

Microdata for Macro Economists: An Introduction to the **Living Standards Measurement Study**

Talip Kilic

Senior Program Manager

Development Data Group | World Bank

tkilic@worldbank.org

Structural Transformation and Economic Growth (STEG) Virtual Lecture

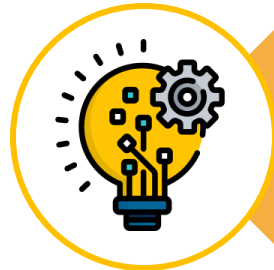
04/12/2024

Overview of today's lecture



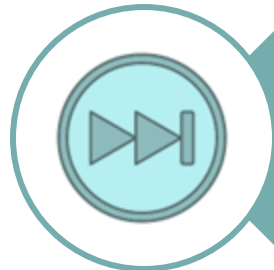
1) The Living Standards Measurement Study (LSMS)

- What does the LSMS do?
- Key features



2) Methodological Innovations

- Time use
- Labor measurement
- Crop yields



3) Looking Forward

- Data integration
- Gearing up for the next phase of our program on longitudinal surveys: **Resilient Futures**



1

The Living Standards Measurement Study (LSMS) Program

A. What does the LSMS do?

B. Key Features



A. What does the LSMS do?



What is the Living Standards Measurement Study (LSMS)?



LSMS is the World Bank's flagship household survey program focused on:

- Strengthening household **survey systems** in countries
- Improving the **quality of microdata** to better inform development policies



It was created in 1980 in response to a need for policy **relevant data**

- Initial focus was on **poverty**
- **Expanded to other areas** such as labor, climate, human capital, etc.
 - » To allow policy makers to understand drivers, inter-relationships

Our workstreams



DATA PRODUCTION

Supporting the **design, implementation, and dissemination** of surveys - households, farms, firms, and facilities



METHODS AND TOOLS

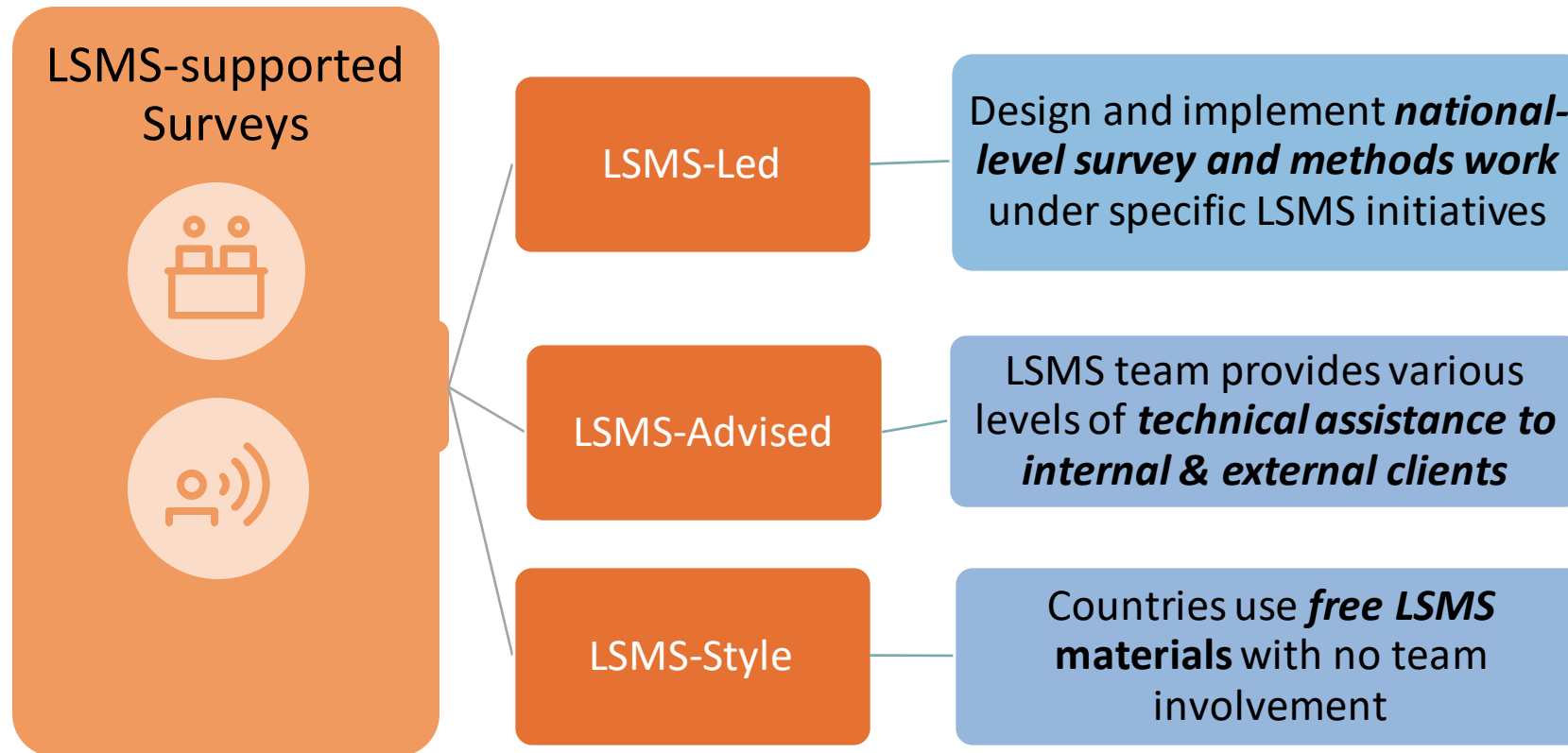
Improving methods and tools for survey data collection and analysis - through field experiments and rigorous research



POLICY RESEARCH

Conducting and promoting research to inform evidence-based development policies

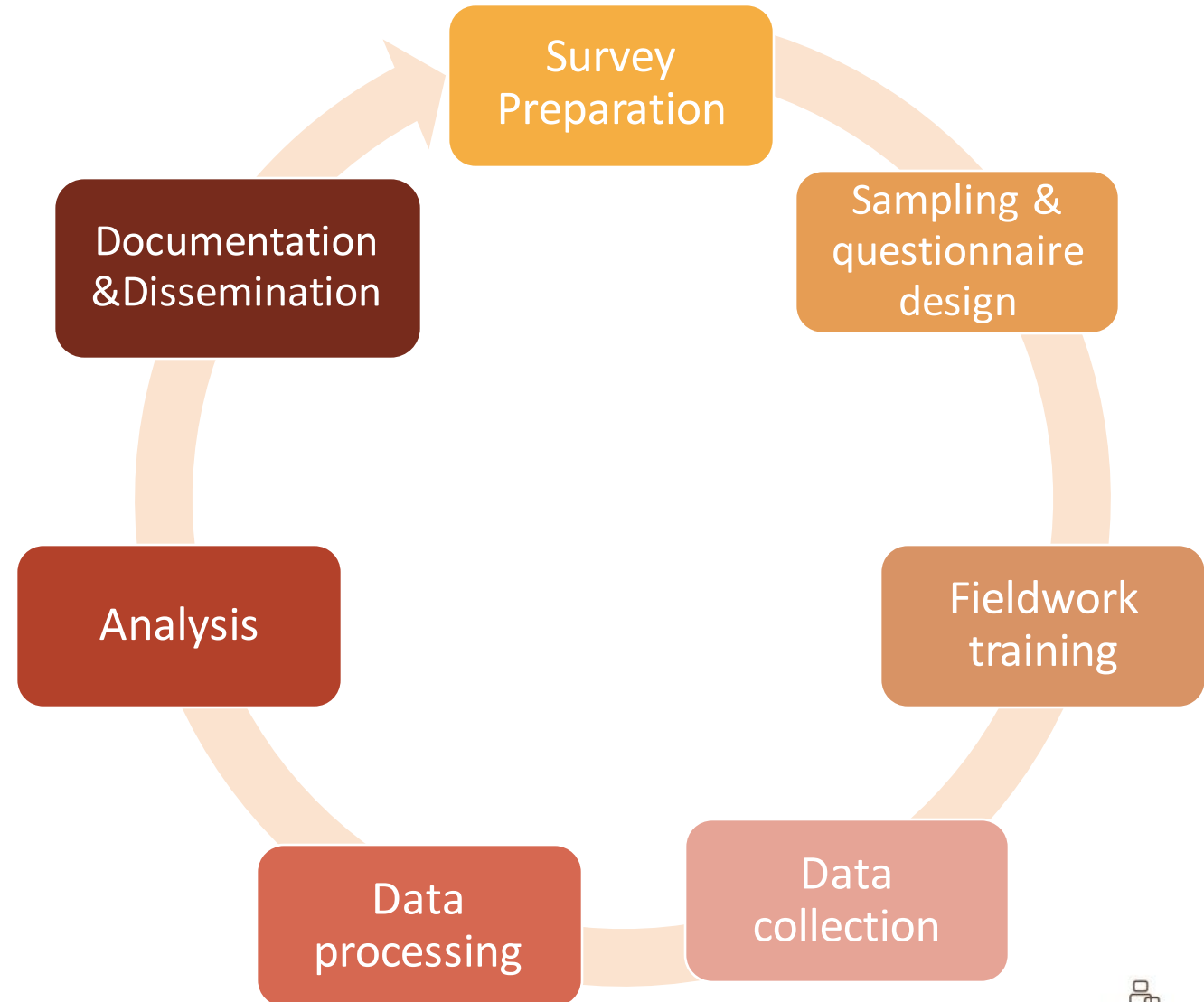
Types of “LSMS Surveys”



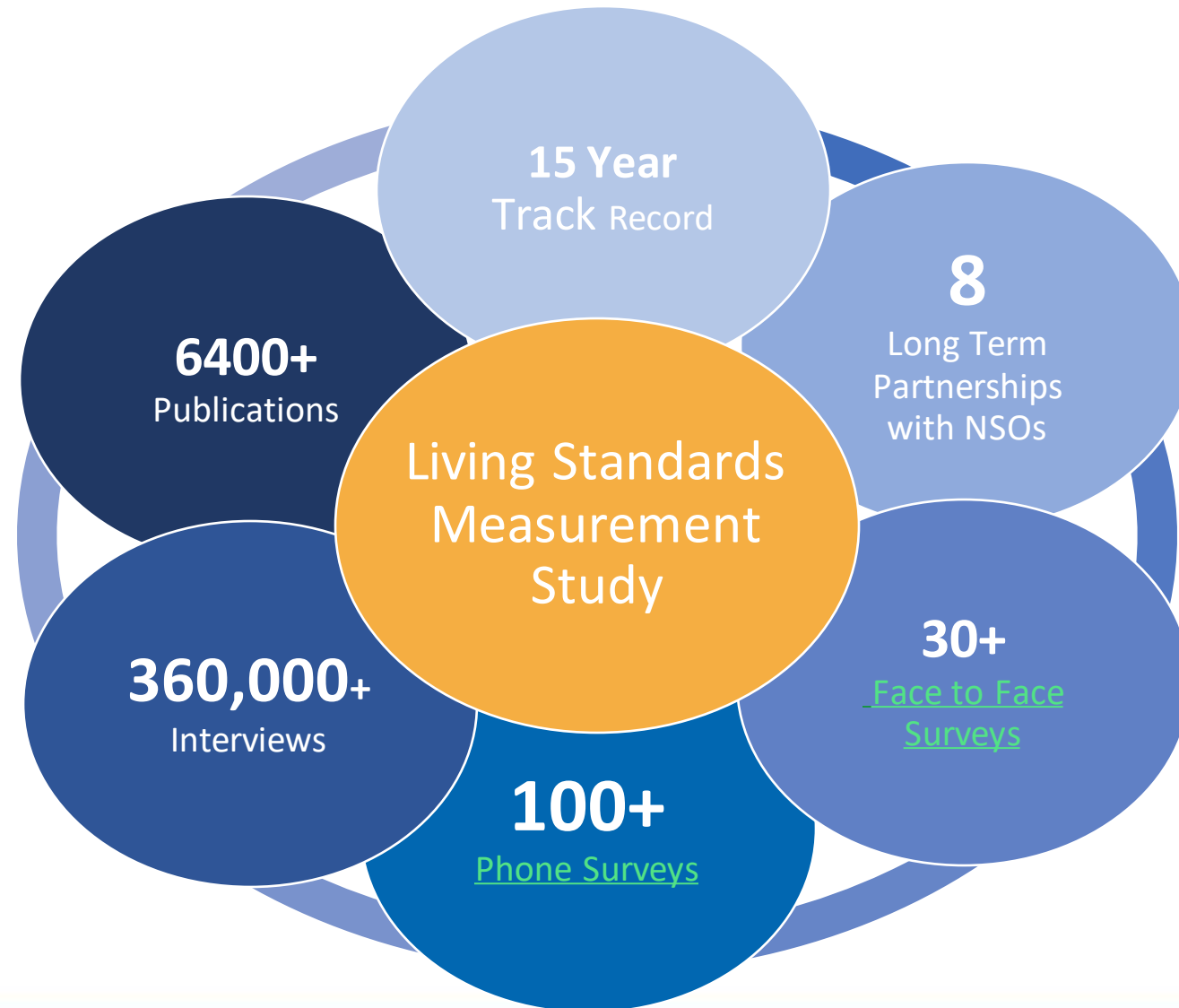
Scope of our expertise

Working with World Bank teams and NSO staff on survey life cycle from beginning through end

Provide Technical Assistance & Advice Across Survey Cycle



A special focus on longitudinal survey systems



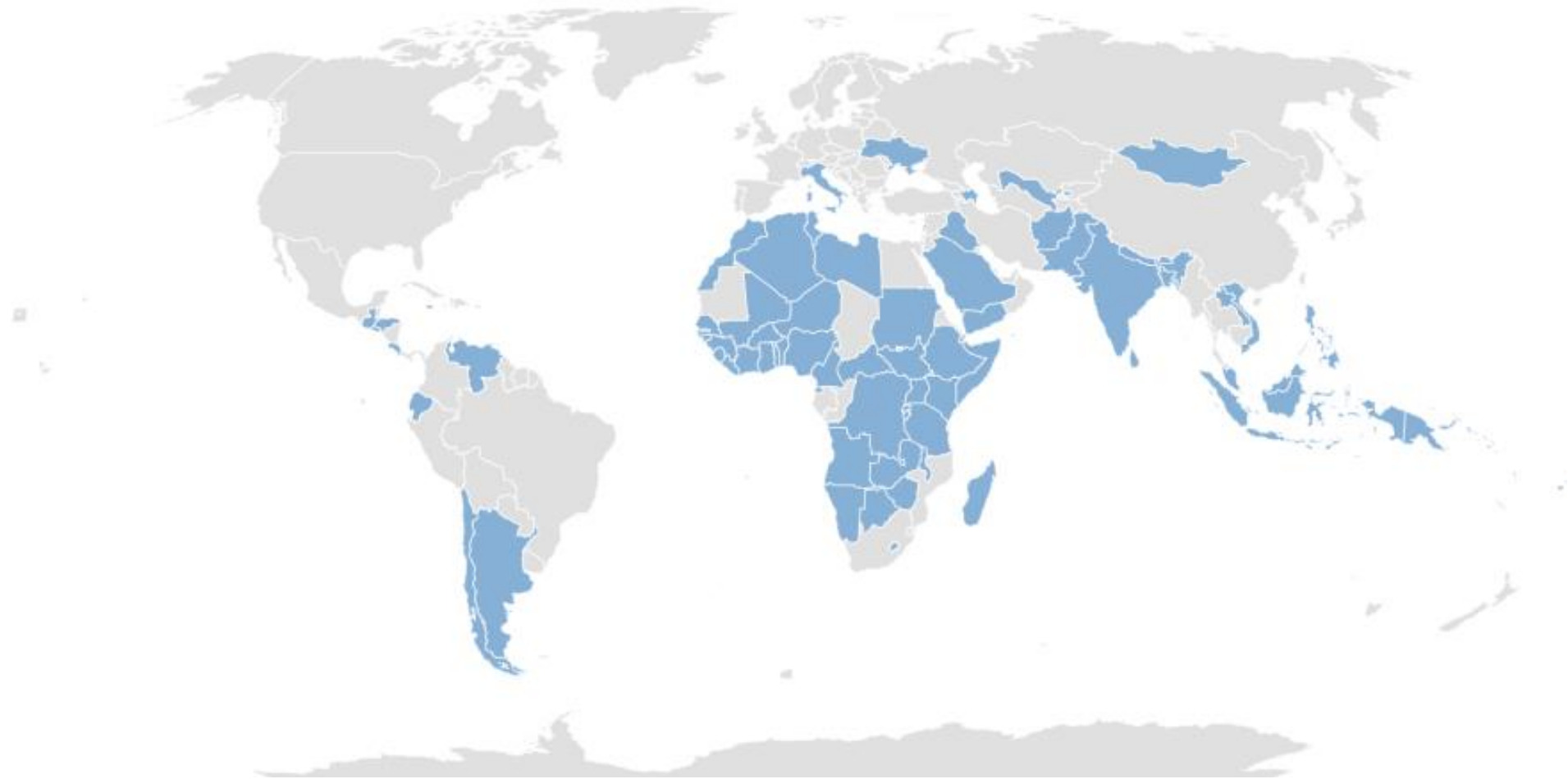
LSMS guidebooks as global public goods

The LSMS team has been pushing the methodological frontier:

- » **By developing and validating improved methods** for survey data collection and integration with other data sources.
- » **By disseminating these methods through guidelines** on several topics, such as labor, energy, agriculture, disability, etc.



