

## Making the most of imperfect data: a critical evaluation of standard information collected in farm household surveys

Simon Fraval, James Hammond, Jannkie Wichern, Simon J. Oosting, Imke J. M. De Boer, Nils Teufel, Mats Lannerstad, Katharina Waha, Tim Pagella, Todd S. Rosenstock, Ken E. Giller, Mario Herrero, David Harris, Mark T. Van Wijk

### Angaben zur Veröffentlichung / Publication details:

Fraval, Simon, James Hammond, Jannkie Wichern, Simon J. Oosting, Imke J. M. De Boer, Nils Teufel, Mats Lannerstad, et al. 2019. "Making the most of imperfect data: a critical evaluation of standard information collected in farm household surveys." *Experimental Agriculture* 55 (2): 230–50. <https://doi.org/10.1017/s0014479718000388>.

# MAKING THE MOST OF IMPERFECT DATA: A CRITICAL EVALUATION OF STANDARD INFORMATION COLLECTED IN FARM HOUSEHOLD SURVEYS

By **SIMON FRAVAL**<sup>†‡¶¶</sup>, JAMES HAMMOND<sup>†§</sup>, JANNIKE WICHERN<sup>¶</sup>,  
SIMON J. OOSTING<sup>‡</sup>, IMKE J. M. DE BOER<sup>‡</sup>, NILS TEUFEL<sup>†</sup>,  
MATS LANNERSTAD<sup>††</sup>, KATHARINA WAHA<sup>‡‡</sup>, TIM PAGELLA<sup>§§</sup>,  
TODD S. ROSENSTOCK<sup>§</sup>, KEN E. GILLER<sup>¶</sup>, MARIO HERRERO<sup>‡‡</sup>,  
DAVID HARRIS<sup>§</sup> and MARK T. VAN WIJK<sup>†</sup>

<sup>†</sup>*International Livestock Research Institute (ILRI), P.O. Box 30709, Nairobi 00100, Kenya,*  
<sup>‡</sup>*Animal Production Systems group, Wageningen University & Research, P.O. Box 338, 6700*  
*AH, Wageningen, The Netherlands,* <sup>§</sup>*International Centre for Research on Agroforestry (ICRAF),*  
*Nairobi 00100, Kenya,* <sup>¶</sup>*Plant Production Systems group, Wageningen University & Research,*  
*P.O. Box 430, 6700 AK, Wageningen, The Netherlands,* <sup>††</sup>*Independent Consultant, 115 23*  
*Stockholm, Sweden,* <sup>‡‡</sup>*Agriculture and Food, Commonwealth Scientific and Industrial Research*  
*Organisation (CSIRO), 306 Carmody Road, St Lucia, QLD 4067, Australia and* <sup>§§</sup>*School of*  
*Environment, Natural Resources and Geography, Bangor University, Deiniol Road, Bangor,*  
*Gwynedd LL57 2UW, UK*

## SUMMARY

Household surveys are one of the most commonly used tools for generating insight into rural communities. Despite their prevalence, few studies comprehensively evaluate the quality of data derived from farm household surveys. We critically evaluated a series of standard reported values and indicators that are captured in multiple farm household surveys, and then quantified their credibility, consistency and, thus, their reliability. Surprisingly, even variables which might be considered ‘easy to estimate’ had instances of non-credible observations. In addition, measurements of maize yields and land owned were found to be less reliable than other stationary variables. This lack of reliability has implications for monitoring food security status, poverty status and the land productivity of households. Despite this rather bleak picture, our analysis also shows that if the same farm households are followed over time, the sample sizes needed to detect substantial changes are in the order of hundreds of surveys, and not in the thousands. Our research highlights the value of targeted and systematised household surveys and the importance of ongoing efforts to improve data quality. Improvements must be based on the foundations of robust survey design, transparency of experimental design and effective training. The quality and usability of such data can be further enhanced by improving coordination between agencies, incorporating mixed modes of data collection and continuing systematic validation programmes.

## INTRODUCTION

Smallholder agricultural production in sub-Saharan Africa (SSA) remains a vital source of sustenance, revenue and employment. In the context of forecasted

<sup>¶¶</sup>Corresponding author. Email: [simon.fraval@wur.nl](mailto:simon.fraval@wur.nl)

population growth and rapid economic development in many African nations, rural communities will have new opportunities and heightened challenges. Effective, inclusive and poverty-reducing rural transformation is not an inevitable outcome in future scenarios for SSA or the broader global rural population (IFAD, 2016). Rather, pro-poor and equitable rural transformation requires well-designed and executed policies and interventions guided by knowledge drawn, at least in part, from the rural communities themselves.

Household surveys are one of the most commonly used tools for generating insights into rural communities (Christiaensen, 2017). These tools are used in place of more detailed studies because they are relatively cost effective. The surveys rely heavily on farmer assessment and recall in place of (more detailed) external monitoring and measurement. The utilisation of low-cost farmer recall enables them to be deployed quickly and at scale which is vital for obtaining representative samples of rural communities and regions. Household surveys can be used for *ex-ante* and *ex-post* analyses. *Ex-ante* applications can be focused on strategic planning purposes, involving prioritisation, characterisation and simulations. *Ex-post* assessments measure the effect of some ‘change’; typical examples include evaluation of new technologies and practices (e.g. those related to cropping, livestock production, land management, natural resource management), or changes to policies and infrastructure (e.g. new roads, market interventions). *Ex-post* assessments will often assess effects on productivity, decisions (farm management, investments, marketing, off-farm activities) and livelihoods (income, nutrition, equity).

Evidence generated to develop and evaluate policies and interventions should be representative of the population of interest, as well as of sufficient quality. Such evidence needs to be founded on a statistically robust sampling protocol that is of sufficient size and designed to minimise sampling error (biases in respondent selection contrary to the population composition) and coverage error (biases from an incomplete sampling frame). The total measurement error of household survey data also consists of random and systematic error, caused by the implementation process of a household survey. Random error can be thought of as instances where repeated measures result in randomly inconsistent values, and systematic errors are errors that are not caused by chance but rather are consistently over or under-reported in a given measurement and observation context. The stages where error can be introduced in a household survey include designing the data collection tool, training enumerators, soliciting households to participate (which can result in unit non-response error) and collecting information from the farmers – often based on estimation and remembering past events. As detailed by Weisberg (2005), there are also specific aspects of survey design, mode of collection, data management and analysis that can introduce random and systematic error.

There is a continuous drive to improve the quality of rural statistics, with a particular focus on reducing random and systematic errors. The central statistical bureaus of many medium and high-income countries have the resources to continuously improve processes and meet domestic data needs and international reporting commitments. For low-income countries, where resources and capacity

for agricultural and rural development are constrained, international partners have taken a more active role, for instance, by providing guidelines and training (e.g. FAO, 2017a; UNFCCC, 2012; OECD, 2009; UN, 2005, as well as the CGIAR and the World Bank). Efforts to improve statistics have addressed the full breadth of issues from experimental design, survey design, enumerator training, data management, analysis and open data policies. The efforts to standardise survey design and indicators are most relevant for this present study because they represent ‘best-practice’ and a process of ongoing improvement. Three comparable survey designs that are multi-topic, multi-purpose and have been internationally applied are the World Bank’s Living Standards Measurement Survey program (LSMS; World Bank, 2017), the Integrated Modelling Platform for Mixed Animal Crop systems (IMPACTlite; Rufino *et al.*, 2013) and the Rural Household Multiple Indicator Survey (RHOMIS; Hammond *et al.*, 2017). Each of these tools aims to improve the consistency and quality of data collection across sites and within sites.

Despite the importance of data quality underlying rural statistics, there are relatively few studies that systematically evaluate data quality. Nevertheless, contributions have been made towards identifying sources and implications of poor data quality. For instance, in a special issue on data quality in Africa, Jerven and Johnson (2015) concluded that limited resources tend to reduce the quality of statistics and that there are risks of bias at many levels. Kilic and Sohnesen (2015) found that survey length has a statistically significant effect on data quality, regardless of topic and question type – potentially mediated through respondent and enumerator fatigue. Finn and Ranchhod (2017) explored methods to detect household survey data fabrication and the implications of fabrication on statistical inference. Pica-Ciamarra *et al.* (2012) reported the perceptions that users (researchers, government departments, etc.) have of the quality of livestock statistics. The effects of gender bias, recall length and respondents’ fatigue on response accuracy have also been explored (Beegle *et al.*, 2012; de Nicola and Giné 2014).

