RAM-OP: A rapid assessment method for assessing the nutritional status, vulnerabilities, and needs of older people in emergency and development settings

Experiences from a field trial in a rural setting

Pascale Fritsch (HelpAge International), Katja Siling (Valid International), Ernest Guevarra (Valid International), and Mark Myatt (Brixton Health)

Introduction

Older people remain a neglected group in humanitarian responses even as their numbers are growing as life-expectancy increases in much of the developed and developing world. This neglect is particularly obvious in the nutrition sector where the overwhelming majority of activities target children and women of childbearing age. Older people are vulnerable to malnutrition but their nutritional status and needs are very seldom assessed and almost never addressed. There is no support for assessing the needs of older people in SMART surveys [SM]. Survey methods such as the Multiple Indicator Cluster Survey (MICS) and the Demographic and Health Survey (DHS) methods collect either no data or very limited demographic data for persons aged 50 years or older [M1,D1]. There exists, therefore, a need for a simple, cheap, and rapid tool to assess the needs of older people and to enable humanitarian actors to advocate, plan, and deliver relevant and comprehensive responses to need in older people.

In January 2014, HelpAge International, Valid International Ltd., and Brixton Health, funded by the Humanitarian Innovation Fund (HIF), began developing a novel method for assessing the nutritional and other needs of older people in emergency and development settings. The **Rapid Assessment Method for Older People** (RAM-OP) is intended to offer a simple, rapid, low cost, accurate, and reliable survey method for assessing the nutritional status, vulnerabilities, and needs of older people.

The RAM-OP method will use a two-stage spatial sample with a small (e.g. $m = 16$ clusters) and spatially even first-stage sample and a small (i.e. $n \leq 200$) overall sample. Modern computationally intensive data analysis procedures will be used. Free and open-source data analysis software will be developed. Guidelines, articles, and training material will be produced. The focus of the work developing RAM-OP is on older people but elements of the work will also be applicable for assessing the nutritional status, dietary diversity, IYCF practices, food security and other indicators in other groups.

The first field trial of the RAM-OP method was undertaken in Addis Ababa (Ethiopia) in early 2014. Testing involved performing two surveys using the SMART method and a prototype of the RAM-OP method simultaneously in the same population and comparing the two surveys in terms of resource requirements and estimates of indicator levels. The results of the first field trial have been described elsewhere [R1]. This trial found that SMART and RAM-OP are functionally similar methods (i.e. the two methods return similar results with similar precision) and that the RAM-OP method was considerably cheaper than the SMART method. The results of this trial informed the design of a second trial which was undertaken in Kibaha District, Pwani (Coastal) Region of Tanzania in August and September 2014.

This article describes key elements of the RAM-OP method, the methods and results of the second field trial of the RAM-OP method in Tanzania, and compares the results of the two field trials.

Survey Setting

Kibaha District is an administrative district in the Pwani (Coastal) Region of Tanzania and is located to the west of Dar es Salaam. It is a predominantly rural area with a population of approximately seventy thousand persons according to the 2012 census. Older people (i.e. people aged 60 years and older) represent about eight per cent of the district population (i.e. about six thousand people). The principal town in Kibaha District is Mlandizi with a population of about seventeen thousand persons.

Kibaha District is partitioned into eleven electoral and administrative divisions (wards) and contains more than fifty villages each of which are surrounded by numerous sub-villages or hamlets. Wards and villages are administered by elected executive committees. Villages are large in terms of their geographic extent. Some villages have an easily identifiable centre with a marketplace, civic and commercial building, and housing areas but many are geographically extensive clusters of scattered dwellings whose boundaries are difficult to identify without the aid of local informants.

Kibaha District was chosen as the location for the work described here because of the need to test the RAM-OP method in a rural setting and because HelpAge International has a long-term working partnership with the Good Samaritan Social Service Trust (GSSST), a local NGO working with older people at the community level in the health and livelihood sectors in Kibaha District.

Sampling

Sample sizes were calculated using standard formulae to achieve a 95% confidence interval of plus or minus two percentage points on a 5% estimate for GAM assuming a design effect of 1.5. The overall sample size required for the SMART survey was $n = 684$. This was increased to $n = 690$ to be collected as $m = 30$ clusters of $n = 23$ respondents. The overall sample size required for the RAM-OP survey was $n = 192$ to be collected as $m = 12$ clusters of $n = 16$ respondents. The difference in the two required sample sizes is due to RAM-OP using a more efficient estimator for GAM than the estimator used by SMART (see *Box CP* for details of the different approaches and *Box SS* and *Figure SS* for details of the sample size calculations) and because of the improved efficiency of the sampling methods used by the RAM-OP survey method.

Both surveys used two-stage cluster sampling :

First-stage samples

In first-stage sampling for the SMART survey, thirty (30) primary sampling units (villages) were selected using population proportionate sampling (PPS) from a complete list of communities in the survey area [SM].

In first-stage sampling for the RAM-OP survey, twelve (12) primary sampling units (villages) were selected systematically from a complete list of communities in the survey area sorted by location (i.e. urban vs. rural) and electoral ward. This procedure, known as *implicit stratification*, should select a sample that is reasonably evenly distributed across the survey area [IA]. This procedure also tends to spread the sample properly among important sub-groups of the population such as rural / urban / peri-urban, administrative areas, ethnic / religious sub-populations, and socio-economic groups and often improves the precision of estimates made from survey data [IA,IB].

Population data from the 2012 Tanzania census was used for the first-stage PPS sample (SMART) and for weighting during data analysis (RAM-OP). Sampling locations were identified using a map of Kibaha District with the assistance of GSSST staff and local government officials.

Second-stage samples

The second-stage sample (i.e. for selection of the respondents in each sampled primary sampling unit (PSU) for the SMART survey was taken using the *EPI* sampling strategy [SM,SA].

The second-stage sample for the RAM-OP survey was taken using systematic sampling of dwellings in the villages (or parts of villages) organised as ribbons of dwellings and a random walk (*EPI3*) sampling strategy in villages (or parts of the villages) organised as clusters of dwellings. The *EPI3* method selects the first household to be sampled using the *EPI* strategy (as with SMART) with subsequent households selected by choosing a random direction and selecting the third nearest house in that direction [SA]. This method has been shown to give results as good as simple random samples and to be better than the unmodified *EPI* strategy when a wide range of indicators is being assessed [SA].

The final sample size for the RAM-OP survey was $n = 196$. The final sample size for the SMART survey was $n = 702$.

The sampling methods used for the SMART survey were approved by the SMART Initiative through the medium of anonymous posting of questions on EN-NET which were answered by members of the SMART team who provided direct answers to questions as well as electronic documents related to sampling.

Survey implementation

Permissions to carry out the surveys were obtained from all relevant national and local bodies.

Ethical clearance, including a data transfer agreement, was sought and obtained from the Ethical Review Committee of the Tanzania National Institute for Medical Research.

Permissions were also sought and obtained from the Regional Commissioner for Pwani (Coastal) Region, the District Administrative Secretary for Kibaha District, and the District Medical Officer for Kibaha District. Each ward committee was informed about the survey by a letter from the District Medical Officer and a preliminary visit or telephone call from GSSST. Sampling schedules were shared with Ward Executive Officers who provided much useful information about ward boundaries to survey teams. Village Executive Officers were informed by telephone prior to each team's visit and assisted survey teams during data collection (i.e. introducing teams to the village chairpersons and guiding survey teams with regard to the location of village boundaries).