78 countries supported over 2019-2024





B. Key Features of Our Surveys



Key Features

- 1 **Multi-topic and multi-level**
- 2 Nationally-**representative** probability sample
- 3 Panel
- 4 **Multi-purpose**
- 5 **Geo-referenced** and **interoperable** with geospatial, admin and census/survey data
- 6 Food and Non-Food **Consumption** and **Expenditures**
- 7 **Individual** Disaggregation
- 8 Computer-Assisted Personal Interviewing
- 9 **Objective measurement**
- 10 **Public Access**

Key Features

1

Multi-topic and multi-level

- » Goes beyond measuring money metric welfare measures
- » Captures correlates and multidimensionality of wellbeing
- » Analytical tool to study behavior, understand phenomena, analyze linkages

2

Nationally-**representative** probability sample

3

Panel

4

Multi-purpose

5

Geo-referenced and **interoperable** with geospatial, admin and census/survey data

6

Food and Non-Food **Consumption** and **Expenditures**

7

Individual Disaggregation

8

Computer-Assisted Personal Interviewing

9

Objective measurement

10

Public Access



Household

- Dwelling GPS Coordinates
- Demographics
- Education
- Health
- Housing and Utilities
- Food and Non-Food Consumption
- Off-Farm Earnings
- Asset Ownership
- Anthropometry
- Food Security
- Safety Nets
- Shocks
- Migration



Agriculture

- Plot GPS Coordinates & GPS-Based Area Measurement
- **Parcels:** Tenure, Ownership
- **Plots:** Physical Attributes, Labor & Non-Labor Input Use
- **Crops:** Cultivation, Production (Plot-Crop-Level), & Disposition (Crop-Level)
- Ag Asset Ownership & Use
- Extension Services
- Livestock Ownership & Production



Community

- Demographics
- Infrastructure
- Facilities
- Access to Services
- Facilities
- Collective Action
- Natural Resource Management
- Community Organizations
- Prices

Key Features

1 Multi-topic and multi-level

2 **Nationally-representative** probability sample

- » Mindful of non-sampling errors and sustainability (Low capacity of data-deprived countries)
- » Representative: national, urban/rural, other urban, main regions
- » Sample size usually anchored in available data on consumption expenditures
- » Smallish sample, higher resolution estimates pursued through data integration, SAE

3 Panel

4 **Multi-purpose**

5 **Geo-referenced** and **interoperable** with geospatial, admin and census/survey data

6 Food and Non-Food **Consumption** and **Expenditures**

7 **Individual** Disaggregation

8 Computer-Assisted Personal Interviewing

9 **Objective** measurement

10 **Public Access**

Key Features

1 Multi-topic and multi-level

2 Nationally-representative probability sample

3 Panel

» Unique system of longitudinal surveys designed to improve the understanding of household and individual welfare and living standards.

4 Multi-purpose

5 **Geo-referenced** and **interoperable** with geospatial, admin and census/survey data

6 Food and Non-Food **Consumption** and **Expenditures**

7 **Individual** Disaggregation

8 Computer-Assisted Personal Interviewing

9 **Objective measurement**

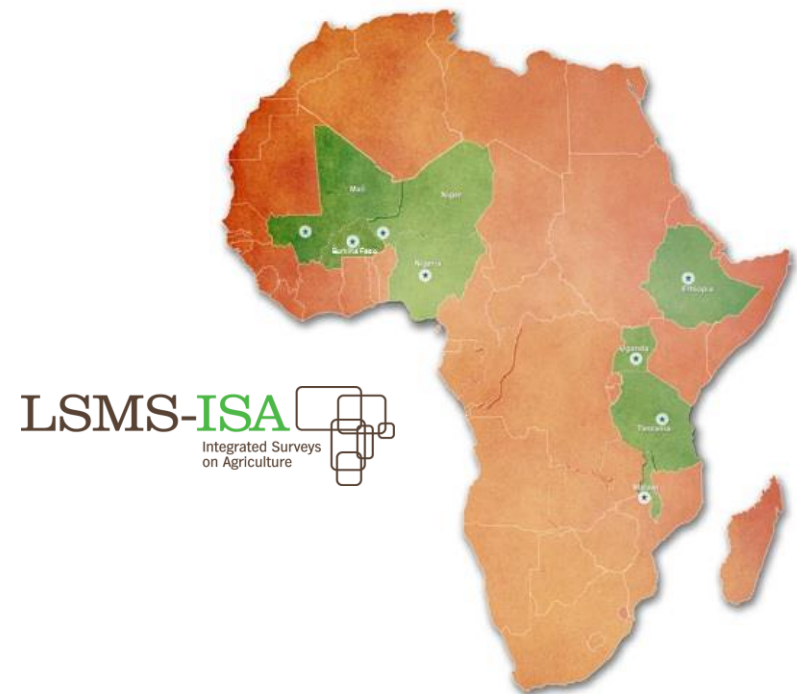
10 **Public Access**

LSMS-ISA (2008-2024)

LSMS-Integrated Surveys on Agriculture (LSMS-ISA)-supported surveys have emerged as the benchmark for nationally-representative panel data collection in Sub-Saharan Africa.

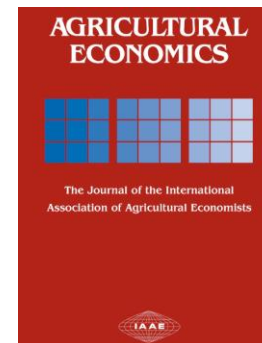
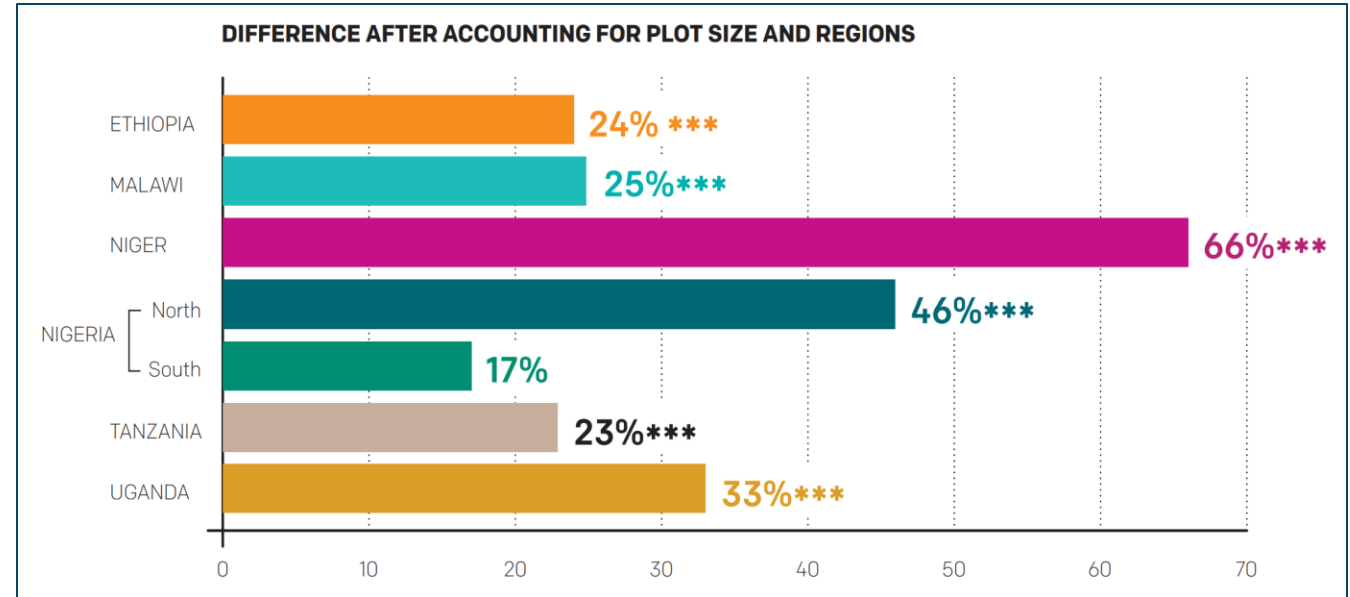
Since 2008:

- » 8 countries
- » 33 face-to-face panel survey rounds
- » 360,000+ interviews
- » 6,400+ publications



Gender differences in agricultural productivity

- Nationally-representative, plot-level survey data from LSMS-ISA revealed that on average, **female-managed plots are 24 to 66% less productive** – with significant within- and cross-country variation.
- **Gender gap in productivity driven by lower levels of fertilizer, labor, and high-value crop cultivation on female-managed plots but also lower returns to these inputs on female-managed plots.**



Ten policy recommendations in [Levelling the Field](#) report and [Agricultural Economics SI on Gender and Agriculture](#)

Key Features

- 1 Multi-topic and multi-level
- 2 Nationally-**representative** probability sample
- 3 Panel
- 4 Multi-purpose**
 - » Measure and monitor poverty and other SDGs
 - » Analytical tool
 - » Use for IE
- 5 **Geo-referenced** and **interoperable** with geospatial, admin and census/survey data
- 6 Food and Non-Food **Consumption** and **Expenditures**
- 7 **Individual** Disaggregation
- 8 Computer-Assisted Personal Interviewing
- 9 **Objective** measurement
- 10 **Public Access**

IE activities piggybacked onto LSMS-ISA

USAID Feed-the-Future Population Based Surveys

- Operating in USAID priority countries
- Geographically focused in USAID Zone of Influence
- Focused on monitoring poverty, nutrition, food security and women's empowerment

LSMS-ISA operating in some of those priority countries but ..

- Small sample size in Zol
- Missing some thematic content

Built on infrastructure in Nigeria and Tanzania of LSMS-ISA panels

- Oversampling in Zol
- Adjusting questionnaires

Key Features

- 1 Multi-topic and multi-level
- 2 Nationally-**representative** probability sample
- 3 Panel
- 4 Multi-purpose
- 5 Geo-referenced and interoperable with geospatial, admin and census/survey data**
 - » Geo-referencing → integration with third-party geospatial data (LSMS-ISA example)
 - » Use of common administrative frame → integration with administrative/census data
 - » Common questions, sampling frames across censuses, surveys
- 6 Food and Non-Food **Consumption** and **Expenditures**
- 7 **Individual** Disaggregation
- 8 Computer-Assisted Personal Interviewing
- 9 **Objective measurement**
- 10 **Public Access**

LSMS-ISA geospatial variables

Provide **Random Off-Set**, EA-Level Coordinates

- Average household-level coordinates in an EA
- Apply a random offset of 0-2 km in urban, 2-5km in rural areas



Journal of Development Economics

Available online 1 July 2022, 102927

In Press, Journal Pre-proof



Regular article

Privacy protection, measurement error, and the integration of remote sensing and socioeconomic survey data ☆

Jeffrey D. Michler ^a, Anna Josephson ^a, Talip Kilic ^b, Siobhan Murray ^b

Show more

+ Add to Mendeley Share Cite

<https://doi.org/10.1016/j.jdeveco.2022.102927>

Get rights and content

Under a Creative Commons license

Open access

Highlights

- Statistical disclosure limitation distorts public use data to preserve privacy.
- Matching public use data with remote sensing weather data may generate distortion.
- We explore how spatial anonymization methods introduce measurement error.
- Spatial anonymization methods have limited to no impact on estimates.
- Estimates do vary by choice of remote sensing product.

LSMS-ISA geospatial variables

Distance

Plot distance to household

Household to nearest main road

Household to major agricultural market

Household to headquarters of district of residence

Household to nearest city or town with +20,000

Household to nearest border post

Climatology

Annual mean temperature

Mean temperature of wettest quarter

Mean annual precipitation

Precipitation of wettest quarter

Precipitation of wettest month

LSMS-ISA geospatial variables

Landscape

Land cover class

Density of agriculture

Population density

Agro-ecological zone

Rainfall (TS)

Survey year annual rainfall

Survey year wettest quarter rainfall

Survey year timing of start of wettest quarter

LSMS-ISA geospatial variables

Soil & Terrain

Elevation

Slope

Terrain roughness

Topographic wetness index

Landscape-level soil characteristics

Phenology

Average total change in greenness within primary ag season

Average timing of onset of greenness increase

Average timing of onset of greenness decrease

Average EVI value at peak of greenness

Total change in greenness in survey year

Timing of onset of greenness increase in survey year

Timing of onset of greenness decrease in survey year

Maximum EVI value in survey year

Specific crop season NDVI crop season aggregates

Research on extreme weather impacts made possible by integration with geospatial and admin data

Food Policy 69 (2017) 68–81

Contents lists available at ScienceDirect

Food Policy

journal homepage: www.elsevier.com/locate/foodpol

Smallholder productivity and weather shocks: Adoption and impact of widely promoted agricultural practices in Tanzania

Aslihan Arslan^a, Federico Belotti^{b,*}, Leslie Lipper^c

^aInternational Fund for Agricultural Development, Via Paolo di Dono, 44, 00142 Rome, Italy
^bDepartment of Economics and Finance, University of Rome Tor Vergata, Via Columbia, 2, 00133 Rome, Italy
^cIndependent Science Partnership Council of CGIAR, Viale delle Terme di Caracalla, 00133 Rome, Italy

ARTICLE INFO

Article history:
 Received 26 February 2015
 Received in revised form 17 February 2017
 Accepted 7 March 2017
 Available online 22 March 2017

KEYWORDS
 Technology adoption
 Productivity analysis
 Climate change
 Panel data
 Tanzania

1. Introduction

Sub-Saharan Africa (SSA) remains the world's most food-insecure region characterized by high levels of child mortality and poverty, low levels of human and physical capital, and poor infrastructure (FAO, 2009). The economies of countries in SSA heavily depend on a smallholder-based agricultural sector, which in turn makes them particularly vulnerable to climate change (Zervopoulos et al., 2008; Barrios et al., 2008; TZC, 2009). An estimated 90 percent of the population depends on rain-fed crop production and pastoralism to meet its basic food needs (Patt and Winkler, 2007). In the absence of a policy agenda to increase economic growth and decrease poverty while maintaining the natural resource base, the negative effects of climate change on crop production will be especially pronounced in SSA. Forecasts for SSA show that *ceteris paribus*, rice, wheat, and maize yields are likely to decline in the next thirty years by 15 percent, 34 percent, and 10 percent, respectively (Nelsson et al., 2009).

^{*} Corresponding author.
 E-mail address: federico.belotti@unroma2.it (F. Belotti).

Environment and Development Economics (2021), 1–18
 doi:10.1017/S1355770X20000455

EDE

RESEARCH ARTICLE

Droughts and floods in Malawi: impacts on crop production and the performance of sustainable land management practices under weather extremes

Nancy McCarthy,¹ Talip Kilic,^{2*} Josh Brubaker,¹ Siobhan Murray,² and Alejandro de la Fuente³

¹Lead Analytics, Washington, DC, USA; ²Development Data Group, The World Bank, Washington, DC, USA and ³Poverty and Equity Global Practice, The World Bank, Washington, DC, USA
^{*}Corresponding author. E-mail: tkilic@worldbank.org

(Submitted 29 August 2019; revised 31 March 2020, 8 July 2020; accepted 30 August 2020)

Abstract
 Climate change is predicted to increase the frequency of extreme weather events, increasing the vulnerability of smallholder farmers dependent on rain-fed agriculture. We evaluate the extent to which farmers in Malawi suffer crop production losses due to extreme weather, and whether sustainable land management (SLM) practices help shield crop production losses from extreme events. We use a three period panel dataset where widespread floods and droughts occurred in separate periods, offering a unique opportunity to evaluate impacts using data collected immediately following these events. Results show that crop production outcomes were severely hit by both floods and droughts, with average losses ranging between 32–48 per cent. Legume intercropping provided protection against both floods and droughts, while green belts provided protection against floods. However, we find limited evidence that SLM adoption decisions are driven by exposure to weather shocks; rather, farmers with more productive assets are more likely to adopt.

Keywords: climate change; crop production; sustainable land management; Malawi
JEL classification: D01; Q12; Q25; Q54

POLICY RESEARCH WORKING PAPER 9666

RECURRENT CLIMATIC SHOCKS AND HUMANITARIAN AID

Impacts on Livelihood Outcomes in Malawi

Nancy McCarthy
 Talip Kilic
 Joshua Brubaker
 Alejandro de la Fuente
 Siobhan Murray

WORLD BANK GROUP
 Development Economics
 Development Data Group
 May 2021

Linking Georeferenced LSMS-ISA Household and Plot Locations with Geospatial Data on Climate and WFP Aid Data

Research on integration of large-scale surveys and satellites for high-res agricultural monitoring

- [Azzari et al. \(2021\)](#) address operationally-relevant research questions on the integration of survey and satellite data in the context of **maize area mapping in Malawi and Ethiopia**
- 10-m resolution crop area and maize area maps for Malawi and Ethiopia for each agricultural season from 2016 to 2019 on **World Bank Development Data Hub**




Article

Understanding the Requirements for Surveys to Support Satellite-Based Crop Type Mapping: Evidence from Sub-Saharan Africa

George Azzari ¹, Shruti Jain ¹, Graham Jeffries ², Talip Kilic ^{3,*} and Siobhan Murray ³

¹ Atlas AI, Palo Alto, CA 94301, USA; george@atlasai.us (G.A.); shruti@atlasai.us (S.J.)

