



OPEN

DATA DESCRIPTOR

The Rural Household Multiple Indicator Survey, data from 13,310 farm households in 21 countries

Mark van Wijk^{1,25} , James Hammond^{1,25}, Leo Gorman^{2,25}, Sam Adams¹, Augustine Ayantunde¹, David Baines³, Adrian Bolliger^{4,5}, Caroline Bosire¹, Pietro Carpena³, Sabrina Chesterman⁶, Amon Chinyophiro⁷, Happy Daudi⁸, Paul Dontsop⁹, Sabine Douxchamps , Willy Desire Emera¹, Simon Fraval^{1,10}, Steven Fonte¹¹, Lyda Hok , Henry Kiara¹, Esther Kihoro¹, Luke Korir¹, Christine Lamanna⁶, Chau T. M. Long¹³, Godfrey Manyawu¹, Zia Mehrabi¹⁴, Dejene K. Mengistu¹⁵, Leida Mercado¹⁶, Katherin Meza^{11,17}, Vesalio Mora¹⁸, Jacob Mutemi¹⁹, Mary Ng'endo⁶, Paulin Njingulula⁹, Chris Okafor⁹, Tim Pagella²⁰, Phonepaseuth Phengsavanh²¹, James Rao¹, Randall Ritzema^{1,22}, Todd S. Rosenstock⁶, Tom Skirrow³, Jonathan Steinke¹⁵, Clare Stirling²³, Jose Gabriel Suchini¹⁶, Nils Teufel¹, Peter Thorne¹, Steven Vanek¹¹, Jacob van Etten , Bernard Vanlauwe⁹, Jannike Wichern²⁴ & Viviane Yameogo¹

The Rural Household Multiple Indicator Survey (RHoMIS) is a standardized farm household survey approach which collects information on 758 variables covering household demographics, farm area, crops grown and their production, livestock holdings and their production, agricultural product use and variables underlying standard socio-economic and food security indicators such as the Probability of Poverty Index, the Household Food Insecurity Access Scale, and household dietary diversity. These variables are used to quantify more than 40 different indicators on farm and household characteristics, welfare, productivity, and economic performance. Between 2015 and the beginning of 2018, the survey instrument was applied in 21 countries in Central America, sub-Saharan Africa and Asia. The data presented here include the raw survey response data, the indicator calculation code, and the resulting indicator values. These data can be used to quantify on- and off-farm pathways to food security, diverse diets, and changes in poverty for rural smallholder farm households.

¹International Livestock Research Institute (ILRI), PO Box 30709, Nairobi, 00100, Kenya. ²Alan Turing Institute, British Library, 96 Euston Road, London, NW1 2DB, England. ³TreeAID, Brunswick Court, Brunswick Square, Bristol, BS2 8PE, UK. ⁴International Centre for Tropical Agriculture (CIAT), Asia Regional Office, Hanoi, Vietnam. ⁵Hanns R. Neumann Stiftung, Kota Liwa, 34812, Lampung Barat, Indonesia. ⁶World Agroforestry (ICRAF), PO Box 30677-00100, UN Avenue, Nairobi, Kenya. ⁷NASFAM, African Unity Avenue, Lilongwe, Malawi. ⁸Tanzania Agricultural Research Institute – Naliendeke, Mtwara, Tanzania. ⁹International Institute of Tropical Agriculture (IITA), P.O. Box 30772-00100, Nairobi, Kenya. ¹⁰Animal Production Systems Group, Wageningen University & Research, P.O. Box 338, Wageningen, 6700 AH, The Netherlands. ¹¹Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, 80523, USA. ¹²Faculty of Agronomy and Centre of Excellence on Sustainable Agricultural Intensification and Nutrition, Royal University of Agriculture, Phnom Penh, 12401, Cambodia. ¹³Western Highlands Agriculture and Forestry Science Institute (WASI), Buon Ma Thuot, Vietnam. ¹⁴The UBC School of Public Policy and Global Affairs, University of British Columbia, Vancouver, British Columbia, V6T 1Z2, Canada. ¹⁵Bioversity International, Rome, Italy. ¹⁶CATIE - Centro Agronómico Tropical de Investigación y Enseñanza, Turrialba, Costa Rica. ¹⁷Grupo Yanapai, Calle Tripoli 365, Miraflores, Lima, Peru. ¹⁸Ministerio de Agricultura y Ganadería, Siquirres, Costa Rica. ¹⁹Lutheran World Relief, Nairobi, Kenya. ²⁰School of Natural Sciences, Bangor University, Bangor, UK. ²¹National Agriculture and Forestry Research Institute (NAFRI), Vientiane, Lao PDR, Laos. ²²Olivet Nazarene University, One University Avenue, Bourbonnais, IL60914, USA. ²³International Maize and Wheat Improvement Centre (CIMMYT), Sustainable Intensification Program, Mexico, DF, Mexico. ²⁴Plant Production Systems group, Wageningen University, P.O. Box 338, 6700 AH, Wageningen, The Netherlands. ²⁵These authors contributed equally: Mark van Wijk, James Hammond and Leo Gorman. ✉e-mail: m.vanwijk@cgiar.org



Fig. 1 Key indicator groups (modules) generated by all RHoMIS applications.