Seventeen enumerators were trained for one week in Mlandizi town. The training was done in both Kiswahili and in English. Training covered sampling and other field procedures, practising the questionnaire (in Kiswahili), measuring mid upper arm circumference (MUAC), and performing a visual acuity test using the 'tumbling E' method. The last day of training was a field test in which survey teams were sent to the field under survey conditions. At the end of the week, fourteen enumerators were recruited (four teams of two for SMART and three teams of two for RAM-OP). Two data entry clerks were also recruited and trained.

Data collection for both surveys was completed in six days (64 person-days for SMART and 24 person-days for RAM-OP). Data entry and checking was completed in seven days (18 person-days for SMART and 4 person-days for RAM-OP).

No serious logistic problems were encountered. Local authorities were cooperative and the survey teams were welcomed by all respondents. A total of thirty villages were visited (24 for SMART, 12 for RAM-OP, 6 for both SMART and RAM-OP). All of the eleven wards in Kibaha District were covered by both surveys. A total of n = 702 (target n = 690) older people were interviewed for the SMART survey. A total of n = 196 (target n = 192) older people were interviewed for the RAM-OP survey.

Data Entry and Checking

The data for both surveys were entered into identically structured *EpiData v3.10* databases by dedicated data-entry staff. Interactive checks for range and legal values were applied. All data were double-entered and validated (verified) using a record-by-record and variable-by-variable comparison. Errors and discrepancies were resolved by consulting data-collection forms. Any errors that could not be resolved resulted in obviously or potentially erroneous values being censored (i.e. set to missing).

Data related to survey costs were collected and entered into an *OpenOffice Calc* spreadsheet.

Data Management

Data management consisted of creating indicators from the collected survey data. This was done for both the SMART and RAM-OP data using the same purpose-written scripts. Scripts were written using the *R Language for Data Analysis and Graphics* and managed using the *R-AnalyticFlow* scientific workflow system. This approach allowed for modular development of data management and data analysis code, and provided tools for the documentation, testing, and debugging of scripts. *Box ADL*, for example, shows the contents of a workflow node that was used to calculate an activities of daily living (ADL) score [AD].

Estimating indicator levels

Estimating indicator levels from the survey data consisted of estimating proportions and means for a variety of indicators (*Table 1*). The indicator sets used in the two survey rounds differed slightly from each other. This is because a key activity of the RAM-OP development project is to identify and test indicators and questionnaire components suitable for use in small sample needs-assessment surveys of older people.

The SMART survey data were analysed using the *Taylor Linearised Deviation* approach, as implemented in the *CSAMPLE* module of *EpiInfo*, to calculate confidence intervals for proportions and means [TL].

The RAM-OP survey data were analysed using a *blocked weighted bootstrap* estimator [BW]:

> **Blocked :** The block corresponds to the PSU or cluster.
>
> **Weighted :** The RAM-OP sampling procedure does not use population proportional sampling to weight the sample prior to data collection as is done with SMART type surveys. This means that a posterior weighting procedure is required. We used a "roulette wheel" algorithm (see *Figure BBW*) to weight (i.e. by population) the selection probability of PSUs in bootstrap replicates.

A total of *m* PSUs are sampled *with-replacement* from the survey dataset where *m* is the number of PSUs in the survey sample. Individual records within each PSU are then sampled *with-replacement*. A total of *n* records are sampled *with-replacement* from each of the selected PSUs where *n* is the number of individual records in a selected PSU. The resulting collection of records replicates the original survey in terms of both sample design and sample size but samples PSUs proportional to population sizes (PPS). A large number of replicate surveys are taken. The work reported here used $r = 400$ replicate surveys because this number of replicates was found to balance stability of estimates with speed of calculation. The required statistic (e.g. the mean of an indicator value) is applied to each replicate survey. The reported estimate consists of the 50th (point estimate), 2.5th (lower 95% confidence limit), and the 97.5th (upper 95% confidence limit) percentiles of the distribution of the statistic across all replicate surveys.

The blocked weighted bootstrap procedure is outlined in *Figure BBW*. The bootstrap approach is computer-intensive but allows estimation of the sampling distribution of almost any statistic using only simple computational methods.

Data related to survey costs were entered and analysed using an *OpenOffice Calc* spreadsheet.

Comparing the SMART and RAM-OP survey methods - Results

The estimates of proportions from SMART and RAM-OP surveys in each pilot were examined using simple scatter-plots (*Figure 1A* and *Figure 1C*). The two survey methods produced broadly similar results for all indicators in both survey rounds.

Differences in the precision of the estimates of proportions were examined using histograms plotting the differences between the half-widths of the 95% confidence intervals of the estimates returned by the SMART and RAM-OP surveys (*Figure 1B* and *Figure 1D*). In this analysis, negative values indicate better precision was achieved by the SMART survey, zero values indicate no difference in precision between the SMART and RAM-OP surveys, and positive values indicate better precision was achieved by the RAM-OP survey. The precision achieved by the SMART and RAM-OP surveys in the Ethiopian round of surveys were broadly similar. The precision achieved by the RAM-OP survey in the Tanzanian round of surveys was generally worse than the precision achieved by the SMART survey.

The prevalence of undernutrition is a key indicator for the SMART and RAM-OP methods. Prevalence of undernutrition was estimated using the classic method (see *Box CP*) with the SMART survey data. Prevalence of undernutrition was estimated using the PROBIT method (see *Box CP*) with the RAM-OP survey data. The precision of the classical estimator (SMART) and the PROBIT estimator (RAM-OP) for GAM, MAM, and SAM prevalence is summarised in *Table 2*. Both methods returned prevalence estimates with useful precision (i.e. *relative precision* was better than that specified in the sample size calculation for the SMART surveys for the majority of indicators in all SMART and RAM-OP surveys). The classical estimator outperformed the PROBIT estimator for GAM and MAM prevalence. The PROBIT estimator outperformed the classical estimator for SAM prevalence. The performance of the PROBIT estimator improved considerably between the two survey rounds. This is probably due to the use of a longer and broader MUAC strap being used in the Tanzanian round of surveys [R1].

The existence of systematic differences (bias) between estimates of indicator levels between the two survey methods was investigated by plotting the difference between the SMART and RAM-OP estimates against the mean of the SMART and RAM-OP estimates (*Figure 2A* and *Figure 2B*). The solid line on these plots marks the mean difference [BA]. The dotted lines on these plots mark the 95% *limits of agreement* [BA]. No systematic differences between the SMART and RAM-OP estimates were observed in either round of surveys. Non-systematic (random) differences were larger in the Tanzanian round of surveys than in the Ethiopian round of surveys.

The results (i.e. means and 95% confidence intervals) returned by the two survey methods in the two survey rounds for *quantitative* variables are shown in *Table 3*. Both methods on both rounds returned similar results with useful precision.