² Farmers Business Network, Rockland, ME 04841, USA; gjeffries@farmersbusinessnetwork.com

³ Development Data Group, World Bank, Washington, DC 20433, USA; smurray@worldbank.org

* Correspondence: tkilic@worldbank.org

Abstract: This paper provides recommendations on how large-scale household surveys should be conducted to generate the data needed to train models for satellite-based crop type mapping in smallholder farming systems. The analysis focuses on maize cultivation in Malawi and Ethiopia, and leverages rich, georeferenced plot-level data from national household surveys that were conducted in 2018–20 and integrated with Sentinel-2 satellite imagery and complementary geospatial data. To identify the approach to survey data collection that yields optimal data for training remote sensing models, 26,250 in silico experiments are simulated within a machine learning framework. The best model is then applied to map seasonal maize cultivation from 2016 to 2019 at 10-m resolution in both countries. The analysis reveals that smallholder plots with maize cultivation can be identified with up to 75% accuracy. Collecting full plot boundaries or complete plot corner points provides the best quality of information for model training. Classification performance peaks with slightly less than 60% of the training data. Seemingly little erosion in accuracy under less preferable approaches to georeferencing plots results in the total area under maize cultivation being overestimated by 0.16–0.47 million hectares (8–24%) in Malawi.

 **check for updates**

Citation: Azzari, G.; Jain, S.; Jeffries, G.; Kilic, T.; Murray, S. Understanding the Requirements for Surveys to Support Satellite-Based Crop Type Mapping: Evidence from Sub-Saharan Africa. *Remote Sens.* **2021**, *13*, 4749. <https://doi.org/10.3390/rs13234749>

Received: 8 September 2021
Accepted: 8 November 2021
Published: 23 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: agriculture; maize; crop type mapping; Sentinel-2; household surveys; training data; Malawi; Ethiopia

1. Introduction

Agriculture is an integral part of livelihoods in sub-Saharan Africa, where it can contribute up to 69% of household income in rural areas [1]. As such, improving the productivity of smallholder farmers has been a long-standing goal in many African countries aiming to eliminate poverty and food insecurity.

To monitor progress towards national and international development goals related to agricultural productivity, countries need accurate, crop-specific measures of area under cultivation, production and yields, not only at the national level but with sufficient within-country disaggregation to guide the targeting and evaluation of policies and programs intended to promote agricultural and rural development, and resilience against disasters and extreme weather events.

With the commencement of the European Space Agency's Sentinel-2 mission in 2015 and the subsequent surge in the public availability of high-resolution satellite imagery, research has shown the feasibility of satellite-based monitoring of agricultural outcomes in smallholder farming systems [2–5]. Recent advances in satellite imagery and remote sensing techniques have the potential to provide timely insights into conditions on the ground and can fill gaps in agricultural monitoring and statistics [9].

Satellite-based approaches to mapping agricultural outcomes, such as crop-specific estimates of cultivated areas and yields, require data for training and validating the underlying remote sensing models. The quality and spatial resolution of satellite-based estimates

Remote Sens. **2021**, *13*, 4749. <https://doi.org/10.3390/rs13234749>

<https://www.mdpi.com/journal/remotesensing>

Headline findings

- Collecting a **complete plot boundary** is preferable to competing approaches to georeferencing plot locations in large-scale household surveys.
- Seemingly-small erosion in maize classification accuracy under less preferable approaches to georeferencing plot locations **consistently results in total area under maize cultivation to be overestimated** - in the range of **0.16 to 0.47 million hectares** (8 to 24 percent).
- **Georeferencing the complete set of plot corners is a second-best strategy**, can approximate full plot boundaries and can in turn train models with comparable performance.
- Classification performance peaks with **~60% of the training data** under preferred and second-best approaches to georeferencing plot locations.
- **If only a single GPS point** can be collected, that location should be **near the plot centroid** rather than at the plot corner. With large datasets, the performance could be comparable to that of complete plot boundaries.
- **No plot observations should be excluded** from model training based on a minimum plot area threshold. And **optical features alone** can provide sufficient signal to maximize prediction quality.

Open access datasets

10-m resolution crop area and maize area maps for Malawi and Ethiopia for each agricultural season from 2016 to 2019 on **World Bank Development Data Hub**

High-Resolution Crop And Maize Area Mapping For Ethiopia

Published (Ver.2216456)

Linked to the research conducted under the Methods and Tools Component of the 50x2030 Initiative (<https://www.50x2030.org/>), this data deposit includes 10-meter spatial resolution maps for (i) areas cultivated with any crops, and (ii) areas cultivated with maize across Ethiopia for each rainy season during the period of 2016-2019. The maps are a product of the analyses conducted by Azzari et al. (2021), as part of the collaboration between the World Bank and Atlas AI, in support of one of the objectives of the 50x2030 Initiative to create guidelines for the collection of minimum-required survey data for training and validating remote sensing models for high-resolution crop type mapping and crop yield estimation. Azzari et al. (2021) integrate Sentinel-2 satellite imagery and complementary geospatial data with georeferenced plot-level data from national household surveys that were conducted by the Malawi National Statistical Office and the Central Statistical Agency of Ethiopia during the period of 2018-2020 in order to identify the optimal approach to collecting survey data for training a machine learning model to identify areas cultivated with maize. The best performing model estimated by Azzari et al. (2021) has been used to generate the 10-meter spatial resolution maps that are being made available here. For more information, please see the accompanying Basic Information Document and Azzari et al. (2021). Less...

Overview | **Data & Resources** | Additional Information | Citations

Basic Information Document for crop and maize area mapping
Resource Type: **Documentation** Data Classification of File: **Public**
[Download](#) | [Preview](#)

Ethiopia maize mask for 2016
Pixels with probability of crop cultivation greater than or equal to 40 percent and probability of maize cultivation greater than or equal to 50 percent.
Resource Type: **Download** Data Classification of File: **Public**
[Go to Resource](#)

<http://bit.ly/ethiopiamaps>

High-Resolution Crop And Maize Area Mapping For Malawi

Published (Ver.2216386)

Linked to the research conducted under the Methods and Tools Component of the 50x2030 Initiative (<https://www.50x2030.org/>), this data deposit includes 10-meter spatial resolution maps for (i) areas cultivated with any crops, and (ii) areas cultivated with maize across Malawi for each rainy season during the period of 2016-2019. The maps are a product of the analyses conducted by Azzari et al. (2021), as part of the collaboration between the World Bank and Atlas AI, in support of one of the objectives of the 50x2030 Initiative to create guidelines for the collection of minimum-required survey data for training and validating remote sensing models for high-resolution crop type mapping and crop yield estimation. Azzari et al. (2021) integrate Sentinel-2 satellite imagery and complementary geospatial data with georeferenced plot-level data from national household surveys that were conducted by the Malawi National Statistical Office and the Central Statistical Agency of Ethiopia during the period of 2018-2020 in order to identify the optimal approach to collecting survey data for training a machine learning model to identify areas cultivated with maize. The best performing model estimated by Azzari et al. (2021) has been used to generate the 10-meter spatial resolution maps that are being made available here. For more information, please see the accompanying Basic Information Document and Azzari et al. (2021).

Reference: Azzari, G., Jain, S., Jeffries, G., Kilic, T., and Murray, S. (2021). "Understanding the Requirements for Surveys to Support Satellite-Based Crop Type Mapping: Evidence from Sub-Saharan Africa." World Bank Policy Research Working Paper No. 9609, LSMS Washington, DC: World Bank. Less...

Overview | **Data & Resources** | Additional Information | Citations

Basic Information Document for crop and maize area mapping
Resource Type: **Documentation** Data Classification of File: **Public**
[Download](#) | [Preview](#)

Malawi maize mask for 2016
Pixels with probability of crop cultivation greater than or equal to 40 percent and probability of maize cultivation greater than or equal to 60 percent.

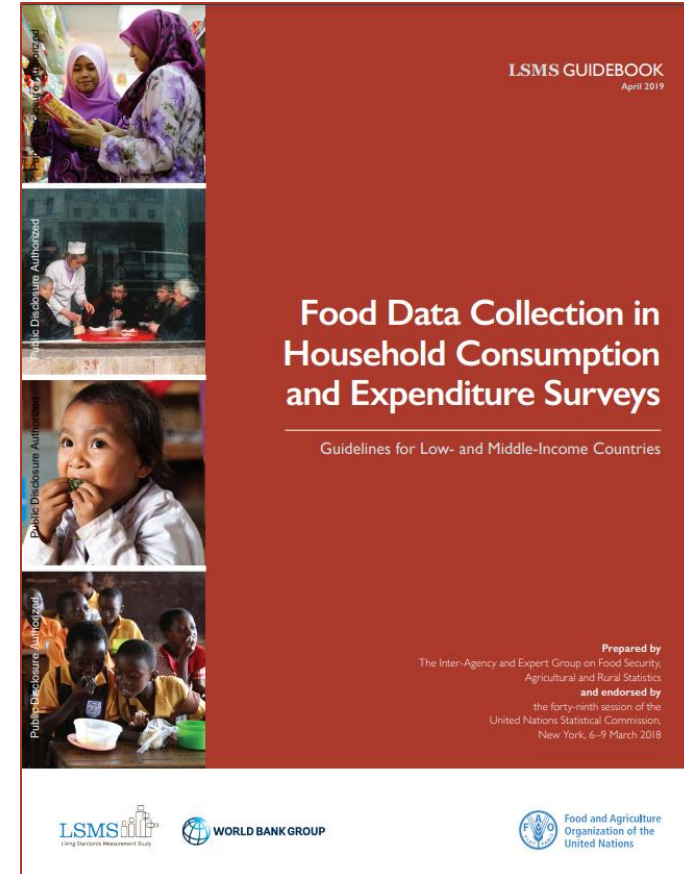
<http://bit.ly/malawimaps>

Key Features

- 1 Multi-topic and multi-level
- 2 Nationally-**representative** probability sample
- 3 Panel
- 4 **Multi-purpose**
- 5 **Geo-referenced and interoperable** with geospatial, admin and census/survey data
- 6 Food and Non-Food Consumption and Expenditures**
 - » Comprehensive modules to create total household consumption aggregates
 - » Survey design does matter for food consumption measurement
 - » Using non-standard units for food consumption measurement
 - » On-going methodological research on intra-household and non-food consumption measurement
- 6 Food and Non-Food **Consumption** and **Expenditures**
- 8 Computer-Assisted Personal Interviewing
- 9 **Objective measurement**
- 10 **Public Access**

Best practices for food consumption data collection in LMICs

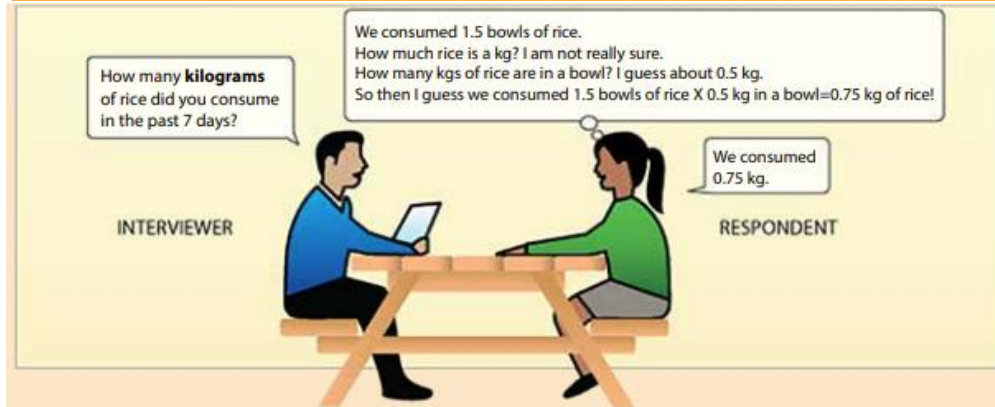
- Main issues: Diary = fatigue; Recall = telescoping & omission
- Research broadly points to **equivalence of recall and diary** in terms of accuracy of poverty estimates, but recall comes at a much **lower cost**
 - Beegle et al (2012), Gibson et al. (2015), Backiny-Yetna et al. (2017), Brzozowska et al. (2017), Sharp et al. (2022) + Recent work in Tanzania, Zimbabwe, Saudi Arabia
- Official World Bank/LSMS recommendation: Use 7-recall period food consumption measurement in LMICs



Using non-standard units for food consumption measurement

FORCING STANDAR UNITS

More burden on the respondent, less consistency in conversion factors

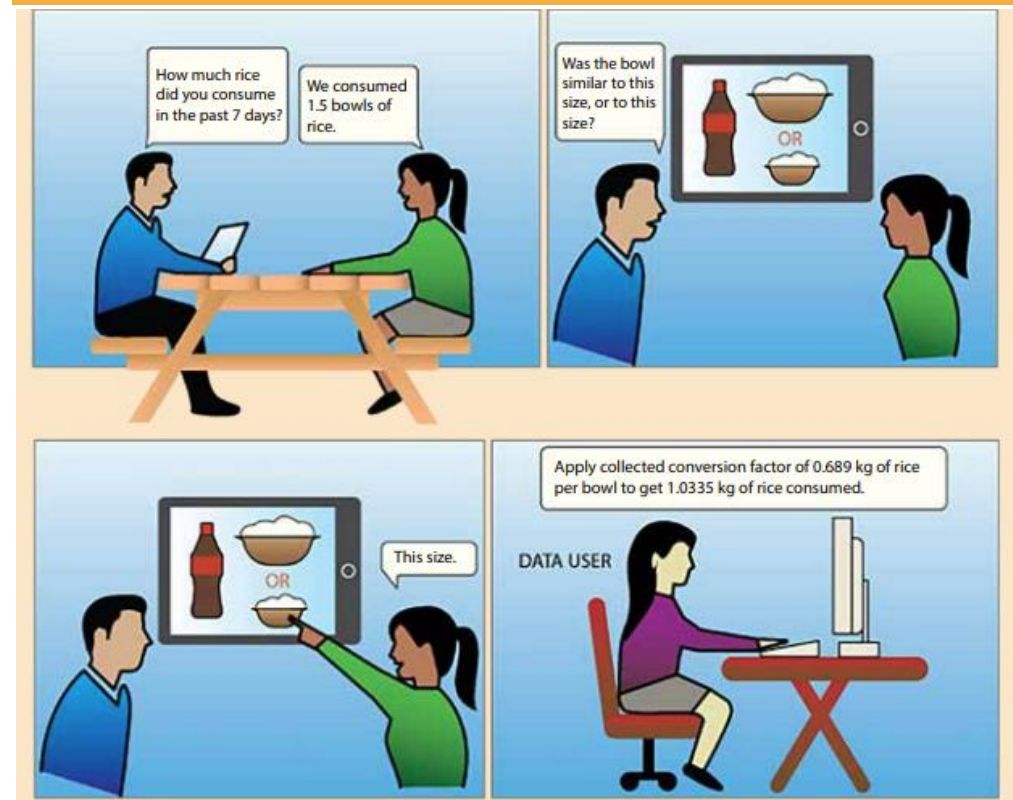


Practical guidance on non-standard units and conversion factors to minimize measurement error in food consumption.