Background & Summary

Agriculture is the most important livelihood option for most rural households in low- and middle-income countries¹. Smallholder farm households in these locations produce food not only for themselves, but, in many countries, produce the majority of the national or even the regional food supply^{1,2}. Smallholder farm households are also highly diverse³, varying in land area, amount of livestock present, crops grown, and farm management strategies. The importance of understanding the diversity and dynamics of rural households is increasingly crucial, given the diverse effects of global changes in climate, population growth, urbanization, and food demand^{4,5}.

Achieving the sustainable development goals (in particular the goals of no poverty and zero hunger, but others too) requires more intensified sustainable food production and development of rural economies. Targeted investment to make progress in agricultural development requires understanding the links between farming practices, livelihood practices, and the effects on farm performance and household welfare. Reliable indicators at farm-household level of both farm performance and household welfare are therefore needed to better understand and model these linkages, and to inform the design and implementation of interventions by governments, donors, and international agencies, across a wide range of differing geographies and socio-economic dimensions⁶.

The lack of standardization of agricultural household surveys, especially in international ‘agriculture for development’ research, has resulted in a proliferation of survey tools and indicators leading to datasets which are often badly documented, incoherent, and with limited interoperability. An example of the consequences of this situation is the study of Frelat *et al.*³, which brought together a series of different household survey datasets, but had a hard time defining a common indicator of food security that could be quantified across all these datasets. The current state of affairs limits our ability to compare outcomes across studies and to draw general conclusions on the effectiveness of interventions and the trade-offs between outcomes, which may be shaped by household structure, farm management and the wider social-environmental context^{3,5}. Efforts like the CGIAR’s Big Data Platform have also recognized this situation, and try to define common layouts for household surveys and sets of ontologies underpinning the information to be collected in household surveys⁷.

In contrast, RHoMIS (Rural Household Multiple Indicator Survey; www.rhomis.org) is a standardized household survey approach designed to rapidly characterise a series of key indicators across the spectrum of agricultural production and off farm activities, alongside market integration, nutrition, food security, poverty and greenhouse gas (GHG) emissions⁸ (Fig. 1). It includes a modular survey tool which takes 40–60 minutes to administer per household, a digital platform to store and aggregate incoming data as well analysis code to quantify indicators and visualize results. Optional modules can be bolted-on. The tool has been systematically designed to enable the quantification of interactions between different components and outcomes of agricultural systems, including productivity, and human welfare at the farm and household level, and it has been widely adopted by research organisations and development partners⁸. Such a streamlined, modular approach has resulted in a strong reduction in costs⁹ compared to traditional households surveys in the field (which in other approaches typically take 2–3 hours per household¹⁰) and of the subsequent data analysis and reporting¹¹.

A harmonized dataset has been developed from all the applications of RHoMIS that took place during the years 2015, 2016, 2017 and the first three months of 2018, resulting in a dataset collected from 13,310 farm households across 21 low- and middle-income countries. The overall database (available at the Harvard Dataverse RHoMIS data repository) consists of the raw data (the 758 variables mentioned above; see subdirectory ‘data to share\rhomis_full_data.csv’) and 41 indicators calculated based on the information provided by these variables (see subdirectory ‘data to share\rhomis_indicators.csv’). The raw data and indicators have already been used for a



Fig. 2 Geographical location of the observations included in the RHoMIS dataset.

wide range of studies at site level^{8,12–17}, for regional analyses^{18,19}, and for continental analysis²⁰. Different aspects of smallholder households have been analysed, including gender equity^{19,21}, dietary diversity^{17,18}, nutritional gaps²⁰, poverty and GHG emissions in relation to production intensification⁸, subsistence- versus market-orientated strategies²⁰, and on-farm vs. off-farm activities²⁰. RHoMIS is an on-going initiative, and we welcome interested parties to the community of practice (see www.rhomis.org for up-to-date information and downloadable survey questionnaires). Records continue to be submitted to the central data repository: in the latter part of 2018 more than 10,000 households were additionally interviewed, and their information added to the database. Further releases will be made public in the near future.

Methods

Basic characteristics and geographic coverage. The countries in which survey data included in the current database were collected, are summarized in the Online-only Table 1, together with key metadata: the research or development project in which the survey was applied, the lead organisation implementing the survey, the number of households surveyed, and a brief summary of the sampling strategy. The locations of the surveyed households are shown in Fig. 2, demonstrating the geographical breadth of the dataset. The dataset also includes detailed documentation and metadata ('MetaData RHoMIS survey applications.docx'), describing the aims of each project in which RHoMIS was applied, the sampling strategy used, and additional data collected outside the core set of RHoMIS variables (i.e., topics beyond those presented in Fig. 1). These additional, project specific data that are outside the core RHoMIS variables but are available on request from the corresponding author of this manuscript, always in consultation with the representative of the organisation which executed the RHoMIS application in the field.

