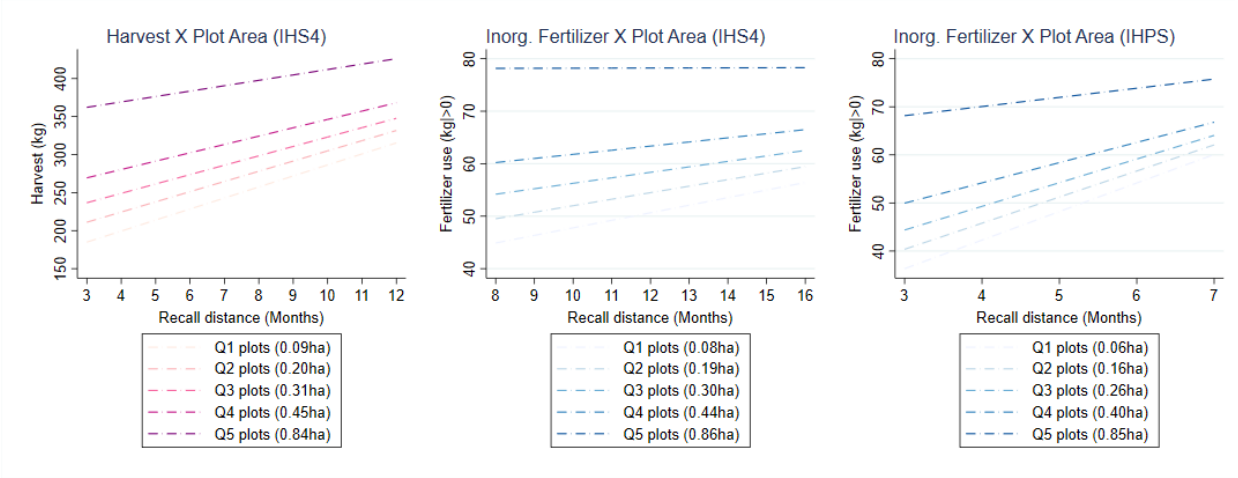

Figure 16. Impact on key productivity measures, plot-level

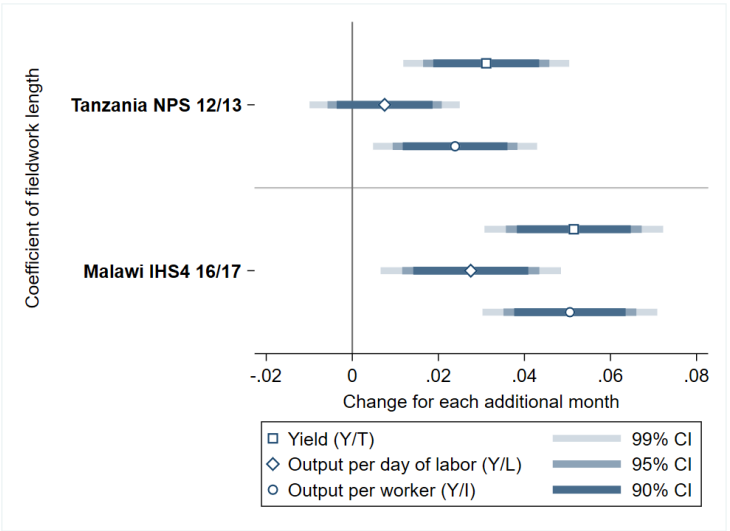

Table 6. Selected productivity measures

Productivity measure	**Construction**	**Individual effects**	**Expected recall effect**	**Estimated recall effect**
Yield (Y/T)	Production per plot/farm (Y) / Land area per plot/farm (T)	Y↑, T∅	Positive	Positive
Output per day worked (Y/L)	Production per plot (Y) / Total labor days per plot/farm (L)	Y↑, L↑	Ambiguous	Positive
Output per worker (Y/I)	Production per plot (Y) / household members per plot/farm (I)	Y↑, I ∅	Positive	Positive

References

Arthi, V., Beegle, K., De Weerdt, J., Palacios-López, A., 2018. Not your average job: Measuring farm labor in Tanzania. Journal of Development Economics 130, 160–172.

