

1. TRAINING SUPERVISORS AND ENUMERATORS

1.1 Pretesting, March 2011

A pretest of twenty households for HBS 2011/12 was undertaken for one week from 17–23 March 2011. The pretest took place in Kibaha District.

- 10 interviews in Kibaha Maili moja centre - an urban EA
- 10 interviews in Pangani village - a rural EA.

A total of **20 households** were interviewed by ten interviewers and ten supervisors (supervisors were all members of the TWG). The resulting recommendations for changes to the questionnaire were few. The pretest gave an insight into areas where, even experienced, interviewers found challenges with definitions and coding responses. The pretest also identified more mundane, but equally important, issues relating to skips and design. The results fed quickly into the re-draft of the pilot questionnaire. Emily Poskett (DFID) and Thomas Danielwitz (World Bank) observed a pretest interview.

To test the questionnaires the team recorded a number of issues:

- Skip errors
- Questions the respondent could not understand or answer simply
- Questions the interviewer had problems reading or accurately recording the answer
- Other issues

To enable a systematic recording of issues arising in the field, the form shown in Annex 2 (developed and tested by the US Census Dept) was used to record the issues. The results of the pretest were outlined in a report consisting of two sections:

1. Issues that **need input from TC** to make amendments for the pilot version of the questionnaire
2. Issues where changes will be made by the HBS secretariat as it is clear what needs to be changed.

It was not the aim of the pretest to enter the data and examine the output. However prior to the pre-test work on the programming of the HBS data entry program (in CSPro) began to ensure the program was at a good stage of development by the time the pilot was completed.

1.2 Piloting, May 2011

The pilot training took place in Morogoro from May 2-15 2011. It was attended by:

- 50 Enumerators
- 25 Supervisors
- 21 Regional Survey Managers (RSMs)
- 11 Technical Working Group members

One main aim of the pilot was to have **all RSMs in the same location at the same time so they would all receive the same training and have the same understanding of the core tasks and concepts**. The large number of enumerators used for the main survey (400) meant that the Training of Enumerators (TOE) took place in six zones throughout Tanzania mainland and the RSMs would not have an opportunity to meet in one location.





Members of the TWG were the trainers and the workload was divided between the team depending on the subject specialty of each member.

The training consisted of:

1. Classes
2. Mock interviews
3. Real exercises in the field interviewing households in rural and urban Morogoro.



1.3 Training of Enumerators, September 2011

The training of enumerators began on 2nd September 2011. Each training session lasted for 14 days and finished on 16th September. Thereafter, each region did the Listing Exercise of households (from September 18th for five days) before starting fieldwork.

Seven training centers were used:

1. Dar es Salaam
2. Kibaha
3. Morogoro
4. Mbeya
5. Dodoma
6. Musoma
7. Moshi

In total 515 participants were trained (400 enumerators, 100 supervisors and 15 editors). Each center had four trainers and a Manager. Staff from DfID and UNICEF observed some training sessions and provided their feedback to NBS.

2. SAMPLING AND WEIGHTING FOR HBS 2011-12¹

2.1 Sampling Frame for 2011-12 HBS

The sampling frame for the 2011-12 HBS is based on the data and cartography from the 2002 Tanzania Population and Housing Census. A stratified multi-stage sample design was used for this survey. The primary sampling units (PSUs) selected at the first stage are the enumeration areas (EAs), which are small operational areas defined on maps for the 2002 Census enumeration. The EAs had an average of 133 households each (155 for rural EAs and 94 for urban EAs), which was an effective size for conducting a new listing of households. There was a total of 52,375 EAs in the 2002 Tanzania Census frame (33,947 rural EAs and 18,428 urban EAs).