The total number of households in the data set is 13,310. The data are all based on single cross-sectional survey applications. In most applications, households were chosen randomly within the sites in which the different implementing projects were working, but some sampling strategies were aimed at evaluations of project interventions (see Online-only Table 1, sampling strategy). Three country-wide applications are available in the current database: two led by the iNGO TreeAID in Burkina Faso, where households were selected across the major agro-ecological zones of the country, and one led by the International Livestock Research Institute in Tanzania that sampled cattle owning households, randomly selected from animal health service records. As is clear from Online-only Table 1, the sample sizes differ per application. It is also of key importance to refer to the metadata information file to assess the representativeness of each of the RHoMIS applications. In previous analyses that aimed for statistical inferences valid for farm type, households were grouped by their farm characteristics²⁰, or if an analysis aimed for spatial representativeness, observations were spatially clustered and resampled or weighted by the local population (e.g. village) densities¹⁷. The varied sampling procedures followed by the individual applications make it essential to weight and/or re-sample households in any analysis making use of this combined RHoMIS dataset for valid statistical inferences.

The questionnaire. The RHoMIS questionnaire is a set of carefully, expertly designed modules that are administered digitally using the ODK software platform (<https://opendatakit.org/>)⁸. The survey is designed to be both flexible enough to suit local contexts and sufficiently standardised to permit rapid deployment, analysis and comparison between multiple sites, without the need for costly post-survey harmonization. The data package that is made available consists of two parts: the dataset itself (containing the raw data and the indicator results) and secondly, the series of documents and analysis code files underpinning the raw data collection effort and the subsequent indicator quantification. We have made the survey available in easily readable pdf format ('RHoMIS for printing_v1.3.pdf'). The questionnaires and their variable names have been linked to a supporting set of data extraction and analysis tools written in R (<https://cran.r-project.org/>), also included in the package. The majority of questions in the survey are used for the estimation of a series of pre-defined indicators that include:

- The Household Food Insecurity of Access Scale²² for measuring the frequency and severity of hunger (this indicator in more recent applications of RHoMIS is replaced by the FIES – Food Insecurity Experience Scale indicator; <http://www.fao.org/in-action/voices-of-the-hungry/background/en/>).

- The Household Dietary Diversity Score²³, providing an indicator of household dietary adequacy; this indicator was adapted to cover both the bad and good seasons
- The Probability of Poverty Index^{24,25}, an asset-based scoring system to estimate the likelihood that a household is in poverty.
- The Potential Food Availability indicator for quantifying the ability of a household to feed itself through both on-farm and off-farm activities^{3,4}.

These indicators are combined with a comprehensive inventory of agricultural crops and livestock production characteristics, including yields, the use of products (consumption, sale, etc.), product sale prices, input use, and an assessment of off-farm incomes. The data captured in the RHoMIS tool place the farm household along a continuum of household and farm characteristics, performance indicators, and welfare indicators (see Fig. 1), enabling in-depth analyses of individual indicators, but also integrative analyses of how indicators co-vary and how on-farm and off-farm livelihood strategies correlate to food security, poverty, and dietary diversity. The questionnaire is organised into seven sections wherein respondents are asked to provide information on the previous 12 months' farming and non-farming activities (Table 1). We calculated the indicator values using custom code, available under the subdirectory 'R scripts' at the Harvard Dataverse RHoMIS data repository. The setup of the code is explained in detail in 'RHoMIS Data Processing Doc.docx', and a pdf and an excel ODK definition file explaining each variable, are also supplied with the data.

The data collection efforts conformed with the principles of the 1964 WMA declaration of Helsinki. Ethical approvals for the survey applications was obtained by the internal ethical review committees of the different institutes (e.g. the Internal Review Ethics Committee (IREC) of the International Livestock Research Institute) or for those partners without an internal ethical committee, by ethical evaluation by the senior management at each organization after careful evaluation of the content, methodology, and with oral informed consent statement built-in to the survey. Survey participants were not particularly vulnerable, data was processed in anonymized form, and survey participants had the possibility to skip questions. Explicit oral informed consent was obtained from all survey participants prior to survey enumeration and documented as the opening question in the RHoMIS survey upon informing survey participants of the study's purpose. If consent was denied, survey enumeration was terminated. Permission for obtaining oral rather than written consent from survey respondents was granted by the Internal Review Ethics Committee (IREC) of the International Livestock Research Institute, implementing research organizations and local agricultural officers, given literacy limitations among the target populations.