Ashour, M., Billings, L., Gilligan, D., Hoel, J.B., Karachiwalla, N., 2017. Do beliefs about agricultural inputs counterfeiting correspond with actual rates of counterfeiting? Evidence from Uganda. International Food Policy Research Institute.

Backiny-Yetna, P., Steele, D., Yacoubou Djima, I., 2017. The impact of household food consumption data collection methods on poverty and inequality measures in Niger. Food Policy 72, 7–19.

Beegle, K., Carletto, C., Himelein, K., 2012a. Reliability of recall in agricultural data. Journal of Development Economics 98, 34–41.

Beegle, K., De Weerdt, J., Friedman, J., Gibson, J., 2012b. Methods of household consumption measurement through surveys: Experimental results from Tanzania. Journal of Development Economics 98, 3–18.

Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S., 2011. Measurement errors in surveys. John Wiley & Sons.

Bold, T., Kaizzi, K.C., Svensson, J., Yanagizawa-Drott, D., 2017. Lemon technologies and adoption: measurement, theory and evidence from agricultural markets in Uganda. The Quarterly Journal of Economics 132, 1055–1100.

Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement error in survey data, in: Handbook of Econometrics. Elsevier, pp. 3705–3843.

Brzozowski, M., Crossley, T.F., Winter, J.K., 2017. A comparison of recall and diary food expenditure data. Food Policy 72, 53–61.

Carletto, C., Jolliffe, D., Banerjee, R., 2015. From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. The Journal of Development Studies 51, 133–148.

Carletto, G., Savastano, S., Zezza, A., 2011. Fact or Artefact: The Impact of Measurement Errors on the Farm Size - Productivity Relationship. Journal of Development Economics 103. https://doi.org/10.1016/j.jdeveco.2013.03.004

D'Alessio, G., 2017. Measurement errors in consumption surveys and the estimation of poverty and inequality indices.

De Groote, H., Traoré, O., 2005. The cost of accuracy in crop area estimation. Agricultural Systems 84, 21–38.

De Mel, S., McKenzie, D.J., Woodruff, C., 2009. Measuring microenterprise profits: Must we ask how the sausage is made? Journal of development Economics 88, 19–31.

Deininger, K., Carletto, C., Savastano, S., Muwonge, J., 2012. Can diaries help in improving agricultural production statistics? Evidence from Uganda. Journal of Development Economics 98, 42–50.

Desiere, S., Jolliffe, D., 2018. Land productivity and plot size: Is measurement error driving the inverse relationship? Journal of Development Economics 130, 84–98.

Dillon, A., Gourlay, S., McGee, K., Oseni, G., 2019. Land measurement bias and its empirical implications: evidence from a validation exercise. Economic Development and Cultural Change 67, 595–624.

Engle-Stone, R., Sununtnasuk, C., Fiedler, J.L., 2017. Investigating the significance of the data collection period of household consumption and expenditures surveys for food and nutrition policymaking: Analysis of the 2010 Bangladesh household income and expenditure survey. Food policy 72, 72–80.

Gaddis, I., Siwatu, G.O., Palacios-Lopez, A., Pieters, J., 2019. Measuring Farm Labor: Survey Experimental Evidence from Ghana. World Bank Policy Research Working Paper.

Global Strategy to improve Agricultural and Rural Statistics, 2017. Handbook on the Agricultural Integrated Survey (AGRIS).

