

Root for the Tubers

Extended-Harvest Crop Production and Productivity Measurement in Surveys

Talip Kilic

Heather Moylan

John Ilukor

Clement Mtengula

Innocent Pangapanga-Phiri



WORLD BANK GROUP

Development Economics
Development Data Group
October 2018

Abstract

To document the relative accuracy of methods for microdata collection on root and tuber crop production, an experiment was implemented in Malawi over a 12-month period, randomly assigning cassava-producing households to one of four approaches: daily diary-keeping, with semi-weekly supervision visits; daily diary-keeping, with semi-weekly supervisory phone calls; two six-month recall interviews, with six months in between; and a single 12-month recall interview. Lapses in diary-keeping can underestimate true production, albeit to a lesser degree compared to recall. And the comparisons between the diary variants and the variation in underestimation by recall period are unclear ex ante. The analysis reveals that compared to traditional diary-keeping, the household-level annual cassava production is 295 kilograms higher, on average, (and assumed as closer to the truth) under diary-keeping with phone calls. This

effect corresponds to 28 percent of the average traditional diary-keeping production estimate. Although the difference between the estimates based on six-month recall and traditional diary-keeping is statistically insignificant, 12-month recall underestimates annual production, on average, by 516 kilograms and 221 kilograms, respectively, compared to diary-keeping with phone calls and traditional diary-keeping. While the recall-based approaches both underestimate true production, six-month recall does so to a lesser extent. The evidence additionally demonstrates likely gross overestimation in international and ministerial statistics on cassava yields in Malawi. For improved microdata on root and tuber crop production, the adoption of (i) diary-keeping with phone calls (particularly if deployed in a broader mobile phone-based survey) or (ii) six-month recall, as a second-best alternative, is recommended.

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

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

**Root for the Tubers:
Extended-Harvest Crop Production and Productivity Measurement in Surveys**

Talip Kilic[‡], Heather Moylan[‡], John Ilukor⁺,
Clement Mtengula[†], and Innocent Pangapanga-Phiri^{#1}

JEL Codes: C83, Q12.

Keywords: Root Crops, Tuber Crops, Tree Crops, Extended-Harvest Crops, Cassava, Production Measurement, Yield Measurement, Harvest Diaries, Recall, Crop Cutting, Household Surveys, Malawi, Sub-Saharan Africa.

¹ [‡]Senior Economist, Living Standards Measurement Study (LSMS), Survey Unit, Development Data Group (DECDG), The World Bank, Rome, Italy; tkilic@worldbank.org. [†]Corresponding Author. Survey Specialist, LSMS, Survey Unit, DECDG, The World Bank, Rome, Italy; hmoylan@worldbank.org. ⁺Survey Specialist, LSMS, Survey Unit, DECDG, The World Bank, Kampala, Uganda; jilukor@worldbank.org. [†]Independent Consultant, Zomba, Malawi; cmtengula@gmail.com. [#]Lecturer, Lilongwe University of Agriculture and Natural Resources, Lilongwe, Malawi; phiriinnocent@gmail.com.

1. Introduction

Agriculture is the backbone of the economy in rural areas that house nearly 70 percent of the population in low-income countries. The importance of agriculture for development is recognized recently during the formulation of the Sustainable Development Goal (SDG) Target 2.3, which requires doubling of agricultural productivity and incomes of small-scale food producers. In Africa, the average share of rural household income tied to agriculture could be as high as 69 percent (Davis et al., 2017), and the research has shown that compared to non-agricultural growth, agricultural growth translates into higher rates of poverty reduction (Dorosh and Thurlow, 2016).

Yet, despite the importance of agriculture for livelihoods, obtaining accurate and timely estimates on crop production and yields remains a significant challenge. The weaknesses in agricultural data in turn undermine the ability to gauge the performance of the sector; assess the extent to which interventions translate into gains in production, productivity and welfare, both within and across agricultural households; and ultimately, anchor policies in evidence (Carletto et al., 2015). In the context of root, tuber and tree crops, the primary concern around the accuracy of prevailing approaches to data collection in large-scale household and farm surveys that underlie official statistics is arguably the reliance on respondent recall.

For cassava, the crop that is at the center of our analysis and that could be planted to hedge against the risk of seasonal crop failure and/or food insecurity during the lean season, harvesting typically takes place over an extended period and as needed. Irrespective of using a given agricultural season, or a more extended time frame, such as 12 months, as a reference period, respondents are likely to underestimate production as they attempt to recall the accounts of continuous harvests that take place throughout the year and often in small quantities (Friedman et al., 2017; de Nicola and Gine; 2014).² And in the specific case of using an agricultural season as the reference period, the recall-based data on production may be incomplete if cassava harvests take place outside the reference season. Other factors that complicate the prospects of reliable production data solicited by household and farm surveys include using non-standard measurement units in farmer-reported crop production³; and control of plots by household members different from the respondents.⁴

² “Rule of thumb error” may too plague reporting under long recall periods, when respondents cease trying to enumerate each harvest but use a rule of thumb to estimate them (Friedman et al., 2017). However, there is no obvious direction of bias associated with rule of thumb error, with household-specific temporal patterns of cassava harvest playing a significant role in its potential manifestation.

³ Common units in which cassava production can be reported include sack, heap and piece, whose size/weight can vary dramatically across farmers – based on locality and cassava variety-specific harvest calendars – and even for the same farmer, as in the case of a farmer reporting two 50-kilogram sacks, of which one could have been 80 percent full while the other could have been overflowing.

⁴ Recall-based interviews are often conducted in a specific time frame that revolves less around the schedule of plot managers, resulting in a greater potential for relying on proxy respondents.

The studies assessing the relative accuracy of recall vis-à-vis diary- or crop cutting-based approaches to data collection on crop production and yields are few and far between.⁵ Deininger et al. (2012) compare recall- versus diary-based estimates of cassava production, captured as part of a national crop diary operation integrated into the Uganda National Household Survey (UNHS) 2005/06. The documented magnitude of discrepancies between the two methods are especially high for cassava. Specifically, the authors show that recall underestimates both the household-level incidence of cassava cultivation and the value of cassava production by 28 percent and 24 percent, respectively. And Deininger et al. (2012) acknowledge that these already large discrepancies may constitute a lower bound given the 6-month duration of the diary operation and the fact that cassava can be harvested over a longer period. In the context of Malawi, at the macro-level, the extent of discrepancies among the international databases, the ministerial data sources, and the national household surveys in terms of reported cassava production and yields further underscore the need for methodological research for improving the statistics on root, tuber and tree crops in general, and cassava in particular.

Considering the importance of cassava for food security and the absence of best practices in accurate survey data collection on extended-harvest crop production and yields, we present the results of a household survey experiment that randomly assigned sampled households in top cassava-producing districts in Malawi to one of four approaches to cassava production measurement over a 12-month period. These methods were: (1) daily diary-keeping, with semi-weekly in-person supervision visits (D1) – the traditional gold standard in survey data collection; (2) daily diary-keeping, with semi-weekly supervisory phone calls (D2); (3) 6-month recall-based data collection in 2 visits that were 6 months apart (R1); and (4) 12-month recall-based data collection in a single visit (R2) – the current approach in Malawi and arguably the most cost effective practice for household and farm surveys collecting information on cassava production in low- and middle-income countries. Beyond the socio-economic and agricultural data collected on the sampled households, we leverage (i) the GPS-based areas of the cassava plots cultivated by the households, and (ii) the crop cutting data tied to a 5x5m sub-plot that was placed at random within a randomly-selected cassava plot of each household (regardless of the survey treatment) and that was harvested and weighed at the time of each household’s preference during the 12-month fieldwork.

The contributions of this work are fourfold. First, the results from this study provide another point of empirical evidence regarding the relative accuracy and cost-effectiveness of recall-based methods vis-à-vis their diary-based counterparts for household-level annual cassava production estimation. The findings feed into the guidance we provide on survey design for data collection on

⁵ The body of work investigating the accuracy of recall vis-a-vis diary extends beyond agricultural data. See, for instance, Beegle et al. (2012); Backiny-Yetna et al. (2017); Brzozowski et al. (2017); Friedman et al. (2017); and Troubat and Grunberger (2017) for the work on consumption data. Though less relevant for extended-harvest crops, Desiere and Jolliffe (2018) and Gourlay et al. (2017) document, in Uganda and Ethiopia, respectively, large (upward) biases in recall-based data on seasonal crop production and yield in comparison to crop cutting.

root crop production. Second, within the diary domain, we further assess the relative accuracy of both options, and the scale-up feasibility of the daily diary keeping with supervisory phone calls. These insights are relevant also for the set-up of mobile phone-based data collection platforms that aim to collect broader data from human subjects at higher frequency. Third, we compare the annual cassava yield estimates obtained under the four survey treatments to the household-specific annual crop cutting-based cassava yield estimates obtained by extrapolating the crop cut sub-plot yields to the entire household area under cassava cultivation. The discussion of the interrelationships among the different yield estimates within our study, as well as their comparisons to the referenced cassava yields at the national- and international-levels, present a sobering need for convergence on a common understanding of what widely-varying cassava yields capture. Finally, this is the first comprehensive study undertaken on the topic in Malawi; a country in which the food security role of cassava cannot be underestimated in the face of intensifying extreme weather events that have adversely affected seasonal agricultural production in the recent past. Our study spans five districts across all three regions of Malawi. The selected districts exhibit remarkable differences in terms of the extent to which cassava production is geared towards home consumption versus market sales; the length of the cassava harvest period (and varieties); and the plot-level intensity of cassava cultivation (in part depending on whether cassava is intercropped with other seasonal crops).

Under the assumption that true cassava production is underestimated due to incomplete record-keeping even in diaries that are implemented over an extended period, the analysis reveals that compared to D1, the annual household cassava production was 295 kilograms higher, on average, under D2, corresponding to 28 percent of the D1 mean, and that the traditional gold standard is outperformed by a competing diary variant in terms of capturing cassava production as comprehensively as possible. Further, we document that the average difference between R1- and D1-based estimates was not statistically significant, but that R2 – the most cost-effective practice in large-scale surveys – underestimated annual production, on average, by 221 and 516 kilograms, compared to D1 and D2, respectively. And while both recall variants underestimated annual production by a significant margin compared to D2, R1 did so to a lesser extent.