Data processing and indicator calculations. A standard set of scripts has been created in the R software environment and used to process the raw data provided in this dataset, and for the calculation of the various indicators. An overview of the data processing and indicator calculations is given in Fig. 3, differentiating between outside information used (e.g. energy content of food items, in dark brown), the different data products produced that are included in this dataset (in light brown) and the processing steps (in blue). We now explain the procedure in more detail. The starting point are the 'raw collected' data. We have limited the data cleaning step (step 2) to only correcting obviously unrealistic values, such as when it was clear that the recorded value in combination with the reported unit generated an impossible indicator value (say 1.5 kg of maize yield from a 1 ha field; here it is clear that the unit should be tons of maize; or similarly if a production of 3,000 tons from a 1 ha field is reported, it is clear that the unit should be kg). However, the most important step in the cleaning process was unifying crop names, livestock species names, and the crop and livestock product names (step 2 & 3 in Fig. 3). This cleaned package from all survey implementations was then merged into a single raw data csv-file, supplied in the data package (step 6 in Fig. 3).

Conversion factors use in the indicator calculations, for example area units, monetary units, or energetic content of foods, are based on literature resources (step 7 in Fig. 3). Within the calculations of some indicators, e.g. income, value of farm production, or crop and livestock productivity, intermediate output variables were also generated: farm gate prices of farm products, and crop and livestock production values per individual crop or livestock species (step 10 in Fig. 3). These are included in the data package. The final product of the indicator calculations is the merged RHoMIS indicator results file, step 13 in Fig. 3. The annotated R code, for performing and documenting all the processing and calculation steps illustrated in Fig. 3, is also supplied. The individual indicator calculations are explained in detail in the 'Explanations_of_Calculations_and_Outputs.xlsx' file.

With the current data management system we follow the FAIR principles: Findability, Accessibility, Interoperability, and Reusability²⁶. By using standardized data-labelling and data processing approaches across all current and future RHoMIS applications we follow the principle of 'assisting machines in their discovery and exploration of data through application of more generalized interoperability technologies and standards at the data/repository level'. The FAIR principles are adhered to via (i) the extended metadata and documentation available at project and survey level (e.g. 'MetaData RHoMIS survey applications.docx'); (ii) the publicly available survey and processing software; as well as (iii) the standardized approach of the core RHoMIS survey.

Data Records

The RHoMIS data²⁷ can be found at Harvard Dataverse, <https://doi.org/10.7910/DVN/9M6EHS>.

All 758 variables of the survey data are described in the file 'Raw Data code book.xlsx'. The variables included in indicator results file are described in 'RHoMIS Data Processing Doc.docx'. We have also included the RHoMIS survey in ODK definition and pdf format to further facilitate the interpretation of the variable names. The RHoMIS survey is continuously updated; the newest version can be found at www.rhomis.org.

Indicator Group	Questions
Metadata	Informed consent; location of the survey (village, district, region, country); project beneficiary; local monetary unit; GPS coordinates of where the interview took place
Household demographics	Respondent sex; whether respondent is household head; age of respondent; type of household (single female; single male; couple; couple - spouse living away); age and sex of household members
Land use/tenure	Total land area cultivated; amount of land owned; access to common land; land rented out/in
Crop production and use	Crops/vegetables/fruits/trees grown, area grown; production per year/season; use of products (sold, consumed, stored, processed, given away); price of crop products; use of crop residues
Agricultural inputs and management	Use of mineral fertilizer, manure, pesticides; irrigation: type, area and/or amount used; intercropping; agroforestry; medicine use for livestock
Livestock	Livestock species kept on farm; production per year/season; use of products (sold, consumed, stored, etc.); price of livestock products
Wild foods	Relative importance of wild foods for food security; months of the year collected; most important wild food items
Food security and hunger	9 standard questions of Hunger and Food Insecurity Access Scale (HFIAS); number of months with hunger; identification of worst and best month in terms of food security
Household dietary diversity	10 food groups of the Minimum Dietary Diversity Score for Women in the worst, best and last month; source of each food group (farm-based, purchased, gathered or a combination of sources)
Off-farm income	Relative importance of off-farm income in overall income; use of income (reinvestment in farm, household (non-food) expenses, food purchase)
Credit, aid and debts	Household access to credit; dependency on/receipts of aid in the last 12 months, from institutions or neighbours; household debt
Probability of Poverty Index (PPI)	10 standard questions, country-specific (see https://www.povertyindex.org/)
Gendered control of produce and income	Per produce and income item: who takes decisions on sale, spending, and consumption

Table 1. General overview of the information (captured by the 758 variables shared in the dataset) captured, organised by indicator group (see Fig. 1) in RHoMIS.