Goldstein, M., Udry, C., 1999. Agricultural innovation and risk management in Ghana. Unpublished, final report to IFPRI.
Gollin, D., 2019. Farm size and productivity: Lessons from recent literature. IFAD Research Series 34, 2018.
Gourlay, S., Kilic, T., Lobell, D., 2017. Could the Debate Be Over? Errors in Farmer-Reported Production and Their Implications for the Inverse Scale-Productivity Relationship in Uganda. World Bank Policy Research Working Paper.
Groves, R.M., Lyberg, L., 2010. Total survey error: Past, present, and future. Public opinion quarterly 74, 849–879.
Kasprzyk, D., 2005. Chapter 9. Measurement error in household surveys: sources and measurement, in: Household Surveys in Developing and Transition Countries. United Nations. Statistical Division, and National Household Survey Capability Programme.
Keita, N., Carfagna, E., 2009. Use of modern geo-positioning devices in agricultural censuses and surveys: Use of GPS for crop area measurement, in: Bulletin of the International Statistical Institute, the 57th Session, 2009, Proceedings, Special Topics Contributed Paper Meetings (STCPM22), Durban.
Kilic, T., Moylan, H.G., Ilukor, J., Mtengula, C., PANGAPANGA-PHIRI, I., 2018. Root for the Tubers: Extended-Harvest Crop Production and Productivity Measurement in Surveys. World Bank Policy Research Working Paper.
Kilic, T., Zezza, A., Carletto, C., Savastano, S., 2017. Missing(ness) in Action: Selectivity Bias in GPS-Based Land Area Measurements. World Development 92, 143–157. https://doi.org/10.1016/j.worlddev.2016.11.018
Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology 49, 764–766. https://doi.org/10.1016/j.jesp.2013.03.013
Liedholm, C., 1991. Data collection strategies for small-scale industry surveys. GEMINI.
Meyer, B.D., Mok, W.K.C., Sullivan, J.X., 2015. Household Surveys in Crisis. Journal of Economic Perspectives 29, 1–29.
Moore, J.C., Stinson, L.L., Welniak, E.J., 2000. Income measurement error in surveys: A review. Journal of Official Statistics-Stockholm- 16, 331–362.
Roberts, John M., Brewer, Devon D., 2001. Measures and tests of heaping in discrete quantitative distributions. Journal of Applied Statistics 28 (7), 887–896.
Schøning, P., 2005. Handheld GPS equipment for agricultural statistics surveys: Experiments on area-measurements done during fieldwork for the Uganda Pilot Census of Agriculture, 2003. Statistisk sentralbyr\a a.
Schündeln, M., 2018. Multiple Visits and Data Quality in Household Surveys. Oxford Bulletin of Economics and Statistics 80, 380–405.
Schwarz, N., 2007. Cognitive aspects of survey methodology. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition 21, 277–287.
Seymour, G., Malapit, H., Quisumbing, A., 2017. Measuring time use in development settings. The World Bank.
Sudman, S., Bradburn, N.M., 1973. Effects of time and memory factors on response in surveys. Journal of the American Statistical Association 68, 805–815.
Troubat, N., Grünberger, K., 2017. Impact of survey design in the estimation of habitual food consumption: A study based on urban households of Mongolia. Food Policy 72, 132–145.
Zezza, A., Carletto, C., Fiedler, J.L., Gennari, P., Jolliffe, D., 2017. Food counts. Measuring food consumption and expenditures in household consumption and expenditure surveys (HCES). Introduction to the special issue. Food Policy 72, 1–6.
Zezza, A., Federighi, G., Adamou Kalilou, A., Hiernaux, P., 2016. Milking the data: Measuring milk off-take in extensive livestock systems. Experimental evidence from Niger. Food Policy 13.

Appendix

A. Supplemental Tables

Table A.1. Difference between raw and corrected recall length variables (months)

Variable	Mean	Min	Median	Max	SD	N
			Tanzania NPS 2012/13			
Recall length, harvest	1.010	0	1	6	1.040	3,079
			Malawi IHS4 2016/17			
Recall length, harvest	0.419	0	0	6	0.653	7,110
Recall length, planting	0.303	0	0	6	0.599	10,392
			Malawi IHPS 2016/17			
Recall length, harvest	0.399	0	0	6	0.696	3,633
Recall length, planting	0.350	0	0	7	0.658	2,566

Table A.2. GPS plot area and error in self-reported plot area.

	Tanzania NPS 2012/13		Malawi IHS4 2016/17		Malawi IHPS 2016/17	
Quintiles of plot area	Rel. error (%)	Plot area, GPS (ha)	Rel. error (%)	Plot area, GPS (ha)	Rel. error (%)	Plot area, GPS (ha)
1st (bottom quintile)	68%	0.09	60%	0.08	118%	0.06
2nd	30%	0.27	29%	0.19	46%	0.16
3rd	9%	0.52	9%	0.30	20%	0.26
4th	-5%	1.04	-6%	0.44	2%	0.40
5th (top quintile)	-18%	4.35	-24%	0.85	-20%	0.82

Table A.3. Linear regression results number of plots and parcels per household.