Finally, compared to crop cutting – which is the traditional gold standard for *seasonal* crop yield measurement and should be understood as an upper bound for *extended-harvest* crop yield realized on the farm, the average household-level annual cassava yields (kilograms per hectare) were underestimated by each survey treatment – ranging from an underestimation of 25 percent under D2 to 47 percent under R2 – as expected and described below. Yet, the latest available average national cassava yield estimates from FAOSTAT and the Malawi Ministry of Agriculture, Irrigation and Water Development (MoAIWD) were at least twice as much as the average crop cutting-based annual cassava yield estimate across our study districts.

Overall, the considerable variation in production and yield estimates by survey treatment underscores the need for adopting improved survey methods to collect the required data for more accurately capturing the contribution of cassava farming to production and welfare outcomes. The findings, together with the cost comparisons, support the use of (i) D2, particularly if deployed as part of a broader mobile-phone based survey effort, or (ii) R1, as a second-best alternative.

The paper is organized as follows. Section 2 covers the country context. Section 3 discusses the data. Section 4 lays out the empirical strategy. Section 5 presents the results. Section 6 concludes.

2. Country Context

Malawi is heavily dependent on agriculture. The agricultural sector makes up 29 percent of the country's gross domestic product. The household-level incidence of participation in agriculture is 83 percent at the national level, and 93 percent in rural areas (NSO, 2017). Davis et al. (2017) report that the share of household income stemming from agriculture stands at 65 percent. However, much of the agricultural production is for subsistence: the average value of crop sales as a share of the value of overall crop production stands at 18 percent (Carletto et al., 2017).

Cassava is an important staple crop that could provide up to 15 percent of the calories consumed at the household level in Malawi.⁶ The labor and non-labor input requirement of cassava cultivation is lower compared to seasonal crops, such as maize, and the crop is known for its adaptability to varying climatic and soil conditions (Kabambe, 2011). While the crop is grown throughout the country, the central and northern region districts along Lake Malawi (i.e. the cassava belt) lead cassava production and gear much of their cassava production towards home consumption. These districts predominantly grow bitter cultivars that constitute an essential component of diets and that can be harvested, depending on the variety, within a 12- to 18-month period. The preference for bitter varieties is anchored in the desire to protect the crop against thieves and animals, and in the tradition to produce *kondowole* – fermented cassava – from these cultivars (Kambewa and Nyembe, 2008). Elsewhere in Central and Southern Malawi, while the caloric contribution of cassava consumption is not nearly as high, sweeter cassava varieties are grown, over harvest periods of 6 to 12 months, predominantly for commercial purposes (sold either in raw form or as processed snack food) and/or as a last resort food security measure during the lean season or in the face of unforeseen shocks (Moyo et al., 2004). Typically, when non-cassava belt districts in southern Malawi rely on cassava as a staple food, they produce *makaka*, a non-fermented cassava flour formed from dried cassava chips. The spatial differences in cultivated cassava varieties are also at the heart of spatial differences in production and yield (Kabambe, 2011).

⁶ Based on the authors' calculations using the Integrated Household Panel Survey (IHPS) 2016 data.

With an increase in the frequency of unforeseen extreme weather events that threaten seasonal crop production and thereby rural livelihoods (McCarthy et al., 2017), cassava cultivation is garnering more attention, as also evident in the National Agricultural Policy (MoAIWD, 2016). The promotion of cassava production and value addition in Malawi requires a change in the approach to public spending on agriculture, of which an average of 50 percent was geared towards maize production during the period of 2006-2013 (FAO, 2014). To inform such process, one needs reliable time series data on cassava cultivated area, production and yields. Currently, there are at least two sources of annual, national-level cassava statistics in Malawi.⁷ Agricultural Production Estimates Survey (APES), which is conducted annually by the Ministry of Agriculture, Irrigation and Water Development (MoAIWD), is the first source. APES approach to cassava yield estimation is based on crop cutting, although the operational documentation on the APES cassava crop cuts, specifically their supervision, timing and processing, is not available. The latest publicly available APES report for the 2015/16 season reports a national cassava yield of 17,564 kilograms per hectare (kg/ha). The same report puts the figure for the 2014/15 season at 18,042 kg/ha.⁸

FAOSTAT is the second source of cassava statistics in Malawi. Figure 1 shows the FAOSTAT national yield and harvested area estimates from 2005 to 2014 (i.e. the latest year for which the estimates are available). The exact details on the computation of the FAOSTAT-based statistics are not available, and they are presumably at least in part a function of the information provided by the MoAIWD in response to the annual FAOSTAT questionnaire. If true, the increase in the national cassava yield from 14,300 kg/ha in 2005 to 22,504 kg/ha in 2014, as well as the parallel surge in the harvested area from 153,687 hectares to 222,750 hectares during the same period would be remarkable.⁹ However, given the lack of immediately-accessible documentation on the generation of the APES and FAOSTAT estimates (and disregarding for now the discrepancies between them), it is not clear whether these yields capture potential, attainable, economic or actual yields, as defined later in section 3.4.

3. Data

To investigate the relative accuracy and cost-effectiveness of a range of survey methods vis-à-vis their gold standard counterparts in the areas of cassava production and productivity measurement, as well as cassava variety identification, the Malawi National Statistical Office (NSO), in collaboration with the World Bank Living Standards Measurement Study (LSMS) and the CGIAR Standing Panel on Impact Assessment (SPIA), implemented a randomized household survey

⁷ A third source is the Integrated Household Survey (IHS), which is conducted by the National Statistical Office (NSO). Previous rounds of the IHS that could be used for computing national estimates of cassava production and yield are 2004/05 (IHS2), 2010/11 (IHS3), and 2016/17 (IHS4).

⁸ The time series information on cassava production and yield was not available on the MoAIWD website at the time this paper first appeared online.

⁹ The available statistics in fact date back to 1964, supportive of the upward trend observed for the 2005-2014 period.

experiment known as CVIP: Methodological Experiment on Measuring Cassava Production, Productivity, and Variety Identification. The CVIP fieldwork started in July 2015 and ended in August 2016. The data were collected on Android tablets, using a computer-assisted personal interviewing (CAPI) platform that was designed with the World Bank *Survey Solutions* CAPI software.

The CVIP target universe included households cultivating cassava in five major cassava producing districts in northern, central, and southern Malawi, namely Nkhatabay, Nkhotakota, Lilongwe, Zomba and Mulanje.¹⁰ The selection of these five districts ensured representation from each agro-ecological zone and Agricultural Development Division (ADD), with the exception of Karonga ADD.¹¹ Based on the last National Census of Agriculture and Livestock (NACAL) 2006/07, Nkhatabay and Nkhotakota were in fact the two districts with the highest percent of households involved in cassava cultivation, production and area under cultivation.

Following the district selection, the district agricultural development officers, the assistant agricultural development officers and the crop specialists in each of the five districts were consulted. The consultations aimed to obtain the APES-based cassava production estimates for the MoAIWD Extension Planning Areas (EPAs) within each district, such that cassava-producing EPAs were identified. Following the identification of these EPAs, the NSO cartography department identified the census enumeration areas (EAs) in each EPA.

Out of the universe of EAs in cassava-producing EPAs, 9 EAs were sampled in each district, with probability proportional to the EA-level count of households as captured in the 2008 Population and Housing Census (PHC). In each sampled EA, a household listing exercise was conducted in June 2015 to identify the universe of households cultivating cassava at that time. Out of that list, 28 households were sampled at random in each EA, and 7 sampled households were randomly assigned to 1 of 4 survey treatment that differed in the approach to data collection on production, as explained in section 3.1. Although the target was to have 1,260 households overall and 315 households per treatment arm, the final sample contained 1,218 households.¹²

¹⁰ To select the districts, the key CVIP team members from the World Bank LSMS and the NSO embarked on five missions throughout 2014 to consult with the experts on agriculture and cassava production from the MoAIWD Department of Crops and Department of Agricultural Research Services, the Lilongwe University of Agriculture and Food Security, the Chancellor College Cassava Value Addition Project, and the International Institute of Tropical Agriculture (IITA). The topics discussed included, but were not limited to, production estimates for cassava at the national- and district-levels; planting and harvesting times for cassava varieties; and the approaches to cassava production and yield estimation as well as the challenges associated with each.

¹¹ Although the Karonga ADD is also known for cassava production, due to the location of the NSO headquarters in Zomba, the experiment did not cover the northernmost districts so that regular supervision trips by NSO management team members were feasible from a cost and time management perspective.

¹² The universe of households in one EA in each of Lilongwe and Mulanje contained fewer than 28. Further, due to a poor agricultural season and the need for 19 sampled households in Lilongwe and Mulanje to start harvesting cassava earlier than our first visit, CVIP started out with 1,241 households. An additional 21 households were lost due to

Each EA was assigned to a specific enumerator.¹³ As noted above, the fieldwork ran for 12 months. During the first month, all sampled households were visited for the first time. In this first visit, the enumerators administered a light multi-topic questionnaire to all sampled households in a given EA, collecting basic information on household demographics, economic activities, dwelling attributes, farm and durable asset ownership, along with detailed information on all gardens and (within-garden) plots owned and/or cultivated.¹⁴ These questionnaire modules did not vary by survey treatment, and were based on the comparable Malawi Integrated Household Survey (IHS) questionnaire modules. In addition, irrespective of the survey treatment, the enumerator obtained the plot areas and the outlines for the cassava plots cultivated by the CVIP households, using a handheld Garmin eTrex 30 GPS unit.¹⁵ The respondent reported the area of each cassava plot,¹⁶ and irrespective of the survey treatment, one cassava plot was selected at random in each household for crop cutting, as explained in section 3.2. During the first visit, the enumerator also laid down a 5x5m crop cutting sub-plot, following a strict randomization protocol, on the selected plot in each household.¹⁷

3.1. Survey Treatments

CVIP implemented two diary- and two recall-based survey treatments. The enumerators scheduled visits to the CVIP households in strict observance of the requirements of each survey treatment arm (in terms of timing and frequency).¹⁸ For all questions regarding agriculture and cassava

migration and the respondent died in 2 1-member households. There was no other loss of sampled households due to refusal and all diary households kept their diaries for as long as they had cassava standing on their plots during the fieldwork period.