Comparisons between the results of the SMART and RAM-OP surveys were performed using a *two-sample z-test*. 95% confidence intervals between indicator levels estimated with the SMART and RAM-OP surveys were also calculated (*Box ZCI)*. Statistically significant differences in estimates of indicator levels were found between the SMART and RAM-OP surveys for two indicators in the Ethiopian round of surveys. These were “household has an improved sanitation facility” and “respondent reports a normal or good appetite for food”. Supervision of survey teams in the field revealed initial confusion with regard to the definition of “improved sanitation facility” which was corrected on the third day of data collection. The appetite indicator was collected in a simple (i.e. compared to methods used in clinical assessment of older people) and non-standard way. No statistically significant differences in estimates of indicator levels were found between the SMART and RAM-OP surveys in the Tanzanian round of surveys.

Comparing the SMART and RAM-OP survey methods - Costs

Table 4 shows the costs for the SMART and RAM-OP surveys in the Tanzanian round of surveys. The RAM-OP survey was considerably cheaper than the SMART survey. The costs of the SMART and RAM-OP surveys in the Ethiopian round of surveys have been reported previously (the RAM-OP survey costs were about two-thirds of those for the SMART survey). It should be noted that some costs were associated with the research aspect of the work and this is likely to have resulted in the cost difference between the two methods being under-estimated.

Progress and next steps

Indicators

We have selected and tested a wide range of indicators useful for assessing the nutritional status, vulnerabilities, and needs of older people in emergency and development settings. The questionnaires used in the pilots surveys are too long for routine field use. We are now in the process of rationalising the core RAM-OP indicator set and producing a RAM-OP questionnaire suitable for routine field use.

Software

Software for entering, checking, analysing, and reporting data is under development. The initial release of software will work with data collected using the core RAM-OP indicator set and questionnaire.

The data entry and checking software will consist of template files (data entry screen layouts, data files, and scripts for interactive data checking) for use with *EpiData v3.10*. These files may be edited allowing the addition and removal of variables as needed. Double-entry validation (verification) is provided by *EpiData v3.10*. The data analysis software will be able to read data in a variety of formats (e.g. text files, EpiInfo, EpiData, STATA, SAS, SPSS, dBase, Minitab, Systat, &c.) and from SQL databases. This will allow users to enter and check data using their own software.

The data analysis system will be open source (i.e. general public license) and is designed to be easily customisable. It is based on the *R Language for Data Analysis and Graphics* and the *R-AnalyticFlow* scientific workflow system. We have tested this type of data analysis software in a variety of settings (i.e. Sierra Leone, Sudan, Ghana, Bangladesh, India, Ethiopia, and Sudan) with a variety of clients (i.e. ministries of health in Sierra Leone and Sudan; UNICEF in Sierra Leone and Sudan; international NGOs in Ghana, Ethiopia, Bangladesh, and India; and local NGOs in Ethiopia, India, and Bangladesh) for RAM (child health, program M&E, nutritional surveillance), RAM-OP, and S3M (mapping) applications. The data analysis software will produce a standard report containing tables, figures, and boilerplate text (e.g. for methods, guidelines for interpreting findings, technical appendices, and references) that can be easily imported for editing and formatting into standard word-processing software. We are currently designing the report format

Documentation

We have started work on a RAM-OP guidebook which will cover all aspects of RAM-OP surveys including sample design, sampling, data collection and measurement, data-entry, data-analysis and reporting. We expect to have a working draft available in early 2015. We will organise a workshop in London mid 2015 to present the RAM-OP method and tools to the humanitarian community and to promote its use. Similar events will take place in Ethiopia and Tanzania.

Plans for future activities

HelpAge International and Valid International Ltd. are members of a consortium led by Plan International which aims to test the RAM method in different age groups (i.e. children aged less than five years and older people) and in various humanitarian contexts (e.g. displacement, acute emergencies, slow onset emergencies). A funding proposal has been submitted to the R2HC (Research for Health in Humanitarian Contexts) program. If this application is successful, RAM-OP will be tested using a sampling design of $m = 16$ clusters of $n = 12$ respondents. Testing of RAM for children aged under five years will follow the process used to test RAM-OP (i.e. at least two rounds of surveys with comparison against the SMART method). HelpAge International is planning to implement further RAM-OP surveys in 2015 and to build their internal capacity to implement RAM-OP surveys when the guidebook becomes available.

Conclusions

In the field trials reported here the RAM-OP survey provided comparable results to the SMART survey at about two-thirds of the cost of the SMART survey. Larger (i.e. up to 50%) cost savings may be possible.

The RAM-OP survey in the Ethiopian round of surveys used a sample size of n = 320 taken as twenty clusters each of sixteen persons (average). The RAM-OP survey in the Tanzanian round of surveys used a sample size of n = 196 taken as twelve clusters each of sixteen persons (average). The observed reduction in precision between the two rounds indicates that a larger sample is required if RAM-OP surveys are to provide similar precision to SMART surveys. It may be sufficient to increase the size of the first-stage sample whilst decreasing the size of the second-stage sample and taking (e.g) a sample consisting of sixteen clusters of twelve people. This will be tested in a further round of surveys.

Box ADL : Example of indicator creation script (activities of daily living (ADL) score)

```
###############################################################################
#                                                                             #
#  Katz ADL score                                                             #
#                                                                             #
#    The Katz ADL score is described in :                                     #
#                                                                             #
#      Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW (1963). Studies    #
#      of illness in the aged. The Index of ADL: a standardized measure of    #
#      biological and psychosocial function. JAMA, 1963, 185(12):914-9        #
#                                                                             #
#      Katz S, Down TD, Cash HR, Grotz, RC (1970). Progress in the development #
#      of the index of ADL. The Gerontologist, 10(1), 20-30                   #
#                                                                             #
#      Katz S (1983). Assessing self-maintenance: Activities of daily living, #
#      mobility and instrumental activities of daily living. JAGS, 31(12),    #
#      721-726                                                                #
#                                                                             #
###############################################################################

###############################################################################
#
#  Recode ADL (activities of daily living) score data
#
#    ADL is scored :
#
#      1 = Independence
#      0 = Needs assistance or supervision
#
#    summed for six activities / dimensions to create the ADL score.
#
ADL01 <- recode(svy$a4, "1=0; 2=1; 9=0; NA=0")     # Bathing
ADL02 <- recode(svy$a5, "1=0; 2=1; 9=0; NA=0")     # Dressing
ADL03 <- recode(svy$a6, "1=0; 2=1; 9=0; NA=0")     # Toileting
ADL04 <- recode(svy$a7, "1=0; 2=1; 9=0; NA=0")     # Transferring (mobility)
ADL05 <- recode(svy$a8, "1=0; 2=1; 9=0; NA=0")     # Continence
ADL06 <- recode(svy$a9, "1=0; 2=1; 9=0; NA=0")     # Feeding

###############################################################################
#
#  Create ADL score (sum of individual dimension scores)
#
scoreADL <- ADL01 + ADL02 + ADL03 + ADL04 + ADL05 + ADL06

###############################################################################
#
#  Severity of dependence (from Katz ADL score)
#
#    1 = Independent
#    2 = Partial dependency
#    3 = Severe dependency
#
severityADL <- recode(scoreADL, "0:2=3; 3:4=2; 5:6=1")
```