Table 1. Distribution of Population in 2002 Tanzania Census Frame by Region, Rural and Urban Stratum

Region	Total		Rural		Urban	
	Population	% National	Population	% Region	Population	% Region
Dodoma	1,684,561	5.0%	1,472,571	87.4%	211,990	12.6%
Arusha	1,253,082	3.7%	850,632	67.9%	402,450	32.1%
Kilimanjaro	1,347,098	4.0%	1,064,778	79.0%	282,320	21.0%
Tanga	1,623,252	4.8%	1,324,969	81.6%	298,283	18.4%
Morogoro	1,709,273	5.1%	1,237,420	72.4%	471,853	27.6%
Pwani	867,831	2.6%	677,139	78.0%	190,692	22.0%
Dar es Salaam	2,460,824	7.3%	150,607	6.1%	2,310,217	93.9%
Lindi	779,451	2.3%	654,313	83.9%	125,138	16.1%
Mtwara	1,124,663	3.3%	898,298	79.9%	226,365	20.1%
Ruvuma	1,095,468	3.3%	925,236	84.5%	170,232	15.5%
Iringa	424,374	1.3%	399,918	94.2%	24,456	5.8%
Mbeya	2,053,205	6.1%	1,634,081	79.6%	419,124	20.4%
Singida	1,079,691	3.2%	932,900	86.4%	146,791	13.6%
Tabora	1,701,617	5.1%	1,486,126	87.3%	215,491	12.7%
Rukwa	722,768	2.2%	587,824	81.3%	134,944	18.7%
Kigoma	1,296,588	3.9%	1,094,881	84.4%	201,707	15.6%
Shinyanga	1,538,060	4.6%	1,359,660	88.4%	178,400	11.6%
Kagera	1,605,400	4.8%	1,514,148	94.3%	91,252	5.7%
Mwanza	1,919,584	5.7%	1,418,625	73.9%	500,959	26.1%
Mara	1,356,202	4.0%	1,103,847	81.4%	252,355	18.6%
Manyara	1,005,102	3.0%	866,186	86.2%	138,916	13.8%
Njombe	639,114	1.9%	540,464	84.6%	98,650	15.4%
Katavi	999,547	3.0%	878,974	87.9%	120,573	12.1%
Simiyu	2,173,672	6.5%	2,050,905	94.4%	122,767	5.6%
Geita	1,131,524	3.4%	1,017,426	89.9%	114,098	10.1%
Total	33,591,951	100.0%	26,141,928	77.8%	7,450,023	22.2%

¹ Sincere thanks to Mr David Megill, International consultant and Mrs Sylvia Meku, NBS for their work on creating the sample and weights for HBS 2011-12

Tanzania mainland was divided administratively into 25 regions², identified in Table 1. Each region was divided into districts, which were further divided into wards. For the 2002 Census the wards were classified by type of residence as urban, rural or mixed, and all the EAs within a ward were assigned the same classification. The EAs in mixed wards were later individually assigned to the rural and urban strata using the EA coding scheme. The EAs with codes of 300 or higher in mixed wards were assigned to the urban stratum, since they are part of small towns. Table 1 shows the distribution of the population by region, rural and urban strata, based on the 2002 Tanzania Census.

It can be seen in Table 1 that the largest region was Dar es Salaam, with 7.3 percent of the population, and the smallest region was Iringa, with 1.3 percent of the population. In reference to type of residence, at the national level 77.8 percent of the population was classified as rural and 22.2 percent as urban.

Table 2. Distribution of EAs and Households in 2002 Tanzania Census Frame by Region, Rural and Urban Strata