¹³ The experiment used 15 enumerators, and each enumerator covered 3 EAs. EAs were divided among the staff in such a way that enumerators could reside within one of their EAs or a nearby district center, and bicycle to their assigned households.

¹⁴ A garden (*munda*) is defined as a continuous piece of land that is not split by a river or a path wide enough to fit an ox-cart or vehicle. A garden can be made up of one or more plots. A plot is continuous piece of land on which a unique crop or a mixture of crops is grown, under a uniform, consistent crop management system. It must be a continuous piece of land and must not be split by a path of more than one meter in width. Plot boundaries are defined based on the crops grown and the operator.

¹⁵ After walking the perimeter of a given plot with the plot manager to identify the boundaries, the enumerators re-paced the perimeter and measured the area with the GPS unit.

¹⁶ The allowable units for self-reported plot areas were square meters, acres and hectares. The majority of respondents reported in acres, with approximately 5 percent reporting in hectares or square meters.

¹⁷ Further, the enumerator obtained a leaf sample from each sub-plot for objective variety identification based on DNA fingerprinting. Each leaf sample was associated with a barcode that was scanned by the Survey Solutions CAPI platform to enable the linkage between the lab results and the household survey data. The assessment of the relative accuracy of farmer-reported variety names and attributes, in comparison to DNA fingerprinting-based cassava variety identification, is the subject of a separate research study.

¹⁸ Given the within-EA proximity of CVIP households to one another, it is possible that the enumerators and the respondents may have seen one another on a more frequent basis than the intended research design. For instance, when an enumerator performed his/her semi-weekly visits to the D1 households or was living day-to-day life in his/her EA of residence, he/she might have inevitably interacted socially with the other CVIP households nearby. While we have

production, the enumerators attempted to interview the plot manager(s) or household member(s) most knowledgeable about decision-making on the gardens and plots owned and cultivated by the household. Table 1 provides an overview of the incentives that were provided to the CVIP households and that are detailed below.

3.1.1. Diary Arms

The first group of households (D1) were given a diary to record cassava harvest on a daily-basis, across a 12-month period, under the supervision of an enumerator that visited them twice a week. Harvest diaries were also filled by the plot manager(s). If any plot manager was illiterate, the enumerator assisted the respondent in identifying another literate household member to assist them in weighing the cassava and filling the diary. The enumerator visits were meant to encourage good record keeping; record information on harvests that had been omitted by the respondent; field questions on the specifics of diary keeping; assist with the weighing of the fresh harvest if it coincided with the visit; and solicit information on the likely timeline for crop cutting on the selected cassava plot. To facilitate D1 household diary keeping, each respondent was provided a total cash incentive of approximately US\$7, distributed in two installments.¹⁹

The second group of households (D2) were too given a diary to record their cassava harvest on a daily-basis for 12 months. However, instead of in-person monitoring visits, the D2 households received calls on their mobile phones twice a week. The inclusion of this treatment arm was intended to assess the feasibility and accuracy of using mobile phones to collection production data on a frequent basis as a potential cost-saving alternative to avoid having to pay resident enumerators to perform in-person visits to households. Additionally, it allowed for centralized supervision with upper management able to check on the quality of the call center operators' performance at any time. The phone calls were initiated from a call center that was set up at the NSO Headquarters in Zomba and that was supervised daily by the NSO CVIP fieldwork coordinator. The initial visit to all households within the first weeks of fieldwork served as the baseline survey for the D2 households. At this visit, the enumerators clearly outlined the purpose of the project and the expectations in terms of recording all cassava harvested and the importance of keeping their phone charged to receive the bi-weekly calls. Additionally, enumerators distributed the mobile phones, SIM cards, and solar chargers at this visit and demonstrated how to use the different components.

no data on the extent of these interactions, the enumerators were given a firm instruction to not discuss any aspect of CVIP data collection outside the scheduled visits to the sampled households. Throughout the fieldwork, the CVIP fieldwork coordinator and team leaders followed up with the enumerators regularly also to ensure that the CVIP-related interactions with the households were taking place in accordance with the intended research design.

¹⁹ The first installment was provided when the crop cut sub-plot was laid on the selected plot. The second installment was provided at the end of the survey.

The call center began calling the D2 households within the first 1 to 2 weeks of fieldwork, and the enumerators began bi-weekly visits to the D1 households within the same time frame. Given the relatively small sample size, the length of each phone call (with an average of approximately 5 minutes), and the required frequency of the calls per week, two phone operators ran the call center dividing the D2 households among themselves. The calls were meant to serve the same objectives of in-person visits, and at each call, the operators asked respondents to report to them the last recorded harvest in their diary. The call center carefully tracked which households were harvesting at a given time to ensure that they did not miss the opportunity to remind respondents to fill their data. Each respondent could additionally call his/her assigned enumerator and the associated supervisor, whenever needed. Furthermore, when diaries were collected at the end of each month, the enumerators sat down with the D2 respondents to review their diaries.

Furthermore, since only 42 percent of rural households in Malawi had been estimated at the time of the CVIP design to own a working mobile phone, each D2 household was given a mobile phone with a solar charger.²⁰ This package was valued at approximately US\$29 (US\$16 for phone, and US\$13 for the solar charger). This was a critical decision to incentivize participation and ensured that all households were treated equally irrespective of their mobile phone ownership. The phones distributed to the D2 households were kept as simple as possible to make sure the respondents retained the asset for the duration of the survey while still maintaining adequate battery life and durability. Further, each D2 household was provided a total airtime incentive of approximately US\$7, distributed in monthly installments. These decisions were in line with the recommendations of Dabalen et al. (2016) and were taken to ensure that the phones remained active and that the respondents stayed motivated.

Finally, each D1 and D2 household was provided with sacks and a hanging weighing scale to record the fresh cassava harvest in their diaries in kilograms. During the first interview, the diary respondents were trained in the use of the diary²¹ and the weighing scale. Each scale was valued at approximately US\$10 and was left with the household at the end of the fieldwork period. The sacks provided to the diary households were valued at approximately US\$4. Overall, the total value of the incentive package was approximately US\$20 for the D1 households, and US\$53 for the D2 households, as shown in Table 1.

3.1.2. Recall Arms

The third group of households (R1) were visited twice over the course of the 12-month fieldwork, with 6 months in between visits. At each visit, the R1 households were asked to report cassava

²⁰ According to the authors' calculations based on the Fourth Integrated Household Survey (IHS4) 2016/17, only 11 percent of households in Malawi have access to electricity.

²¹ Enumerators distributed paper production diaries leading up to the start of each calendar month. Following the collection of the paper diaries from the prior month, the supervisors for each district entered the data into the Survey Solutions CAPI platform. A second data entry and verification of the paper diaries took place in September 2016.

production at the plot-level, as in the case of Malawi IHS, but for the past 6 months. The R1 households received their first visit in February 2016 to collect plot-level cassava production information for the period of August 2015–January 2016. The R1 households were visited for a second time in July/August 2016 to collect the comparable harvest information for the period of February 2016–July 2016.

The fourth group of households (R2) were visited once, at the end of the 12-month fieldwork in each EA (i.e. in July/August 2016), and were asked to report fresh cassava production at the plot-level for the past 12 months, in line with the prevailing approach to household and farm surveys collecting information on production of tree and root crops, as in the case of the Malawi IHS.

Similar to the cash/airtime incentives provided to the diary households, each R1 and R2 household was provided a cash incentive of approximately US\$7. Half of the cash incentive was paid at the end of the first visit, while the remaining half was paid in the final visit to the recall households.

3.2. Crop Cutting

Although a well-implemented diary is the gold standard for *household-level production* measurement, crop cutting has been recognized as the gold standard for *plot-level yield* measurement for seasonal crops since the 1950s by the Food and Agriculture Organization of the United Nations (FAO). Besides the cost- and supervision-intensive nature of the exercise, several concerns have been raised regarding the accuracy of the method. Even if one places only one random crop cutting sub-plot within the sampled plot, the resulting yield estimate may carry a sampling error if the yields exhibit within-plot heterogeneity. Further, Fermont and Benson (2011) mention the following additional possible sources of error: (1) More thorough-harvesting of crop cut sub-plots vis-à-vis the typical farmer harvesting practices, (2) possible rounding of crop cut production estimates obtained through scales; (3) using faulty or inappropriate scales; (4) omitting to net out the weight of the measurement container from the measured production; (5) including plants that fall outside of the sub-plot; and (6) non-random placement of crop cut sub-plots. The CVIP crop cutting exercise attempted to overcome these potential concerns through random sub-plot placement; the use of well-monitored weighing scales of acceptable precision; and intensive training of enumerators on the sub-plot placement, harvesting and weighing. In each CVIP household, regardless of the survey treatment, 1 cassava plot was selected at random for crop cutting, and a 5x5m subplot was set up, also at random, in line with the international best practices.²²

Further, the crop cut sub-plot was cordoned off until the respondent was ready to harvest this area of the plot. Each EA had a designated crop cutting assistant that paid periodic visits to the households to check on the integrity of the sub-plots, and to inform the enumerator regarding the

²² Refer to Appendix I for CVIP Crop Cutting Protocol.

expected timing of the sub-plot harvest. There were no instructions given to the households on timing of the crop cut. They were, however, requested to harvest the cassava within the sub-plot whenever *they* were ready and were asked to contact the enumerator and/or the crop cutting assistant beforehand to arrange a time. For districts where households typically harvest their cassava at once, the crop cutting exercise was then done at the time of the main harvest. In districts where cassava is harvested over several weeks or months, the respondents would simply inform the enumerator when they were ready to harvest that piece of land. Together with the enumerator and the crop cutting assistant, the respondents would harvest the 5x5 sub-plot, and weigh the cassava using the same type of weighing scale as distributed to diary households. All crop cutting production was reported in KGs, with details collected on cassava lost prior to harvest and the number of plants harvested.