Furthermore, the methodological validation programme of the LSMS Integrated Survey on Agriculture (LSMS-ISA) programme has systematically assessed the deviation of espoused volumes and areas from higher quality measurements (Gibson *et al.*, 2015; Kilic *et al.*, 2013; Zezza *et al.*, 2014; these are akin to the pioneering validation-based improvements made by the USDA-NASS since the 1950s as detailed in Fecso (2011)). The LSMS-ISA – along with other ‘donor-funded surveys’ – has also provided opportunities to test new methods that can then improve the quality of data collected in national statistical programmes (Jerven and Johnson, 2015).

The objective of this study was to further our understanding of the quality of rural statistics by critically evaluating a series of reported values and indicators captured in panel farm household surveys. We do this by assessing the credibility and reliability of information commonly collected in farm household surveys. The results of this assessment are essential in view of the use of survey data in the scientific literature and more practical, policy formulation and agricultural development planning. Based on our results, we suggest ways in which data collection approaches could be improved and the impact of low-quality data can be minimised.

## METHODS

We first describe the household survey data we used in the analyses, and then describe in detail the analyses we have performed focusing on credibility, consistency and reliability.

### *Farm household survey panel data*

Our analysis draws on three comparable multi-topic survey tools: IMPACTlite (Rufino *et al.*, 2013), RHOMIS (Hammond *et al.*, 2017) and LSMS-ISA (World Bank, 2017). IMPACTlite was developed in the context of a large-scale climate change mitigation and adaptation research programme. The IMPACTlite tool was designed to better understand the implications of mitigation and adaptation strategies ‘on livelihoods, food security and the environment’ (Rufino *et al.*, 2013, p. 3).

RHOMIS was developed in response to the general challenges caused by the ‘inefficient multiplicity of survey instruments’ (Carletto *et al.*, 2013, p. 30), and in particular inspired by efforts to conduct cross-dataset analyses of farm household surveys in SSA (Frelat *et al.*, 2016). The tool was designed to capture information efficiently and systematically, allowing the analyst to link farm management to issues of livelihoods, poverty, food security and gender. The indicators which can be calculated from the survey are generally widely validated and internationally recognised. The scope of the survey was defined in relation to the Sustainable Development Goals, specifically SDGs 1, 2, 5 and 13 (no poverty, zero hunger, gender equality and climate action), but the scope is also of relevance to the assessment of Climate Smart Agriculture principles, and Sustainable Intensification. Data collation and analysis are also components of RHOMIS. There are two overall purposes of the RHOMIS tool: to provide a rapid characterisation of farm systems, for use in ex-ante or ex-post analyses, and second, through the building of a large, harmonised dataset from many sites, to permit identification of general principles which can guide the design of rural development interventions. Data from IMPACTlite (2012) and RHOMIS (2015 and 2016) sample the same households and so form panel datasets over three sites, namely, Lushoto, Tanzania ( $n = 149$ ), Wote, Kenya ( $n = 160$ ) and Nyando, Kenya ( $n = 161$ ).

The LSMS-ISA tool was developed with a specific focus on Africa with the intention of improving the quality of rural statistics and building the capacity of local statistics offices. The core purpose of a LSMS-ISA implementation is to ‘improve the understanding of the links between agriculture, socioeconomic status and non-farm income activities’ (World Bank, n.d). LSMS-ISA has been implemented in several countries and collected as panel datasets. In this study, the analysis is limited primarily to Uganda ( $n = 2374$ ) for the surveys held in 2009/10, 2010/2011 and 2011/12. Analysis of LSMS-ISA data from Tanzania ( $n = 3265$ ) and Ethiopia ( $n = 4000$ ) from 2010/11 are also included.

The sampling approach differed between the two panel datasets. IMPACTlite and RHOMIS sampled villages in a  $10 \times 10$  km grid across multiple locations. The household members most aware of farm activities were interviewed in RHOMIS,

and in IMPACTlite other household members contributed to specific sections where necessary. The LSMS-ISA for Uganda, in contrast has been designed to be nationally representative (UBOS, 2007; UBOS, n.d.; UBOS, 2002). The household head was interviewed and in his/her absence, a ‘usual member of the household’ capable of responding was interviewed (UBOS, n.d).

The formulation of questions and mode of data collection also differed in each survey (summarised in Table 1). Perhaps most notably, LSMS-ISA revisits households on a seasonal basis within a 12-month period, whereas IMPACTlite and RHOMIS were conducted only once with multiple recall periods. Surveys incorporated questions on household demographics, farm characteristics, product marketing, income and household diet diversity (in the case of IMPACTlite and RHOMIS, this was calculated based on Swindale and Bilinski (2006)). All variables assessed in this study were answered by all respondents. In addition, zero values (for land holdings, maize yield and livestock) were crosschecked with other sections of the surveys to identify potential item non-responses – all zero values were corroborated.

### *Data analyses*

We use the household data described to assess their credibility (in terms of inaccuracies) and reliability (measurement precision; Alwin, 2007; Evans, 1995). We first assess the credibility of observations in one survey round, which also gives us insight into systematic errors. Credibility (identifying inaccurate observations) in this context is concerned with whether values fall outside acceptable bounds. We then assess the consistency of measurements between two panel rounds with two household survey instruments that are similar in complexity and are focused on single site applications, i.e. RHOMIS and IMPACTlite. For a more robust assessment of consistency, we also model the reliability of the LSMS-ISA dataset using three rounds of survey data. This measure of reliability better accounts for survey round specific systematic errors, but does not distinguish between true population scale temporal volatility, random error and non-survey round based systematic error.

To conclude our analysis, we assess the implications of varying levels of reliability on required sample sizes. Although all of these analyses give insight into the possible existence of systematic errors, none of the methods above will allow us to really quantify these. For that mixed method approaches are needed, e.g. really measured crop yields or GPS-based field size estimates, where one can quantify the deviation between farmer recall-based information and ‘reality’. This lies outside the scope of this study.

### *Credibility analysis*

Crop yields and market prices were used to assess the credibility of farmer reported and estimated values. As indicated in Table 1, we calculated crop yields as a composite of farmer reported harvest volumes and area planted, and market prices could be enumerated as the unit price or a composite of total value and volume sold. Due to the limited availability of secondary data, crop productivity and market prices were

Table 1. Characteristics, question formulation and relevant period of survey tools.

		ImpactLite	RHOMIS	LSMS-ISA
Locations		Kenya, Tanzania	Kenya, Tanzania	Uganda
Survey rounds		1	1 (ImpactLite revisit)	3 revisits
Observations		470	470	2374
Representativeness		Development domains	Development domains	Nationally
Duration		2.5–3 h	45 min to 1 h	Unknown
Crop harvest	Question formulation	Subplot harvest by crop and	Crop harvest – local volume units	Subplot harvest by crop
	Relevant period	Seasonal	Seasonal and annual	Seasonal – revisits
Crop area planted	Question formulation	Area planted by crop	Area planted by crop – local area units	Area planted by crop
	Relevant period	Seasonal	Seasonal and annual	Seasonal - revisits
Price	Question formulation	Price per kg	Price per kg/tonne/unit or total value	Value of yield in plot
	Relevant period	Seasonal	Annual	Seasonal – revisits
Household head age	Question formulation	Date of birth all members	Age of head	Date of birth all members
	Relevant period	As of interview date	As of interview date	As of interview date
Household size	Question formulation	Full household roster	By age category	Full household roster
	Relevant period	> 1 season	> 3 months per year	12 months
Livestock holdings	Question formulation	Full list converted to TLU	Full list converted to TLU	Full list converted to TLU
	Relevant period	Current holdings	Current holdings	Current holdings
Land owned	Question formulation	Parcel size	Total land owned	Parcel size
	Relevant period	Annual	Annual	Annual
Off-farm income	Question formulation	Full income register	Proportion of total	Full income register
	Relevant period	Annual/Monthly/week	Annual	Annual
Household diet diversity	Question formulation	Itemised food list open question	12 or 8 category list prompted	61 item food list
	Relevant period	Seasonal	Seasonal	7 day recall

only assessed for maize (*Zea mays* L.), quantifying the yield ( $\text{kg ha}^{-1}$ ) and the farm-gate price per kilogram for each farm household. Yields calculated from farmer reported harvest volumes and area planted were compared with historical yield estimates (from fertilised crop trials and government monitored plots) from the Global Yield Gap Atlas (GYGA, n.d.). Historical yield estimates compiled in the GYGA formed the basis for setting lower credible bounds. The threshold was set at 10% of the average historical GYGA yield for the same climate-zone and country. Simulated water-limited yield potential formed the basis for setting credible upper bounds. It is unlikely that enumerated yields exceed the simulated potential. Historical, potential and survey reported yields were compared on a country and climate-zone basis (using the GYGA climate zones). The historical yields in Uganda, for example, ranged from  $0.7 \text{ tonnes ha}^{-1}$  to  $1.31 \text{ tonnes ha}^{-1}$  (a summary of used thresholds is provided in Table S1, available online at <http://dx.doi.org/10.1017/S0014479718000388>).

