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

HEATHER MOYLAN
Survey Specialist
Living Standards Measurement Study
Development Data Group – Survey Unit – World Bank
hmoylan@worldbank.org

Co-Authors: **TALIP KILIC, JOHN ILUKOR, CLEMENT MTENGULA and INNOCENT PANGAPANGA PHIRI**

ICID-SITES-IFAD Conference on International Development

Rome, Italy – October 4, 2018

The Root of the Measurement Problem

- Challenges in producing reliable agricultural statistics for smallholder systems, despite the importance of the sector
- Root and tuber crops, including cassava, are no exception - notwithstanding their role in hedging against seasonal crop failure and/or food insecurity during the lean season
- Primary concern around smallholder statistics on production and yields of roots and tubers: Reliance on farmer-recall in surveys
    - Crop harvests throughout the year, as needed, often in small quantities
    - With the typical reference period of 12 months or an agricultural season, farmer-recall expected to underestimate production
- Other concerns include:
    - Use of non-standard measurement units
        - Different measurement units of varying sizes along the value chain, crops appearing in different conditions
        - Development of conversion factors to express product-condition-NSU combos in KG equivalent terms in its infancy
    - Control of plots by household members different than respondents
- Clear need to develop improved approaches to microdata collection on root and tuber production in large household and farm surveys

Our Contribution

- Provides 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

- Assesses the relative accuracy of implementing diary-visit versus diary-phone options, and the scale-up feasibility of the daily diary keeping with supervisory phone calls

- Compares the annual cassava yield estimates obtained under diary and recall survey treatments to the crop cutting-based cassava yield estimates and their national and international counterparts

- First comprehensive study undertaken on continuous crop production in Malawi

CVIP: Methodological Experiment on Measuring Cassava Production, Productivity, and Variety Identification

Funding
- LSMS Minding the (Agricultural) Data Gap Research Program, funded by UKAid
- Global Strategy to Improve Agricultural and Rural Statistics

Objectives
- Test subjective approaches to measurement vis-à-vis objective methods for land area; cassava production and productivity; & cassava variety identification
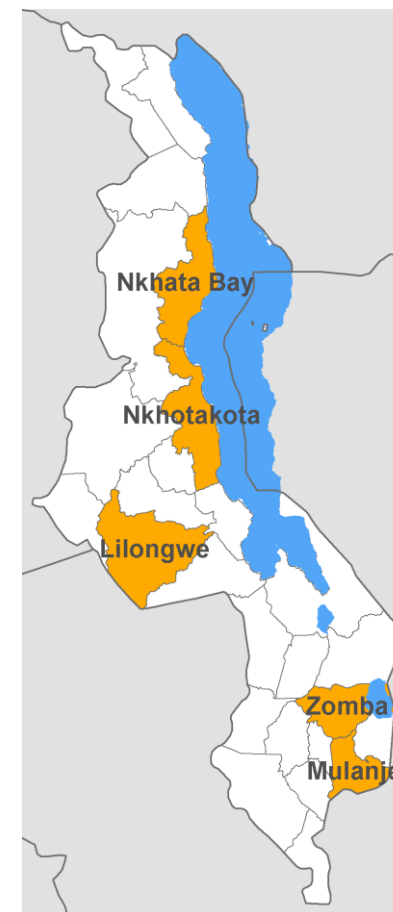
Technical Partnerships
- National Statistical Office (Implementing Agency)
- CGIAR Standing Panel on Impact Assessment (Variety Identification)

CVIP: Context

- **Along lakeshore in North & Central Regions**: Cassava is used as a staple food crop along the shores of Malawi (Nkhatabay, Nkhotakota)

- **Non-cassava belt areas in Southern & Central**: Cassava is mainly grown for sale or use as a snack (Lilongwe, Zomba, Mulanje)

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**	

CVIP: Methods

Methods Tested:

Cassava Production	• Crop-cutting (5mx5m subplot) with balance scales
	    • One plot/household for all 1,241 households
	• Crop diaries with enumerator visits twice/week (**D1**)
	• Crop diaries with telephone calls twice/ week (**D2**)
	• Farmer-reported harvest (two visits, 6-month recall) (**R1**)
	• Farmer-reported harvest (single visit, 12-month recall) (**R2**)
	    • Prevailing approach in Malawi Integrated Household Survey

Land area	• GPS measurement • Farmer-reported area
Variety	• DNA fingerprinting of leaf samples • Farmer-reported varieties & attributes w/ photo aid
CAPI	• Questionnaires administered on *Survey Solutions*

5
Districts

45
Enumeration Areas

1217
Households

∼ 305
Households per treatment

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.