Region	Total		Rural		Urban	
	No. EAs	No. Hhs.	No. EAs	No. Hhs.	No. EAs	No. Hhs.
Dodoma	2,217	381,140	1,732	330,711	485	50,429
Arusha	2,147	284,964	1,254	177,940	893	107,024
Kilimanjaro	2,313	298,262	1,632	227,471	681	70,791
Tanga	2,284	360,498	1,599	292,583	685	67,915
Morogoro	2,956	385,148	1,750	270,609	1,206	114,539
Pwani	1,389	201,281	919	155,284	470	45,997
Dar es Salaam	6,721	603,393	181	37,688	6,540	565,705
Lindi	1,338	191,449	1,010	158,175	328	33,274
Mtwara	2,073	297,757	1,408	237,918	665	59,839
Ruvuma	1,470	233,129	1,080	192,957	390	40,172
Iringa	670	98,786	611	92,718	59	6,068
Mbeya	3,046	496,926	2,103	390,286	943	106,640
Singida	1,500	219,217	1,199	185,364	301	33,853
Tabora	2,200	293,663	1,728	244,715	472	48,948
Rukwa	1,078	149,952	799	120,249	279	29,703
Kigoma	1,761	238,783	1,273	201,134	488	37,649
Shinyanga	2,122	264,101	1,693	222,012	429	42,089
Kagera	2,038	350,093	1,873	326,937	165	23,156
Mwanza	2,594	337,775	1,595	226,451	999	111,324
Mara	2,026	248,570	1,398	195,088	628	53,482
Manyara	1,529	196,447	1,215	162,473	314	33,974
Njombe	1,044	153,553	800	128,341	244	25,212
Katavi	1,500	170,066	1,244	143,978	256	26,088
Simiyu	2,940	326,855	2,660	301,137	280	25,718
Geita	1,419	185,158	1,191	161,865	228	23,293
Total	52,375	6,966,966	33,947	5,184,084	18,428	1,782,882

Table 2 shows the distribution of the total number of EAs and households in the 2002 Tanzania Census frame by region and stratum.

² Njombe is now part of Iringa. Katavi is now part of Rukwa. Simiyu is now part of Shinyanga and Mwanza. Geita is now part of Shinyanga, Mwanza and Kagera.

Table 3 presents the average number of households per EA and the average number of persons per household in the 2002 Tanzania Census frame, by region, rural and urban stratum. It can be seen that the average number of households was higher for the rural EAs (153) than for the urban EAs (97). The average number of persons per household was 4.8 at the national level, and was considerably higher for the rural areas (5.0) than the urban areas (4.2).

Table 3. Average Number of Households per EA and Average Number of Persons per Household in 2002 Tanzania Census Frame by Region, Rural and Urban Stratum

Region	Total		Rural		Urban	
	Hhs./EA	Persons/hh.	Hhs./EA	Persons/hh.	Hhs./EA	Persons/hh.
Dodoma	172	4.4	191	4.5	104	4.2
Arusha	133	4.4	142	4.8	120	3.8
Kilimanjaro	129	4.5	139	4.7	104	4.0
Tanga	158	4.5	183	4.5	99	4.4
Morogoro	130	4.4	155	4.6	95	4.1
Pwani	145	4.3	169	4.4	98	4.1
Dar es Salaam	90	4.1	208	4.0	86	4.1
Lindi	143	4.1	157	4.1	101	3.8
Mtwara	144	3.8	169	3.8	90	3.8
Ruvuma	159	4.7	179	4.8	103	4.2
Iringa	147	4.3	152	4.3	103	4.0
Mbeya	163	4.1	186	4.2	113	3.9
Singida	146	4.9	155	5.0	112	4.3
Tabora	133	5.8	142	6.1	104	4.4
Rukwa	139	4.8	150	4.9	106	4.5
Kigoma	136	5.4	158	5.4	77	5.4
Shinyanga	124	5.8	131	6.1	98	4.2
Kagera	172	4.6	175	4.6	140	3.9
Mwanza	130	5.7	142	6.3	111	4.5
Mara	123	5.5	140	5.7	85	4.7
Manyara	128	5.1	134	5.3	108	4.1
Njombe	147	4.2	160	4.2	103	3.9
Katavi	113	5.9	116	6.1	102	4.6
Simiyu	111	6.7	113	6.8	92	4.8
Geita	130	6.1	136	6.3	102	4.9
Total	133	4.8	153	5.0	97	4.2