Table 1 : Indicators calculated reported at each round of SMART and RAM surveys

Category	Indicator	Round: Ethiopia	Round: Tanzania
Demography	Age in years	●	●
	Lives alone	●	●
	Size of household	●	●
Diet	Meal frequency	●	●
	Dietary diversity	●	●
	Protein rich food consumed	○	●
	Plant sources of vitamin	●	●
	Animal sources of vitamin A	●	●
	Any sources of vitamin A	●	●
	Iron rich foods consumed	●	●
	Calcium rich food	●	●
	Zinc rich foods	○	●
	Vitamin B1 rich foods	○	●
	Vitamin B2 rich foods	○	●
	Vitamin B3 rich foods	○	●
	Vitamin B6 rich foods	○	●
	Vitamin B12 rich foods	○	●
	Vitamin B-complex rich foods	○	●
Anthropometry	GAM prevalence	●	●
	MAM prevalence	●	●
	SAM prevalence	●	●
	MUAC	●	●
	Bilateral pitting oedema	●	●
	Previously screened	●	●
Disability	Activities of daily living (ADL)	●	●
	Bathing	●	●
	Dressing	●	●
	Toileting	●	●
	Transferring (mobility)	●	●
	Urinary and faecal continence	●	●
	Feeding	●	●
	Difficulty hearing	●	●
	Low vision / blind	○	●
	Housebound / bed-ridden	●	●
Chronic Disease	Chronic disease	●	●
	Diabetes	●	○
	Hypertension	●	○
	High cholesterol	●	○
	Pain / arthritis	●	○
	TB	●	○
	Respiratory (non-TB)	●	○
	HIV / AIDS	●	○
	Takes drugs	●	●
	Acute disease	●	●
	Arthritis	○	●
	Stomach pain	○	●
	Fever	○	●
	Diarrhoea	○	●
	Fall	○	●
	Other	○	●
	Accessed care / drugs	●	○

Category	Indicator	Round: Ethiopia	Round: Tanzania
Mental health	Summary score (K6)	●	●
	Psychological distress	●	●
Income	Has a source of income	●	●
	Agriculture	●	
	Livestock	●	●
	Fishing	●	
	Unskilled labour	●	
	Skilled labour	●	●
	Salaried work	●	
	Sales (charcoal / bricks)	●	●
	Sales (building material)	●	
	Sales (handicrafts)	●	
	Petty trading	●	●
	Commercial trading	●	
	Pension / allowance	●	●
	Sale (food aid)	●	
	Gift / remittance	●	●
	Alms	●	
	Begging	●	
	Savings	○	●
	Cash-transfer	○	●
	Other	●	●
AID / relief	Cash-transfer (as above)	○	●
	Ration (self)	●	●
	Ration (family)	●	●
	Ration (any)	●	●
	Non-food relief items	○	●
WASH	Improved drinking water	●	●
	Probable safe drinking water	●	●
	Improved sanitation facility	●	●
	As above and not shared	●	●
Miscellany	Has a carer	●	●
	Has childcare role	●	●
	Good appetite	●	●
	Can chew food	●	●
	Eats alone	●	●
	Shares food	●	●
	Active in the community	○	●
	Community role	○	●
	Feast day yesterday	○	●
	Fasting day yesterday	○	●

Indicators were sometimes changed between rounds because a key activity of the RAM-OP development project was to identify and test indicators and questionnaire component suitable for use is small sample needs-assessment surveys of older people.

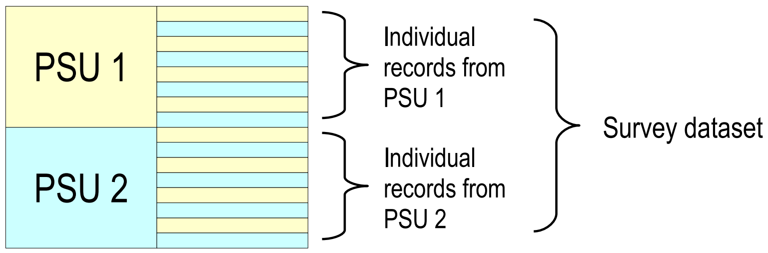
The symbols used in this table are:

- ● Indicator present
- ○ Indicator not present

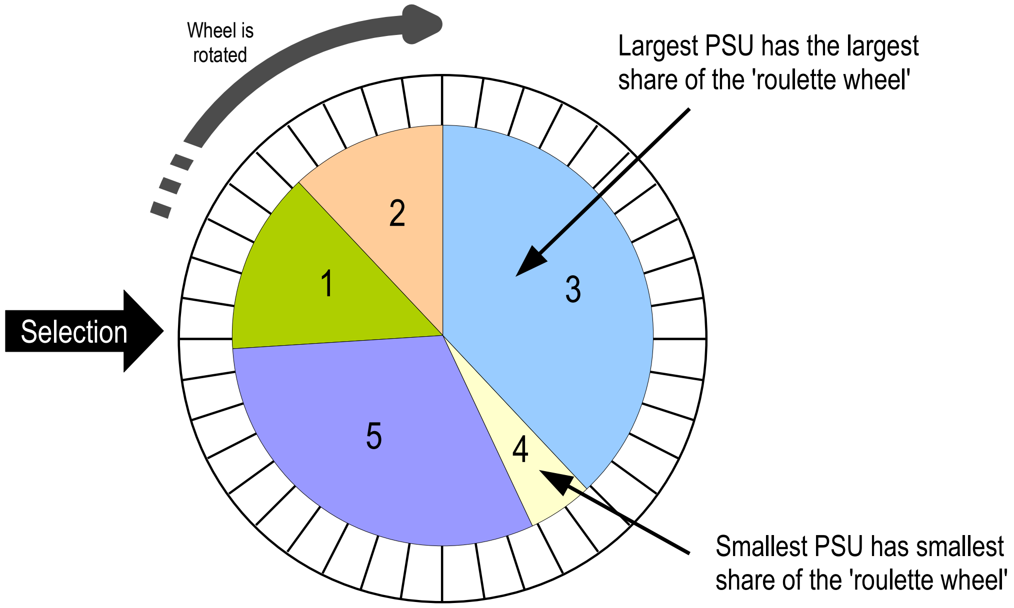
The Ethiopia round of surveys sampled adults aged 50 years and older. The Tanzania round of surveys sampled adults aged 60 years and older. This reflected the intervention groups targeted by HelpAge International's local partners.