Farm-gate prices were compared with the average price for each location (i.e. Lushoto, Wote, Nyando, Kampala, western Uganda, etc.) and survey tool as well as the wholesale market prices in major cities (Kampala in Uganda and Nairobi in Kenya and Tanzania; sourced from FAO, 2017a). In this component of our credibility analysis, we assume a high degree of market integration, where there is a close association with farm gate price, regional price and market price. Prices in the surveys were also averaged across seasons to give an annual average. Lower limits were set at 10% of the average survey prices for a given location; upper limits were set at the maximum wholesale market price. A summary of price thresholds is provided in Table S2. The statistics on wholesale market prices also have errors associated with them, this analysis only provides information about the uncertainty surrounding farmer estimates rather than an absolute benchmarking of data quality.

To assess the consequences of data credibility for more complex, constructed indicators, we examine the commonly used indicators of food self-sufficiency and potential food availability (FA) (as detailed in Frelat *et al.*, 2016). The FA indicator is a quantification of the potential kilocalories available for each male adult equivalent per day consumed from farm production, and from cash obtained through the sale of farm produce and off-farm income, where all income is converted to a calorific value based on the cost of a local staple crop. For our calculations of FA, we used the median farm gate price for each location and time period. Results of these calculations can be used to perform a combined data quality assessment of information obtained on crop and livestock production, sales, consumption and off-farm income. Two problems with this composite indicator are commonly encountered. First, an underestimation of the calorie availability at the lower end of the scale, suggesting an extreme level of starvation. Although this may be a true representation of some households, it can also be an indication of missing information on income or food consumption. Second, there can be a substantial over-estimation of consumption of crop and livestock products for a large number of households (i.e. food self-sufficiency), indicating problems with yield, consumption or household size data. The lower bound threshold for credible FA was set at 1250 kcal per male adult equivalent per day, which is



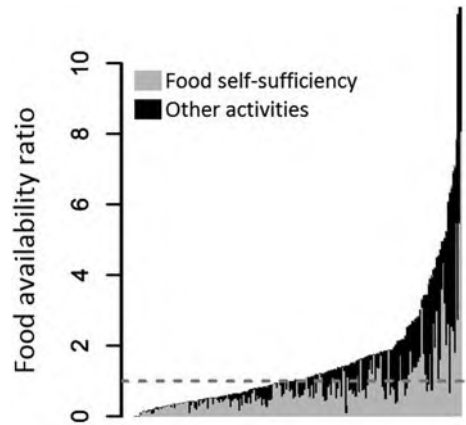


Figure 1. Food availability, food self-sufficiency and household energy needs: an example of unreliable values. Dashed line represents a case where 2500 kcal are provided for each male adult equivalent (Source: Rosenstock *et al.*, 2017; based on 200 households in northern Ghana).

below the basal metabolic rate for adult males (approximately 1590 kcal for a 60 kg male; FAO, 2001). Two upper bounds for credible food self-sufficiency were set (i) 3500 kcal per adult equivalent per day, representing the average intake of developed nations (OECD and FAO, 2017) and (ii) 5000 kcal, which is double the approximate requirement for an adult male.

The results from Rosenstock *et al.* (2017) provide an example of extremes in FA and food self-sufficiency for households in northern Ghana. This case is represented here in Figure 1 as the ratio of FA, where the value 1 represents a case where 2500 kcal are provided for each male adult equivalent (indicated with a horizontal dotted line). Also, represented is the ratio of FA sourced directly from farm production (the grey bars). Instances of apparent starvation are increasingly severe as the ratio decreases below 1, which eventually declines below the basal metabolic rate for adult males. Over-estimated consumption is apparent in households that have more energy sourced directly from the farm than is required (grey bars larger than 1).

### *Consistency and reliability analyses*

In the consistency and reliability analyses, we included variables that we would expect to be (i) highly consistent (age of the household head), (ii) relatively stationary in East Africa over the whole population over short time periods, including household size, productive assets (land owned and livestock holdings) and crop yields and (iii) those that may be more variable (off-farm income, FA and food self-sufficiency). Age of household head is expected to be highly consistent after accounting for the time elapsed between survey rounds and whether there was a change in household head. Household size was expected to be relatively stationary in East Africa given that death rates have been estimated to be less than 1% per annum (CIA, 2016a) and the rate of urbanisation estimated to be less than 5.5% per annum (CIA, 2016b). Productive assets are also expected to be relatively stationary due to their livelihood and cultural

value – for both land (Jayne *et al.*, 2016) and livestock (Thornton *et al.*, 2015). Livestock holdings, however, are expected to be more temporally variable than land holdings due to their role in financing large expenditures, cultural utility (i.e. bride-wealth) and exposure to climatic and disease risks (ibid.). Similarly, crop yields are expected to be temporally stable (at a population level) in the absence of extreme weather events (Gollin, 2006). During the periods of observation, there were instances of extreme weather events, with a severe drought impacting northern Kenya and north-eastern Uganda (potentially impacting <0.5% of households in LSMS-ISA Uganda) and some evidence of increased extreme precipitation events in western Kenya, but to our knowledge, this did not affect the sampled households (Gebrechorkos, 2018). Climatic conditions were consistent over the two survey rounds in Tanzania (Fraval *et al.*, 2018).

We explored the consistency of data collected in farm household surveys between two points in time, comparing, respectively, IMPACTlite (2012) with RHOMIS (2015/16), and LSMS-ISA (2009/10 and 2011/12). Summary statistics of these changes between initial survey and revisit are provided in Table S3. In the absence of survey round specific biases, the correlations in these consistency results would provide a measure of reliability (Alwin, 2007). As this is not the case, we can only interpret the strength of correlation as a measure of consistency, rather than reliability. Spearman’s correlation was used to assess association, as it is less sensitive to extreme non-credible values.

Reliability was more formally modelled using the core variables and derived indicators quantified from LSMS-ISA Uganda (2009/10, 2010/11, 2011/12), excluding some non-credible values. We used an approach described by Shrout and Fleiss (1979) that calculates intraclass correlation (ICC). In this specification, we assume the following linear model:

$$x_{ij} = \mu + b_j + w_{ij} \quad (1)$$

where  $x_{ij}$  is the  $i$ th survey round ( $I = 1, 2, 3$ ) of  $j$  household  $j = (1, \dots, n)$ ;  $\mu$  is the population mean;  $b_j$  is the difference from  $\mu$  to the  $j$ th household’s mean across the survey rounds;  $w_{ij}$  is the residual, equal to the sum of the effects of survey round, survey round-household interaction and error. The ICC is then estimated as follows:

$$\text{ICC} = \frac{\text{MSB} - \text{MSW}}{\text{MSB}} \quad (2)$$

where MSB is the mean square between (sum of square total/obs) and MSW is mean square within, calculated as

$$\text{MSW} = \frac{\text{SS}_{\text{round}} + \text{SS}_{\text{resid}}}{df_{\text{round}} + df_{\text{resid}}} \quad (3)$$

For this analysis, some non-credible observations were excluded as they had a disproportionate influence on the linear models. Observations with off-farm income above \$US 60,000 in one survey round were excluded ( $n = 1$ ), as were maize yields above 15 tonnes ha<sup>-1</sup> ( $n = 33$ ), livestock holdings (TLU > 100,  $n = 3$ ), land owned

Table 2. Summary of analysis, datasets and variables.