In order to examine the variability of the EAs by size, Table 4 shows the distribution of the EAs for different ranges of the number of households, by rural and urban stratum. This table also shows the standard deviation of the number of households by EA, which is a measure of the variability in size. It can be seen that the standard deviation for the rural EAs (62.6) is considerably higher than that for the urban EAs (36.9). Most of the EAs were within the range of 60 to 250 households, with the rural EAs concentrated in the higher end of this range. The maximum number of households in an EA was 906 for the rural stratum and 653 for the urban stratum. The minimum size of an EA was one household for both the urban and rural strata, but there were only 186 EAs with less than 30 households in the entire frame. Given the selection of the sample EAs systematically with probability proportional to size (PPS) within each stratum, such small EAs had a very small probability of selection.

Table 4. Distribution of EAs in 2002 Tanzania Census Frame by Size (Number of Households), Rural and Urban Stratum

Number of Households in EA	Number of EAs		
	Total	Rural	Urban
1-29 hhs.	186	91	95
30-59 hhs.	3,032	1,083	1,949
60-79 hhs.	6,268	1,949	4,319
80-99 hhs.	8,020	3,268	4,752
100-119 hhs.	7,752	4,356	3,396
120-149 hhs.	9,775	7,245	2,530
150-199 hhs.	10,491	9,384	1,107
200-249 hhs.	4,484	4,284	200
250-299 hhs.	1,562	1,505	57
300-349 hhs.	499	486	13
350-399 hhs.	168	165	3
400-499 hhs.	102	97	5
500+ hhs.	36	34	2
Total	52,375	33,947	18,428
Mean	133.0	152.7	96.7
Minimum	1	1	1
Maximum	906	906	653
Std. Dev.	61.1	62.6	36.9

Following the selection of the sample EAs at the first sampling stage, a new listing of households was conducted in each sample EA. At the second sampling stage households were selected from the listing for each sample EA. The units of analysis for the 2011/12 Tanzania HBS were the individual households and the persons in these households.

2.2. Stratification of the Sampling Frame for the 2011-12 HBS

In order to increase the efficiency of the sample design for 2011/12 Tanzania HBS, it was important to divide the sampling frame of EAs into strata that were as homogeneous as possible. The first stage sample selection was carried out independently within each explicit stratum. The nature of the stratification depends on the most important characteristics to be measured in the survey, as well as the domains of analysis; the strata should be consistent with the geographic disaggregation to be used in the survey tables. It is also desirable to order the EAs within each stratum by certain criteria that are correlated with key survey variables, in order to provide further implicit stratification when systematic selection is used.

The first level of stratification corresponds to the geographic domains of analysis defined for the 2011/12 HBS. The sampling design has three major geographic domains:

- Dar es Salaam
- Other urban
- Rural

It was recommended to order the EAs in the frame for the other urban and rural domains by region in order to provide an implicit stratification by region. The separate frame of EAs for Dar es Salaam Region was ordered by urban and rural residence; there are not many rural EAs in this region, so a small proportional sample of rural EAs would be selected for Dar es Salaam. Given that the sample EAs were selected systematically with PPS, this ordering of the sampling frame also automatically provided a proportional allocation of the sample EAs in each region based on the total number of households in the frame. In this

case the rural and urban part of each region could also be treated as explicit strata for the calculation of sampling errors for the estimates of key indicators, as described later in this report. Within the rural and urban part of each region, the EAs were further sorted by district, ward and EA codes to ensure that the sample was geographically representative.

In the case of the 2007 HBS, the listing sheet included information that was used for a second stage stratification of the households by high, medium and low socioeconomic groups based on assets ownership criteria. In determining whether to use this approach for the 2011/12 HBS there were several issues to be considered:

- In some countries they have found that the quality of the listing data for stratifying the households is variable, given that the listing is conducted in a relatively short period of time. This results in some misclassification of the households by socioeconomic stratum.
- This procedure would increase the cost of the listing, and complicate the second stage sampling procedures and the calculation of the weights, since the weights will vary by the number of households listed in each second stage stratum within an EA.
- The distribution of the households by socioeconomic stratum would vary by EA, thus increasing the variability in the weights.
- Given the homogeneity of the household characteristics within an EA, most of the households may belong to one stratum (high, medium or low), so it is difficult to find a sufficient number of households in the other strata.
- The effectiveness of this approach also depends on the proportion of households in each socioeconomic stratum.