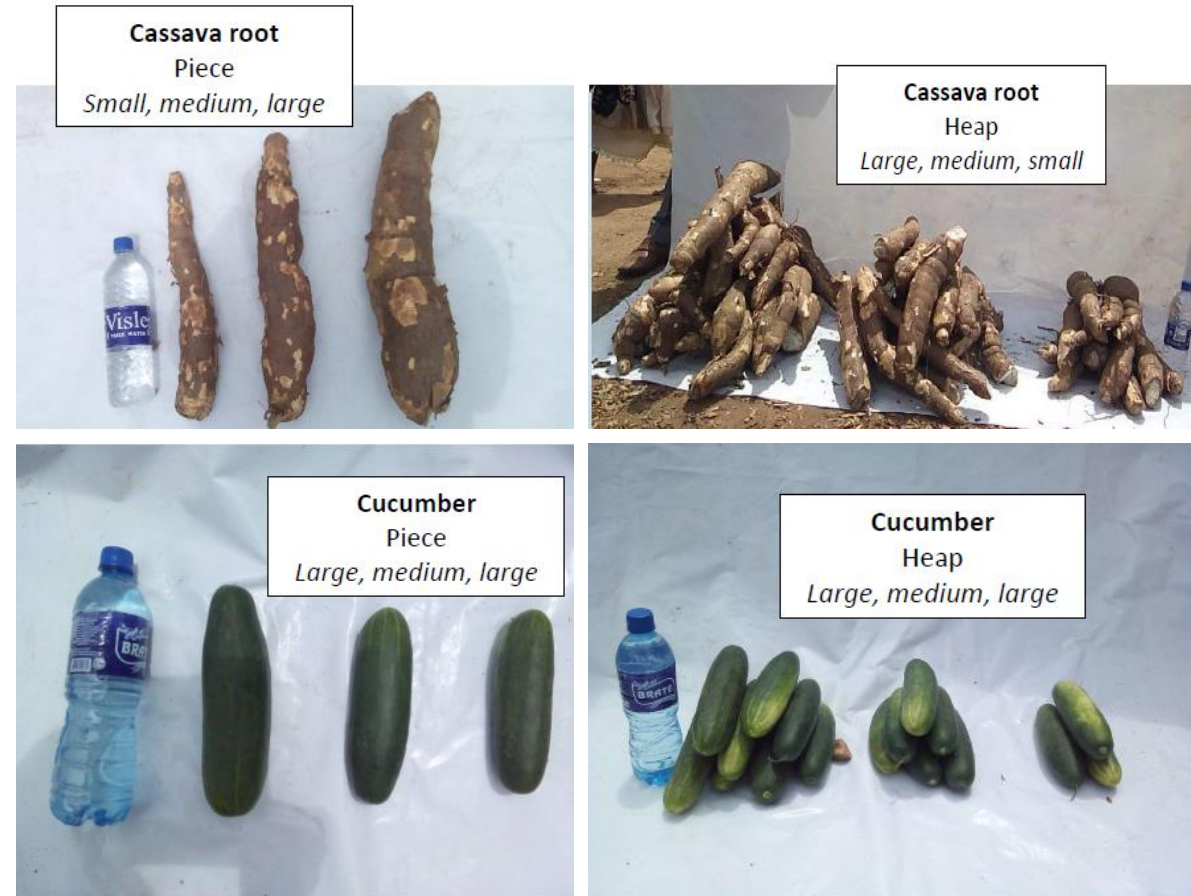
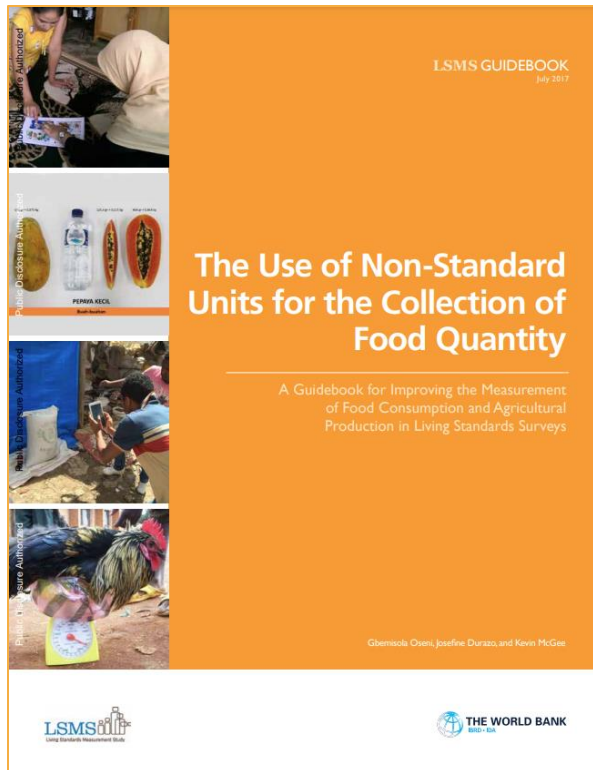
ALLOWING NON-STANDARD UNITS

Simplifies respondent's role, conversion factors are consistent



Using non-standard units for food consumption measurement

Practical guidance on non-standard units and conversion factors to measure food consumption



Key Features

- 1 Multi-topic and multi-level
- 2 Nationally-**representative** probability sample
- 3 Panel
- 4 **Multi-purpose**
- 5 **Geo-referenced and interoperable** with geospatial, admin and census/survey data
- 6 Food and Non-Food **Consumption** and **Expenditures**
- 7 Individual disaggregation**
 - » Traditionally done for education, health, labor, anthropometrics (0-59 months)
 - » Expanded focus in the recent past on **individual-specific interviews**, limiting the reliance on **proxy respondents**
- 8 Computer-Assisted Personal Interviewing
- 9 **Objective measurement**
- 10 **Public Access**



Living Standards Measurement Study Plus

Project Overview (2016-2023)

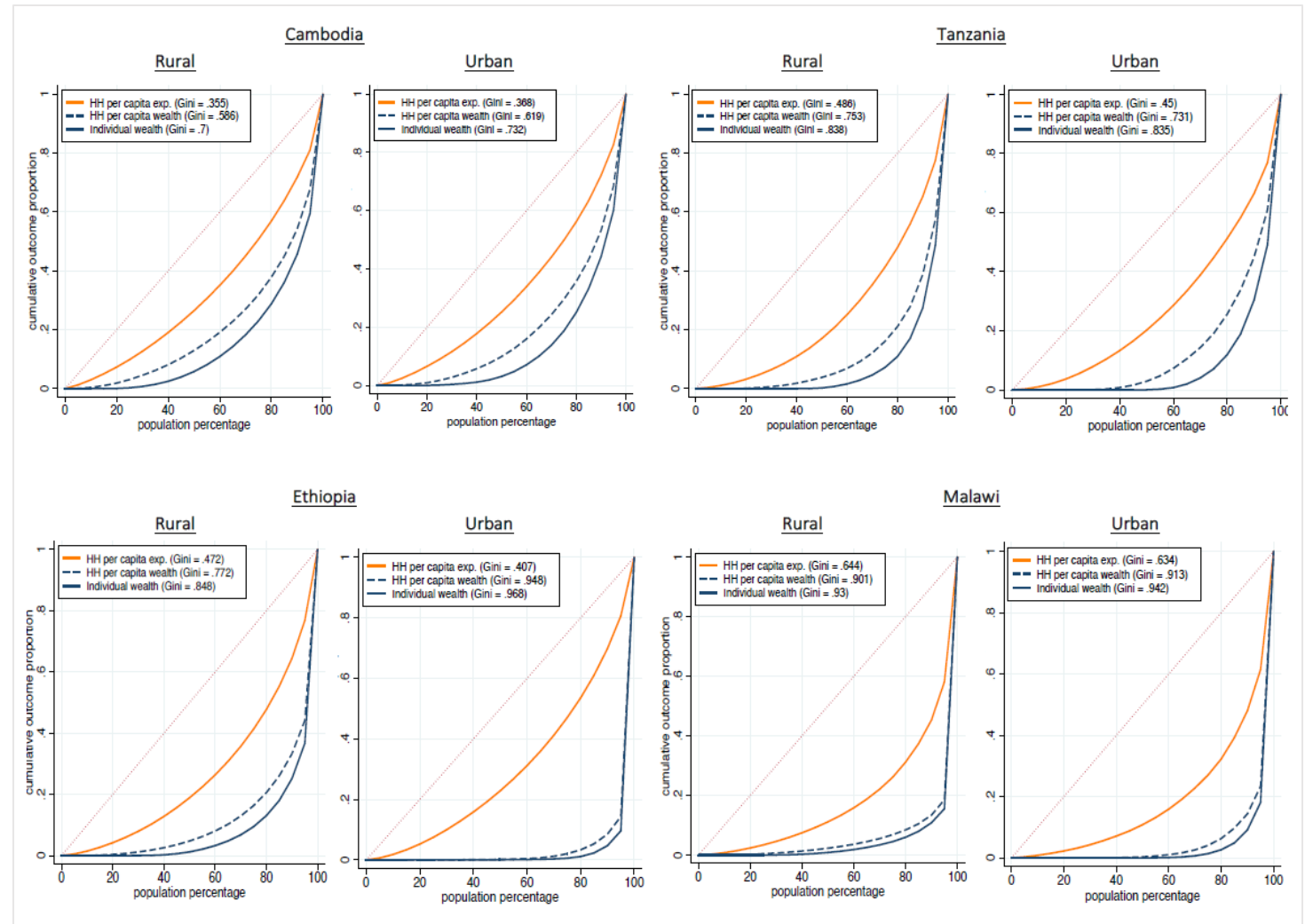
- Operationalized best practices on individual-disaggregated survey data collection on **asset ownership** and **employment**
 - Integrated into national surveys in 6 IDA countries through an individual-level questionnaire
 - Aimed to interview all adult household members in private, with enumerator-respondent gender match and simultaneous intra-household interviews (80%+ response rate)
- Informed **Strengthening Gender Statistics Project (2021-present)** to scale up recommendations as part of Statistical Capacity Projects in 12 IDA countries

National surveys supported by LSMS+

	Malawi*	Tanzania*	Ethiopia*	Cambodia*	Nepal	Sudan
Survey	2016 Malawi Integrated Household Panel Survey	2019/20 Tanzania National Panel Survey	2018/19 Ethiopia Socioeconomic Survey	2019/20 Cambodia Socioeconomic Survey	2022 Nepal Phone Survey	2022 Sudan Labor Market Panel Survey
Implementing Agency	National Statistical Office	National Bureau of Statistics	Central Statistics Agency	National Institute of Statistics	Central Bureau of Statistics	Central Bureau of Statistics
Fieldwork Period	April '16 – Dec '17	Feb '19 – Jan '20	June – August '19	Oct – Dec '19	May – June '22	Sep – Nov '22
LSMS+ Sample Size	2,508	900	7,200	1,512	756	5,000
Asset Classes Included in Data Collection	Agricultural and dwelling land, financial accounts, mobile phones	Agricultural and dwelling land, financial accounts, mobile phones	Agricultural and dwelling land, financial accounts, mobile phones, livestock	Agricultural and dwelling land, financial accounts, mobile phones, apartments/condos, consumer durables	Agricultural and dwelling land, financial accounts, mobile phones	Agricultural and dwelling land, financial accounts, mobile phones, livestock, consumer durables
Other Topics of Individual-Disaggregated Data Collection	Employment, non-farm enterprises, education, health, food insecurity	Employment, non-farm enterprises, education, health	Employment, non-farm enterprises, education, health	Employment, non-farm enterprises, education, health, time use, internal and international migration	Employment, non-farm enterprises, education, health, time use, internal and international migration	Employment, non-farm enterprises, education, health, time use, internal and international migration

Inequality revisited with LSMS+ data

Self-reported, asset-level data on asset ownership and valuation allow for revisiting inequality analysis in **Cambodia, Ethiopia, Malawi and Tanzania**:



Source: [Hasanbasri et al. \(2022\). "Individual Wealth Inequality: Measurement and Evidence from Low- and Middle-Income Countries."](#) (Forthcoming in *Review of Income and Wealth*)

Inequality revisited with LSMS+ data

Self-reported, asset-level data on asset ownership and valuation allow for revisiting inequality analysis in **Cambodia, Ethiopia, Malawi and Tanzania**:

	Within-group component of Theil T (1)	Between-group component of Theil T (2)	Share of overall wealth inequality attributable to within-household inequality $[(2) / ((1)+(2))] * 100$
Cambodia			
<i>Individual wealth specification:⁽¹⁾</i>			
(A)	0.36 [0.02]	0.92 [0.05]	28.1%
(B)	0.33 [0.02]	0.95 [0.06]	25.8%
(C)	0.35 [0.02]	0.79 [0.06]	30.7%
(D)	0.30 [0.02]	0.82 [0.04]	26.8%
Tanzania			
<i>Individual wealth specification:⁽¹⁾</i>			
(A)	0.66 [0.03]	1.28 [0.05]	34.0%
(B)	0.62 [0.04]	1.28 [0.04]	32.6%
(C)	0.42 [0.02]	1.14 [0.04]	26.9%
(D)	0.37 [0.02]	1.10 [0.04]	25.2%
Ethiopia			
<i>Individual wealth specification:⁽¹⁾</i>			
(A)	0.39 [0.32]	3.9 [0.49]	8.9%
(B)	0.58 [0.33]	3.6 [0.45]	14.1%
(C)	0.40 [0.11]	4.3 [0.51]	8.6%
(D)	0.58 [0.09]	3.9 [0.20]	13.0%
Malawi			
<i>Individual wealth specification:⁽¹⁾</i>			
(A)	0.57 [0.12]	3.4 [0.46]	14.4%
(B)	0.47 [0.11]	3.3 [0.22]	12.5%
(C)	0.55 [0.12]	3.3 [0.46]	14.3%
(D)	0.47 [0.13]	3.3 [0.3]	12.5%

Notes:

(1) Bootstrapped standard errors (250 repetitions) in brackets.

(2) For individual wealth specification, (A) = No imputation, self-reported values for joint owners; (B) = No imputation, max value for joint owners; (C) = Missing values imputed through multiple imputation, self-reported values for joint owners; (D) = Missing values imputed through multiple imputation, max value for joint owners.

Inequality revisited with LSMS+ data

Self-reported, asset-level data on asset ownership and valuation allow for revisiting inequality analysis in **Cambodia, Ethiopia, Malawi and Tanzania:**

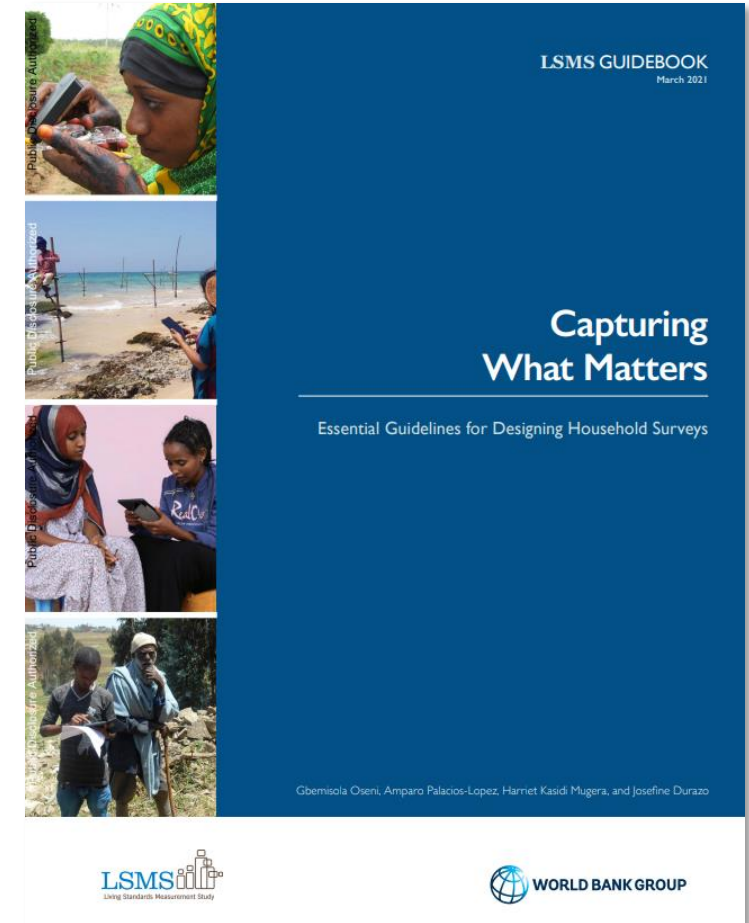
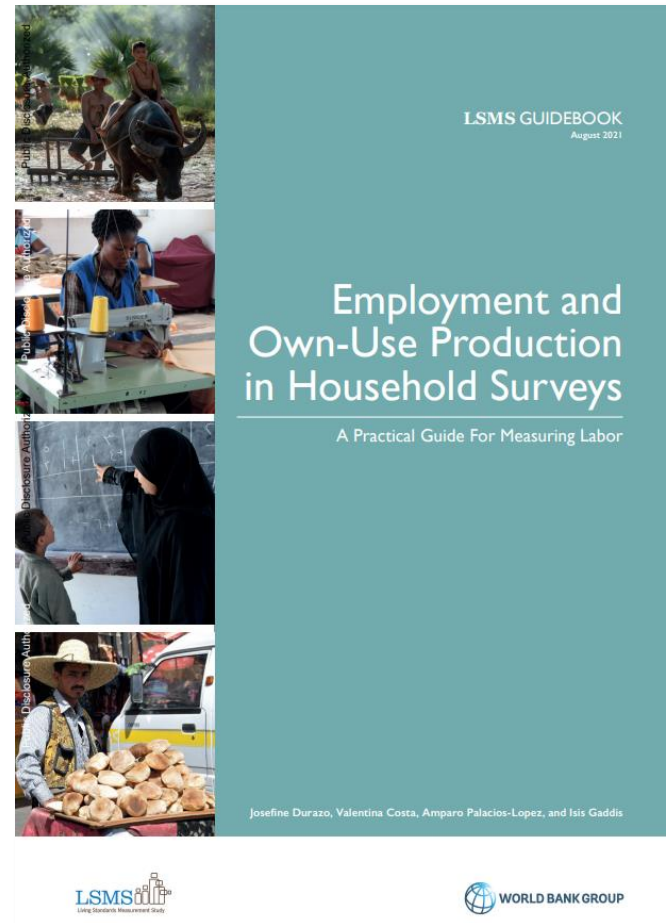
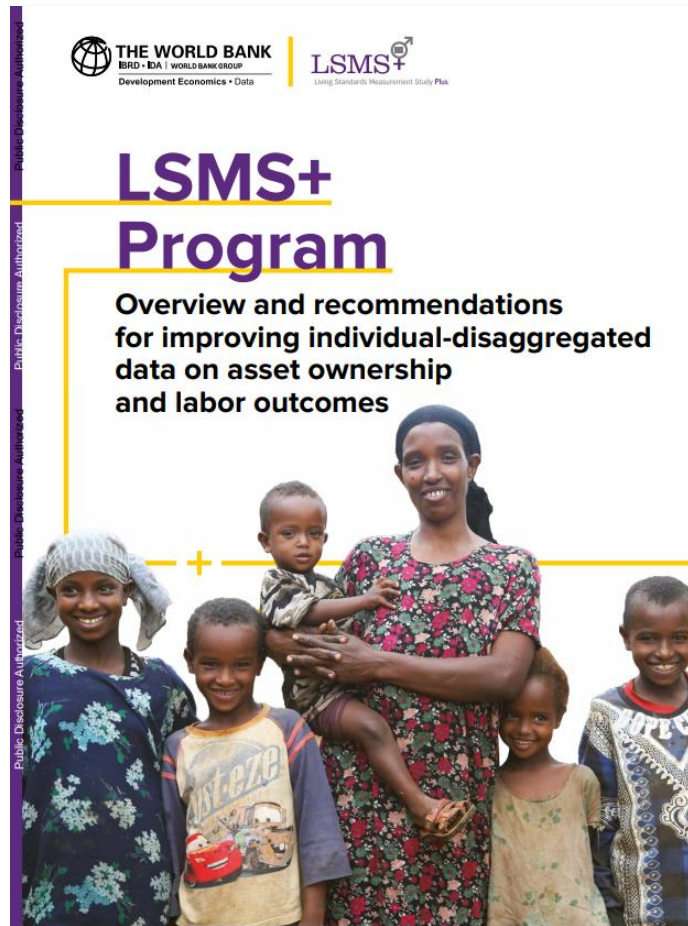
	Factor shares by individual asset class:								
	Non-residential land	Residential land	Financial accounts	Mobile phones	Livestock	Motorcycle	Car	Tractor	Other ⁽¹⁾
Cambodia									
<i>Indiv. wealth specification:⁽¹⁾</i>									
(A)	0.33	0.54	0.002	0.003	0.015	0.04	0.06	0.02	-
(B)	0.34	0.54	0.002	0.003	0.02	0.03	0.05	0.01	0.005
(C)	0.33	0.55	0.002	0.003	0.02	0.03	0.05	0.01	0.005
(D)	0.33	0.57	0.002	0.003	0.01	0.03	0.04	0.01	0.005
Tanzania									
<i>Indiv. wealth specification:⁽¹⁾</i>									
(A)	0.22	0.77	0.01	-	-	-	-	-	-
(B)	0.22	0.77	0.01	-	-	-	-	-	-
(C)	0.20	0.78	0.01	-	-	-	-	-	-
(D)	0.21	0.78	0.01	-	-	-	-	-	-
Ethiopia									
<i>Indiv. wealth specification:⁽¹⁾</i>									
(A)	0.01	0.78	0.05	0.01	0.15	-	-	-	-
(B)	0.01	0.82	0.05	0.01	0.12	-	-	-	-
(C)	0.01	0.89	0.02	0.003	0.08	-	-	-	-
(D)	0.01	0.93	0.01	0.002	0.05	-	-	-	-
Malawi									
<i>Indiv. wealth specification:⁽¹⁾</i>									
(A)	0.39	0.61	-	-	-	-	-	-	-
(B)	0.41	0.59	-	-	-	-	-	-	-
(C)	0.37	0.63	-	-	-	-	-	-	-
(D)	0.40	0.60	-	-	-	-	-	-	-