Survey tool	Country	Year	Credibility (one round)	Consistency (two rounds)	Reliability (three rounds)
ImpactLite	Kenya & Tanzania	2012	x	a	
RHOMIS	Kenya & Tanzania	2015/16	x	a	
LSMS-ISA	Uganda	2009/10		b	c
LSMS-ISA	Uganda	2010/11	x		c
LSMS-ISA	Uganda	2011/12		b	c
LSMS-ISA	Tanzania	2010/11	x		
LSMS-ISA	Ethiopia	2010/11	x		

x = used in credibility assessment.

a, b, c = panel dataset, where two or three survey rounds are used in an analysis.

Table 3. Credibility of maize yield data: Comparing enumerated yields with historical yields and water-limited potential yields, by survey tool\* (proportion of households).

	Less than 10% of historical yields	Greater than potential yield and less than double	Double potential yield	Within bounds
ImpactLite 2012	0.43	0.00	0.01	0.56
RHOMIS 2015/16	0.12	0.00	0.00	0.88
LSMS-ISA 2010/11	0.03	0.04	0.03	0.90

\*Impact lite and RHOMIS include sites from Kenya and Tanzania, LSMS-ISA is limited to Uganda.

(ha > 100;  $n = 2$ ) and FA (FA > 1 500 000 kcal,  $n = 2$ ). The reliability analysis was implemented using the psych package in R (Revelle, 2017). This analysis resulted in a reliability estimate ranging between 0 (low reliability) and 1 (high reliability) together with a 95% confidence interval of this estimate. The three methodological steps and associated datasets, locations and years are summarised in Table 2.

The relationship between sample reliability, effect size and sample size was simulated using the ‘pwr’ package in R (Champely, 2016), where reliability is mediated through the effect size ( $ES = \text{population ES} \times \sqrt{r}$ ; Kanyongo *et al.*, 2007). Using the pwr package, we simulated both a paired (panel data) and two-sample (Randomised Control Trial type of data) *t*-tests to quantify detectable differences for the core variables and derived indicators of which we have quantified reliability and uncertainty estimates. The simulated *t*-tests assumed a Type II error rate of 20% (Power of 0.8) and a Type I error rate of 5% ( $\alpha$  of 0.05).

## RESULTS

### *Credible bounds of core variables and derived indicators*

Yields and prices were highly variable in each of the three farm household survey tools. This variability across a wide range of crops could reflect agro-climatic or management differences, market volatility as well as biases and errors introduced by the survey tool, enumerator or the respondent (Mathiowetz *et al.*, 2001; UN, 2005). This section provides a summary of non-credible values for yields (Table 3), prices (Table 4), FA and food self-sufficiency (Table 5).

Table 4. Credibility of maize price data: Comparing enumerated prices with average survey prices and wholesale market prices by survey tool\* (proportion of households).

	Less than 10% of average prices of survey	Greater than maximum wholesale price and less than double	Double maximum wholesale price	Within bounds
ImpactLite 2012	0.03	0.50	0.03	0.43
RHOMIS 2015/16	0.02	0.09	0.05	0.84
LSMS-ISA 2010/11	0.01	0.11	0.04	0.84

\*Impact lite and RHOMIS include sites from Kenya and Tanzania, LSMS-ISA is limited to Uganda.

Table 5. Credibility of food availability (FA) and food self-sufficiency (FSS) by survey tool (proportion of households).

	Food availability less than 1250 kcal	FSS above OECD average and less than double	FSS double 2500 kcal	Within bounds
ImpactLite 2012	0.39	0.05	0.04	0.52
RHOMIS 2015/16	0.15	0.06	0.04	0.75
LSMS-ISA Uganda 2010/11	0.12	0.07	0.14	0.66
LSMS-ISA Tanzania 2010/11	0.20	0.06	0.07	0.67
LSMS-ISA Ethiopia 2010/11	0.40	0.02	0.07	0.51

\*Impact lite and RHOMIS include sites from Kenya and Tanzania.

Comparing our calculated maize yields with historical yield statistics compiled in the GYGA gives a reference point to assess yield estimates for each of the three survey tools. IMPACTlite had the highest proportion of households with crop yields less than 10% of GYGA historical maize yields, followed by RHOMIS and then LSMS-ISA (Table 3). On the other extreme, LSMS-ISA had the most substantial proportion of yields exceeding the simulated water-limited potential yield for the region – occurring in 7% of households. A total of 4% of households in LSMS-ISA 2010/11 Uganda exceeded potential maize yields and an additional 3% were double this potential yield (Table 3).

Exceeding the simulated potential yield is possible, but unlikely – even historical yields in optimal growing conditions were never more than 50% of simulated potential yields (results not shown). It is more difficult, however, to assess the credibility of calculated yields at the lower end of the scale, which are far more prevalent. Comparing maize yields from the same sites from both IMPACTlite (2012) and RHOMIS (2015/16) data, 62% of households that had yields less than 1 tonne ha<sup>-1</sup> had those low yields in both surveys. A total of 27% purportedly produced at least an additional tonne per ha in the RHOMIS (2015/16) survey.

The farm-gate price per kg of maize was compared with the median price for each data collection instance as well as the market wholesale price (sourced from FAO, 2017b). The average market wholesale price of maize in Kampala, Uganda during the year of data collection was 20 US cents kg<sup>-1</sup>, with a minimum of 11 cents. Approximately 15% of LSMS-ISA (2011/12) Uganda crop prices exceeded the maximum wholesale price, and 1% was below the lower threshold. There were also such potential credibility issues in the IMPACTlite and RHOMIS datasets,

with maize prices exceeding maximum wholesale prices in 53 and 14% of cases, respectively (Table 4). Similar to yield, the lower range on price is more difficult to assess. There are product quality aspects, market timing and geographical differences also influencing price (particularly for LSMS-ISA in Uganda which aims to be nationally representative). These aspects could result in farm gate prices well below the regional average.

Non-credible values in core variables will propagate through to composite indicators, such as food self-sufficiency and potential FA. Specifically, non-credible values in farm production, area planted, product marketing, off-farm income and household size can be compounded in these indicators. Table 5 shows the proportion of households that have non-credible FA and self-sufficiency values (defined by being below half or more than double the energy demands of the household). Instances of non-credible FA values exist in all survey implementations, but more so in IMPACTlite. LSMS-ISA Uganda and LSMS-ISA Tanzania had the most non-credible food self-sufficiency values. LSMS-ISA Tanzania and Ethiopia are included in this table as an example of the variability in data quality generated with the same survey tool. LSMS-ISA Tanzania and Ethiopia appear to provide much lower quality FA and food self-sufficiency estimates as compared to LSMS-ISA Uganda.

#### *Consistency of core variable measurements and derived indicators over time*

After accounting for the time elapsed between the two surveys and excluding households where the household head had changed (LSMS-ISA,  $n = 168$ ) or household head gender was different between rounds (IMPACTlite 2012-RHOMIS 2015/16,  $n = 67$ ), 84% of households in LSMS-ISA Uganda were within 1 year of the expected age, given the age provided in the initial survey and time elapsed between surveys. By contrast, only 47% were within the expected range of 1 year in IMPACTlite-RHOMIS. The relationship between the successive surveys, as shown by Spearman's correlation coefficient in Figure 2a, is strong in both LSMS-ISA ( $r = 0.99$ ) and IMPACTlite-RHOMIS ( $r = 0.93$ ). However, there were instances of substantial differences, with 1% of households in LSMS-ISA, and 8% of households in IMPACTlite-RHOMIS being 10 years greater or less at revisit than the expected age.

Household size was highly correlated between the 2009/10 and 2011/12 surveys in Uganda ( $r = 0.9$ ) and moderately correlated in IMPACTlite-RHOMIS ( $r = 0.51$ ; Figure 2b). The majority of households in each dataset, however, remained within one adult equivalent of the initial visit. For instance, 62% of households remained within one adult equivalent in LSMS-ISA and 55% in IMPACTlite-RHOMIS (results not shown). There were isolated cases of extreme increases in each dataset, with the maximum increase in LSMS-ISA being 11 adult equivalents and 14 in IMPACTlite-RHOMIS (Figure 2b).