Figure BBW : The blocked weighted bootstrap used to make estimates from RAM-OP data

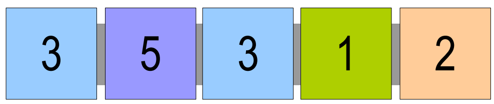
Data are collected using a two-stage cluster design:



PSUs are selected from the survey dataset *with-replacement* and with probability proportional to population size using a *roulette wheel* algorithm :



PSUs are selected *with-replacement* and proportional to population size :



Individual records are selected *with-replacement* from within each PSU to create a *replicate* survey. The estimator is applied to the replicate survey and the result recorded. This process is repeated many times. The estimate of the indicator value is made from the distribution of the results from each replicate survey which is the *empirical sampling distribution* of the indicator :

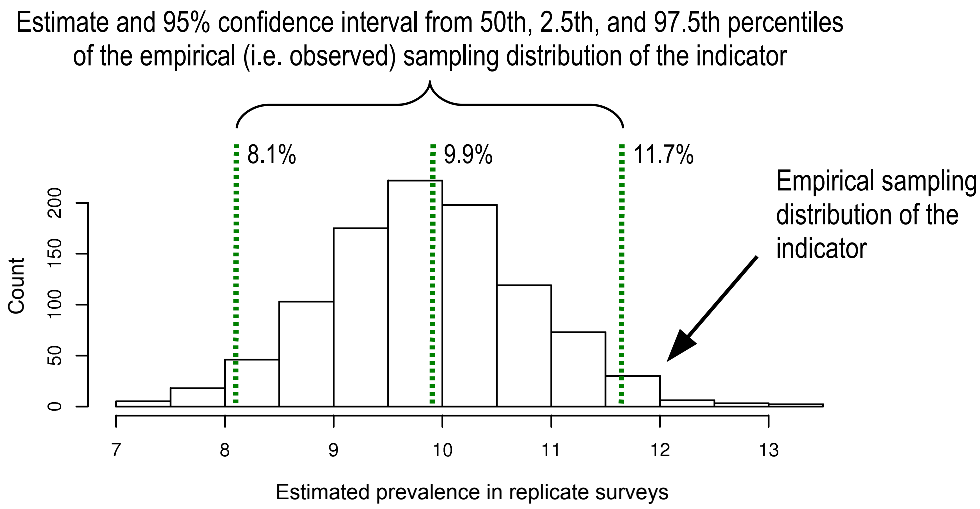


Figure 1 : Comparison of estimates and precisions for all indicators returning proportion from SMART and RAM-OP surveys in the two survey rounds

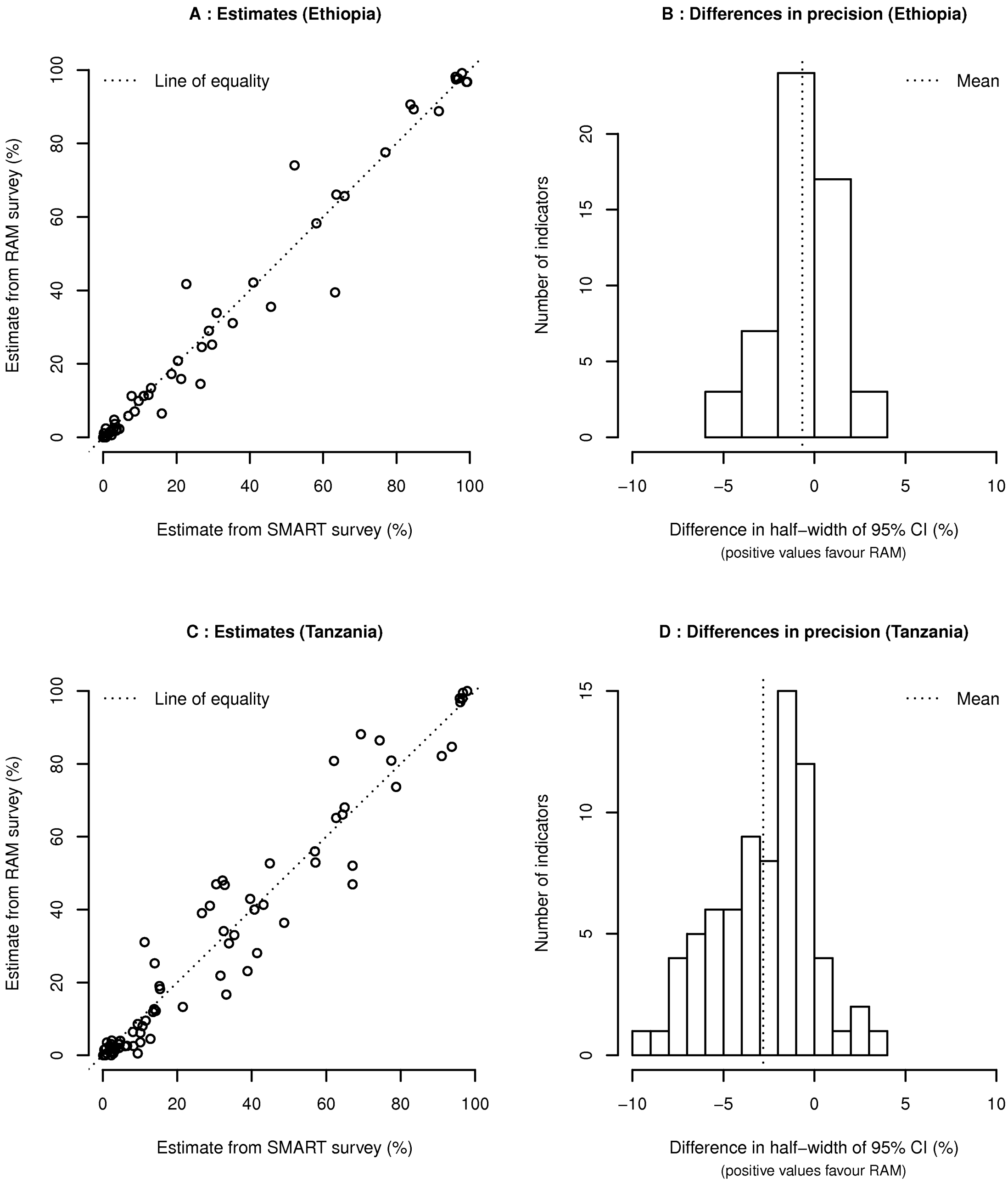
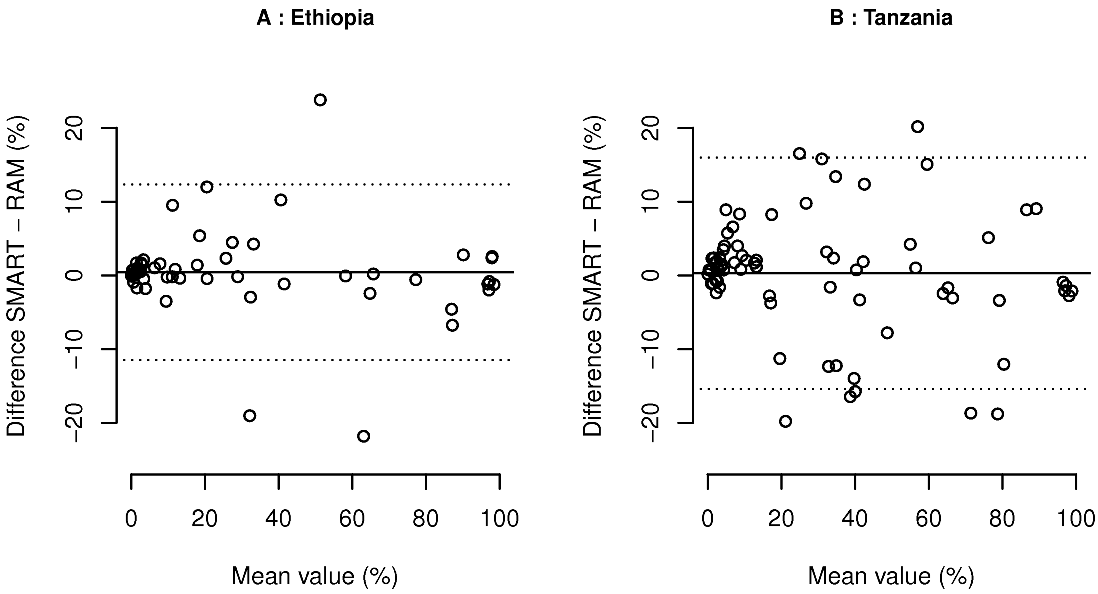