Notes:

(1) "Other" in the case of Cambodia includes computers, tuk tuks, boats and bicycles.

(2) (-) = no valuation data collected for that particular asset class.

Guidance on (individual-level) survey data collection

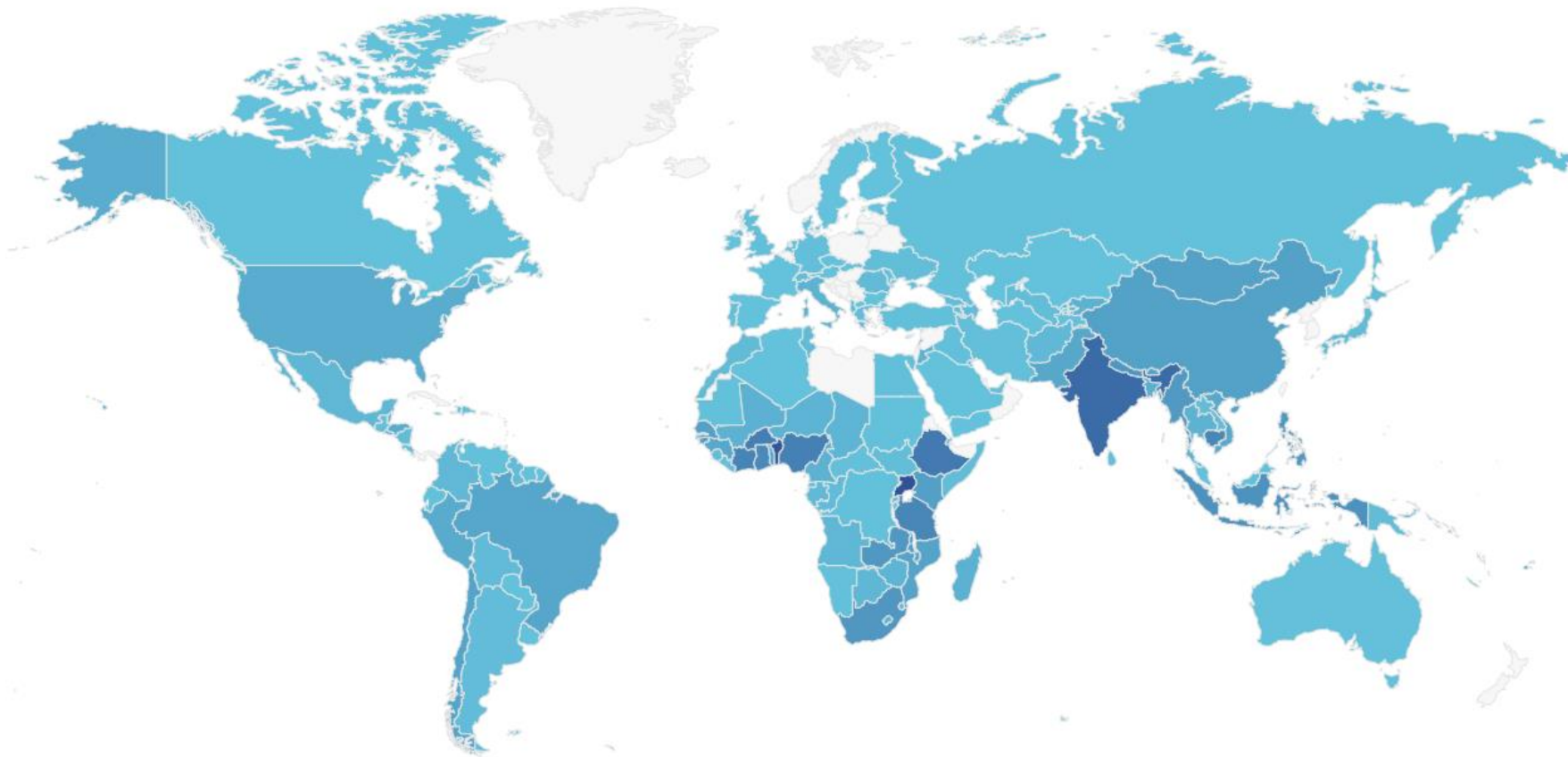


Key Features

- 1 Multi-topic and multi-level
- 2 Nationally-**representative** probability sample
- 3 Panel
- 4 **Multi-purpose**
- 5 **Geo-referenced and interoperable** with geospatial, admin and census/survey data
- 6 Food and Non-Food **Consumption** and **Expenditures**
- 7 Individual Disaggregation
- 8 Computer-Assisted Personal Interviewing (CAPI)**
 - » **World Bank Survey Solutions CAPI Software**: Automated routing, immediate data validation, platform for survey management, collection and use of paradata.
- 9 Objective measurement
- 10 Public Access

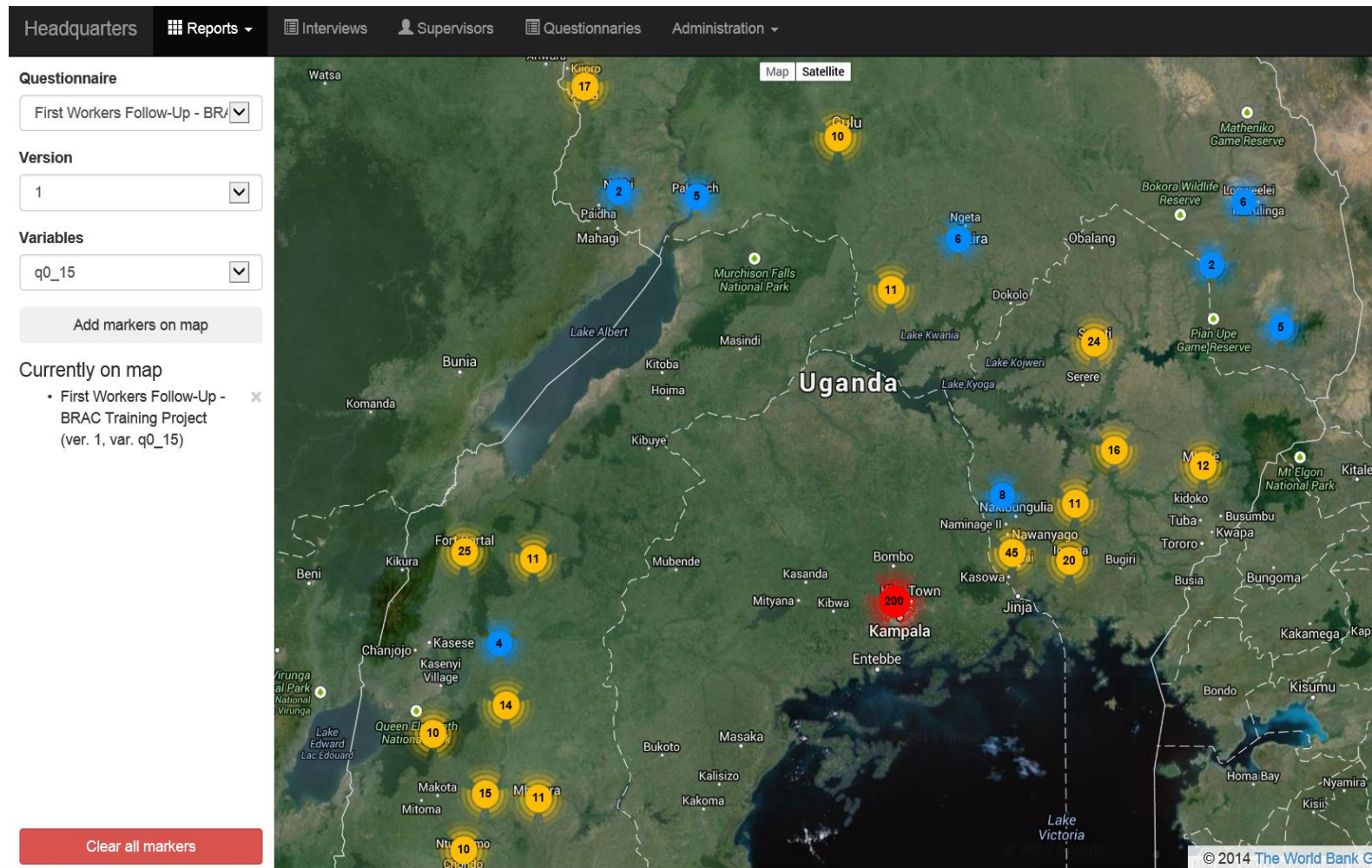


Globally trusted platform for data collection





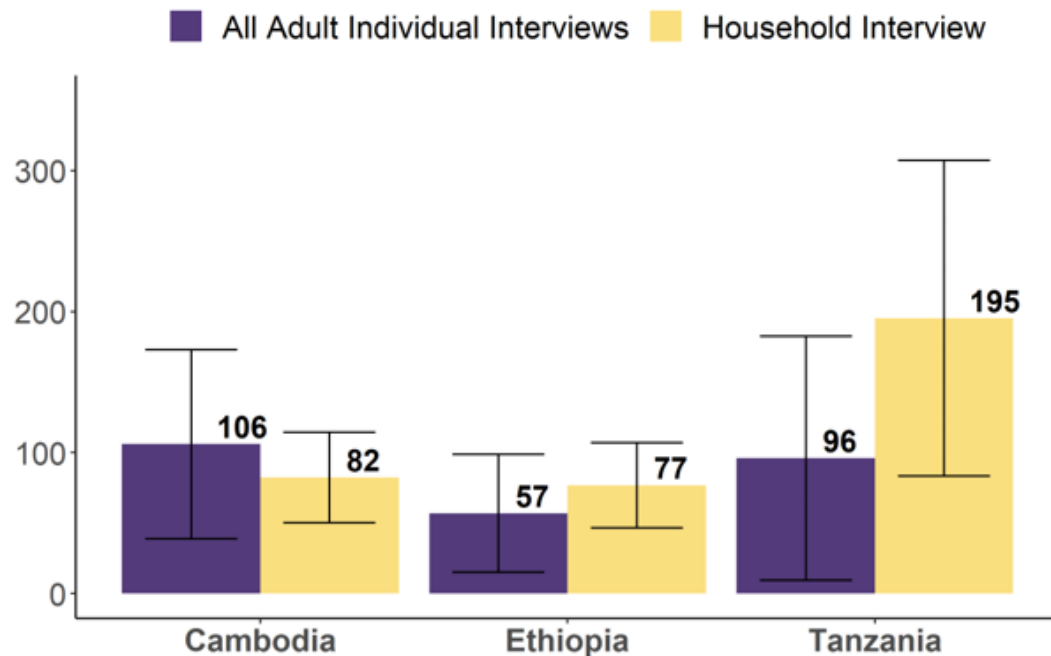
Using GIS capabilities for quality assurance



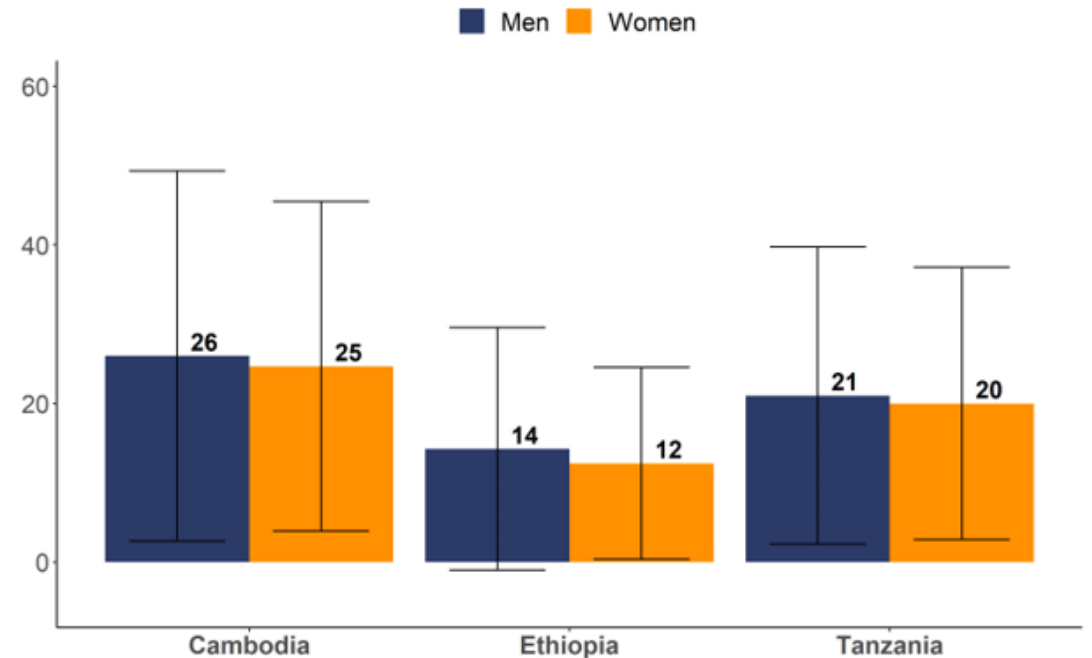
Insights from Survey Solutions paradata

Average Duration (Mins) of Adult Individual Questionnaire Modules: Tanzania NPS 2018/19

1a. Average Time Spent at Household



1b. Average Time per Adult Individual Interview



Source: Hasanbasri, Kilic, Koolwal and Moylan (2023). "Using Paradata to Assess Respondent Burden and Interviewer Effects in Household Surveys Evidence from Low- and Middle-Income Countries" World Bank Policy Research Working Paper No. 10456. Forthcoming in *Statistical Journal of the IAOS*.