	Tanzania NPS 12/13		Malawi IHS4 2016/17			Malawi IHPS 2016/17		
Dependent Variable	Plots listed	Plots cultivated	Plots listed	Plots cultivated	Parcels listed	Plots listed	Plots cultivated	Parcels listed
Recall length, planting	-0.0336***	-0.00854	-0.0530***	-0.0510***	-0.0346***	-0.102**	-0.148***	-0.0530*
	(0.00662)	(0.00630)	(0.00543)	(0.00519)	(0.00434)	(0.0410)	(0.0417)	(0.0280)
Observations	3,225	3,225	6,742	6,742	6,742	1,786	1,786	1,786
Pseudo R-squared	0.190	0.158	0.253	0.245	0.183	0.274	0.253	0.220
Controls	Full controls	Full controls	Full controls	Full controls	Full controls	Full controls	Full controls	Full controls
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS

Table A.4. Regression results number of plots listed per household, quadratic specification

	Tanzania NPS 12/13	Malawi IHS4 2016/17	Malawi IHPS 2016/17
Dependent Variable	Plots Listed	Plots listed	Plots listed
Recall length, season	-0.0220	0.0184	-0.405
	(0.0379)	(0.0544)	(0.384)
Recall length^2	-0.000385	-0.00277	0.0258
	(0.00128)	(0.00208)	(0.0311)
Observations	3,225	6,742	1,786
Adjusted R-squared	0.190	0.253	0.274
Controls	Full controls	Full controls	Full controls
Estimator	OLS	OLS	OLS
Joint F-Test	12.98	49.43	3.28

Table A.5. Linear regression error in self-reported plot area relative to GPS area measurement, log-level OLS

	Tanzania NPS 12/13	Malawi IHS4 2016/17	Malawi IHPS 2016/17

Dependent Variable	Ln(Error in Self-Reported Plot Area (Self-Reported Area – GPS Area))					
	Error>0	Error<0	Error>0	Error<0	Error>0	Error<0
Recall length, planting	-0.00632	-0.00213	0.0257***	0.00377	-0.0242	-0.0113
	(0.00877)	(0.0105)	(0.00534)	(0.00534)	(0.0305)	(0.0314)
Observations	1,612	1,411	7,416	5,738	2,008	1,281
Adjusted R-squared	0.070	0.132	0.184	0.146	0.028	0.339
Controls	Full controls	Full controls	Full controls	Full controls	Full controls	Full controls
Estimator	OLS	OLS	OLS	OLS	OLS	OLS

Table A.6. Linear regression results for maize harvest per plot

	Tanzania NPS 12/13	**Malawi IHS4 2016/17**	**Malawi IHPS 2016/17**
Dependent Variable	Maize harvest (kg)	Maize harvest (kg)	Maize harvest (kg)
Recall length, harvest	4.901**	11.97***	15.85**
	(1.983)	(1.961)	(6.996)
Observations	2,979	7,094	2,261
Adjusted R-squared	0.188	0.398	0.446
Controls	Full controls	Full Controls	Full Controls
Estimator	OLS	OLS	OLS

Table A.7. Linear regression results for maize harvest per plot, log-level specification

	Tanzania NPS 12/13	**Malawi IHS4 2016/17**	**Malawi IHPS 2016/17**
Dependent Variable	Ln(Maize harvest (kg))	Ln(Maize harvest (kg))	Ln(Maize harvest (kg))
Recall length, harvest	0.0211***	0.0418***	0.0768**
	(0.00650)	(0.00727)	(0.0300)
Observations	2,798	6,770	2,176
Adjusted R-squared	0.254	0.259	0.374
Controls	Full controls	Full controls	Full Controls
Estimator	OLS	OLS	OLS

Table A.8. Regression results for maize harvest per plot with quadratic recall length term