Table 2 : Precision of the classical estimator (SMART) and the PROBIT estimator (RAM-OP) for GAM, MAM, and SAM prevalence

	Estimated Prevalence ± Half-width of 95% CI[1]			
	Ethiopia		**Tanzania**	
Indicator	**SMART**	**RAM**	**SMART**	**RAM**
GAM prevalence	4.39% ± 1.65%	2.28% ± 2.49%	3.28% ± 1.47%	2.29% ± 1.83%
MAM prevalence	3.69% ± 1.60%	2.26% ± 2.50%	3.13% ± 1.38%	2.26% ± 1.85%
SAM prevalence	0.70% ± 0.65%	0.05% ± 0.10%	0.14% ± 0.28%	0.01% ± 0.07%

[1] Symmetrical errors are presented here to allow comparison between the two method. The classical estimator in SMART uses an *approximate* method and coverage of the 95% CI is *nominal*. The PROBIT estimator in RAM-OP uses an *exact* method and coverage of the 95% CI will be close to 95%.

Figure 2 : Systematic and random variation of estimates for all indicators returning proportions from SMART and RAM surveys in the two survey rounds



The solid line on these plots marks the mean difference. The dotted lines on these plots mark the 95% limits of agreement [BA].

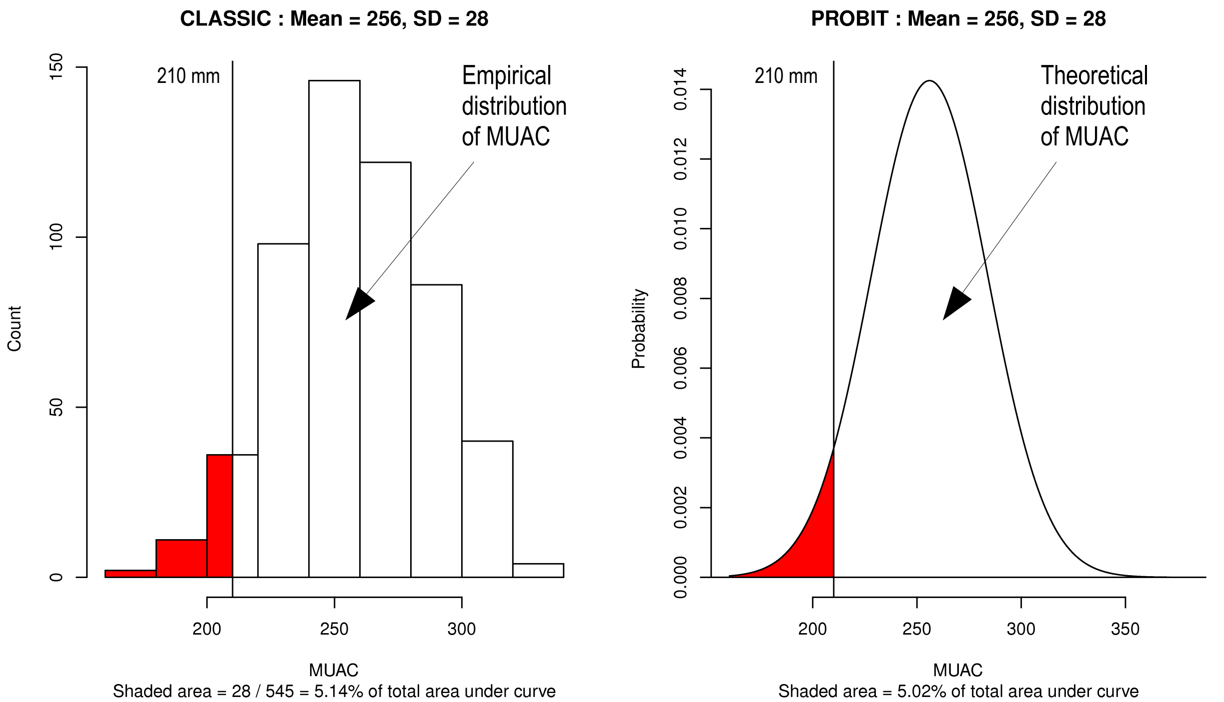
Box CP : Classic and PROBIT estimators of GAM prevalence

The estimates of GAM, MAM, and SAM made from the SMART survey data used the classic estimator. For (e.g.) GAM:

$$prevalence = \frac{number\ of\ respondents\ with\ MUAC < 210\,mm}{total\ number\ of\ respondents}$$

The estimates of GAM, MAM, and SAM made from the RAM-OP survey data used a PROBIT estimator. This is a *model-based* approach. The PROBIT function is also known as the *inverse cumulative distribution* function. This function converts parameters of the distribution of an indicator (e.g. the mean and standard deviation of a *normally distributed* variable) into cumulative percentiles. This means that it is possible to use the normal PROBIT function with estimates of the mean and standard deviation of indicator values in a survey sample to estimate the proportion of the population falling below a given threshold. For example, for data with a mean MUAC of 256 mm and a standard deviation of 28 mm the output of the normal PROBIT function for a threshold of 210 mm is 0.0502 meaning that 5.02% of the population are estimated to fall below the 210 mm threshold.

Both the classic and the PROBIT methods can be thought of as estimating area:



The prevalence of GAM and SAM are estimated directly. The prevalence of MAM is estimated as the difference between the estimates of GAM and SAM.

The principal advantage of the PROBIT approach is that the required sample size is usually smaller than that required to estimate prevalence with a given precision using the classic method (see *Box SS*) [P1,P2].

The PROBIT method assumes that MUAC is a normally distributed variable. If this is not the case then the distribution of MUAC can be transformed towards normality. This was done with data from the RAM-OP surveys reported here.