Insights from Survey Solutions paradata

Table 3: Average Cost Estimates of Three LSMS+ Surveys

Average Cost (in 2019 USD)	Cambodia LSMS+ 2019/2020	Ethiopia ESS 2018/2019	Tanzania NPS 2019/2020
<ul style="list-style-type: none"> Per Minute: Average Duration of a Household Interview Average Cost of a Household Interview Average Duration of an Individual Interview Average Cost of an Individual Interview 	<p>\$0.87</p> <p>82 minutes \$71</p> <p>25 minutes \$22</p>	<p>\$1.71</p> <p>77 minutes \$131</p> <p>13 minutes \$23</p>	<p>\$3.94</p> <p>120 minutes \$472</p> <p>20 minutes \$80</p>

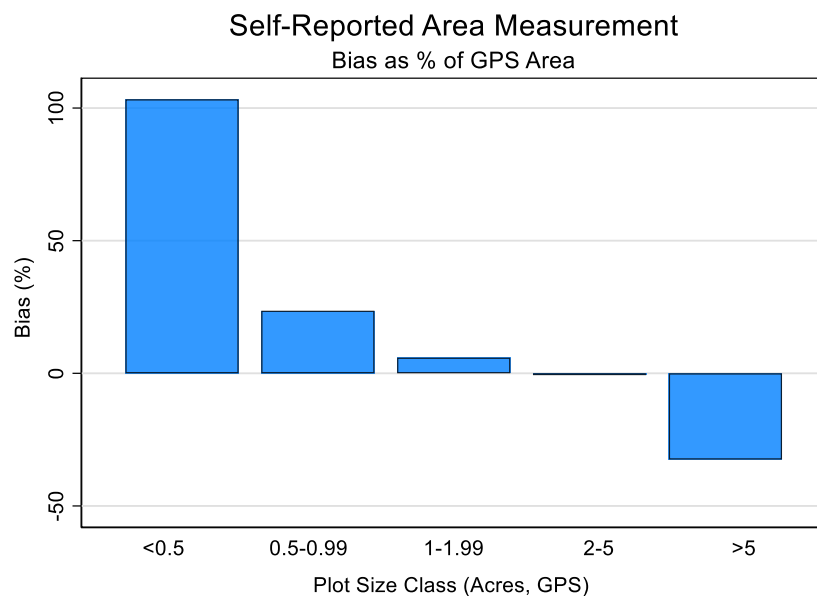
Source: Hasanbasri, Kilic, Koolwal and Moylan (2023).” [Using Paradata to Assess Respondent Burden and Interviewer Effects in Household Surveys Evidence from Low- and Middle-Income Countries](#)” World Bank Policy Research Working Paper No. 10456. Forthcoming in *Statistical Journal of the IAOS*.

Key Features

- 1 Multi-topic and multi-level
- 2 Nationally-**representative** probability sample
- 3 Panel
- 4 **Multi-purpose**
- 5 **Geo-referenced and interoperable** with geospatial, admin and census/survey data
- 6 Food and Non-Food **Consumption** and **Expenditures**
- 7 Individual Disaggregation
- 8 **Computer-Assisted Personal Interviewing**
- 9 Objective measurement**
 - » **Use of GPS** for georeferencing households, plots, facilities + land area measurement
 - » On-going work on objective measurement, augmented with imputation
 - Crop cutting for yield estimation
 - Mobile phones for high frequency data capture (e.g., tree and root crop production)
 - Sensors for data collection on soil quality and climate
 - Accelerometers for physical activity tracking
 - DNA fingerprinting for crop variety identification
- 10 **Public Access**

Land Area

- **Self-reported land areas:** Common but potentially error-prone
- Farmers systematically over-estimate small plot areas, while under-estimating areas of larger plots



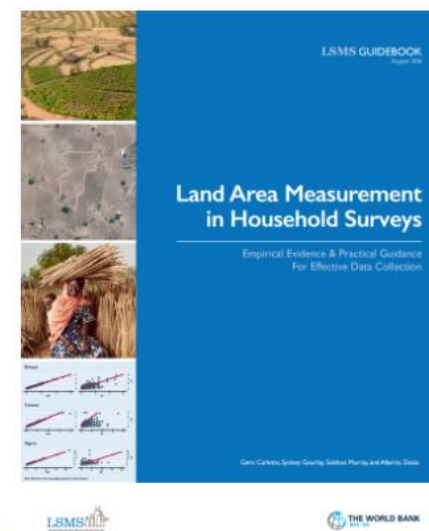
Systematic measurement error in farmer reporting alters the conclusions of research.

Fact or artifact. Calogero Carletto, Sara Savastano, Alberto Zezza. *Journal of Development Economics*, 2013.

From Guesstimates to GPStimates. Calogero Carletto, Sydney Gourlay, Paul Winters; *Journal of African Economies*, 2015.

Cheaper, Faster, and More Than Good Enough. Calogero Carletto, Sydney Gourlay, Siobhan Murray, Alberto Zezza. *Survey Research Methods*, 2017.

Missing(ness) in Action. Talip Kilic, Alberto Zezza, Calogero Carletto, Sara Savastano. *World Development*, 2017.



Key Features

- 1 Multi-topic and multi-level
- 2 Nationally-**representative** probability sample
- 3 Panel
- 4 **Multi-purpose**
- 5 **Geo-referenced and interoperable** with geospatial, admin and census/survey data
- 6 Food and Non-Food **Consumption** and **Expenditures**
- 7 Individual Disaggregation
- 8 **Computer-Assisted Personal Interviewing**
- 9 **Objective measurement**
- 10 Public Access**
 - » Target: data release within 6 months
 - » Deadline: data release within 12 months
 - » Including geo-variables

The Living Standards Measurement Study (LSMS)

The World Bank's flagship household survey program since 1980.

[Home](#)
[About LSMS](#)
[Work Areas](#)
[Projects](#)
[Rome Center](#)
[Publications](#)
[Training](#)
[News & Events](#)


LSMS DATA CATALOG

Access Our Data 

LSMS data are publicly available and open access: explore the full datasets in the World Bank's Microdata Catalog



HIGH-FREQUENCY MONITORING

DATA

High-Frequency Phone Surveys

LSMS is supporting high-frequency phone surveys in Burkina Faso, Ethiopia, Malawi, Mali, Nigeria, Tanzania and Uganda to track responses to and socio-economic impacts of COVID-19 and economic shocks.



A unique system of longitudinal surveys designed to improve the understanding of household and individual welfare, livelihoods, and smallholder agriculture in Africa.

VIDEO

LSMS-ISA: 15 years of impact in development

The LSMS-ISA is a unique system of longitudinal surveys designed to improved the understanding of livelihoods, welfare and smallholder agriculture in Africa.

[Home](#) / [Central Data Catalog](#) / [Living Standards Measurement Study \(LSMS\)](#)

Living Standards Measurement Study (LSMS)

[about](#)[← Central Data Catalog](#)

The LSMS is a research project that was initiated in 1980. It is a response to a perceived need for policy relevant data that would allow policy makers to move beyond simply measuring rates of unemployment, poverty and health care use, for example, to understanding the determinants of these observed social sector outcomes.

Years



Showing 1-15 of 338,988 variables

Study view

Variable view

Relevance



Countries/Economies



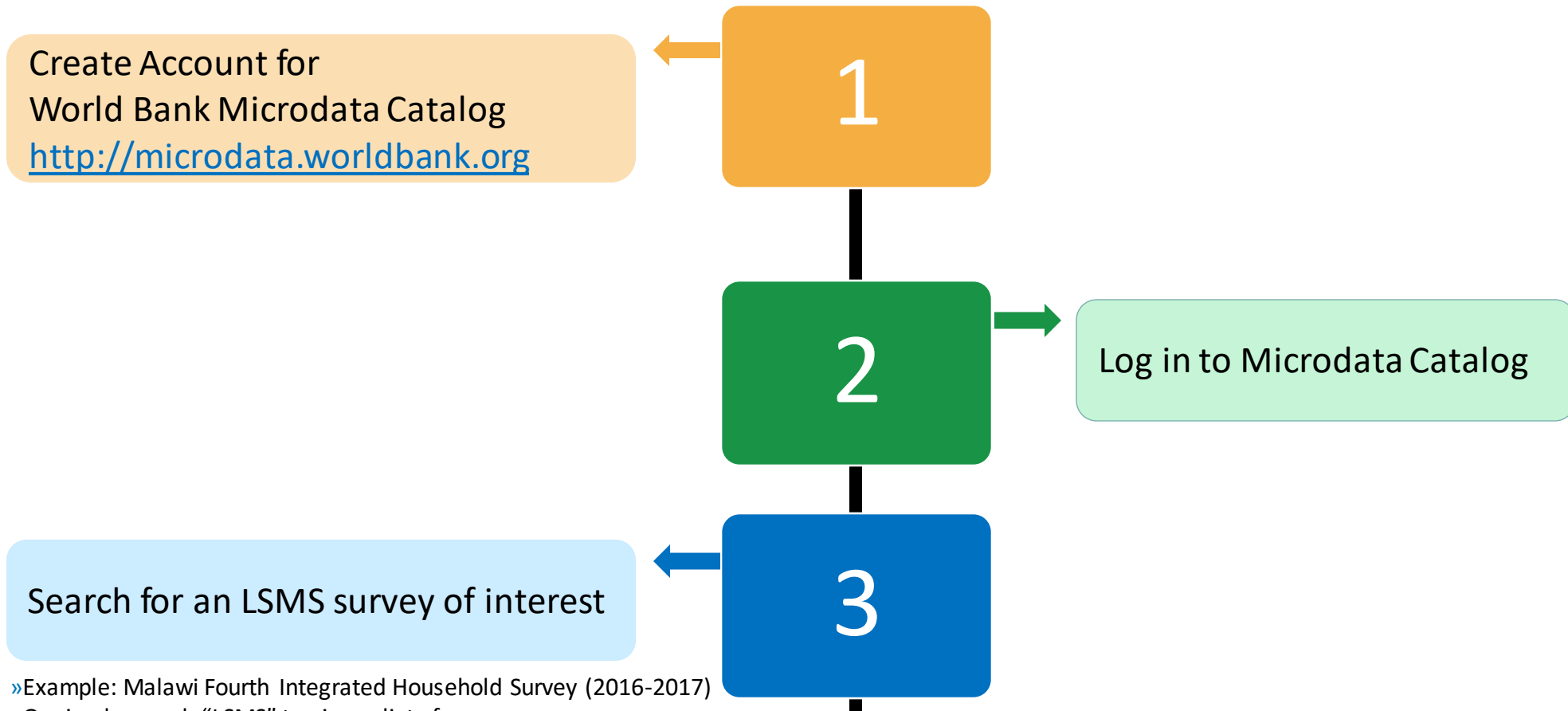
Living Standards Measurement Study (LSMS) ×

Reset search

License

<http://microdata.worldbank.org>

How to Access



- »Example: Malawi Fourth Integrated Household Survey (2016-2017)
- »Or simply search “LSMS” to view a list of surveys
- »File structure corresponds to the questionnaire modules
- »Variable names correspond to the question numbering



2

Methodological Innovations



A. Time Use



B. Labor Inputs



C. Crop Yields





A. Time Use



Time Use

Motivation

- Time use data are key to understanding economic decision-making among men and women, and highlighting inequities within and across households
- Surveys in low- and middle-income countries rely on recall-based interview-based diaries or stylized questions to measure time use
- Dearth of evidence on how recall may affect time use data accuracy, particularly in contexts with low literacy and numeracy



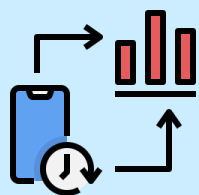
Our Study

Survey experiment in Malawi to examine how innovative techniques could sidestep measurement errors in recall-based time use survey methods

- Households randomly assigned to one of two treatment arms in each community
 - **Traditional time-use diary with a 24-hour recall**
 - » Adapted from IFPRI WEAI Time Use Module
 - **Self-administered smartphone-based pictorial time use diary – TimeTracker**
 - » Revamped the app originally developed for low-literacy settings
- One adult man and one adult woman in each household
- 11-day continuous reporting for *TimeTracker* vs. repeated (3) measures for recall, distributed across the same reporting period in each community
- Both arms received a stylized 7-day recall time use module in the final interview
- **Multi-disciplinary effort:** App development, illustration design and training + Survey design, implementation and analysis



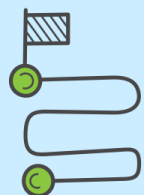
Time Use Contributions



Document relative accuracy of recall-based approaches to time use data collection vis-à-vis (highly successful) real time data collection



Extend a nascent literature on smartphone-based time use data collection in low- and middle-income countries



Advance the piloting of improved approaches to time use data collection that can be considered for adoption in large-scale household surveys

POLICY RESEARCH WORKING PAPER

10695

Recording the Time Divide

A Comparative Study of Smartphone- and Recall-Based Approaches to Time Use Measurement

*Talip Kiliç
Gayatri Koolwal
Wilbert Drazzi Vundru
Thomas Daum
Hannes Buchwald
Greg Seymour
Peter Mvula
Alister Munthali
Monice Kachinjika*

 **WORLD BANK GROUP**
Development Economics
Development Data Group
February 2024

Time Use

TimeTracker App

- Android app originally developed by the University of Hohenheim and the Institute for Applied Science at the University of Media, Stuttgart (Daum et al., 2018) and enhanced for this study with LSMS support
- Real-time recording of time use, allowing for simultaneous activities
 - » 84 percent of activities had at least one other activity conducted simultaneously
- Smartphones received by respondents only have the app installed; one-to-one match in activity categories with recall treatment arm
- Separate sets of images for men and women respondents + scrolling captions in Chichewa

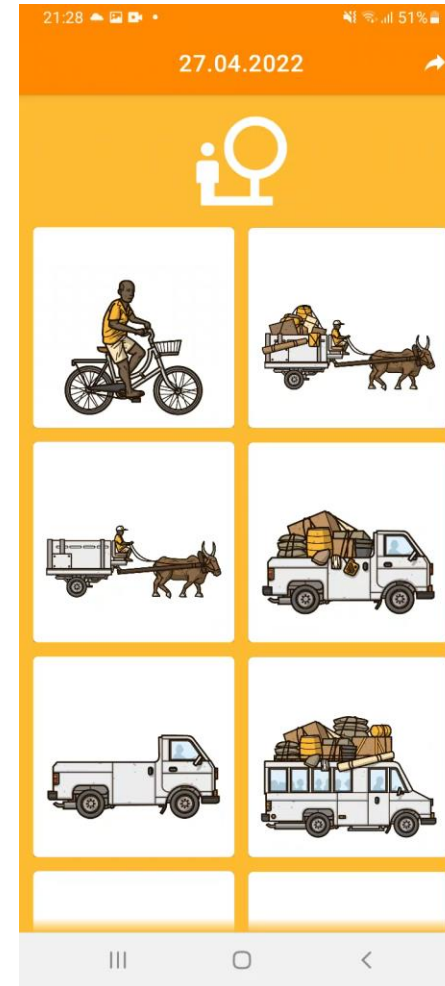


Time Use

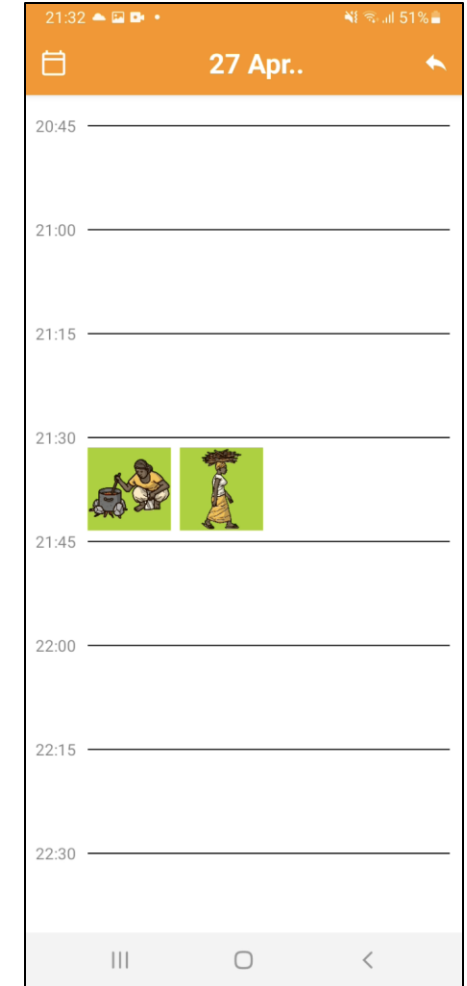
TimeTracker App (2)

- Data are downloaded as excel files, transferred from the application to a laptop via local Wi-Fi network
- Allows enumerators/ researchers to cross-check data in the field
- Ease of adding/removing activities/illustrations- adaptable to other contexts

Main screen



Calendar view



Time Use

Key takeaways

- Participation rates in smartphone arm are greater across several categories of employment, unpaid work, while reported time estimates are higher in 24-hour diary – due to the minimum 15-minute increments in which activities can be reported
- Gender gaps in unpaid work remain large in smartphone sample, but narrower somewhat in household resource collection and care
- Smartphone arm captures more time in evening hours, and multitasking, esp. for women
- 7-day recall overreports relative to both 24-hour recall and smartphone – implications for interpreting standard stylized instrument commonly used in labor force modules