Land owned in LSMS-ISA survey rounds had a higher correlation ( $r = 0.68$ ) than IMPACTlite-RHOMIS ( $r = 0.55$ ; Figure 2c). Isolated cases of extreme changes in

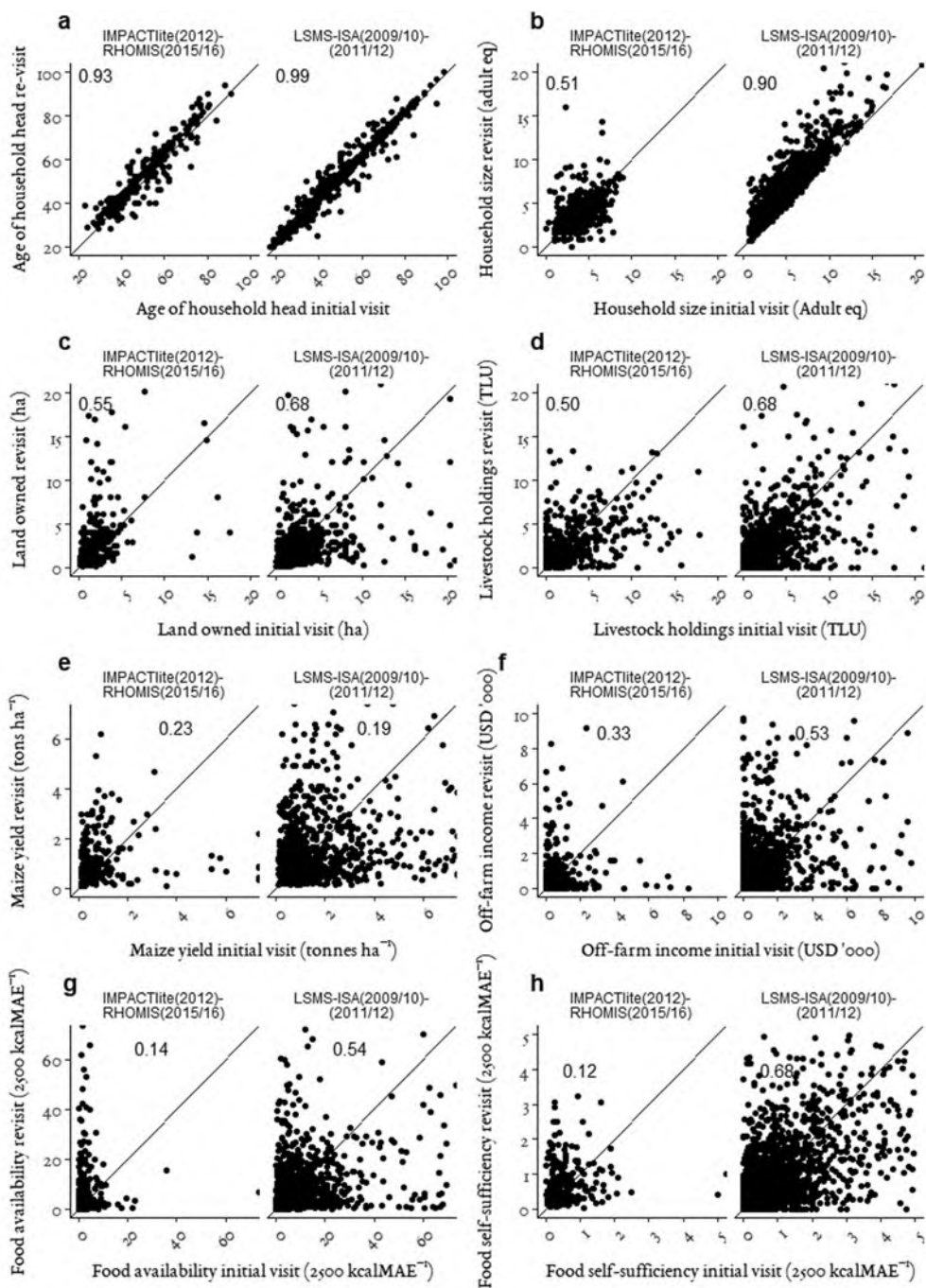


Figure 2. Consistency between IMPACTlite (2012)-RHOMIS (2015/16) in Kenya and Tanzania and LSMS-ISA (2009/10–2011/12) in Uganda\* \*Spearman's correlation coefficient (initial visit to revisit) indicated on each respective plot.

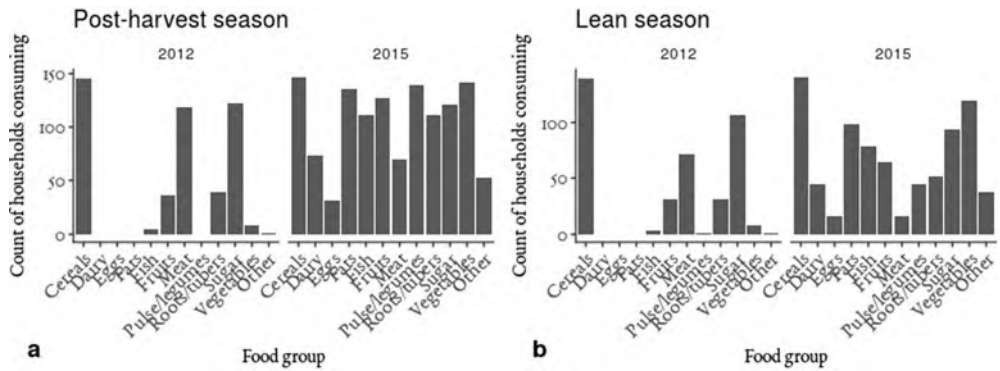


Figure 3. Diet diversity by category and period in IMPACTlite-RHOMIS Tanzania.

land owned were present in LSMS-ISA. Livestock holdings in LSMS-ISA Uganda had a Spearman's correlation coefficient of 0.68 between survey rounds. The level of association between rounds was lower in IMPACTlite-RHOMIS ( $r = 0.50$ ; Figure 2d). There was a similar level of correlation for maize yields in IMPACTlite-RHOMIS ( $r = 0.23$ ) and LSMS-ISA ( $r = 0.19$ ; Figure 2e).

Off-farm income was moderately correlated between rounds in LSMS-ISA ( $r = 0.53$ ) and less so in IMPACTlite-RHOMIS ( $r = 0.33$ ; Figure 2f). Changes in off-farm income of \$5000 or more occurred in 6% of the households in Uganda and 2% in IMPACTlite-RHOMIS. Households in LSMS-ISA, span a wide geographical range with varying proximity to urban locations which may explain such outliers.

FA at initial visit and revisit had a moderate association in LSMS-ISA ( $r = 0.54$ ) and less so in IMPACTlite-RHOMIS ( $r = 0.14$ ; Figure 2g). There were instances of outliers in both survey comparisons; these few cases, however, could be realistic given large changes in on-farm and off-farm income. Food self-sufficiency followed a similar pattern to FA, with LSMS-ISA having a having a greater level of correlation between survey rounds ( $r = 0.68$ ) when compared to IMPACTlite-RHOMIS ( $r = 0.12$ ; Figure 2h).

Household diet diversity in the IMPACTlite-RHOMIS surveys also provides a notable case of inconsistency between panel rounds. For example, median increases in diet diversity range from three food categories in the lean season, and up to six food categories in the Tanzanian post-harvest season. As desirable as leaps in diet diversity are, it is unlikely to observe such a change over a short space of time in these communities (IFAD, 2016). Figure 3 shows the differences between the survey rounds in both periods for Tanzania as an example, where the same applies for IMPACTlite-RHOMIS in Kenya. The initial visit has instances where common food categories (e.g. fats and oils) are supposedly not consumed at all. The likely causes of these differences relate to survey design and duration. IMPACTlite enumerated a wide range of food items (not food groups) asked as an open question. Furthermore, these questions came at the end of a 3-hour interview, potentially resulting in respondent



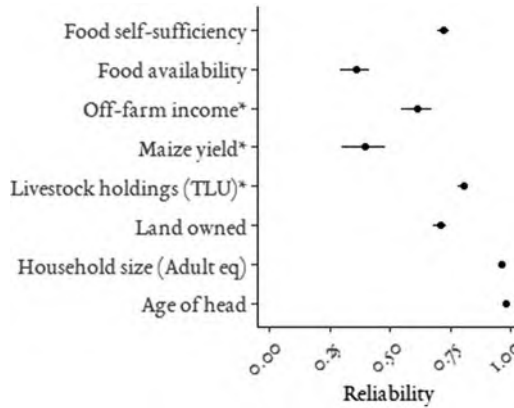


Figure 4. Reliability of initial visit variables in the Living Standards Measurement Survey, Uganda; output from intraclass correlations with 95% Confidence Intervals. \*Limited to households that were > 0 in each survey round. 27% of households had off farm income in all three rounds, 75% cultivated maize, 60% kept livestock.

fatigue. RHOMIS, on the other hand, asked about these food groups specifically and was completed within an hour.