Box SS : Sample size calculations for classic and PROBIT estimators of GAM prevalence

The sample size for the SMART survey was calculated to achieve a 95% confidence interval of plus or minus two percentage points on a 5% estimate for GAM assuming a design effect (DEFF) of 1.5 using the standard formula:

$$n = DEFF \times \frac{p\,(1 - p)}{(e \div 1.96)^2} = 1.5 \times \frac{0.05 \times (1 - 0.05)}{(0.02 \div 1.96)^2} = 684$$

The sample size for the RAM-OP surveys was calculated to yield a similar precision on a similar prevalence estimate using the following process:

> Previous experience surveying older people in Chad and Ethiopia indicated that we could reasonably expect MUAC to be approximately normally distributed with a mean of about 270 mm and a standard deviation (SD) of about 40 mm. These parameters correspond to a prevalence of about 6.7% using a MUAC < 210 mm case-definition for GAM. We used these parameters in the sample size calculation*.
>
> A mean MUAC of 264.3 mm with the same SD (i.e. 40 mm) corresponds to a prevalence of about 8.7% (i.e. a 2% difference from 6.7%) using a MUAC < 210 mm case-definition for GAM. This was found by 'trial and error' using a spreadsheet package as shown in *Figure SS*. The difference between these two means:
>
> $$e = 270 - 264.3 = 5.7$$
>
> is the half-width of the 95% confidence interval for the mean MUAC that corresponds to a half-width of the 95% confidence interval of the PROBIT estimate of GAM prevalence of about plus or minus two percentage points. The sample size required for this level of precision is about:
>
> $$n = \frac{SD^2}{(e \div 1.96)^2} = \frac{40^2}{(5.7 \div 1.96)^2} = 189$$
>
> This was rounded up to n = 192 because 192 has many whole number divisors and this simplifies spreading the sample amongst survey PSUs (clusters).
>
> No design effect was specified because any design effect was expected to be close to one due to the use of implicit stratification in the first-stage sample and segmentation and random walks in the second-stage sample [IA,IB,SA].

This RAM-OP sample size calculation was supported by work done prior to the RAM-OP project on testing the PROBIT estimator with anthropometric data from children showing the PROBIT estimator with a sample size of n = 192 performing at least as well as the classic estimator with a sample of n = 544 (i.e. the largest sample size specified in the SMART manual) [P1,SM].

A sample size of n = 192 also guarantees a precision of plus or minus ten percentage points or better on estimates of proportions at survey design effects of up to 2.0.

It is proposed that future RAM-OP surveys use a standard minimum sample size of n = 192.

* In the Tanzanian round of surveys mean = 270 mm and SD = 38 mm. In the Ethiopian round of surveys mean = 273 mm and SD = 40 mm.

Figure SS : Sample size calculator for a PROBIT estimator

PROBIT sample size by 'trial & error' calculator (Microsoft Excel formulas shown)

	A	B	C
1	Previous Surveys	Mean MUAC	270.0
2		SD	40.0
3		Prevalence (PROBIT)	=NORM.DIST(210,C1,C2,1)
4	Trial & error	Mean MUAC	**ENTER TEST MUAC IN THIS CELL**
5		Prevalence (PROBIT)	=NORM.DIST(210,C4,C2,1)
6		e	=ABS(C1-C4)
7		n	=C2^2/(C6/1.96)^2

Trial & error : Working MUAC down (prevalence up)

	A	B	C
1	Previous Surveys	Mean MUAC	270.0
2		SD	40.0
3		Prevalence (PROBIT)	6.7%
4	Trial & error	Mean MUAC	264.3
5		Prevalence (PROBIT)	8.7%
6		e	5.7
7		n	189

About
2% difference

Trial & error : Working MUAC up (prevalence down)

	A	B	C
1	Previous Surveys	Mean MUAC	270.0
2		SD	40.0
3		Prevalence (PROBIT)	6.7%
4	Trial & error	Mean MUAC	276.8
5		Prevalence (PROBIT)	4.7%
6		e	6.8
7		n	133

About
2% difference

For the work described here the largest sample size (i.e. $n = 189$ rounded to $n = 192$) was used. The mean half-width of the 95% confidence interval should, therefore, have been less than about two percentage points (in the Tanzania RAM-OP survey it was ± 1.83%).

Box ZCI : Methods used for between survey comparisons of estimates of indicator levels

Comparisons between the results from both surveys were performed using a *two-sample z-test*.

The *z-test* was calculated using the standard errors (SE) reported by the *CSAMPLE* module of *EpiInfo* for the SMART survey data and calculated from estimates of indicator levels for the RAM-OP survey data:

$$SE = \frac{UCL - LCL}{2 \times 1.96}$$

where *LCL* and *UCL* are the 95% upper and lower confidence limits of the bootstrap estimates made from the RAM-OP survey data. Standard errors were pooled:

$$SE_{Pooled} = \sqrt{SE^2_{\text{SMART}} + SE^2_{\text{RAM-OP}}}$$

and the *z-test* calculated as:

$$z = \frac{|p_{\text{SMART}} - p_{\text{RAM-OP}}|}{SE_{\text{Pooled}}}$$

Where p_{SMART} and $p_{RAM\text{-}OP}$ are the point estimates from the SMART and RAM-OP survey data.

95% confidence intervals on the differences between indicator levels the SMART and RAM-OP surveys were also calculated:

$$95\%\,CI = |p_{\text{SMART}} - p_{\text{RAM-OP}}| \pm 1.96 \times SE_{Pooled}$$

Sixty-one between survey comparisons were performed on the data from the Ethiopian pilot and eighty-two between survey comparisons were performed on the data from the Tanzanian pilot.

Without *correction for multiple comparisons* the probability of finding at least one significant test at the $p < 0.05$ significance level over these numbers of tests are approximately:

Ethiopian Pilot (61 comparisons)	**Tanzania pilot (82 comparisons)**
$p_{Positive\ test} = 1 - (1 - 0.05)^{61} = 0.9562$	$p_{Positive\ test} = 1 - (1 - 0.05)^{82} = 0.9851$

even when all of the differences are actually non-significant.

At the $p < 0.05$ significance level we would expect to find three "significant" results in sixty-one tests and four "significant" results in eighty-two tests even when all of the differences are actually non-significant.

To avoid these *false positives* results whilst maintaining the $p < 0.05$ significance level the critical value for p should be reduced from $p < 0.05$ to:

Ethiopian Pilot (61 comparisons)	**Tanzania pilot (82 comparisons)**
$p < 1 - (1 - 0.05)^{1/61} = 0.0008$	$p < 1 - (1 - 0.05)^{1/82} = 0.0006$

This procedure is known as the *Šidák Correction* [SC]. The thresholds for p given by the Šidák Correction are similar to those given by the better known *Bonferroni Correction* [BC]*:*

Ethiopian Pilot (61 comparisons)	**Tanzania pilot (82 comparisons)**
$p \leq \frac{0.05}{61} = 0.0008$	$p \leq \frac{0.05}{82} = 0.0006$

Table 3 : Results returned by the two survey methods in the two survey rounds for quantitative variables