Technical Validation

Each RHoMIS survey application included in the data set has undergone a series of standard data quality evaluation steps to assess the overall quality of the data collected. This did not lead to removal of individual observations, we prefer that the user of the data can make their own informed decisions on this topic. The validation information supplied here is only advice on how the data user might be able to check the quality of the data they want to use and to be able to base decisions on a systematic approach if he or she does decide to remove observations or data points. Note that we used the validation approach (described below) to evaluate each application made available in the current dataset, and that for each application no more than 25% of the data were found to need more attention, which compared to other household survey tools is a good score¹¹.

Before validation, two key quality filters were applied, correcting some of the observations. One is the earlier mentioned ‘correcting obviously unrealistic values’, see the text accompanying Fig. 3 for explanation. The second is correction of farm gate price values for the different types of farm produce, which is used in the calculation of several RHoMIS reported indicators. Especially when there are low numbers of observations of a certain type of farm produce, reported price values can be quite extreme values, even when median values are calculated across the dataset of the survey application. We therefore defined ranges of prices (price reported plus or minus 100%) for each commodity based on FAOSTAT data to especially avoid unrealistically high prices which would affect some of the indicators strongly.

The validation process entailed three steps. In step 1 consists of subjective evaluation by the enumerators themselves. These questions are ‘In your opinion, how easily did you establish rapport with the respondent?’ (with possible answers: ‘Easily’, ‘OK’, ‘Difficult’ and ‘Very difficult’) and ‘How reliable do you think these answers are? Consider the accuracy and willingness to answer.’ (with possible answers: ‘Very reliable’, ‘Reliable’, ‘OK’, ‘Occasional doubts’ and ‘Regular or serious doubts’). This information, in combination with the total survey duration (based on start and end time of survey implementation), provides insight on the overall reliability of the information recorded. For example, if the survey duration is extremely long (e.g. beyond 2 h) it is also likely that data quality will be questionable. Example results of these quality indicators are given in Fig. 4, showing the typical distribution of enumerator observed reliability and survey length of survey. This may serve as a useful norm for the initial quality assessment of new RHoMIS applications. The information collected in individual survey applications where the reliability answers show low scores, or where many survey records show abnormally short or long durations, can be more thoroughly investigated or even rejected.

In step 2 we follow the food availability/self-sufficiency approach^{3,11} to evaluate the overall reliability of the data on production and consumption of farm produce. Calculated livestock productivity and crop yield values are compared to realistic ranges normally encountered in smallholder systems in the agro-ecosystem of interest²⁸. Because the plausible value ranges for these checks are still large, we also scrutinise composite indicators that combine information from a number of survey variables. The two composite indicators we examine for quality assessment are food self-sufficiency and potential food availability³. The food availability indicator represents the total food energy potentially available daily per household member (adjusted to adult male equivalent calorie demand), and is calculated from the reported consumption of farm products, from cash sales of farm products, and from off-farm activities, whereby all income is converted to a calorific value based on the cost of a local staple crop. Results of these calculations can be used to assess the data quality of information on crop

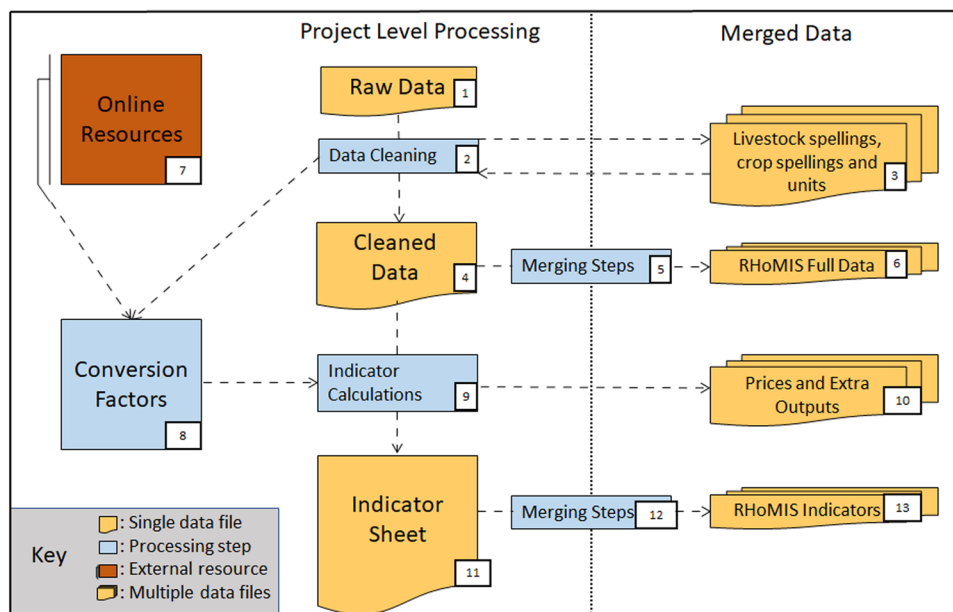


Fig. 3 Overview of the data processing and indicator calculation steps.