	Tanzania NPS 12/13	**Malawi IHS4 2016/17**	**Malawi IHPS 2016/17**
Dependent variable	Maize harvest (kg)	Maize harvest (kg)	Maize harvest (kg)
Recall length	19.57*	52.77***	-21.22
	(10.62)	(12.57)	(43.76)
$(\text{Recall length})^2$	-0.782	-2.494***	3.032
	(0.570)	(0.728)	(3.534)
Observations	2,979	7,094	2,261
Adjusted R-squared	0.188	0.399	0.446
Controls	Full controls	Full Controls	Full Controls
Estimator	OLS	OLS	OLS
Joint F-Test	4.50	21.0	3.1

Table A.9. Regression results for labor inputs, Tanzania NPS 2012/13

Dependent variable	Total person-days	Family person-days	Hired labor (Y/N)	Hired person-days (|>0)	Hired person-days
Recall length	1.626***	1.286***	0.00213	0.244	0.252
	(0.276)	(0.275)	(0.00210)	(0.174)	(0.171)
Observations	6,068	6,068	5,986	1,911	6,068
Adjusted R-squared*	0.263	0.253	0.156	0.180	0.062
Controls	Full controls	Full controls	Full controls	Full controls	Full controls
Estimator	OLS	OLS	Probit	OLS	Tobit

*Pseudo R-squared for Probit and Tobit

Table A.10. Regression results for labor inputs, Malawi IHS4 2016/17

	Total person-days	Family person-days	Hired labor (Y/N)	Hired person-days (|>0)	Hired person-days
Recall length	1.235***	1.216***	0.000141	0.0683	0.0581
	(0.205)	(0.202)	(0.00165)	(0.157)	(0.140)
Observations	10,331	10,331	10,278	1,394	10,331
Adjusted R-squared*	0.387	0.359	0.298	0.316	0.145
Controls	Full controls	Full controls	Full controls	Full controls	Full controls
Estimator	OLS	OLS	Probit	OLS	Tobit

*Pseudo R-squared for Probit and Tobit

Table A.11. Total labor input, log-level specification.

	Tanzania NPS 12/13	**Malawi IHS4 2016/17**
Dependent Variable	Log(Total person-days)	Log(Total person-days)
Recall length	0.0228***	0.0280***
	(0.00388)	(0.00369)
Observations	5,938	10,363
Adjusted R-squared	0.269	0.392
Controls	Full controls	Full controls
Estimator	OLS	OLS

Table A.12. Regression results for labor inputs by activity, log-level specification, Tanzania NPS 2012/13

Dependent Variable	Preparation and planting ln(person-days)	Weeding, fertilizing ln(person-days)	Harvest and post-harvest ln(person-days)
Recall length	0.0184***	0.0245***	0.0144***
	(0.00446)	(0.00442)	(0.00456)
Observations	5,486	5,718	4,616
R-squared	0.225	0.222	0.273
Controls	Full controls	Full controls	Full controls
Estimator	OLS	OLS	OLS

Table A.13. Regression results for labor inputs by activity, log-level specification, Malawi IHS4 2016/17

Dependent Variable	Preparation and planting ln(person-days)	Weeding, fertilizing ln(person-days)	Harvest and post-harvest ln(person-days)
Recall length	0.0249***	0.0307***	0.0253***
	(0.00401)	(0.00410)	(0.00417)
Observations	9,915	9,893	9,763
Adjusted R-squared	0.357	0.414	0.365
Controls	Full controls	Full controls	Full controls
Estimator	OLS	OLS	OLS

Table A.14. Regression results for labor inputs by activity, log-level specification, Malawi IHPS 2016/17

Dependent Variable	Preparation and planting ln(person-days)	Weeding, fertilizing ln(person-days)	Harvest and post-harvest ln(person-days)
Recall length	0.0300	0.104***	0.0142
	(0.0269)	(0.0256)	(0.0118)
Observations	3,585	3,585	3,576
Adjusted R-squared	0.357	0.343	0.306
Controls	Full controls	Full controls	Full controls
Estimator	OLS	OLS	OLS