3.3. Production Measurement and Expected Inter-Relationships

The analysis focuses on household-level annual cassava production computed in kilograms. Since all diary households were provided with a hanging scale, they were expected to record all entries in kilograms on a daily-basis. The production for the diary households is simply the total of the daily records solicited during the 12 months of fieldwork. All diary data was recorded and weighed “fresh”, as respondents weighed the cassava directly after harvesting from the plots. The diary did not collect plot-specific harvest information, but rather information on the total cassava harvest for the household on a specific day.²³

On the recall front, the respondents reported production at the plot-level²⁴; could use standard as well as non-standard measurement units; and could have reported the harvest condition as fresh or dry – following the prevailing practices in household and farm surveys and the approach as part of the Malawi IHS. Consequently, the recall respondents reported production for 74 percent of cassava plots in 50-kilogram bags; 19 percent in kilograms; and 3 percent in 70-kilogram bags. The production on the remaining plots were recorded in pails and pieces (each differentiated as small, medium or large), and in *gondolos* (non-standard, cylindrical baskets that can vary in size and that are often attached to bicycles to carry produce). To convert these non-standard units to kilograms, we used the regional conversion factors readily available from the Malawi IHS

²³ The harvest that was weighed could have also been expressed in terms of non-standard measurement units used in the recall arms, as such the kilogram equivalences obtained for the non-standard measurement units for fresh cassava harvest could have been used as ancillary data for gauging the sensitivity of the recall-based production estimates that would have been reported in non-standard units. In the end, most of the cassava harvests in the diaries ended up being expressed in 50-kilogram bags (alongside their objective weights in kilograms), as the diary households relied on the sacks that were distributed to them.

²⁴ To address any potential for differences arising from the diary reporting at the household-level and recall reporting at the plot-level, the analysis was run on the subsample of households cultivating only one cassava plot and the relationships between the four treatment arms remained the same.

program. In turn, we aggregated the kilogram-equivalent production across the cassava plots to obtain the household-level annual cassava production estimates.

If implemented with the intended level of supervision and respondent participation, diary-keeping is traditionally assumed to be the gold standard in data collection on production (Deininger et al., 2012) and on consumption (Beegle et al., 2012). However, as Beegle et al (2012) highlight, while well-implemented diaries are expected to yield higher and presumably closer to actual levels of consumption, the success of a diary operation is mediated by respondent literacy and motivation, as well as field staff effort, among others.²⁵ Although CVIP attempted to circumvent respondent fatigue through incentives, and relied on highly-trained, -supervised, and -experienced enumerators and call center operators, given the prevalence of diary decay even in 7-day operations, we concede that omission of harvest records must have been unavoidable during the 12-month CVIP fieldwork. Therefore, while diary-keeping is still expected to underestimate *true* production, it is assumed to (i) underestimate to a lesser degree compared to recall, and (ii) offer the best chance to capture annual cassava production as accurately as possible.

In terms of inter-relationships, D1 was set-up to be executed as well as any traditional diary-based field operation but given the long fieldwork duration and the potential respondent burden, imperfect compliance was expected, albeit to a limited degree. However, the D1 versus D2 comparison is not clear *ex-ante*. On the one hand, D2 could underestimate annual cassava production with respect to D1, given the lack of frequent personal interaction with a qualified enumerator, who would nudge for timely recording of harvests; help recall omitted harvests; and field questions. On the other hand, the degree of non-compliance could be more limited among the D2 respondents in comparison to their D1 counterparts, mainly due to the significantly more valuable set of incentives that they received to be part of CVIP.²⁶

In comparison to the recall variants, both diary treatments are expected to reduce two main sources of bias that typically lead to underestimation of annual cassava production. First, since the harvest is captured immediately upon harvest, recall decay should be minimized. Second, diaries should reduce the scope for respondent bias that may arise with recall. Recall-based interviews are often conducted in a specific time frame that revolves less around the schedule of the ideal respondents (i.e. the plot managers) resulting in a greater potential for proxy reporting. Lastly, within the recall domain, it is possible that R1 reduces the scope for recall decay to a greater degree compared to R2, since the R1 respondents are asked to recall over 6, as opposed to 12, months.

²⁵ A specific example of an unsuccessful diary effort is in fact from Malawi, with 40 percent of households in the First Integrated Household Survey (IHS1) 1997/98 having been judged to have incomplete or unreliable expenditure data due to poor diary-keeping.

²⁶ This claim needs to be supported with the available empirical evidence on the impact of gifting on survey data quality.

3.4. Yield Measurement and Expected Inter-Relationships

The annual household-level cassava yield estimates were obtained by dividing household-level annual cassava production (in kilograms), as computed above, with the household-level GPS-based land area cultivated with cassava (in hectares), across all cassava plots that were cultivated by the household and measured with GPS.

Since each household also had a crop cutting sub-plot set up on a randomly selected cassava plot, we took the measured production in kilograms within a 5x5m area; extrapolated it over the total GPS-based land area cultivated with cassava; and divided it by this area measure to obtain the household-level crop cutting-based yield estimate.

In conceptualizing what various yield measures capture, the framework presented by Fermont and Benson (2011) is useful. The authors provide the definitions for *potential yield*; *attainable yield*; *economic yield*; and *actual yield*, as understood by economists; and the definitions for *harvested yield* and *economic* (renamed as *net*) *yield*; as understood by sociologists.

The readapted definitions for the first set are as follows:

1. *Potential yield* is the highest yield for a crop genotype in a specific climatic environment, obtained based on modeling.
2. *Attainable yield* is the yield under the full use of available technology obtained on research stations or as maximum levels from on-farm trials.
3. *Economic yield* is the yield that provides highest returns to investment in addressing biotic and abiotic stress, obtained on on-farm input trials or through economic modeling.
4. *Actual yield* is the yield under the partial use of available technology, based on sub-plot crop cutting.

The readapted definitions for the second set are as follows:

5. *Harvested yield* is the actual yield as defined above net of harvest losses, based on entire field harvests.
6. *Net yield* is the harvested yield net of post-harvest losses during cleaning, threshing, winnowing, and drying, but not inclusive of potential storage losses, based on farmer recall or total harvest weighing.

According to Fermont and Benson (2011), the average yield under these 6 definitions would be in descending order, with *potential yield* as defined by economists being the highest, and *net yield* as defined by sociologists being the lowest.

In view of these definitions, the timing of cassava sub-plot harvests strictly being a function of household preference and the extrapolation of each household's sub-plot harvest over the entire farm area cultivated with cassava, including portions that may have been harvested following a shorter growing period or may go unharvested depending on the household need, the average CVIP crop cutting based yield (i) should be understood as an *upper-bound* for the cassava yield realized on the farm during the 12 months of interest, falling between *economic* and *actual* yields, as defined by economists, and (ii) should be higher, by design, than the average diary and recall yields. Further, the CVIP diary-based yields align more closely with the definition of *harvested yield*, while the recall-based yields are in line with the definition of *net yield*. As such, the CVIP diary-based yields are expected to be higher than their recall-based counterparts to a limited extent by design, beyond the potential under-estimation of production under recall.

3.5. Descriptive Statistics

Before proceeding to the analysis, we must confirm that the randomization of households was successful. Table 2 provides sample means on a core set of household and plot attributes across treatment arms and districts. Taking the overall sample from all 4 treatment arms, only 7 of 40 pairwise comparisons across the attributes between the treatment arms are statistically significantly different from one another at least at the 10 percent level. While the descriptive evidence largely supports the hypothesis that there are no systematic differences between the samples selected for each treatment arm, the regression analysis will nevertheless control for an extensive set of observable attributes that may be correlated with any remaining unobserved heterogeneity that may jointly determine the outcomes and the treatment arm assignment.

Tables 2 and 3 provides empirical evidence to accompany the qualitative discussion of the Malawi context presented in section 2. Overall, on average, households cultivate 1.38 plots with cassava and plots are 0.25 acres. In Nkhatabay, the northernmost CVIP district, the averages for total land area cultivated with cassava and number of cassava plots stand at .31 acres and 2.1 plots, respectively, and are higher than all other districts.²⁷ Patterns of sales versus consumption are also evident across the 5 districts. In terms of the number of households that sell any of their harvested cassava, in Nkhatabay only 12 percent of households are involved in sales as in this area cassava is considered a staple crop for home consumption. In a geographically-close district, Nkhotakota, 20 percent of households sell at least some of their cassava, but this is not significantly different from Nkhatabay, as expected. In Southern Malawi, the rate of sales in Mulanje is not much higher at 22 percent but in Zomba 40 percent of households sell cassava, and in Lilongwe almost all cassava is intended for sales with 96 percent of households selling cassava.

²⁷ These differences are statistically significant at the 1 percent level. While not reported here, of the 1,218 households that inform our analysis, 74 percent had 1 cassava plot, 19 percent had 2 cassava plots, 5 percent had 3 cassava plots, and only 2 percent had 4 or more cassava plots. The comparisons of these shares among the treatment arms do not yield statistically significant differences.

Further, Table 4 presents the breakdown of crop cutting subplots harvested by month and district. Additionally, the average weight (in kilograms) of the cassava harvested from the 5x5 sub-plot is shown. The three Southernmost districts completed most of their crop cutting activities within the first seven months of fieldwork, whereas in Nkhatabay and Nkhotakota, the two districts falling within the cassava belt, the crop cuts were spread out over the course of the 12 months.²⁸ The spatial differences in the timing of the crop cuts were tied to the spatial differences in the objectives of cassava cultivation as well as the cassava varieties that were grown. The weight of the cassava harvested from the subplots was much higher in Nkhatabay, Nkhotakota, and Lilongwe compared to Mulanje and Zomba due to both the varieties cultivated as well as the planting pattern. Referring to Table 3, in Nkhatabay, an average of 46 cassava plants were harvested from the 5x5 sub-plot versus 6 in Mulanje. This is also in line with the fact that the share of cassava plots that were intercropped stood at 39 percent in Nkhatabay versus 96 percent in Mulanje.

Table 5 shows that the average GPS-based area measurement is 0.18 hectares, while the comparable figure based on farmer reporting is 0.36. The difference represents 100 percent of the GPS-based average.²⁹ To delve further into understanding this large gap, Figure 2 shows the mean difference by GPS-based plot area decile. While there is a downward trend in the extent of difference across the plot area deciles, the levels are considerable across the distribution. These findings are in line with those reported by Carletto et al. (2017), based on the analysis of the data from methodological survey experiments in Ethiopia, Nigeria and Zanzibar, as well as the secondary data collected as part of the World Bank Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) initiative.