		SMART	RAM	Differences (SMART - RAM)			
indicator	**Round**	**Estimate of mean (95% CI)**	**Estimate of mean (95% CI)**	**Difference between estimates (95% CI)**	**z***	**p**	**Precision**
Age (years)	Ethiopia	63.1 (62.2; 64.0)	63.5 (62.3; 64.5)	-0.4 (-1.2;1.0)	0.5671	0.5706	-0.15
	Tanzania	71.7 (70.8; 72.7)	70.6 (68.6; 73.0)	+1.1 (-1.3;3.5)	0.8959	0.3703	-1.28
Size of HH	Ethiopia	5.5 (5.2; 5.7)	5.3 (5.4; 5.9)	-0.2 (-0.2;0.2)	0.9609	0.3366	-0.04
	Tanzania	4.5 (4.2; 4.8)	4.4 (3.6; 5.1)	+0.1 (-0.7;0.9)	0.2123	0.8319	-0.40
Meal frequency	Ethiopia	2.7 (2.6; 2.8)	2.8 (2.7; 2.9)	-0.1 (-0.2;0.1)	0.8980	0.3692	-0.06
	Tanzania	2.6 (2.5; 2.8)	2.8 (2.5; 3.1)	-0.1 (-0.5;0.2)	0.9061	0.3649	-0.15
Dietary diversity	Ethiopia	6.0 (5.8; 6.2)	6.2 (5.98; 6.4)	-0.2 (-0.5;0.1)	1.4357	0.1510	-0.01
	Tanzania	4.6 (4.3; 4.8)	4.6 (4.2; 4.9)	-0.0 (-0.5;0.4)	0.1159	0.9077	-0.07
ADL score	Ethiopia	5.6 (5.5; 5.8)	5.7 (5.6; 5.3)	-0.1 (-0.3;0.1)	0.8884	0.3743	-0.03
	Tanzania	5.6 (5.6; 5.7)	5.6 (5.5; 5.8)	-0.0 (-0.2;0.1)	0.4033	0.6867	-0.08
K6 score	Ethiopia	6.2 (5.2; 7.2)	6.1 (5.6; 6.6)	+0.1 (-1.2;1.2)	0.1680	0.8666	+0.50
	Tanzania	10.9 (10.3; 11.7)	11.5 (10.3; 12.7)	-0.6 (-2.0;0.9)	0.7736	0.4392	-0.53
MUAC (mm)	Ethiopia	270.0 (265.1;270.6)	267.7 (264.4;270.6)	+2.3 (-3.1;8.1)	0.7778	0.4367	+1.77
	Tanzania	274.7 (270.1;297.2)	274.3 (266.2;282.1)	0.3 (-8.8;9.5)	0.0727	0.9420	-3.35

* Two sample *z-test* (see *Box ZCI* for details)

Table 4 : Costs for SMART and RAM-OP surveys in the Tanzanian round of surveys

				RAM-OP		SMART		
Activity	**Item**	**Unit**	**Unit Cost**	**Units**	**Cost**	**Units**	**Cost**	**Difference**
Training	person-days (survey staff)	person-day	$26.00	50	$1,300.00	50	$1,300.00	$0.00
	person-days (national supervisor)	person-day	$295.00	5	$1,475.00	5	$1,475.00	$0.00
	person-days (international manager)	person-day	$582.00	6	$3,492.00	6	$3,492.00	$0.00
	Venue hire	days	$23.00	4	$92.00	4	$92.00	$0.00
	Refreshments	person-day	$5.00	60	$300.00	60	$300.00	$0.00
	Stationery	lump sum	$100.00	1	$100.00	1	$100.00	$0.00
	Printing / copying	sheet	$0.30	180	$54.00	180	$54.00	$0.00
	Accommodation	person-day	$12.00	50	$600.00	50	$600.00	$0.00
Survey	Survey enumerators	person-day	$47.00	24	$1,128.00	64	$3,008.00	-$1,880.00
	Data-entry staff	person-day	$47.00	4	$188.00	18	$846.00	-$658.00
	National assistant supervisor	person-day	$58.00	3	v174.00	8	$464.00	-$290.00
	International survey manager	person-day	$582.00	4	$2,328.00	9	$5,238.00	-$2,910.00
	Translators & local guides	person-day	$47.00	3	$141.00	8	$376.00	-$235.00
	Vehicles with drivers & fuel (survey teams)	vehicle-day	$175.00	9	$1,575.00	24	$4,200.00	-$2,625.00
	Vehicles with drivers & fuel (supervision)	vehicle-day	$175.00	3	$525.00	8	$1400.00	-$875.00
	Photocopying of questionnaires	sheet	$0.06	2000	$120.00	7500	$450.00	-$330.00
	Maps	for area	$82.00	1	$82.00	1	$82.00	0.00
Other	Communications	lump sum	$200.00	1	$200.00	1	$200.00	0.00
	Flights	flight	$1,440.00	1	$1,440.00	1	$1,440.00	0.00
				Total	$15,314.00		$25,117.00	-$9,803.00

References

[SM] Golden M, Brennan M, Brennan R, Kaiser R, Mone C, Nathan R, Robinson WC, Woodruff B, Seaman J, Erhardt J (2006). *Measuring Mortality, Nutritional Status, and Food Security in Crisis Situations : SMART Methodology*. SMART Initiative, 41-60

[M1] UNICEF (2006). *Monitoring the situation of children and women. Multiple Indicator Cluster Survey Manual*, Division of Policy and Planning, United Nations Children's Fund, New York, USA

[D1] ICF International (2012). *Survey Organization Manual for Demographic and Health Surveys*. MEASURE DHS, ICF International, Calverton, Maryland , USA

[R1] Fritsch P, Siling K, Myatt M (2014). RAM-OP : *A rapid assessment method for assessing the nutritional status, vulnerabilities, and needs of older people in emergency and development settings. Experiences from a field trial in an urban setting*. Field Exchange, 49:67-71

[IA] Anon. (2005). *Designing Household Survey Samples : Practical Guidelines, Statistics Division (Studies in Methods, Series F, No.98)*, Department of Economic and Social Affairs of the United Nations Secretariat, New York, USA

[IB] Piazza T (2010). *Fundamentals of Applied sampling*, in Marsen PV, Wright JD (eds) (2010). *Handbook of Survey Research (2nd Edition)*, Emerald Group Publishing, Bingley, UK

[SA] Bennett S, Radalowicz A, Vella V, Tomkins A (1994). A *computer simulation of household sampling schemes for health surveys in developing countries*. International Journal of Epidemiology, 23(6):1282-1291

[AD] Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW (1963). *Studies of illness in the aged. The Index of ADL : A standardized measure of biological and psychosocial function*. JAMA, 1963, 185(12):914-9

[TL] Woodruff RS (1971). *A simple method for approximating the variance of a complex estimate,* Journal of the American Statistical Association, 66(334):411-414

[BW] Cameron AC, Gelbach JB, Miller DL (2008). *Bootstrap-based improvements for inference with clustered errors*, Review of Economics and Statistics, 90, 414-427

[BA] Bland JM, Altman DG (1986). *Statistical methods for assessing agreement between two methods of clinical measurement*. Lancet 1986;1:307-10

[P1] Dale NM, Myatt M, Claudine Prudhon C, Briend A (2012). *Assessment of the PROBIT approach for estimating the prevalence of global, moderate and severe acute malnutrition from population surveys*, Public Health Nutrition, July 2012:1-6 (doi:10.1017/S1368980012003345)

[P2] Blanton CJ, Bilukha OO (2013). *The PROBIT approach in estimating the prevalence of wasting : Revisiting bias and precision*, ETE, 10:8:1-8 (doi:10.1186/1742-7622-10-8)

[SC] Šidák, ZK (1967). *Rectangular Confidence Regions for the Means of Multivariate Normal Distributions*, Journal of the American Statistical Association 62(318):626-633

[BC] Bonferroni CE (1935). *Il calcolo delle assicurazioni su gruppi di teste* (in *Studi in Onore del Professore Salvatore Ortu Carboni*). Rome, Italy, 13-60