Table A.15. Regression results for fertilizer inputs, Tanzania NPS 2012/13

Dependent Variable	Organic fertilizer (Y/N)	Inorganic fertilizer (kg)	Inorganic fertilizer (Y/N)	Inorganic fertilizer (kg|>0)
Recall length	-0.00292**	3.722***	-0.00126	2.224
	(0.00117)	(0.940)	(0.00122)	(1.568)
Observations	7,284	7,322	6,949	675
Adj. R-squared*	0.168	0.064	0.281	0.233
Controls	Full controls	Full controls	Full controls	Full controls
Estimator	Probit	Tobit	Probit	OLS

*Pseudo R-squared for Probit and Tobit

Table A.16. Regression results for fertilizer inputs, Malawi IHS4 2016/17

Dependent Variable	Organic Fertilizer (Y/N)	Inorganic Fertilizer (kg)	Inorganic Fertilizer (Y/N)	Inorganic Fertilizer (kg|>0)
Recall length	-0.00176	-0.179	-0.00489*	0.900***
	(0.00210)	(0.300)	(0.00282)	(0.302)
Observations	10,215	10,331	10,229	5,498
Adj. R-squared*	0.153	0.082	0.346	0.234
Controls	Full controls	Full controls	Full controls	Full controls
Estimator	Probit	Tobit	Probit	OLS

*Pseudo R-squared for Probit and Tobit

Table A.17. Regression results for fertilizer inputs, Malawi IHPS 2016/17

Dependent Variable	Organic fertilizer (Y/N)	Inorganic fertilizer (kg)	Inorganic fertilizer (Y/N)	Inorganic fertilizer (kg|>0)
Recall length	0.00256	-1.308	-0.0337***	4.304***
	(0.0105)	(1.098)	(0.00825)	(1.105)
Observations	3,364	3,585	3,421	2,186
Adj. R-squared*	0.147	0.091	0.410	0.338
Controls	Full controls	Full controls	Full controls	Full controls
Estimator	Probit	Tobit	Probit	OLS

*Pseudo R-squared for Probit and Tobit

Table A.18. Regression results for fertilizer inputs, log-level specification

	Tanzania NPS 12/13	**Malawi IHS4 2016/17**	**Malawi IHPS 2016/17**
Dependent Variable		Ln(Inorganic fertilizer use (kg|>0))	
Recall length	0.0187	0.0190***	0.0884***
	(0.0154)	(0.00562)	(0.0248)
Observations	675	5,498	2,185
Adjusted R-squared	0.325	0.234	0.331
Controls	Full controls	Full controls	Full controls
Estimator	OLS	OLS	OLS

Table A.19. Impact of longer recall length (longer fieldwork period) on productivity measures, plot-level

	Tanzania NPS 12/13 – plot level			**Malawi IHS4 2016/17 – plot level**		
Dependent Variable	log(Y/T)	log(Y/L)	log(Y/I)	log(Y/T)	log(Y/L)	log(Y/I)
Fieldwork / Recall length	0.0311***	0.00752	0.0239***	0.0515***	0.0275***	0.0506***
	(0.00748)	(0.00677)	(0.00739)	(0.00803)	(0.00811)	(0.00786)
Observations	2,775	2,774	2,767	7,093	7,087	7,050
Adjusted R-squared	0.240	0.245	0.258	0.289	0.336	0.335
Controls	Full controls	Full controls	Full controls	Full Controls	Full Controls	Full Controls
Estimator	OLS	OLS	OLS	OLS	OLS	OLS

Table A.20. Impact of longer recall length (longer fieldwork period) on productivity measures, farm-level

	Tanzania NPS 12/13 – farm level			**Malawi IHS4 2016/17 – farm level**		
Dependent Variable	log(Y/T)	log(Y/L)	log(Y/I)	log(Y/T)	log(Y/L)	log(Y/I)
Fieldwork / Recall length	0.0224***	0.0186**	0.0251***	0.0531***	0.0310***	0.0299***
	(0.00753)	(0.00743)	(0.00757)	(0.00756)	(0.00759)	(0.00728)
Observations	1,877	1,875	1,871	5,790	5,786	5,763
Adjusted R-squared	0.274	0.270	0.245	0.280	0.308	0.308
Controls	Full controls	Full controls	Full controls	Full Controls	Full Controls	Full Controls
Estimator	OLS	OLS	OLS	OLS	OLS	OLS