4. Empirical Analysis

This section provides a description of the empirical framework for estimating the relative survey treatment effects that CVIP was designed to isolate. Within-EA randomization of the systematically sampled households across treatment arms allows us to estimate the causal effects associated with each treatment arm. The three core specifications are as follows:

$$(1) \text{Production}_i = \alpha_1 + \beta_1 D2_i + \beta_2 R1_i + \beta_3 R2_i + \gamma C + \varepsilon_i$$

$$(2) \text{Yield}_i = \alpha_2 + \beta_1 D2_i + \beta_2 R1_i + \beta_3 R2_i + \gamma C + \varepsilon_i$$

$$(3) \text{Yield}_i = \alpha_3 + \beta_1 D1_i + \beta_2 D2_i + \beta_3 R1_i + \beta_4 R2_i + \gamma C + \varepsilon_i$$

²⁸ In both districts, some households had to be asked to harvest their crop cut subplots during the final visit in the twelfth month of fieldwork, despite the households not expressing a preference for harvest.

²⁹ Though not reported, the difference does not exhibit statistically significant variation by survey treatment - another indication that randomization was successful.

where i represents household; α and ε represent constant and error terms, respectively; and D1, D2, R1 and R2 are identifiers representing a household's assignment to diary-visit, diary-phone, 6-month recall, and 12-month recall, respectively. D1 is the comparison category in Equations 1 and 2, and crop cutting is the comparison category in Equation 3. C is a vector of household attributes that is included with the intention of capturing any remaining unobserved heterogeneity that may be correlated with these controls and that may jointly determine both the dependent variables and household survey treatment assignment.^{30, 31}

As described in Section 3, annual cassava production (kilograms) and cassava yield (kilograms/hectare) are the dependent variables of interest. To limit the influence of potential outliers that were not deemed agronomically plausible upon extensive review of the data, both production and yield estimates were winsorized at the 95th percentile. The sample includes 1,218 households for Equations 1 and 2. To estimate Equation 3, we append the household-level crop cutting-based yield estimates available for 1,123 households onto the data set used for the estimation of Equations 1 and 2. As such, the standard errors are clustered at the household-level for Equation 3, and at the EA-level for Equations 1 and 2. The results from each regression are coupled with the full spectrum of tests of equality of coefficients for complete inter-arm comparisons.

5. Results

Panels A, B, and C of Table 6 report the results from the estimations of Equations 1, 2 and 3, respectively.³² If true cassava production is assumed to be underestimated even by diary-keeping due to imperfect compliance, we find that compared to D1, the annual household cassava production was 295 kilograms higher, on average, under D2, corresponding to 28 percent of the

³⁰ The control variables include: the binary variable identifying whether the household head is female; the household head age in years; the highest grade in school in the household; the binary variable identifying any household non-farm employment in the last 12 months; the binary variable identifying any household paid employment in the last 12 months; household size; household dependency ratio; a principal components analysis based wealth index of dwelling attributes and ownership of consumer durables; the binary variables identifying district residence, with Nkhatabay being the comparison category. The plot-level variables were aggregated at the household-level and are represented as follows: total land area cultivated with cassava in acres; the binary variable identifying whether the household has sold any cassava harvested in the last 12 months. The findings are robust to excluding the control variables from the regressions - another indication that the randomization of households across treatment arms worked according to plan.

³¹ Heterogeneity of impact across different treatment arms was explored through the expansion of equations 1-3 to include interaction terms between survey treatment identifiers and selected control variables, including the length of the harvest period, the involvement in cassava sales, the land area under cassava cultivation, and the highest education level in the household. None of the interaction terms was associated with coefficients that were statistically significant at least at the 10 percent level.

³² Figure 3 provides a complementary, distributional comparison of annual household cassava production by survey treatment.

D1 mean. On the assumptions that the D2 respondents do not misreport daily entries as higher than the cassava they weighed, and that they do not record entries on days that they did not harvest any cassava, the D2-based estimates are taken to be closer to the true production value, despite the commonly-held beliefs around the importance of personal supervision of respondent-kept production diaries.³³ These assumptions are not unreasonable since the enumerators were instructed to make sure that both D1 and D2 respondents understood that they would be receiving the incentives irrespective of how much cassava production is reported on their diaries; that the research team had no expectations whatsoever regarding how much cassava they will produce; and that the research team were interested in capturing their harvests exactly in accordance with the actual events on the farm. These messages were communicated at the initial visit to both diary arms and were emphasized during the semi-weekly visits to the D1 households and the monthly visits to the D2 households for picking up the diaries. These communication efforts intended to eliminate the possibility of social desirability bias that may otherwise have resulted in the diary households recording fake harvests or inflating actual harvest figures.

The monthly production comparisons between the diary arms, shown in Table 8, provide further support the hypothesis that the D2 respondents did a more complete job in reporting cassava harvests throughout the fieldwork period, and particularly during the lean season months of January, February and March, when cassava would be relied upon as a food security measure. Moreover, while the number of D1 and D2 households recording any harvest in the diaries was similar over the first half of the fieldwork, we see fewer D1 households with diary entries in the second half of the fieldwork. Considering the evidence for successful randomization into the survey treatment arms, the evidence points to the higher likelihood of D1 households missing cassava harvest entries in their diaries. While not empirically testable, the significantly more valuable set of incentives that were received by the D2 households (Table 1) may have resulted in better record keeping in comparison to their D1 counterparts.

Coming back to Panel A of Table 6, while R1- and D1-based estimates of production are statistically indistinguishable at the mean, R2, the most common approach to cassava production measurement, underestimates annual production, on average, by 221 kilograms compared to D1, corresponding to 21 percent of the D1 mean. Both recall variants underestimate annual production by a significant margin relative to D2 – though R1 does so to a lesser extent. Furthermore, given the lack of variation in the total household area cultivated by cassava across the survey treatment arms, the yield comparisons among the diary and recall arms presented in Panel B mirror the results emerging from Panel A.

³³ Table 7 breaks down the average number of visits to the diary households by month. Within the D1 domain, there is no evidence of decay in the enumerator visits to the households. Within the D2 domain, apart from the limited follow-up visits in the first quarter to answer specific questions that the respondents approached the enumerators with, the replacement of in-person supervision with phone supervision went according to plan – certainly in the last half of the fieldwork, during which we in fact see the significant reporting differences between the diary arms.

Finally, the average household-level crop cutting-based annual cassava yield estimate stands at 8,958 kg/ha across all sample households (Panel C), while the comparable D1, D2, R1 and R2 estimates are 5,208; 6,218; 5,798; and 4,671, respectively (Panel B). The estimation of Equation 3, as reported in Panel C of Table 6, implies that the diary and recall survey treatments, on average, underestimate cassava yield compared to crop cutting, as expected given the reasoning provided in section 3.4. The extent of foreknown underestimation, as a percentage of the crop cutting based yield, stands at 25 percent under D2, 40 percent under D1, 33 percent under R1 and 47 percent under R2. The inter-relationships between the 4 core treatment arms are the same as those reported in Panel B.

Nevertheless, the average crop cutting-based yield estimate, which should be understood as the *upper-bound* for the cassava yield realized on the farm during the 12 months of interest, is still significantly lower than the FAOSTAT estimates shown in Figure 1, and the MoAIWD APES-based 2015/16 national cassava yield estimate of 17,564 kgs/ha.³⁴ The latest (2014) FAOSTAT estimate of national cassava yield in fact stands at 24,800 kgs/ha – nearly three times the size of the CVIP crop cutting-based yield estimate, and four times the size of the D2-based counterpart.³⁵

6. Conclusions

In household and farm surveys across the low- and middle-income countries, the prevailing (and most cost-effective) approach to data collection on the production of root, tuber and tree crops is to field a single visit and collect information either specific to an agricultural season or with a 12-month recall. In turn, there is worry that the existing survey methods may elicit unreliable and incomplete information on harvests cutting across seasons, taking place over extended periods, and varying by crop variety.

This paper presents the results of a household survey experiment that was implemented in top cassava-producing districts in Malawi and that randomly assigned sampled households to one of four approaches to cassava production measurement over 12 months, namely (1) daily diary-keeping at the household-level for 12-months, with semi-weekly in-person supervision visits (D1) – the traditional gold standard; (2) daily diary-keeping at the household-level for 12 months, with semi-weekly supervisory mobile phone calls (D2); (3) a two-visit recall-based data collection at the plot-level for a 6-month reference period, carried out in two visits 6 months apart during a 12-

³⁴ The MoAIWD APES-based cassava statistics are taken at face value most recently by Kanyamuka et al. (2018) who study the cassava value chain in Malawi and provide recommendations for its improvement. However, the concerns regarding the gross overestimation in cassava production and yields reported by the MoAIWD were in fact raised nearly two decades ago by Russel Freeman in a report that was prepared in 2000 for the International Monetary Fund.

³⁵ Similar discrepancies between crop cutting and FAOSTAT-based estimates are noted by Gourlay et al. (2017) in the case of maize in Uganda.

month period (R1); and (4) a recall-based data collection at the plot-level for a 12-month reference period, administered in a single visit (R2).

The paper provides evidence for the survey practitioners to move away from R2, given the significant underestimation of production and yield estimates, and asserts that the D2 households provided the most accurate information on annual cassava production. The latter is shown to be ensured through sustained participation in diary-keeping, which could have been connected to the relatively more valuable set of in-kind incentives these households received to be part of the experiment – though this hypothesis is not testable. However, as shown in Table 9, D2, while being significantly cheaper compared to the traditional diary operation, it is still more resource- and supervision-intensive compared to the recall variants.³⁶ The D2 unit cost per household is US\$330, compared to US\$469 for D1, US\$186 for R2, and US\$157 for R1.³⁷

Thus, even if D2 may not always be an option for the survey implementing agencies, it would be more feasible if packaged within a broader effort to collect more frequent data through mobile phone calls on a wider range of topics, including but not limited to extended-harvest crop production, following baseline face-to-face interviews. In the case of Malawi NSO, an effort such as the existing Listening to Malawi (L2M) mobile phone survey, as part of the regional Listening to Africa initiative (Dabalén et al., 2016), could provide such platform. In the worst-case scenario, the analysis indicates that R1 is a second-best alternative to D2 and a clear improvement over R2. R1, at the mean, also performs as well as the traditional diary operation, and comes with less than US\$30 additional cost per household compared to R2.³⁸ On the whole, given the contribution of cassava farming to food security and agricultural commercialization, the evidence underscores the need for adopting improved survey methods to collect better data on cassava production and productivity, facilitating a renewed look at the role of cassava farming in production and welfare outcomes.