LSMS
Living Standards Measurement Study

Measurement of Production and Yield

- <u>Production</u> measured as **total kilograms harvested per household** across the 12-month period
  - Diary Arms:
    - Sum of all daily records of "fresh" cassava harvests, across all plots, weighed directly upon harvest

  - Recall Arms:
    - Total production solicited in recall interviews –reported in two visits for R1; single visit for R2
    - Collected at the plot-level (in line with the existing IHS approach) - summed across plots
    - Allow for the use of non-standard harvest units
    - Existing conversion factors vs. alternatives from diary sample (region-, district-, EA-, month-specific)

- Gold-standard for annual household-level production: D1 vs. D2
  - Well-implemented diary is the gold standard for *household-level production* measurement because recall decay and respondent bias are minimized but still depends on
    - Supervision/field staff efforts
    - Respondent participation/motivation
    - Respondent literacy

Measurement of Production and Yield (2)

- Annual household-level <u>yield:</u> measured by dividing household-level annual cassava production (in kilograms) with the household-level GPS-based land area cultivated with cassava (in hectares), across all cassava plots cultivated by the household

- Gold-standard for annual household-level yield: **Crop cutting**
  - Recognized as the gold standard for *plot-level seasonal crop yield* measurement since the 1950s
  - Extrapolated from sub-plot to (GPS-measured) total area under cassava cultivation
  - Should be understood as an upper-bound for the cassava yield realized on the farm
  - Should be higher, by design, than the average diary and recall yields

Estimation of Survey Treatment Effects

$$(1)\ Production_i = \alpha + \beta_1 D2_i + \beta_2 R2_i + \beta_3 R1_i + \gamma C + \varepsilon_i$$

$$(2)\ Yield_i = \alpha + \beta_1 D1_i + \beta_2 D2_i + \beta_3 R2_i + \beta_4 R1_i + \gamma C + \varepsilon_i$$

- i represents household, $\propto$ & ε are constant & error terms, respectively
- Equation 1 – Outcome: Annual Cassava Production (KGs) – Comparison Category: **D1**
- Equation 2 – Outcome: Annual Cassava Yield (KGs/Ha) – Comparison Category: **CC**
- D2, R1 & R2 represent identifiers for diary-phone, 6-month recall & 12-month recall, respectively
- C is a vector of household attributes – inclusion of which has no bearing on our findings
- Standard errors clustered at the EA-level for Equation 1, at the household-level for Equation 2

LSMS
Living Standards Measurement Study

Results

Selected Coefficients from Production and Yield Regressions

	Panel A		Panel B		Panel C	
	Total Production (Kg)		**Yield (Kg/Ha, GPS)**		**Yield (Kg/Ha, GPS)**	
	Mean	Coefficient‡	Mean	Coefficient‡	Mean	Coefficient#
Diary - Visit †	1,072	N/A	5,208	N/A	5,208	-3582***
						(507)
Diary - Phone †	1,391	295***	6,618	1431***	7,717	-2211***
		(80)		(430)		(591)
6-Month Recall †	1,102	37	5,798	561	5,798	-2990***
		(68)		(400)		(434)
12-Month Recall †	844	-221***	4,671	-617***	4,671	-4187***
		(61)		(337)		(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.

Results (2)

Average Cassava Production by Month & Survey Treatment

	Diary - Visit (D1)				Diary - Phone (D2)						6-Month Recall (R1)		
	HHs Reporting Any Harvest		**Monthly Average‡**	**Semi-Annual Average#**	**HHs Reporting Any Harvest**		**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**
	Observations	**%**			**Observations**	**%**							
August	23	0.08	307		17	0.06	295	0.910					
September	145	0.53	317		138	0.50	305	0.830					
October	113	0.41	247	706	117	0.42	311	0.344	603	0.285	584	**0.098**	0.795
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**					
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	641	59	0.21	255	0.146	977	**0.000**	586	0.463	**0.000**
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.