B. Labor Inputs



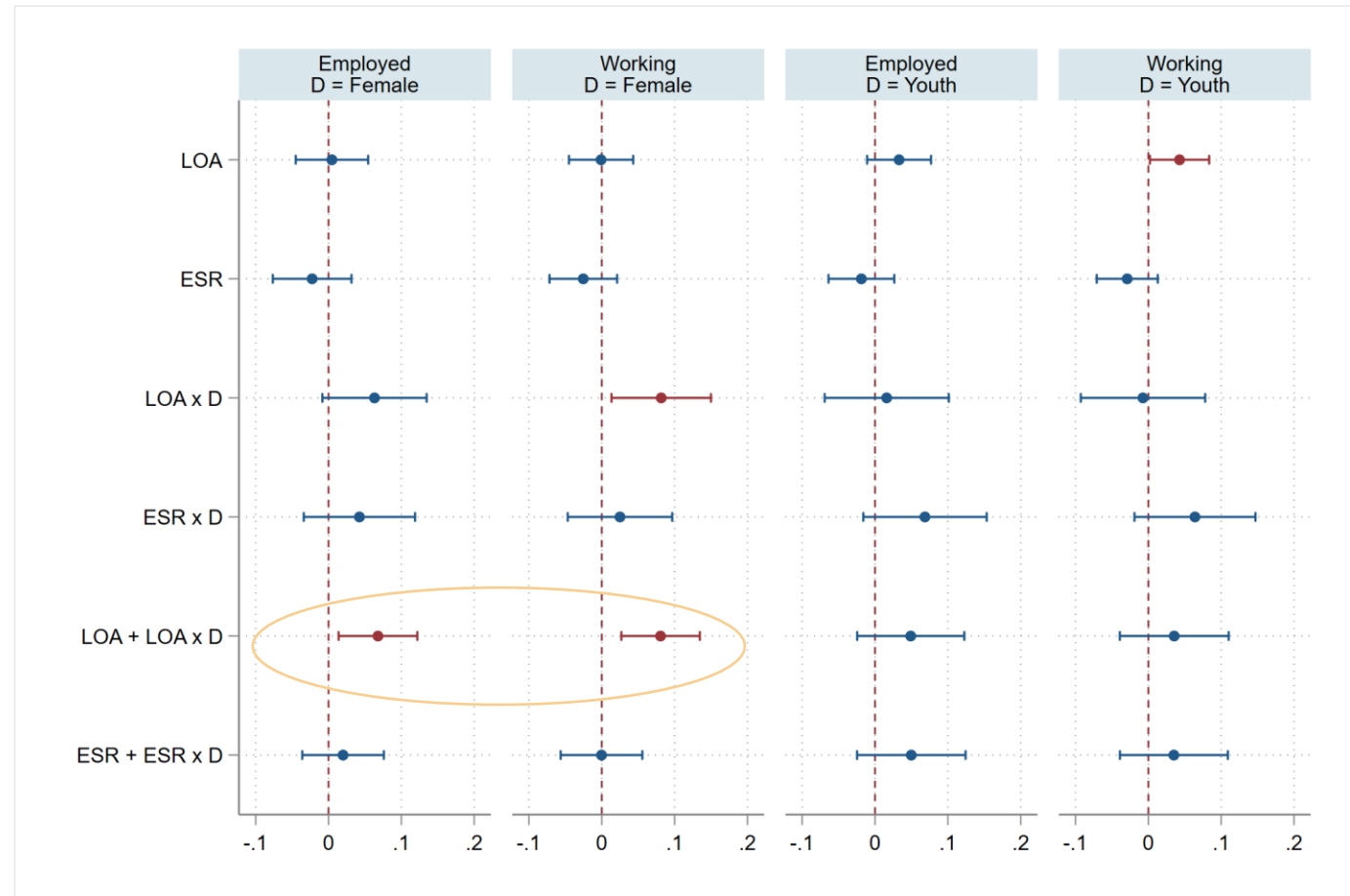
Improving labor measurement for vulnerable groups

- Contreras et al. (2024) conducted a randomized survey experiment in El Salvador that compares two alternative survey methods—a List of Activities survey module (LOA) and enforced self-responses (ESR)—against a traditional household survey, which consists of proxy responses without a LOA module.

Group	Type of Labor Module	Type of Respondent
Control Group	Standard labor module (following 19 th ICLS)	Proxy respondents permitted for unavailable household members. If available, each household member (15-64 years old) responds on his/her own.
List of Activities (LOA)	Standard + LOA	Proxy respondents allowed.
Enforced self-responses (ESR)	Standard labor module (following 19 th ICLS)	Self-reporting is enforced for eligible household members (15-64 years old).

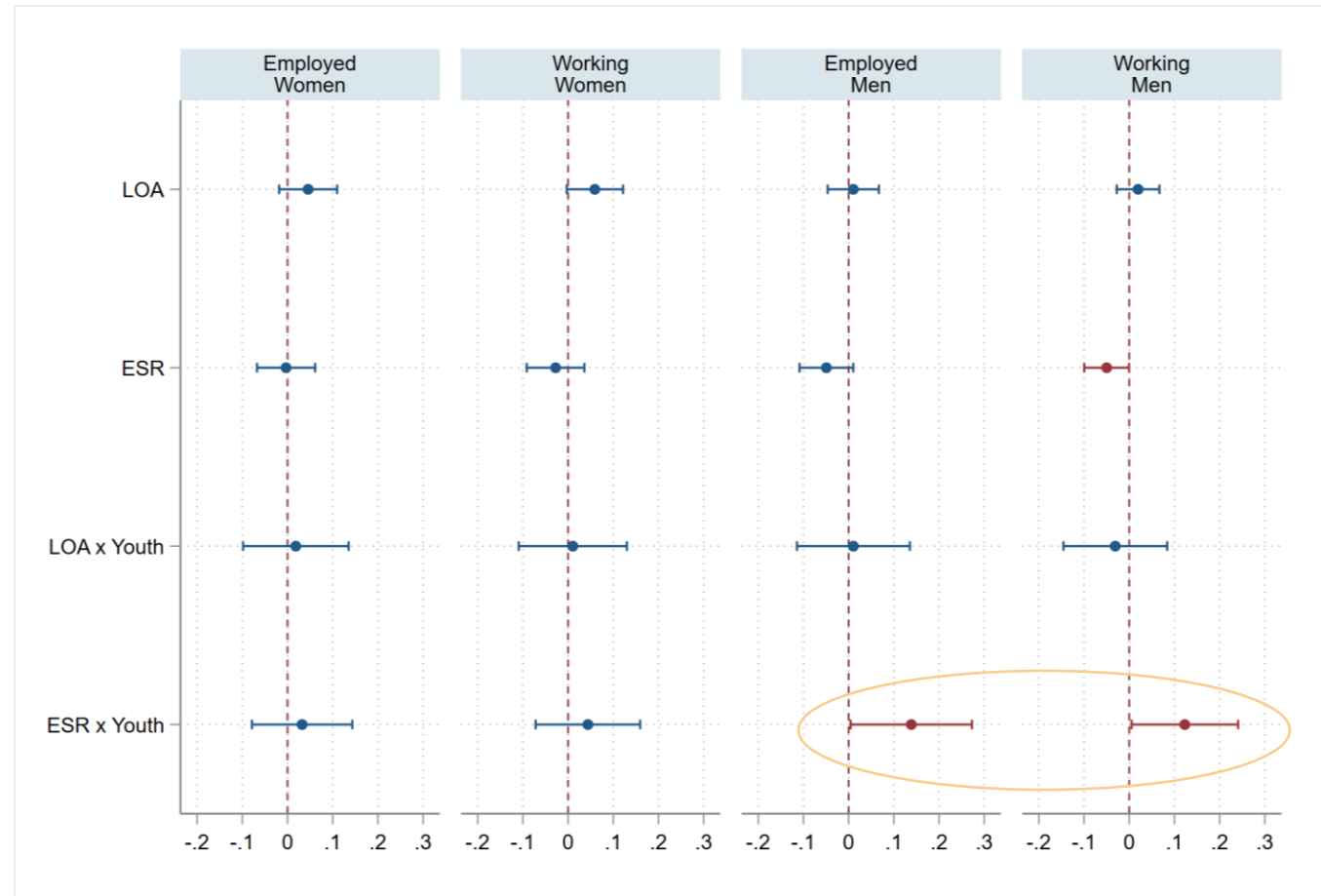
Improving labor measurement for vulnerable groups

Women's employment rate is higher when they are given examples of work activities (LOA) that contribute to their own identification as part of the working population



Improving labor measurement for vulnerable groups

Enforcing self-reporting increases the share of young males reporting to be employed (compared to older males).



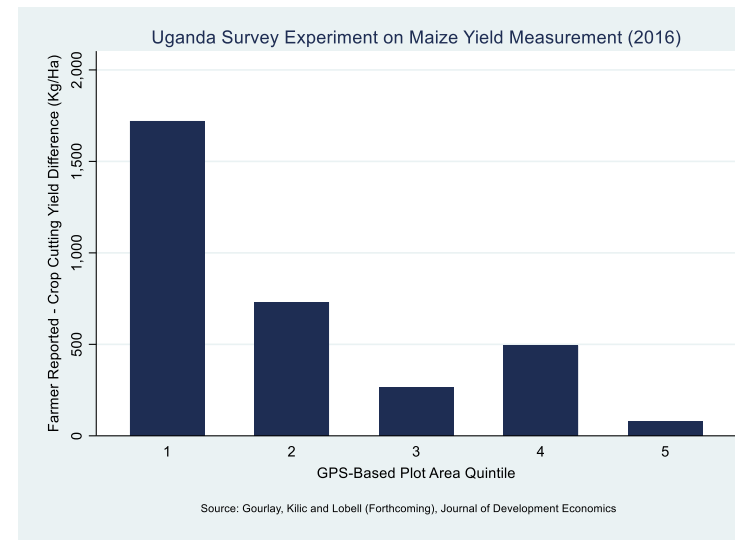
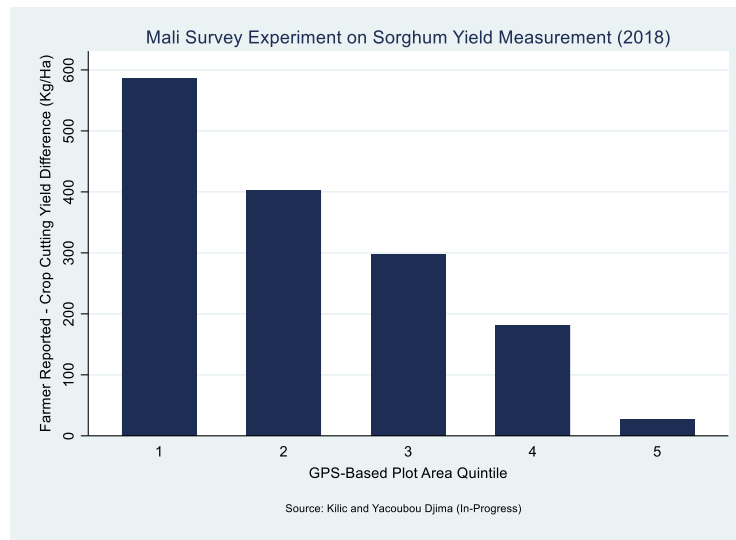
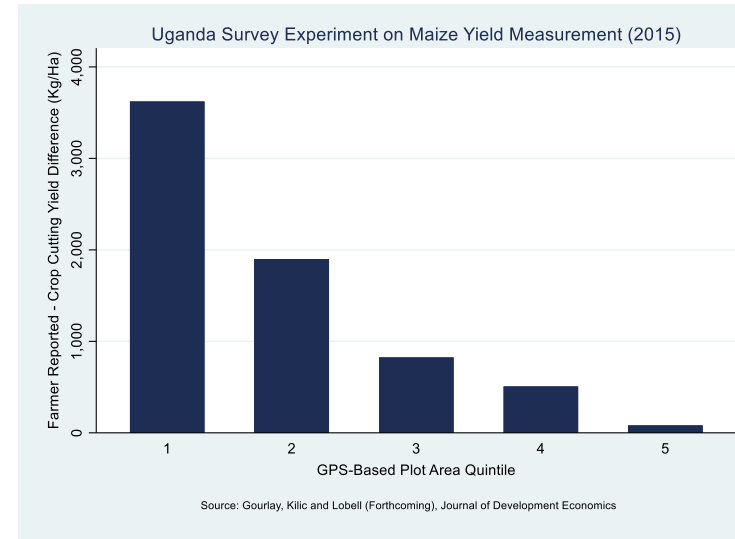
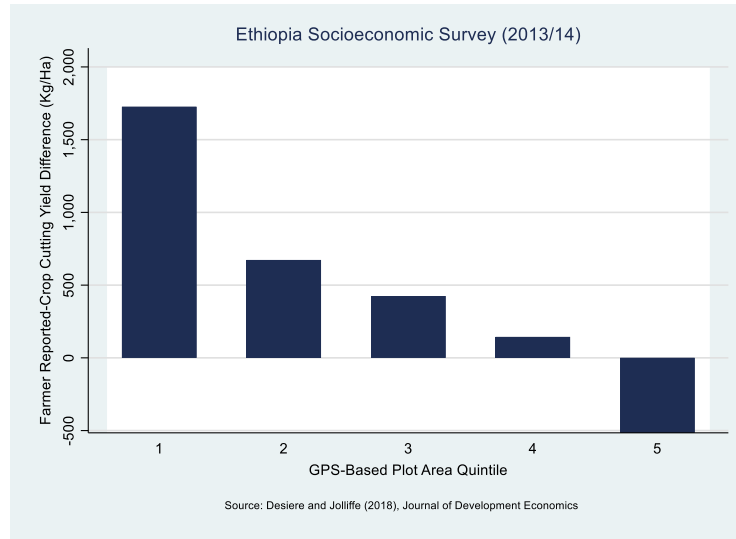


C. Crop Yields



Crop Yields

Seasonal Crop Yields

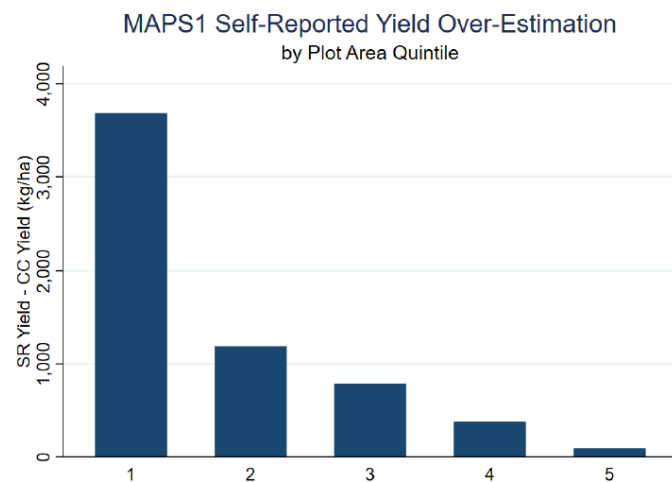


Crop Yields

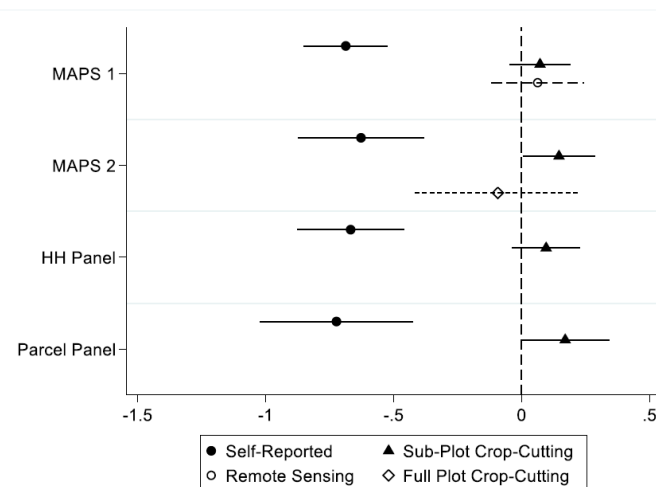
Measurement errors & implications for inference

[A New Spin on an Old Debate? Errors in Farmer-Reported Production and Their Implications for the Inverse Scale-Productivity Relationship in Uganda](#)

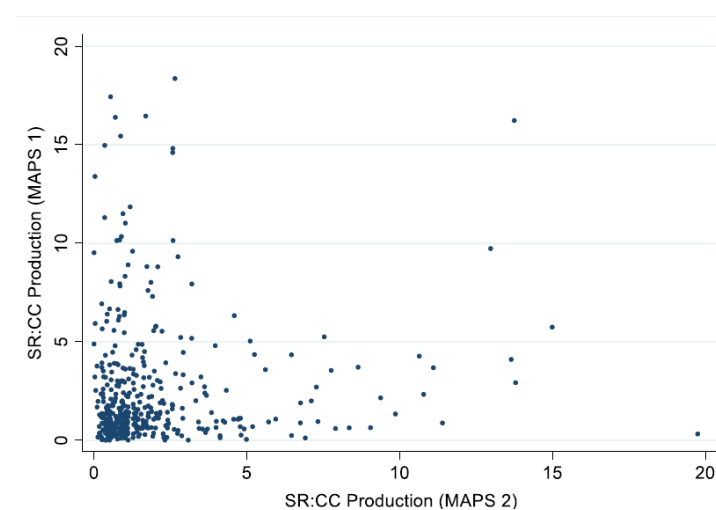
Journal of Development Economics, 141, 2019. (Open-Access)



Selected Plot Area Coefficients w/ 95% Confidence Intervals - All Plots



Consistency of Bias Across MAPS I and II



Motivation: The observation that smaller plots/farms are more productive than their larger counterparts has puzzled researchers for nearly a century. This relationship, if not merely an artifact of measurement error or market failures, has direct implications for the design of poverty reduction and land distribution policies in smallholder systems.