³⁶ In assessing the CVIP costs by treatment arm, Table 9 breaks differentiates between two broad categories: (1) Fixed costs applied to all treatment arms i.e. training, listing, purchasing of core survey equipment; and (2) Variable costs across the four treatment arms including main fieldwork costs i.e. staff time and fuel; incentives provided to the households; call center set-up and maintenance costs. Furthermore, although CVIP was implemented using the World Bank Survey Solutions CAPI software, the costs of the Android tablets were not considered in Table 9, since these tablets had been readily available at the NSO. Likewise, only maintenance costs were considered regarding the bicycles used by the enumerators and the motorcycles used by the supervisors, as these items too had been available at the NSO.

³⁷ The D2 costs were exacerbated, on one hand, by the relatively low number of households requiring semi-weekly calls; the purchase of mobile phones and solar chargers (due to poor electricity access) for all D2 respondents, and were alleviated, on the other hand, given the small number of call center operators and the NSO experience with the Listening to Malawi (L2M) mobile phone survey. In preparation for the CVIP fieldwork, the NSO management worked closely with the two main carriers in Malawi to determine the most cost-effective airtime bundle. An additional cost saving mechanism would have been to provide mobile phones to only those households not currently owning a working device, which may be much more feasible in other settings for low- and middle-income countries.

³⁸ For information on how D2 or R1 can be operationalized in the context of existing household survey types in low- and middle-income countries, Moylan et al. (Forthcoming) can be consulted.

ACKNOWLEDGMENTS

The funding for the design, implementation and analysis of CVIP: Methodological Experiment on Measuring Cassava Production, Productivity, and Variety Identification originated from (1) the World Bank Living Standards Measurement Study (LSMS) “Minding the (Agricultural) Data Gap” methodological research program, funded by UK Aid; (2) the Global Strategy for Improving Agricultural and Rural Statistics, led by the Food and Agriculture Organization of the United Nations; and (3) the CGIAR Standing Panel on Impact Assessment (SPIA). The authors would like to thank Bob Baulch and Sydney Gourlay for their comments on the earlier versions of this paper.

REFERENCES

- Backiny-Yetna, P., Steele, D., and Yacoubou Djima, I. (2017). “The impact of household food consumption data collection methods on poverty and inequality measures in Niger.” *Food Policy*, 72, pp. 7-19.
- Beegle, K., De Weerd, J., Friedman, J., and Gibson, J. (2012). “Methods of household consumption measurement through surveys: experimental results from Tanzania.” *Journal of Development Economics*, 98, pp. 3-18.
- Brzozowski, M., Crossley, T. F., and Winter, J. K. (2017). “A comparison of recall and diary food expenditure data.” *Food Policy*, 72, pp. 53-61
- Carletto, G., Gourlay, S., Murray, S. and Zezza, A. (2017). “Cheaper, faster and more than good enough: is GPS the new gold standard in land area measurement?” *Survey Research Methods*, 11.3, pp. 235-265.
- Carletto, C., Corral, P., and Guelfi, A. “Agricultural commercialization and nutrition revisited: empirical evidence from three African countries.” *Food Policy*, 67, pp. 106-118.
- Dabalen, A., Etang, A., Hoozeveer, J., Mushi, E., Schipper, Y., and von Engelhardt, J. (2016). “Mobile phone panel surveys in developing countries: a practical guide for microdata collection.” *Directions in Development*. Washington, DC: World Bank.
- Davis, B., Di Giuseppe, S., and Zezza, A. (2017). “Are African households (not) leaving agriculture? Patterns of households’ income sources in rural Sub-Saharan Africa.” *Food Policy*, 67, pp. 153-174.
- de Nicola, F., and Gine, X. (2014). “How accurate are recall data? evidence from coastal India.” *Journal of Development Economics*, 106, pp. 52-65.
- Deininger, K., Carletto, C., Savastano, S., and Muwonge J. (2012). “Can diaries help in improving agricultural production statistics? evidence from Uganda.” *Journal of Development Economics*, 98, pp. 42-50.
- Desiere, S. and D. Jolliffe. (2018). “Land productivity and plot size: is measurement error driving the inverse relationship?” *Journal of Development Economics*.
- Dorosh, P., and Thurlow, J. (2016). “Beyond agriculture *versus* non-agriculture: decomposing sectoral growth-poverty linkages in five African countries.” *World Development*.
- Fermont, A., and Benson, T. (2011). “Estimating yield of food crops grown by smallholder farmers: a review in the Uganda context.” International Food Policy Research Institute Discussion Paper No. 1097.
- Food and Agriculture Organization of the United Nations (FAO) (2014). “Analysis of public expenditure in support of food and agriculture in Malawi, 2006-2013.” Lilongwe: FAO.
- Friedman, J., Beegle, K., De Weerd, J., and Gibson, J. (2017). “Decomposing response error in food consumption measurement: implications for survey design from a randomized survey experiment in Tanzania.” *Food Policy*, 72, pp. 94-111.

- Gourlay, S., Kilic, T., and 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 No. 8192.
- Kabambe, V. H. (2011). “Guide cassava production and utilisation in Malawi.” Background paper prepared for the Trustees of Agricultural Promotion Programme.
- Kambewa, P., and Nyembe, M. (2008). “Structure and dynamics of Malawi cassava markets.” Background paper prepared for the Cassava Transformation in Southern Africa Start Up Project, Michigan State University.
- Kanyamuka, J. S., Dzanja, J. K., and Nankhuni, F. J. (2018). “Analysis of the value chains for root and tuber crops in Malawi: the case of cassava.” Feed the Future Innovation Lab for Food Security Policy Research Brief No. 65, Michigan State University. Retrieved from <https://goo.gl/HSy8ih>.
- Malawi Ministry of Agriculture, Irrigation and Water Development (MoAIWD). (2016). *National agriculture policy*. Lilongwe, Malawi: MoAIWD.
- McCarthy, N., Kilic, T., de la Fuente, A., and Brubaker, J. (2017). “Shelter from the storm? household-level impacts of, and responses to, the 2015 floods in Malawi.” World Bank Policy Research Working Paper No. 8189.
- Moylan, H., Kilic, T., Carletto, C., and Zezza, A. (Forthcoming). “Cassava Production and Productivity Measurement in Household Surveys: A Guidebook for Improving Household Survey Data Collection.” Washington, DC: World Bank.
- Moyo et al. (2004). “Cassava and sweet potato yield assessment in Malawi.” *African Crop Science Journal*, 12.3, pp. 295-303.
- National Statistical Office (NSO) (2017). “Fourth Integrated Household Survey (IHS4) 2016/17 survey report.” Zomba, Malawi: NSO.
- Troubat, N., and Grunberger, K. (2017). “Impact of survey design in the estimation of habitual food consumption. A study based on urban households of Mongolia.” *Food Policy*, 72, pp. 132-145.

TABLES

Table 1. Incentives Provided to CVIP Households by Treatment Arm

Incentive	MWK	USD	Diary - Visit (D1)	Diary - Phone (D2)	6-Month Recall (R1)	12-Month Recall (R2)
Weighing Scale	6,780	9.69	X	X		
Sacks	2,544	3.63	X	X		
Mobile Phone	11,000	15.71		X		
Solar Charger	9,000	12.86		X		
Airtime	5,000	7.14		X		
Cash	5,000	7.14	X	X*	X	X
Total Incentive (USD)			20.46	52.61	7.14	7.14
Observations			305	307	304	302

Note: *Cash disbursement to each D2 household was MWK 2,500.

Table 2. Sample Means by Treatment Arm

	Overall	Diary - Visit (D1)	Diary - Phone (D2)	6-Month Recall (R1)	12-Month Recall (R2)	Results from Tests of Mean Differences			
						D1 vs. R1	D1 vs. R2	D2 vs. R1	D2 vs. R2
# of Cassava Plots	1.38	1.37	1.39	1.36	1.38				
Share of Cassava Plots Intercropped	0.5	0.48	0.52	0.52	0.48				
Total GPS-Based Land Area Cultivated with Cassava (Ha)	0.25	0.25	0.26	0.24	0.24				
Total Farmer-Reported Land Area Cultivated with Cassava (Ha)	0.49	0.51	0.53	0.45	0.49			*	
GPS-Based Land Area for Cassava Plot Selected for Crop Cutting (Ha)	0.19	0.19	0.21	0.20	0.18				
Farmer-Reported Land Area for Cassava Plot Selected for Crop Cutting (Ha)	0.41	0.39	0.39	0.35	0.37				
Household Sold Any Cassava†	0.38	0.39	0.37	0.38	0.38				
Household Size	5.63	5.70	5.50	5.80	5.52				
Dependency Ratio	1.5	1.42	1.56	1.56	1.45				
Highest Years of Education	8.32	8.31	8.25	8.23	8.48				
Head of Household: Female †	0.28	0.27	0.25	0.31	0.29				
Head of Household: Age (Years)	45.95	45.64	47.50	45.95	44.71				**
Wealth Index	0.07	0.07	0.08	0.07	0.07			**	**
Household Has Any Self-Employment Income †	0.52	0.57	0.60	0.47	0.44	***	***	***	***
Household Has Any Wage-Employment Income †	1.65	1.63	1.63	1.65	1.68				
Observations	1218	305	307	304	302				

Table 3. Sample Means by District

	Overall	Nkhatabay	Nkhotakota	Lilongwe	Zomba	Mulanje
Total GPS-Based Land Area Cultivated with Cassava (Ha)	0.25	0.31	0.2 ***	0.28	0.2 ***	0.23 ***
# of Cassava Plots	1.4	2.1	1.2 ***	1.7 ***	1.2 ***	1.2 ***
% of Cassava Plots Intercropped	50	39	23 ***	7 ***	81 ***	96 ***
Household Sold Any Cassava†	38	12	20	96 ***	40 ***	22 *
% of Diary-Based Production Allocated to Consumption	56	94	87 ***	9 ***	61 ***	27 ***
Length of Harvest Period (Months)	3.5	6.5	5.8 ***	1.3 ***	2.1 ***	1.7 ***
Harvested # of Cassava Plants in Crop Cut Sub-Plot	26	46	35 ***	31 ***	12 ***	6 ***
Observations	1218	245	243	233	248	249