#### *Reliability of variables in LSMS-ISA Uganda*

Modelling the reliability of variables explored in the consistency analysis provides further insight into the three waves in the LSMS-ISA Uganda case. The model outputs suggest a high degree of reliability for age, household size and livestock holdings (Figure 4); land owned was less reliable than these other stationary variables and maize yield was one of the least reliable variables. It is more difficult to evaluate the reliability estimates of off-farm income, FA and food self-sufficiency. The paucity of information about the temporal stability of these variables (despite efforts to assess the quality of variables such as income – notably by Fisher *et al.* 2010; Juster *et al.*, 2007; Moore *et al.*, 2000; Neri and Ranalli, 2012) make it difficult to identify whether the reliability scores of these three variables are influenced by true population level temporal volatility, but it is clear that the reliability of these variables is low.

The reliability of these variables will ultimately affect inference as it reduces the power of tests (increasing Type II error) and inflates error estimates in multivariate analyses. Additionally, in instances of new studies using existing data for setting required sample sizes, consideration needs to be given to the reliability of available variables and how the proposed study will differ in terms of measurement error; a new study with a coarser measurement tool will require a larger sample than a previous study with more accurate measures. Figure 5 shows the relationship between sample reliability, effect size and sample size for paired and two sample *t*-tests. For example, the sample size required to detect a relatively small effect size (0.2) in a paired test with a Type II error rate of 20% and type I error rate of 5%, will be 220 with a reliability of 0.9 (as we see for household size) and 983 households for a reliability of 0.2 (as we see



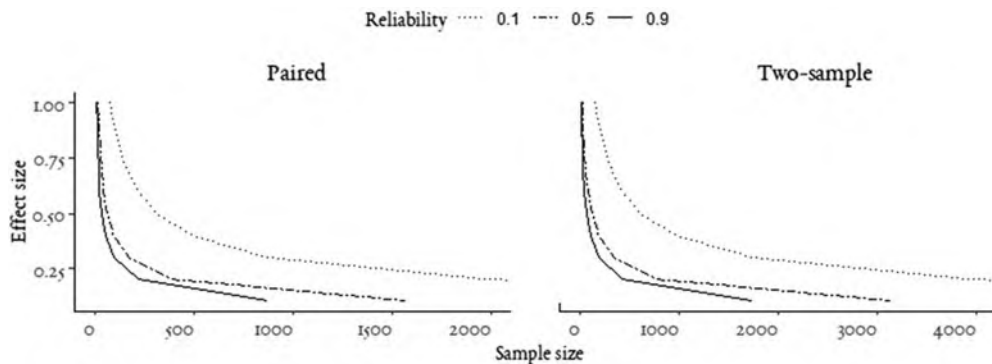


Figure 5. Sample size and effect size given different levels of reliability for  $t$ -test (power = 0.8;  $\alpha$  = 0.05).

for off-farm income and crop yield). These sample sizes will be higher when design effects are incorporated and when two sample tests are needed.

## DISCUSSION

### *Credibility of crop yield and market price*

Using data from three cross-sectional farm household survey approaches, we assessed the credibility and reliability of core variables and derived indicators. This study has identified quality limitations in each survey tool – with LSMS-ISA and RHOMIS staying within credible bounds more frequently than IMPACTlite (Tables 3 and 4). The higher performance of these two survey tools may be due to their innovative data collection strategies – particularly in the case of enumerating cropping activity. In the case of LSMS-ISA, enumerators visit households each cropping season – with the intention of minimising recall error. In the case of RHOMIS, households can quantify harvest volumes in a unit of their choice (such as standard sized sacks) rather than force kg estimates – minimising error due to respondent estimation. These innovations are positive; however, they are not sufficient to eliminate non-credible values.

### *Consistency of variables between two survey rounds*

The inconsistencies identified in this study – including age of household head, household size (Figure 2) and diet diversity (Figure 3) – reinforce the notion that researchers need to consider their data collection strategy for each variable rather than assuming that some are ‘easy’ to enumerate. In the instance of diet diversity, the differences between survey rounds may be explained by an unfortunate combination of question design and survey length. The data collection strategy of IMPACTlite, in this instance, was to ask an open question – ‘what food items did you consume?’ – and allow the enumeration of a detailed list of food items. These questions on diet diversity, however, came at the end of more than two hours of questions and so the

quality of data may have suffered from the farmer (and enumerator) being fatigued (as systematically explored by Kilic and Sohnesen (2015)).

#### *Implications of the lower reliability over three survey rounds*

The lack of reliability of land owned and crop yields have implications in the monitoring the food security status, poverty status and land productivity of households. These variables are also used to answer essential questions of cause and effect and can have a substantial bearing on policy decisions. For instance, the question of whether smaller farms are more productive than larger farms in developing countries has implications for reducing yield gaps and has been an active area of debate. Recently, the robustness of the data underpinning the analysis of this relationship was tested using LSMS-ISA data. In order to test the robustness of underpinning data, farmer reported values were compared against Global Positioning System (GPS)-based land area estimates (Carletto *et al.*, 2011; Kilic *et al.*, 2013). These studies found that despite measurement error in farmer reported values, the inverse productivity relationship was still detected.

The implications of non-credible values and lower reliability are more pronounced in composite indicators such as food self-sufficiency and FA – where the uncertainty propagates from multiple variables. This leads to (i) substantial portions of the survey results to be beyond credible bounds (Table 5); (ii) the need for larger sample sizes so that change and differences between groups can be detected (Figure 5) and (iii) limitations in identifying relationships in multivariate analysis, such as FA and agricultural land use strategies (among households and over time).

#### *Data quality in perspective – sample sizes, continuous improvement and transparency*

Despite this seemingly bleak picture of data quality of variables derived from farm household surveys, there is still cause for optimism. First, the sample sizes needed to detect substantial changes (which are often the changes of interest in agricultural development research) in the variables assessed in this study are in the hundreds, and not in the thousands such as for a Randomised Control Trial (see Figure 5). With tight controls on quality and a less variable population to represent, this sample size can be even smaller. The second cause for optimism is that ongoing developments in data collection strategies and tools are likely to improve data quality and reduce measurement error. There have been many areas of progress in the last 10 years that have improved quality including: harmonised survey tools, in-country capacity building, mixed modes of data collection (e.g. GPS data, phone, SMS; Carletto *et al.*, 2016; Deininger *et al.*, 2011; Leeuw, 2005), quality control protocols (e.g. rapid data quality checks and variable triangulation; Fisher *et al.*, 2010) and non-paper-based collection (Rosenstock *et al.*, 2017).

An example of a new, more systematised household survey is RHOMIS (Hammond *et al.*, 2017). The evaluations in this study show that such targeted data collection does result in highly credible quantification of indicators like food self-sufficiency and FA (e.g. Table 5). RHOMIS furthermore assesses food security by quantifying clusters of

indicators rather than a single indicator (Household Diet Diversity score, the USAID Hunger and Food Insecurity Access scale, FA and the number of hunger months as indicators), thereby allowing for a more integral picture of food security. These benefits do have some limitations when used to follow up on households that have been surveyed with different tools, where [Figure 3](#) shows inconsistencies and thus a potential to increase Type II error – not having the power to identify significant differences between communities and over time.

The third cause for being optimistic about the use of household-based rural statistics is that the survey tools analysed in this study were transparent about sampling and data collection procedures. This practice informs data users of what the sample is representative of, and the nature of the questions asked. Such transparency guides the secondary utilisation of these data and can reduce misuse and misinterpretation.