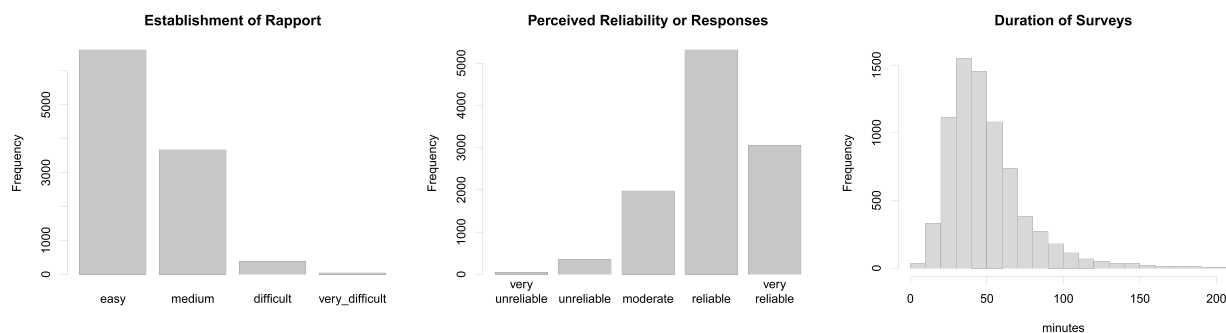


Fig. 4 Overview of the three key aspects considered in step 1 of the data quality evaluation (for explanation see text): subjective enumerator evaluation of data quality through ‘establish rapport’, ‘perceived reliability’, and survey duration.

and livestock production, sales and consumption as well as off-farm income. Two problems with this composite indicator are commonly encountered. First, a considerable number of household records at the lower end of the food availability scale appear to suggest an underestimation of calorie availability, suggesting an extreme level of starvation. This may be a true representation of some households, but it can also be an indication of missing information on income or food consumption. Second, a substantial number of households can also show a substantial over-estimation of consumption of crop and livestock products – indicating possible problems with yield, consumption or household size estimates. We set a lower bound threshold for food availability at 1,250 kilocalories (kcal) per male adult equivalent per day, which is below the basal metabolic rate for adult males (approximately 1,590 kcal for a 60 kg male) and adolescents (1,360 kcal for a 40 kg adolescent male). We set an upper bound for food self-sufficiency (i.e. consumption only) at: (a) 3,500 kcal per male adult equivalent per day, representing the average intake of developed nations²³; and (b) 5,000 kcal, which is double the approximate requirement for an adult male. Observations which fall outside these bounds (e.g. Fig. 5) are used to examine the overall reliability of the survey application and can be used to identify individual survey applications where the data do not appear reliable from a composite perspective. Typically, between 10 and 25% of the records within the total survey sample of an application site may show questionable values of these indicators. These performance values are representative for the uncertainty encountered in recall based cross-sectional farm household surveys, and actually better than the performance of two other widely used sources of farm household information (Fig. 5). All sites included in the databases published in this article fall within this performance range. Further trust in these findings can be developed by triangulating these results with other indicators of food security included in the RHoMIS surveys, for example HFIAS, dietary diversity and the number of months with hunger (see also step 3).

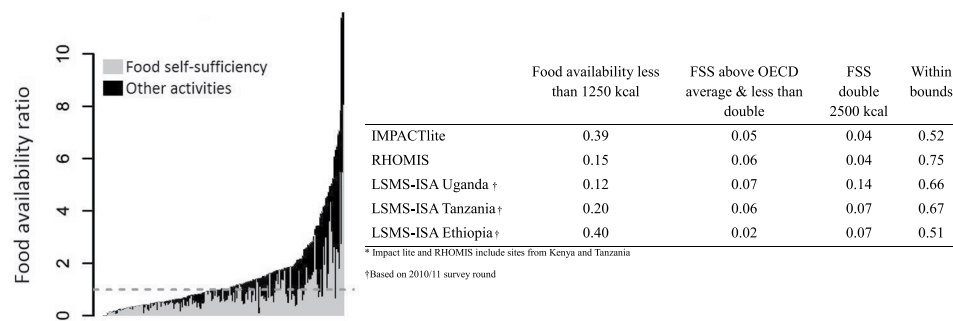


Fig. 5 Step 2 of the data quality evaluation: the identification of systematic under- or over-estimation of production and consumption values based on food self-sufficiency and potential food availability, presented as the relative number of observations. RHoMIS results are compared with results obtained by other survey instruments (ImpactLite and the Living Standards Measurement Survey – Integrated Survey on Agriculture (LSMS-ISA)) in the same countries. Reported elsewhere in detail¹¹.