Table 4. Distribution of Household Crop Cuts by Month & District

	Overall			Nkhatabay			Nkhotakota			Lilongwe			Zomba			Mulanje		
	Obs	%	Monthly Average‡	Obs	%	Monthly Average‡	Obs	%	Monthly Average‡	Obs	%	Monthly Average‡	Obs	%	Monthly Average‡	Obs	%	Monthly Average‡
August	110	0.10	8	0	0.00	0	1	0.00	8	14	0.06	22	61	0.27	5	34	0.14	9
September	298	0.27	12	6	0.03	33	17	0.08	27	37	0.16	29	70	0.30	3	168	0.68	9
October	116	0.10	20	6	0.03	33	17	0.08	29	53	0.23	27	10	0.04	3	30	0.12	6
November	123	0.11	28	17	0.08	30	16	0.08	35	75	0.32	30	5	0.02	2	10	0.04	6
December	94	0.08	31	21	0.10	35	19	0.09	32	51	0.22	32	2	0.01	1	1	0.00	10
January	75	0.07	22	25	0.12	26	23	0.11	30	1	0.00	28	26	0.11	10		0.00	---
February	88	0.08	22	27	0.13	30	24	0.12	35	0	0.00	---	36	0.16	7		0.00	---
March	39	0.03	25	15	0.07	35	16	0.08	25	0	0.00	---	7	0.03	8	1	0.00	3
April	38	0.03	31	14	0.07	40	19	0.09	30	0	0.00	---	5	0.02	6		0.00	---
May	36	0.03	45	22	0.11	53	12	0.06	37	0	0.00	---	2	0.01	9		0.00	---
June	30	0.03	46	16	0.08	33	12	0.06	68	0	0.00	---	1	0.00	15	1	0.00	15
July	76	0.07	47	38	0.18	49	31	0.15	54	0	0.00	---	5	0.02	5	2	0.01	29
Observations	1,123			207			208			231			230			247		

Table 5. Mean & Distributional Differences in Farmer-Reported vs. GPS-Based Cassava Plot Areas
Pooled Cassava Plot Across Treatment Arms (N =1817)

Farmer-Reported Plot Area (Ha)	GPS-Based Plot Area (Ha)	Difference as a % of GPS Mean	P-Value – Mean Difference	P-Value - Distributional Differences
0.36	0.18	100%	0.00	0.00

Table 6. Selected Coefficients from Production and Yield Regressions

	Panel A		Panel B		Panel C	
	Total Production (Kg) Mean	Coefficient‡	Yield (Kg/Ha, GPS) Mean	Coefficient‡	Yield (Kg/Ha, GPS) Mean	Coefficient#
Diary - Visit †	1,072	N/A	5,208	N/A	5,208	-3582*** (507)
Diary - Phone †	1,391	295*** (80)	6,618	1431*** (430)	7,717	-2211*** (591)
6-Month Recall †	1,102	37 (68)	5,798	561 (400)	5,798	-2990*** (434)
12-Month Recall †	844	-221*** (61)	4,671	-617*** (337)	4,671	-4187*** (444)
Comparison Category	Diary - Visit †		Diary - Visit †		Crop Cutting	
Comparison Category Mean	1,072		5,208		8,958	
Controls Included?	YES		YES		YES	
Observations	1,218		1,218		2,345	
R2	0.45		0.36		0.44	
Tests of Equality of Coefficients						
D1 = D2	--		--		0.00	
D1 = R1	--		--		0.35	
D1 = R2	--		--		0.03	
D2 = R1	0.00		0.04		0.06	
D2 = R2	0.00		0.00		0.00	
R1 = R2	0.00		0.01		0.01	

Notes: † denotes a dummy variable. Constant estimated but not reported. ***/**/* denote statistical significance at the 1/5/10 percent level, respectively. ‡ denotes standard errors clustered at the enumeration area-level. # denotes standard errors clustered at the household-level.

Table 7. Average In-person Visits to Diary Households

Month	Diary-Visit (D1)		Diary-Phone (D2)	
	No of eligible HHs*	Avg. # of Visits to HH	No of eligible HHs*	Avg. # of Visits to HH
August	305	2.6	307	1.7
September	279	5.4	286	1.9
October	215	6.6	235	1.3
November	181	5.7	198	0.4
December	159	4.1	170	0.4
January	135	5.1	142	0.5
February	125	4.6	129	0.2
March	111	5.3	120	0.5
April	94	4.6	111	0.5
May	82	3.9	99	0.7
June	72	3.2	89	0.1
July	50		58	
Observations	305	37.2	307	6.6

* Harvest not yet completed

Table 8. Average Cassava Production by Month & Survey Treatment

	Diary - Visit (D1)			Diary - Phone (D2)					6-Month Recall (R1)				
	HHs Reporting Any Harvest Observations	%	Monthly Average‡	Semi-Annual Average#	HHs Reporting Any Harvest Observations	%	Monthly Average‡	Difference Significant?	Semi-Annual Average#	Test of Mean Difference wrt D1	Semi-Annual Average#	Test of Mean Difference wrt D1	Test of Mean Difference wrt D2
August	23	0.08	307	706	17	0.06	295	0.910	603	0.285	584	0.098	0.795
September	145	0.53	317		138	0.50	305	0.830					
October	113	0.41	247		117	0.42	311	0.344					
November	90	0.33	333		103	0.37	408	0.274					
December	79	0.29	407		91	0.33	462	0.587					
January	55	0.20	221	69	0.25	309	0.067	977	0.000	586	0.463	0.000	
February	63	0.23	215	80	0.29	281	0.021						
March	56	0.20	195	69	0.25	295	0.018						
April	48	0.17	202	59	0.21	255	0.146						
May	31	0.11	219	41	0.15	310	0.036						
June	32	0.12	258	46	0.17	242	0.682						
July	21	0.08	343	31	0.11	297	0.667						
Observations	259			262					265				

Note: ‡ For a given month, the average is computed only based on households reporting positive harvest. # Semi-annual averages are not conditional on households reporting positive harvest.

Table 9. Implementation Cost Estimates (in USD) by Treatment Arm

	Diary - Visit	Diary - Phone	6-Month Recall	12-Month Recall
Fixed Fieldwork Costs				
Survey & Questionnaire Design	667	667	667	667
Equipment	7,496	7,496	7,496	7,496
Household Listing	7,069	7,069	7,069	7,069
Training	11,527	11,527	11,527	11,527
Crop Cut Monitors & Assistants	11,303	11,303	11,303	11,303
Variable Fieldwork Costs				
Per Diems & Salaries for Field Staff	60,429	24,429	10,072	5,036
Per Diems & Salaries for Management	11,914	4,816	1,986	993
Fuel & Maintenance for Field Teams	5,592			
Fuel & Maintenance for Management	19,965	8,071	3,328	1,664
Fuel to Travel to field	1,888	4,879	2,252	1,126
Airtime for Field Staff	2,228	901	371	186
Airtime for Management	1,308	529	218	109
Incentives				
Weighing Scale	3,051	3,051		
Sacks	1,145	1,145		
Mobile Phone		4,950		
Solar Charger		4,050		
Airtime		2,160		
Cash	2,250	1,125	2,250	2,250
Call Center				
Facility/ Equipment		36		
Staff		2,058		
Airtime		3,771		
Total Cost	147,832	104,033	58,538	49,425
Cost Per Household	469	330	186	157

FIGURES

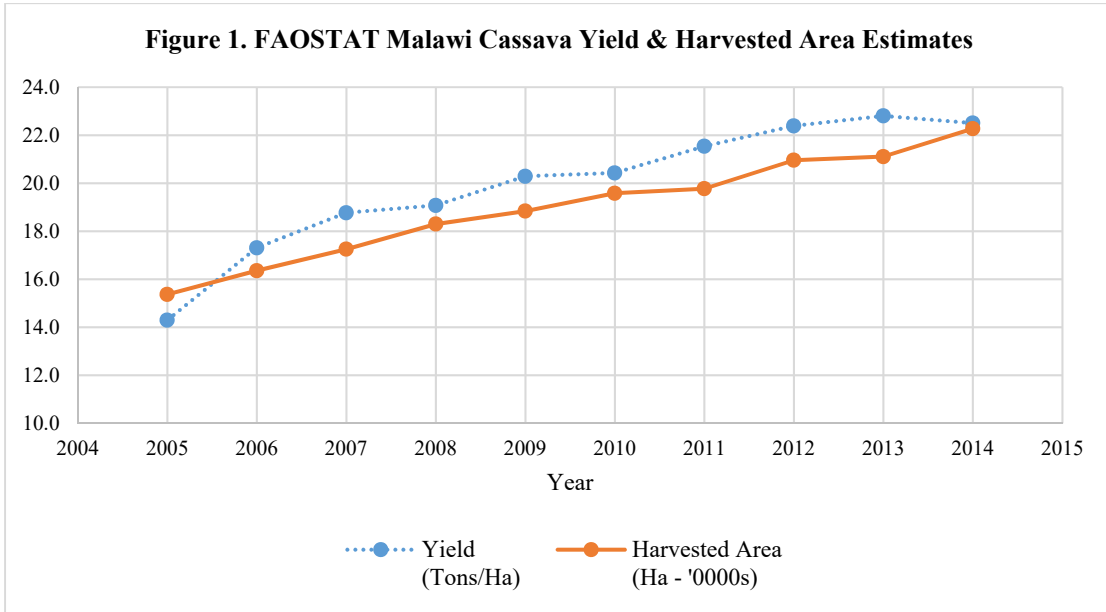


Figure 2. Mean Self-Reported-GPS Plot Area Difference by GPS-Based Plot Area Decile