Research: Using data from a methodological experiment conducted in Eastern Uganda, in which maize production, maize variety, plot area, and soil quality were measured both objectively and subjectively, we propose that this inverse relationship between area and productivity is due to errors in farmer-reported crop production.

Findings: Systematic measurement error in farmer-reported production generates the inverse relationship, primarily due to significant farmer over-estimation of production on small plots. Robust to a range of sensitivity analyses. On average, actual yields may be nearly 1,400 kg/ha lower than previously believed.



3

Looking Forward



A. Data Integration



B. Resilient Futures

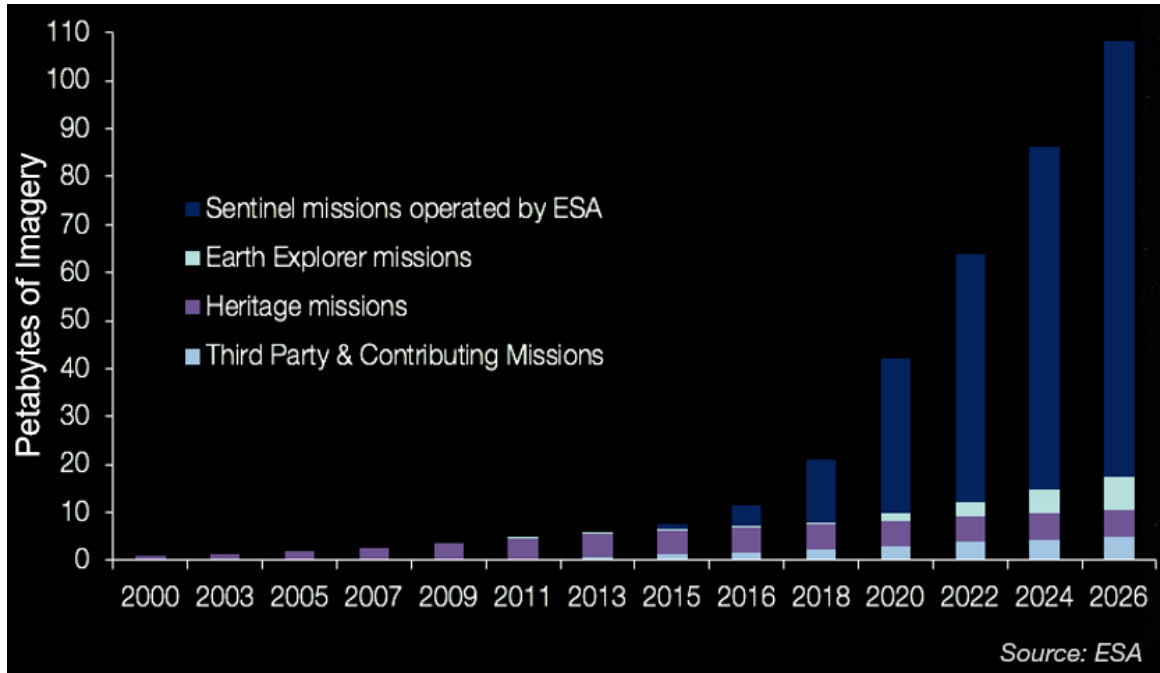


A. Data Integration



Why integrate surveys with alternative data sources? (1)

Increased availability of satellite imagery and adoption of ICT innovations - including cell phones, social media platforms digital transactions, and mobile money - particularly in low- and middle-income settings.



Why integrate surveys with alternative data sources? (2)

- **Result:** Wealth of alternative data sources that are increasingly available and suited to a wide range of secondary uses, though with unique accompanying challenges
- **Alternative data sources are attractive for several reasons:**
 - » **“Big data”**. Wide reach and scope. Growing use of cell phones and social media platforms.
 - » **“Always on”**. Typically collected at much higher frequency vis-a-vis public intent data.
 - » **“Zoomed in”**. Disaggregated information specific to individuals and localities.
 - » **“Potentially less biased”**. May be less prone to social desirability bias than personal interviews.

Why integrate surveys with alternative data sources? (3)

Integrating surveys with alternative data sources can deliver on several fronts:

1) Enhance **analytical value** and **cost-effectiveness** of surveys

2) **Calibrate** and **validate** AI-powered insights to serve humanity as a whole

3) Enhance **spatial disaggregation** and **temporal frequency** of specific indicators

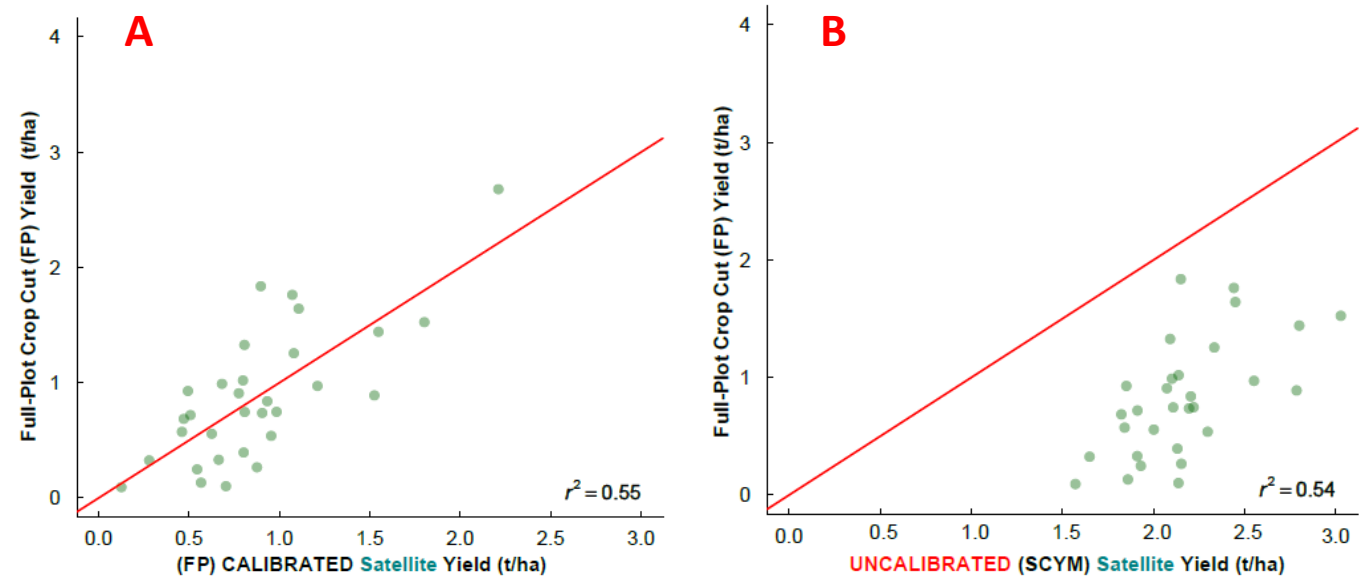
Data Integration

Measuring Agricultural Productivity, with Eyes in the Sky and Boots on the Ground

Can we leverage the latest advances in satellite imagery and remote sensing to generate accurate crop yields in smallholder systems?

Integration of georeferenced LSMS survey data and publicly-available high-resolution Sentinel-2 satellite imagery reveals the importance of calibrating remote sensing models.

Uganda Plot-Level Ground (Crop Cutting, Survey-Based) versus Satellite-Based Maize Yields



Source: Lobell et al. (2019). "Eyes in the Sky, Boots on the Ground", *AJAE*

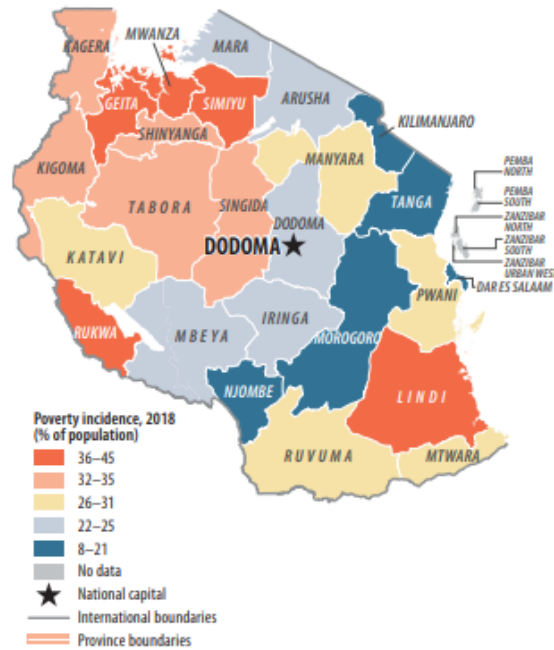
Data Integration

More examples to motivate our future work

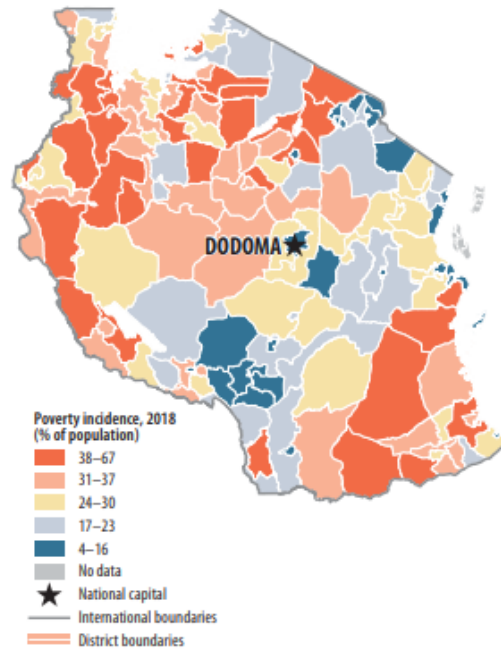
Integration of Tanzania Household Budget Survey Data with Publicly Available Satellite Imagery

Map O.3 Combining satellite imagery with household survey data increases the resolution of the poverty map of Tanzania

a. Poverty map using the Household Budget Survey (20 regions)

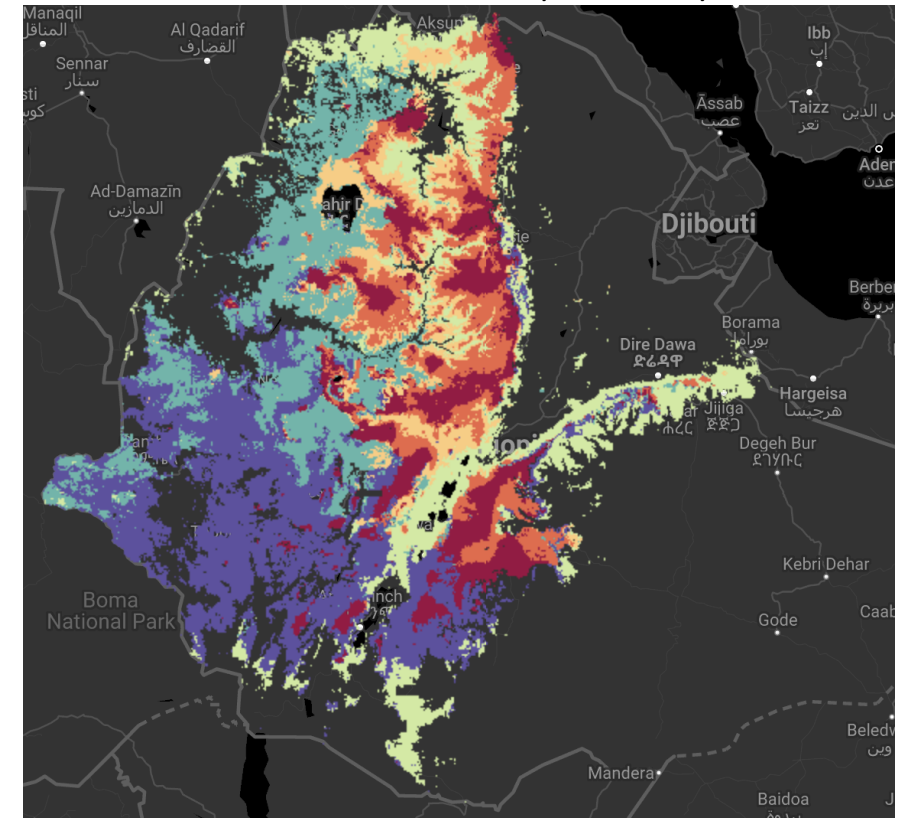


b. Poverty map combining the data in panel a with satellite imagery (169 districts)



Integration of Ethiopia Socioeconomic Survey Data with Publicly Available Satellite Imagery

10-meter Resolution Crop Area Maps





B. Resilient Futures



Resilient Futures: Next Phase of Our Panel Surveys

- Low-income and vulnerable households bear the brunt of the polycrisis era
- Boosting resilience from the bottom-up is critical to support the vulnerable
- To understand resilience, we need to look beyond consumption/income, and also consider **health, education, financial and physical assets, social protection**, and access to improved **energy** sources, **WASH** services and **digital** technologies.
- Despite the importance of resilience for weathering recurrent crises, data are often **inadequate** to design well-targeted resilience-building programs
 - Identifying who will be most affected by shocks is inherently difficult
 - Advancements in remote sensing and geospatial data do not resolve the difficulty of identifying at-risk households and understanding how they cope, adapt or when they cannot
 - Large-scale household surveys used for poverty monitoring are not designed to provide representative and **dynamic** insights regarding resilience to and impacts of climate shocks
- Providing timely support that boosts resilience requires agile data systems that are pre-established and can be activated in response to shocks
 - These systems are not the norm, and most countries lack the capacity and resources to develop them

Resilient Futures: Next Phase of Our Panel Surveys



- Aim to support countries in building **agile data systems** that produce the evidence for designing and monitoring interventions that **strengthen resilience** and **target the most vulnerable**
- Starting in 15 countries with high-levels of socioeconomic and environmental vulnerability, **Resilient Futures** will build state-of-the-art national **longitudinal high-frequency living standards surveys** – conducted both in person and via phone, that
 - Capture levels of and changes in resilience and vulnerability in multiple dimensions
 - Deliver the spatial and temporal granularity to meet the knowledge gaps on localized impacts of shocks
 - Supply ground-truth data to calibrate and validate AI-powered insights
 - Are georeferenced for seamless integration with geospatial climatic and environmental data
 - Provide opportunities for in-situ sensor deployments for highly granular climate data
 - Build on existing longitudinal survey systems supported by the LSMS
- **Target outcomes:**
 - **Increased country and task team capacity** to respond to shocks more quickly and effectively;
 - **Increased financial leverage and development impacts** achieved by countries through the design and implementation of better targeted and proactive interventions

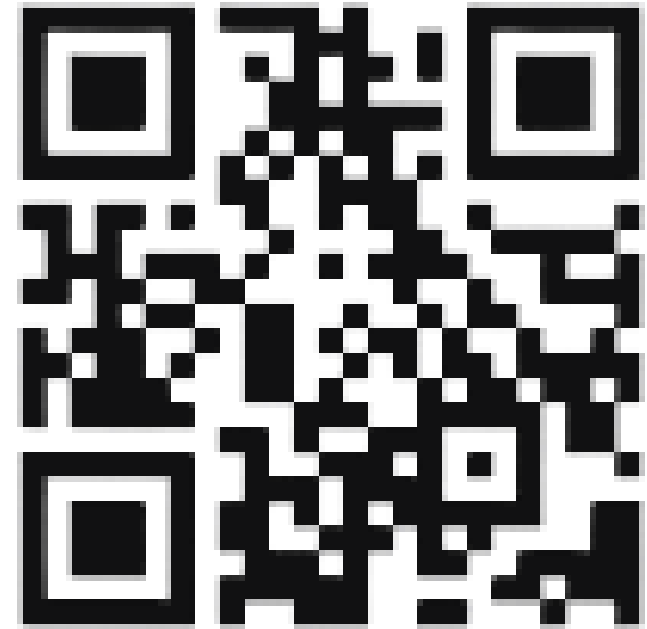
Stay connected with LSMS

GET UPDATES

 Worldbank.org/lsms

 World Bank's Living Standards
Measurement Study

 lsms@worldbank.org



Microdata for Macro Economists: An Introduction to the **Living Standards Measurement Study**

Talip Kilic

Senior Program Manager

Development Data Group | World Bank

tkilic@worldbank.org

Structural Transformation and Economic Growth (STEG) Virtual Lecture

04/12/2024