Competing Malawi Cassava Yield Estimates for 2015/16
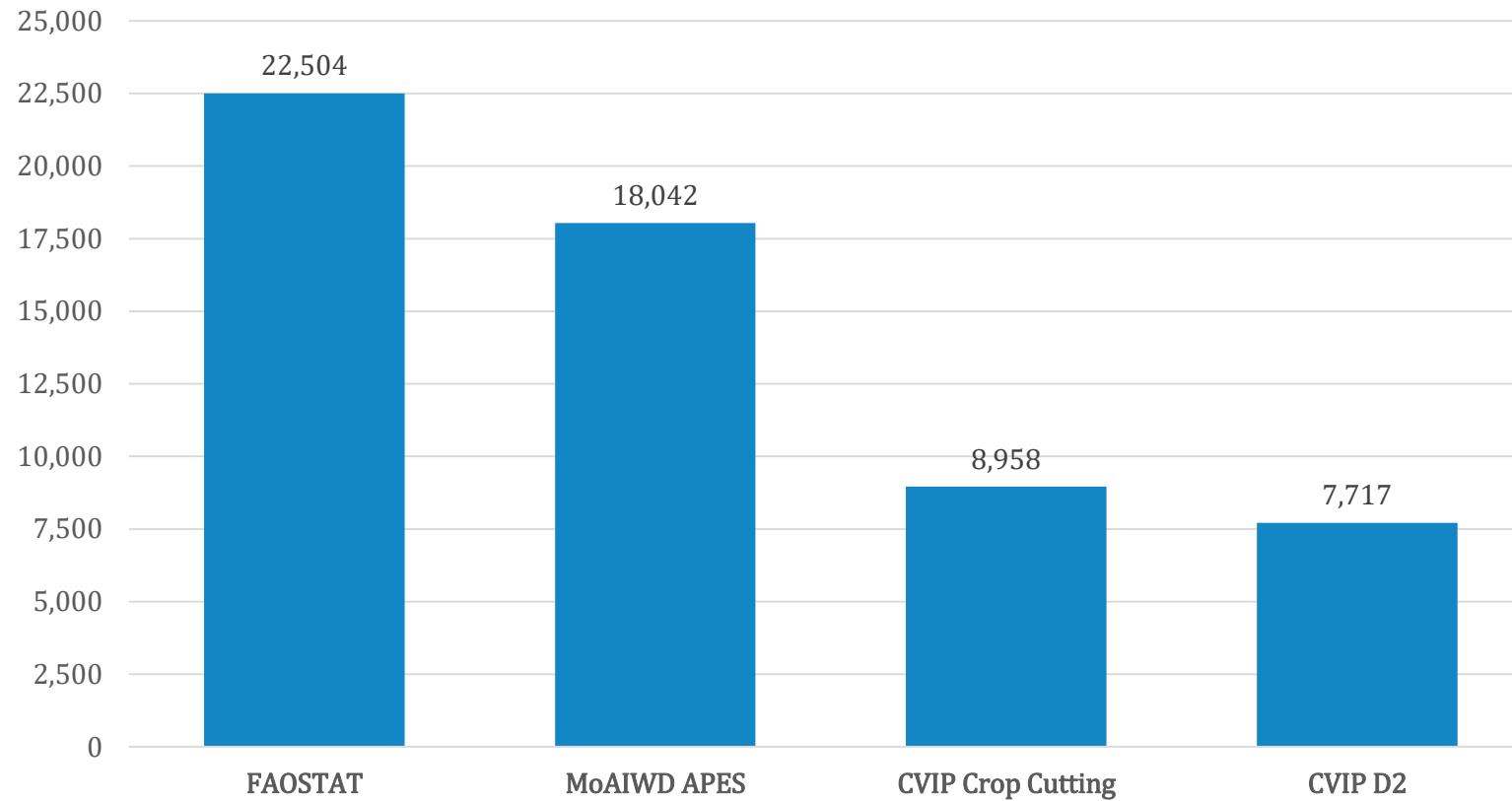
Conclusions

- Given the contribution of cassava farming to food security and agricultural commercialization, evidence underscores critical need for survey practitioners to…
  - Adopt improved survey methods to collect better data on cassava production and productivity
  - Avoid using seasonal or 12-month reference periods, given the significant underestimation of production and yield

- **Recommendation:** Diary-phone
  - Provided the most accurate information on annual cassava production – sustained participation in diary-keeping and potentially connected to in-kind incentives
  - Significantly cheaper than diary-visit, but still more resource- and supervision-intensive compared to recall
  - More feasible if packaged within a broader effort to collect more frequent data through mobile phone calls

- **Second-Best:** 6-Month Recall
  - Performs as well as Diary-Visit (i.e. the traditional gold standard)
  - A viable alternative to existing methods when D2 considered too costly

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

HEATHER MOYLAN
Survey Specialist
Living Standards Measurement Study
Development Data Group – Survey Unit – World Bank
hmoylan@worldbank.org

Co-Authors: **TALIP KILIC, JOHN ILUKOR, CLEMENT MTENGULA and INNOCENT PANGAPANGA PHIRI**

ICID-SITES-IFAD Conference on International Development

Rome, Italy – October 4, 2018