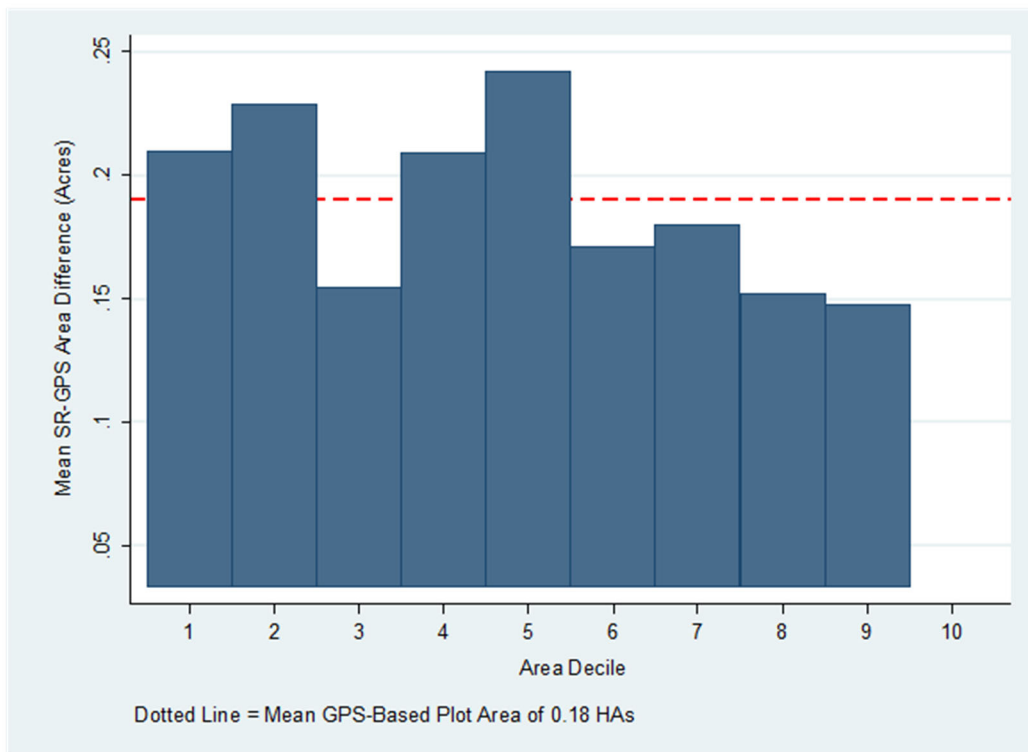
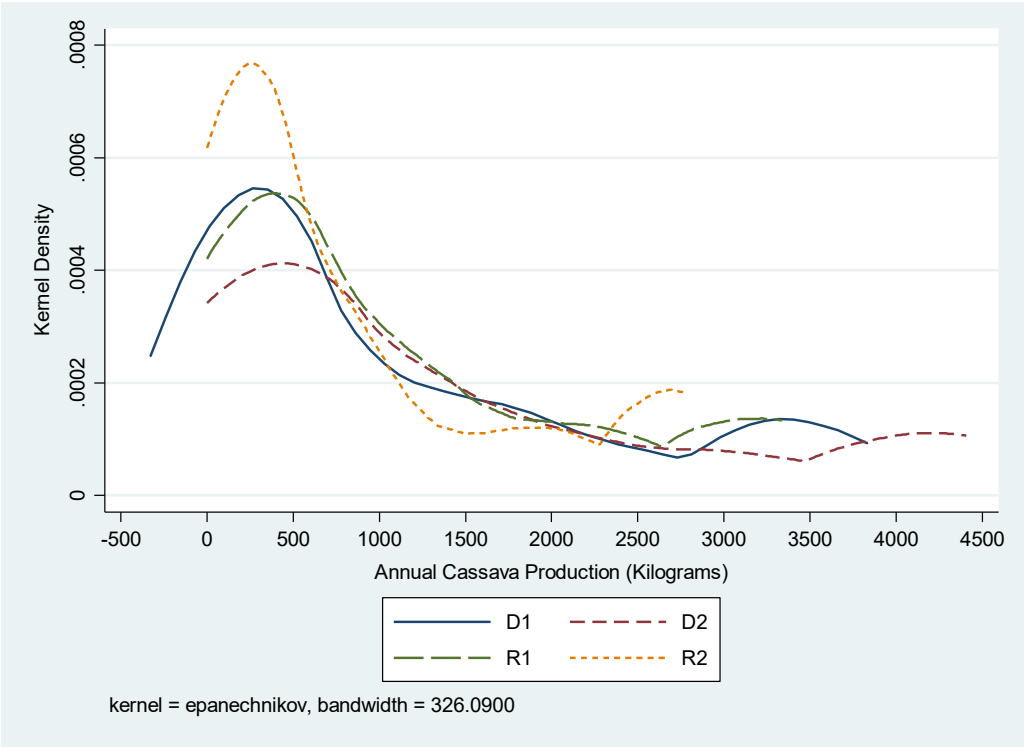


Figure 3. Kernel Density Estimation of Annual Cassava Production by Survey Treatment



Appendix I – CVIP Crop-Cutting Protocol

The following is an excerpt from the CVIP field staff manual.

Crop cutting is a method that allows us to estimate the quantity of production of an entire plot by measuring a small randomly selected section of that plot, and then using this information in conjunction with the area of the entire plot to estimate the total production quantity. We will be conducting crop-cutting on each of the cassava plots that are selected for objective measurement exercises. It is therefore very important that you follow the instructions carefully. You will need the assistance of the Crop Farmer to conduct this exercise.

Crop-Cutting is only conducted on the cassava plot randomly selected by the CAPI program (one per household). There are two aspects to this exercise – the first is conducted with the questionnaire administered in the first visit to the household (the full HH & AG questionnaire for diary households and the shortened HH & AG questionnaire for recall households) and the second is conducted at the time of harvest:

- 1) The first aspect is the selection of a random 5m x 5m crop-cutting subplot within the plot. The 5x5m subplot will be selected using random number tables. This will take place as part of the first visit to all households.
- 2) The second aspect of this exercise is the harvesting of the cassava once it is ready for harvest. This should be done at a time that is convenient for the farmer. It is very important that the farmer does not harvest the land before you arrive – therefore, please coordinate with the farmer and the local crop-cut monitor to learn the time at which he/she would like to harvest and be sure to arrive without delay. The crop will be weighed at the time of harvest.

The materials that you will need for use in this exercise are:

- Compass
- Sticks for Area Demarcation
- Measuring Tape
- Rope (30+ meters per household)
- Writing Materials, e.g. Pen, Pencil, etc.
- Digital Weighing Scale (with batteries)

Compass: This is a device used for capturing geographic bearings in degrees (0^0).

Sticks: These will be used to mark the four corners of the areas selected for crop cutting. Eight sticks will be used to mark the corners of the 4x4 meter subplot and four sticks to mark the corners of the 2x2 meter subplot.

Measuring Tape: This is a distance-measuring instrument marked in metric-units (segments), which will be used to determine the location of the areas in the plot.

Digital Weighing Scale: This will be used to weigh the harvested cassava at the time of harvest.

Writing Materials: These materials can include pen, pencil, etc.

Procedure for Crop Cutting

We will be conducting crop cutting on a 5m x 5m subplot of the cassava plot. Here, we describe in further detail each of the main aspects to the crop cutting exercise.

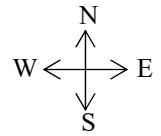
You will construct the 5m x 5m subplot by following steps 1 and 2 below.

1) Crop Cutting Area Selection:

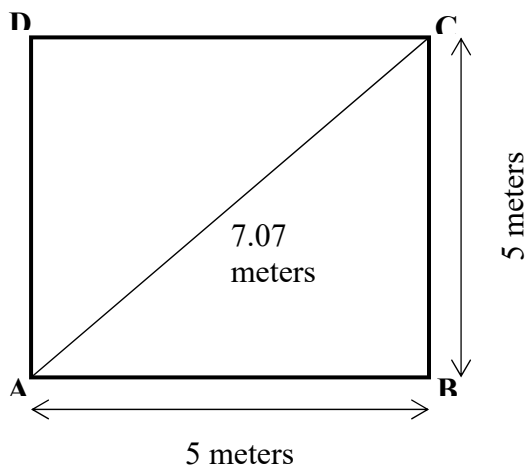
- a. Use Random Number Table #1 to identify the corner from which you will start. Use the first number in the random number table that matches one of the corners of the plot. The corner in which you started the area measurement, the northwest corner, is corner #1. Corner #2 is the next corner of the plot, moving around the plot clockwise.
- b. Measure the distance of the two sides along the selected corner with the measuring tape. Identify which is the longer side and which is the shorter side.
- c. Take the bearing from the start corner down the shorter side. Note this in your notebook.
- d. Use the Random Number Table #2 provided for this household. The first number should be the number of meters that you will walk along the length of the longer side of the plot. If the first number is larger than the length of the side, choose the next random number (and so on, until you find a number that is less than the length of the side). For example, if the length of the longer side is 25 meters and the first random number in the list is 28, move on to the next number.
- e. Beginning at your starting point and continuing along the longer side of the plot, walk the number of meters indicated by your random number.
- f. Turn into the plot so that your bearing is the same as the bearing you measured down the shorter side of the plot. This means you will be entering the plot parallel to the shorter side. Choose the next random number from Random Number Table #2 that is shorter than the length of the shorter side and walk the number of meters indicated by this second random number. You should be walking in a direction that is parallel to the shorter edge of the plot. Walk in a straight line. Try not to veer to the right or left to avoid shrubs or wet spots.
- g. The corner of the crop cutting subplot is located where your foot lands on the last step: this is point A.

2) Crop Cutting Subplot Demarcation:

- a. At point A, insert the first stick firmly into the ground, then turn your face to the east and measure 5 meters *directly* to the east which will be Point B. From Point B, next measure 5 meters *directly* to the north where we put Point C. And, finally, measure 5 meters *directly* to the west for Point D.



- b. Insert sticks exactly at each corner.
- c. Tie a rope around all four sticks. The rope will stay on the subplot until the time of harvest.
- d. In order to make sure that the subplot size is correct, check to make sure that the diagonal line (Line A-C) is 7.07 meters.



Note: If the random numbers obtained from the random table for long and short sides of the plot do not fall in the crop plot area, drop both random numbers and start over again. Each time when one or both random numbers fail to fall in the plot, drop both and start again until both random numbers fall on the plot.

If there is an obstacle in one or more of the crop-cutting subplots, such as a large tree stump, a boulder, large ant hill, etc. re-select the subplot by starting with a new random corner.

If in one or more of the crop-cutting subplots there is cassava damage, DO NOT re-select the subplot. Leave it as is and we will record the damage in the crop-cut questionnaire.