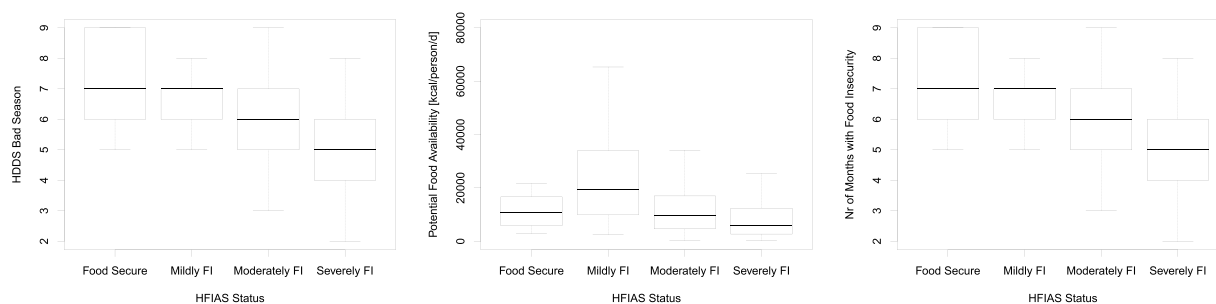


Fig. 6 Relationships encountered between the household dietary diversity score in the bad season (HDDS Bad Season), potential food availability, the number of months with food insecurity, and the Hunger and Food Insecurity Access Scale (HFIAS) status, based on the 2016 Kenya Wote dataset.

In step 3 we cross-checked relationships between different food security indicators. The information collected by RHoMIS underpins 4 different indicators of food security (HFIAS, dietary diversity, number of months with hunger, and potential food availability). Even though these indicators capture different aspects of food security, we do expect strong correlation between them. Figure 6 provides examples of the typical relationships encountered between these variables. A strong deviation of the relationships found in a new RHoMIS application from these typical relationships may necessitate a deeper investigation, and might indicate questionable data quality in one or more of these indicators.

Usage Notes

The objective of RHoMIS is to gather information on the common variables of interest in all agricultural development research, but not to go too deep into any one topic. The overall strategy of RHoMIS is to collect data which permits an overview of the farming system and the main livelihood activities. Based on this information we can identify farm level constraints, deficiencies or successes, and sift meaning from the high degree of variation observed amongst smallholder households. This is in contrast to the design of many impact assessment studies, which collect data on a narrow topic but at a higher resolution, thus permitting evaluation of that specific topic, but limiting the ability to assess the over-arching farming system and rural livelihoods. This dataset can be used to investigate the characteristics of agricultural systems in low- and middle-income countries at one point in time. On- and off-farm strategies can be analysed in relation to a series of welfare indicators representing food security, poverty and gender equity. This can be done both at individual household level, to disentangle the livelihood strategies of specific household types, but also at population level to determine more generic patterns and investigate equity aspects. We emphasize that the results of such analysis should be interpreted within the context of the data set, considering the potential biases and limitations described in this paper. It is especially important that users note the varied sampling procedures followed in the different RHoMIS applications, and recognise that it essential to weight and/or re-sample the households in any analysis before making any across-site statistical inferences. See the Methods section for more details.

Received: 1 November 2019; Accepted: 14 January 2020;