#### *Improving the quality of farm household survey data*

Several steps can be taken to improve the data quality of farm household surveys further. First, researchers can compare a subset of collected data to the ‘truth’, where there is a possibility to collect detailed data from a smaller subset of households (referred to as ‘two method measurement designs’, Little and Rhemtulla, [2013](#)) which is widely accepted to be of far greater reliability and accuracy (as demonstrated by Giller *et al.*, [2011](#)). For instance, plot sizes of a sub-sample might be measured using GPS receivers or through remote sensing, providing a more consistent level of accuracy across households. Second, researchers can analyse data for reliability (as done in this study) and potential instances of data fraud (as discussed in Finn and Ranchhod, [2017](#)) which would highlight issues that could improve overall quality.

Rural SSA is entering a stage of transformation where the opportunities and challenges for rural communities are becoming more pronounced, and at the same time, the means of gaining insight into these communities is broadening. In this setting, the fundamentals of generating fit-for-purpose and representative observations remain a vital basis for informed decision making. For decision makers to make the most of such inherently coarse data, it is essential to have the foundation of robust sampling, quality-centric survey design (questions and length), transparency of experimental design and effective training. The quality and usability of such data can be further enhanced by improving coordination between agencies, incorporating mixed modes of data collection and continuing systematic validation programmes.

*Acknowledgements.* We are grateful to the research teams who were involved in designing and implementing the three surveys assessed in this study. Without their rigor, openness and thorough documentation this study would not have been possible. We also thank the two anonymous reviewers and special edition editor Jens Andersson, whose comments and suggestions greatly improved the quality of this article. This study was made possible by the CGIAR Research Program on Livestock and its donors and by support of the American People provided to the Feed the Future

Innovation Lab for Sustainable Intensification through the United States Agency for International Development (USAID). The views expressed in this paper cannot be taken to reflect the official opinions of these organisations.

## REFERENCES

- Alwin, D. F. (2007). *Margins of Error: A Study of Reliability in Survey Measurement*. New Jersey: John Wiley & Sons, Inc. <http://doi.org/10.1002/9780470146316>.
- Beegle, K., Carletto, C. and Himelein, K. (2012). Reliability of recall in agricultural data. *Journal of Development Economics* 98(1):34–41. <http://doi.org/10.1016/j.jdeveco.2011.09.005>.
- Carletto, C., Gourley, S., Murry, S. and Zezza, A. (2016). *Cheaper, Faster, and More Than Good Enough is GPS the New Gold Standard in Land Area Measurement?* (No. 7759). Washington, DC: World Bank.
- Carletto, C., Savastano, S., & Zezza, A. (2011). *Fact or Artefact: The Impact of Measurement Errors on the Farm Size - Productivity Relationship* (No. 5908). Washington, DC: World Bank. Retrieved from <http://documents.worldbank.org/curated/en/2011/12/15545065/fact-or-artefact-impact-measurement-errors-farm-size-productivity-relationship>
- Carletto, C., Zezza, A. and Banerjee, R. (2013). Towards better measurement of household food security: Harmonizing indicators and the role of household surveys. *Global Food Security*. <https://doi.org/10.1016/j.gfs.2012.11.006>
- Central Intelligence Agency (CIA). (2016a). Death rate: Country Comparison to the World. Retrieved March 20, 2017 from <https://www.cia.gov/library/publications/the-world-factbook/fields/2066.html#tz>
- Central Intelligence Agency (CIA). (2016b). Urbanisation: Country Comparison to the World. Retrieved March 20, 2017 from <https://www.cia.gov/library/publications/the-world-factbook/fields/2212.html#tz>
- Champely, S. (2016). PWR: Basic Functions for Power Analysis. Retrieved from <https://cran.r-project.org/package=pwr>
- Christiaensen, L. (2017). Agriculture in Africa – Telling myths from facts: A synthesis. *Food Policy* 67:1–11. <https://doi.org/10.1016/j.foodpol.2017.02.002>
- de Nicola, F. and Giné, X. (2014). How accurate are recall data? Evidence from coastal India. *Journal of Development Economics* 106:52–65. <http://doi.org/10.1016/j.jdeveco.2013.08.008>.
- Deininger, K., Carletto, C., Savastano, S. and Muwonge, J. (2011). *Can Diaries Help Improve Agricultural Production Statistics? Evidence from Uganda* (No. 5717). Washington, DC: World Bank.
- Evans, B. (1995). On the difference between reliability of measurement and precision of survey instruments. *The Canadian Journal of Program Evaluation* 10(2):17–32.
- Fecso, R. (2011). A review of errors of direct observation in crop yield surveys. In *Measurement Errors in Surveys*, 327–346 (Eds P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz and S. Sudman). Hoboken: John Wiley & Sons, Inc. <http://doi.org/10.1002/9781118150382.ch17>
- Finn, A. and Ranchhod, V. (2017). Genuine fakes: The prevalence and implications of data fabrication in a large South African survey. *The World Bank Economic Review* 31(1):129–157. <https://doi.org/10.1093/wber/lhv054>
- Fisher, M., Reimer, J. J. and Carr, E. R. (2010). Who should be interviewed in surveys of household income?. *World Development* 38(7):966–973. <http://doi.org/10.1016/j.worlddev.2009.11.024>.
- Food and Agricultural Organisation of the United Nations (FAO). (2001). Human energy requirements: Report of a joint FAO/WHO/UNU expert consultation. FAO Food and Nutrition Technical Report Series, 0:96. [https://doi.org/92\\_5\\_105212\\_3](https://doi.org/92_5_105212_3).
- Food and Agricultural Organisation of the United Nations (FAO). (2017a). *Statistical Programme of Work 2016–2017*. <http://www.fao.org/3/a-br622e.pdf>.
- Food and Agricultural Organisation of the United Nations (FAO). (2017b). Food Price Monitoring and Analysis. Retrieved 1 Mar 2017 from <http://www.fao.org/giews/food-prices/tool/public/index.html#/home>.