Published online: 11 February 2020

References

- Herrero, M. *et al.* Farming and the geography of nutrient production for human use: a transdisciplinary analysis. *Lancet Plan. Health* **1**, e33–e42 (2017).
- Samberg, L. H., Gerber, J. S., Ramankutty, N., Herrero, M. & West, P. C. Subnational distribution of average farm size and smallholder contributions to global food production. *Environ. Res. Lett.* **11**, 124010 (2016).
- Frelat, R. *et al.* Drivers of household food availability in sub-Saharan Africa based on big data from small farms. *Proc. Natl. Acad. Sci. USA* **113**, 458–463 (2016).
- Ritzema, R. S. *et al.* A simple food availability analysis across smallholder farming systems from East and West Africa: Is production intensification likely to make farm households food-adequate? *Food Sec.* **9**, 115–131 (2017).
- Waha, K. *et al.* Agricultural diversification as an important strategy for achieving food security in Africa. *Glob. Chang. Biol.* **24**, 3390–3400 (2018).
- Coe, R., Sinclair, F. & Barrios, E. Scaling up agroforestry requires research ‘in’ rather than ‘for’ development. *Curr. Opin. Env. Sust.* **6**, 73–77 (2014).
- Van Wijk, M. *et al.* Towards a core approach for cross-sectional farm household survey data collection: a tiered setup for quantifying key farm and livelihood indicators. Community of Practice on Socio-economic Data report COPSED-2019-001 (CGIAR Platform for Big Data in Agriculture, 2019).
- Hammond, J. *et al.* 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. *Agric. Syst.* **151**, 225–233 (2017).
- Rufino, M. C. *et al.* Developing Generic Tools for Characterizing Agricultural Systems for Climate and Global Change Studies (IMPACTlite – Phase 2). Nairobi (ILRI 2012).
- Vuong, Q.-H. The (ir)rational consideration of the cost of science in transition economies. *Nature Hum. Behav.* **2**, 5 (2018).
- Fraval, S. *et al.* Making the most of imperfect data: a critical evaluation of standard information collected in cross-sectional farm household surveys. *Exp. Agric.* **55**, 230–250 (2019).
- Hammond, J. *et al.* Shea butter: a pro-poor, pro-female route to increased income. In: Rosenstock, T., Nowak, A. & Girvetz, E. (Eds.). *The Climate-Smart Agriculture Papers: Investigating the Business of a Productive, Resilient and Low Emission Future* (Springer International Publishing 2019).
- Fraval, S. *et al.* Livelihoods and food security in an urban linked, high potential region of Tanzania: Changes over a three year period. *Agric. Syst.* **160**, 87–95 (2018).
- Steinke, J. *et al.* Prioritising household-specific options for agricultural development through the Positive Deviance approach. *PlosOne* **14**, e0212926 (2019).
- Steinke, J. *et al.* Household-specific targeting of agricultural advice via mobile phones: Feasibility of a minimum data approach for smallholder context. *Comp. Electr. Agric.* **162**, 991–1000 (2019).
- Bosire, C., Rao, J., Hammond, J., Lukuyu, B. & van Wijk, M. T. Adaptation opportunities for smallholder dairy farmers facing resource scarcity: integrated livestock, water and land management. *Agric. Ecosyst. Env.* **284**, 106592 (2019).
- Beveridge, L. *et al.* Constructing and deconstructing ‘food security’ across scales: Household indicators and lived experiences in the dry corridor of Central America. *Front. Sust. Food Syst.* **3**, 65 (2019).
- Ritzema, R. S. *et al.* Household-level drivers of dietary diversity in transitioning agricultural systems: evidence from the Greater Mekong Subregion. *Agric. Syst.* **176**, 102657 (2019).
- Tavener, K. *et al.* Intensifying Inequality? Gendered Trends in Commercializing and Diversifying Smallholder Farming systems in East Africa. *Front. Sust. Food Syst.* **3**, article 10 (2019).
- Fraval, S. *et al.* Nutritional gaps of rural households in east and west Africa: prevalence and determinants based on rapid indicators. *Front. Sust. Food Syst.* **3**, 104 (2019).
- Van Wijk, M. *et al.* Including gender equity in a survey tool for rural households. In: *A Different Kettle of Fish. Gender Integration in Livestock and Fish Research*, ed. Pyburn, R. (Amsterdam: LM Publishers), 9–17 (2016).
- Coates, J., Swindale, A., Bilinsky, P. *Household Food Insecurity Access Scale (HFIAS) for measurement of food access: indicator guide*, Washington, DC (2007).
- Swindale, A. & Bilinsky, P. *Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide (v.2)*. Washington, D.C. FHI 360/FANTA (2006).
- Desiere, S., Vellema, W. & D’Haese, M. A validity assessment of the Progress out of Poverty Index (PPI)TM. *Eval. Progr. Plan.* **49**, 10–18 (2015).
- Grameen Foundation. *Poverty Probability Index*, <https://www.povertyindex.org/> (2015).
- Wilkinson, M. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
- Van Wijk, M. *et al.* The Rural Household Multiple Indicator Survey (RHoMIS) data of 13,310 farm households in 21 countries. *Harvard Dataverse*. <https://doi.org/10.7910/DVN/9M6EHS> (2019).
- Organisation for Economic Co-operation and Development (OECD), & Food and Agriculture Organisation of the UN (FAO). *OECD FAO Agricultural Outlook 2017–2026* (2017).

Acknowledgements

We are grateful to the research teams involved in designing and implementing all surveys made available in this study. Without their rigor, openness, and thorough documentation this study would not have been feasible. This study was made possible by the CGIAR Research Program on Livestock and its donors, and through the 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). We also thank all donors who made the survey implementations possible through their financial support to the respective projects described in the metadata information file.

Author contributions

J.H. and M.v.W. designed the RHoMIS tool and coordinated the overall work. J.v.E., N.T., T.P. and T.R. gave conceptual input to the RHoMIS tool development. L.G. prepared the final data, indicator calculations and R scripts. S.A. coordinated and supported several of the RHoMIS applications presented in the dataset. A.A., D.B., A.B., C.B., P.C., S.C., A.C., H.D., P.D., S.D., W.E., S.F., S.F., L.H., H.K., E.K., L.K., C.L., C.T.M.L., G.M., K.M., D.M., L.M., K.M., V.M., J.M., M.N., P.N., C.O., P.P., J.R., R.R., T.R., T.S., J.S., C.S., J.S., N.T., P.T., S.V., J.v.E., B.V., J.W. and V.Y. have led the RHoMIS applications that form the RHoMIS dataset made available. M.v.W. drafted the manuscript, with critical input from Z.M., J.H. and S.F. All co-authors contributed to the finalisation of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.v.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2020