- Fraval, S., Hammond, J., Lannerstad, M., Oosting, S. J., Sayula, G., Teufel, N., Silvestri, S., Poole, E. J., Herrero, M. and van Wijk, M. T. (2018). Livelihoods and food security in an urban linked, high potential region of Tanzania: Changes over a three year period. *Agricultural Systems* 160(January 2017):87–95. <https://doi.org/10.1016/j.agry.2017.10.013>
- Frelat, R., Lopez-Ridaura, S., Giller, K. E., Herrero, M., Douxchamps, S., Djurfeldt, A. A., Erenstein, O., Henderson, B., Kassie, M., Paul, B. K., Rigolot, C., Ritzema, R. S., Rodriguez, D., van Asten, P. J. A. and van Wijk, M. T. (2016). Drivers of household food availability in sub-Saharan Africa based on big data from small farms. *Proceedings of the National Academy of Sciences USA* 113:1518384112. doi:10.1073/pnas.1518384112
- Gebrechorkos, S. H., Hülsmann, S. and Bernhofer, C. (2018). Changes in temperature and precipitation extremes in Ethiopia, Kenya, and Tanzania. *International Journal of Climatology* 1–13. <https://doi.org/10.1002/joc.5777>.
- Gibson, J., Beegle, K., De Weerd, J. and Friedman, J. (2015). What does variation in survey design reveal about the nature of measurement errors in household consumption?. *Oxford Bulletin of Economics and Statistics* 77(3):466–474. <http://doi.org/10.1111/obes.12066>.
- Giller, K. E., Tittonell, P., Rufino, M. C., van Wijk, M. T., Zingore, S., Mapfumo, P., Adjei-Nsiahe, S., Herrero, M., Chikowod, R., Corbeels, M., Rowe, E. C., Bajjukya, F., Mwijage, A., Smith, J., Yeboah, E., van der Burg, W. J., Sanogo, O. M., Misiko, M., de Ridder, N., Karanjaf, S., Kaizzi, C., K'ungu, J., Mwale, M., Nwaga, D., Pacini, C. and Vanlauwe, B. (2011). Communicating complexity: Integrated assessment of trade-offs concerning soil fertility management within African farming systems to support innovation and development. *Agricultural Systems* 104(2):191–203. <http://doi.org/10.1016/j.agry.2010.07.002>.
- Global Yield Gap and Water Productivity Atlas. (GYGA, n.d). Retrieved 2017 from [www.yieldgap.org](http://www.yieldgap.org).
- Gollin, D. (2006). *Impacts of International Research on Intertemporal Yield Stability in Wheat and Maize: An Economic Assessment*. Mexico: CIMMYT.
- Hammond, J., Fraval, S., van Etten, J., Suchini, J. G., Mercado, L., Pagella, T., Frelat, R., Lannerstad, M., Douxchamps, S., Teufel, N., Valbuena, D. and van Wijk, M. T. (2017). The Rural Household Multi-Indicator Survey (RHOMIS) for rapid characterisation of households to inform climate smart agriculture interventions: Description and applications in East Africa and Central America. *Agricultural Systems* 151:225–233. doi: 10.1016/j.agry.2016.05.003.
- International Fund for Agricultural Development (IFAD). (2016). *Rural Development Report 2016*. Retrieved from [https://www.ifad.org/documents/30600024/30604583/RDR\\_WEB.pdf/c734d0c4-fbb1-4507-9b4b-6c432c6f38c3](https://www.ifad.org/documents/30600024/30604583/RDR_WEB.pdf/c734d0c4-fbb1-4507-9b4b-6c432c6f38c3).
- Jayne, T. S., Chamberlin, J., Traub, L., Sitko, N., Muyanga, M., Yeboah, F. K., Anseuw, W., Chapoto, A., Wineman, A., Nkonde, C. and Kachule, R. (2016). Africa's changing farm size distribution patterns: The rise of medium-scale farms. *Agricultural Economics* 47:197–214. doi: 10.1111/agec.12308
- Jerven, M. and Johnston, D. (2015). Statistical tragedy in Africa? Evaluating the data base for African economic development. *The Journal of Development Studies* 51(2):111–115. <https://doi.org/10.1080/00220388.2014.968141>.
- Juster, F. T., Cao, H., Couper, M., Hill, D., Hurd, M. D., Lupton, J., Perry, M. and Smith, J. P. (2007). *Enhancing the Quality of Data on the Measurement of Income and Wealth* (No. 151). Ann Arbor.
- Kalkuhl, M., Braun, J. Von and Torero, M. (2016). *Food Price Volatility and its Implications for Food Security and Policy*. Springer Open. <https://doi.org/10.1007/978-3-319-28201-5>.
- Kanyongo, G. Y., Brooks, G. P., Kyei-Blankison, L. and Gocmen, G. (2007). Reliability and statistical power: How measurement fallibility affects power and required sample sizes for several parametric and nonparametric statistics. *Journal of Modern Applied Statistical Methods* 6(1):81–90.
- Kilic, T., Carletto, C., Zezza, A. and Savastano, S. (2013). *Missing (Ness) in Action: Selectivity Bias in GPS-Based Land Area Measurements* (No. 6490). Washington, DC: World Bank. <http://doi.org/10.1016/j.worlddev.2016.11.018>.
- Kilic, T. and Sohnesen, T. P. (2015). *Same Question But Different Answer: Experimental Evidence on Questionnaire Design's Impact on Poverty Measured by Proxies. Review of Income and Wealth*. Washington, DC. <https://doi.org/10.1111/roiw.12343>.
- Leeuw, E. D. De. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics* 21(2):233–255.
- Little, T. D. and Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives* 7(4):199–204. <https://doi.org/10.1111/cdep.12043>
- Mathiowetz, N. A., Brown, C. and Bound, J. (2001). *Measurement Error in Surveys of the Low-Income Population. Studies of Welfare Populations: Data Collection and Research Issues*. (Vol. 1). Washington, DC: The National Academies Press.
- Moore, J. C., Stinson, L. L. and Welniak, E. J. J. (2000). Income measurement error in surveys: A review. *Journal of Official Statistics* 16(4):31–361. Retrieved from <http://www.jos.nu/Articles/abstract.asp?article=164331>.

- Neri, A. and Ranalli, M. G. (2012). *To Misreport or not to Report? The Measurement of Household Financial Wealth* (No. 870). October. Rome: Banca D'Italia. <http://doi.org/10.1162/JEEA.2008.6.6.1109>.
- Organisation for Economic Co-operation and Development (OECD), & Food and Agricultural Organization of the United Nations (FAO). (2017). *OECD-FAO Agricultural Outlook 2017-2026*.
- Organisation for Economic Co-operation and Development (OECD). (2009). *Methods to Monitor and Evaluate the Impacts of Agricultural Policies on Rural Development*. Paris: OECD.
- Pica-ciamarra, U., Morgan, N. and Baker, D. (2012). *Core Livestock Data and Indicators: Results of a Stakeholder Survey*. Rome: FAO.
- Reardon, T., Crawford, E. and Kelly, V. (1994). Links between nonfarm income and farm investment in African households: Adding the capital market perspective. *American Journal of Agricultural Economics* 76(5):1172–1176.
- Revelle, W. (2017). Psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych> Version = 1.7.8.
- Rosenstock, T. S., Lamanna, C., Chesterman, S., Hammond, J., Kadiyala, S., Luedeling, E., Shepherd, K., Derenzi, B. and Wijk, M. T. Van. (2017). When less is more: Innovations for tracking progress toward global targets. *Current Opinion in Environmental Sustainability* 26–27:54–61. <http://doi.org/10.1016/j.cosust.2017.02.010>.
- Rufino, M. C., Quiros, C., Boureima, M., Desta, S., Douxchamps, S., Herrero, M., Kiplimo, J., Lamissa, D., Mango, J., Moussa, A. S., Naab, J., Ndour, Y., Sayula, G., Silvestri, S., Singh, D., Teufel, N. and Wanyama, I. (2013). *Developing Generic Tools for Characterizing Agricultural Systems for Climate and Global Change Studies (IMPACTlite – Phase 2)*. Nairobi: ILRI.
- Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability.1. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 86(2):420–428. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18839484>.
- Swindale, A. and Bilinsky, P. (2006). *Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide (v.2)*. Washington, DC: FHI 360/FANTA.
- Thornton, P. K. and Herrero, M. (2015). livestock farming systems in sub-Saharan Africa. *Nature Publishing Group* 5(9):830–836. <https://doi.org/10.1038/nclimate2754>.
- Uganda Bureau of Statistics (UBOS). (n.d.). *The Uganda national panel survey 2009/10: Basic Information Document*. Kampala.
- Uganda Bureau of Statistics (UBOS). (2002). *2002 Uganda Population and Housing Census Analytical Report. Distribution*. Kampala. Retrieved from <http://www.ubos.org/onlinefiles/uploads/ubos/pdf> documents/2002 CensusPopnSizeGrowthAnalyticalReport.pdf.
- Uganda Bureau of Statistics (UBOS). (2007). *Uganda national household survey 2005/2006*. Kampala. Retrieved August 15, 2016 from [http://www.ubos.org/onlinefiles/uploads/ubos/statistical\\_abstracts/Statistical](http://www.ubos.org/onlinefiles/uploads/ubos/statistical_abstracts/Statistical) Abstract 2014.pdf.
- United Nations Department of Economic and Social Affairs (UN). (2005). *Household Sample Surveys in Developing and Transition Countries. Studies in Methods* (Vol. F). Retrieved from <http://unstats.un.org/unsd/hhsurveys/>.
- United Nations Framework Convention on Climate Change (UNFCCC). (2012). Standard for sampling and surveys for CDM project activities and programme activities. Bonn.
- Weisberg, H. (2005). *The Total Survey Error Approach*. Chicago, IL: The University of Chicago Press.
- World Bank. (2017). Living Standards Measurement Survey. Retrieved 15 Jan 2017 from [www.worldbank.org/lsms](http://www.worldbank.org/lsms).
- World Bank. (n.d) Living Standards Measurement Study-Integrated Surveys on Agriculture. Retrieved September 15, 2018 from <http://surveys.worldbank.org/lsms/programs/integrated-surveys-agriculture-ISA>.
- Zezza, A., Federighi, G., Adamou, K. and Hiernaux, P. (2014). *Milking the Data: Measuring Income from Milk Production in Extensive Livestock Systems Experimental Evidence from Niger* (No. 7114). Rome.