The only advantage of this sampling method is if it is necessary to increase the number of sample households in the high socioeconomic stratum. A random systematic sample of households within each EA will provide a proportional distribution of the sample households by socioeconomic stratum. Therefore it was agreed by HBS TWG to avoid the complication of the socioeconomic stratification of households in the listing at the second sampling stage given the limitations of this approach.

2.3. Sampling Errors and Design Effects for Estimates from 2007 Tanzania HBS Data

In order to study the sample size requirements for the different sampling alternatives and the corresponding expected level of precision for estimates of key indicators by domain for the 2011/12 HBS, it was important to examine the results from the 2007 Tanzania HBS. First we examined the distribution of the sample EAs and households selected for the 2007 HBS by domain, shown in Table 5. There were originally 448 sample EAs selected for this survey, but one sample PSU in Dar es Salaam was later dropped.

Table 5. Distribution of Sample EAs and Households Selected for 2007 HBS by Domain

Domain	No. Sample EAs	No. Sample Households
Dar es Salaam	152	3,648
Other Urban	158	3,792
Rural	137	3,288
Tanzania	447	10,728

In calculating the sampling errors for estimates from the 2007 HBS data, it is important that the variance estimator takes into account the nature of the sample design. The Complex Samples module of SPSS uses a linearized Taylor series variance estimator that incorporates the stratification and clustering in the sample design, as described later in the section on Calculation of Sampling Errors. Therefore this software package was used for tabulating the standard errors and design effects for the poverty rate and the mean expenditure per adult equivalent by domain based on the 2007 HBS data. Table 6 presents the 2007 HBS results for the estimates of the poverty rate by domain, and the corresponding standard errors, coefficient of variation,

design effects, 95% confidence intervals and number of observations (sample households). Table 7 shows similar results for estimates of the mean expenditures per adult equivalent by domain.

*Table 6. Estimates of Poverty Rate from Tanzania 2007 HBS Data by Domain, with Corresponding Standard Errors, Coefficients of Variation, Design Effects, 95% Confidence Intervals and Number of Sample Households**

Domain	Estimate	SE	CV (%)	DEFF	95% Confidence Interval		No. Sample Households
					Lower	Upper	
Dar es Salaam	0.161	0.015	9.0%	1.33	0.133	0.190	3,456
Other Urban	0.242	0.018	7.4%	3.27	0.206	0.277	3,735
Rural	0.374	0.024	6.5%	19.60	0.326	0.421	3,271
Tanzania	0.333	0.019	5.6%	16.44	0.296	0.370	10,462

Table 7. Estimates of Mean Expenditure per Adult Equivalent from Tanzania 2007 HBS Data by Domain, with Corresponding Standard Errors, Coefficients of Variation, Design Effects, 95% Confidence Intervals and Number of Sample Households

Domain	Estimate	SE	CV (%)	DEFF	95% Confidence Interval		No. Sample Households
					Lower	Upper	
Dar es Salaam	34,196	1,027.1	3.0%	1.76	32,178	36,215	3,456
Other Urban	30,690	1,084.8	3.5%	3.56	28,558	32,822	3,735
Rural	23,373	824.0	3.5%	13.57	21,753	24,992	3,271
Tanzania	25,917	631.1	2.4%	8.83	24,677	27,158	10,462

The design effect (DEFF) is defined as the ratio of the variance of a particular estimate based on the actual stratified two-stage sample design and the corresponding variance from a simple random sample of the same size; it is a measure of the relative efficiency of the sample design. The relatively high design effects in Tables 6 and 7, especially for the rural stratum, are due to the large number of sample households selected in each sample EA (24). One reason such a large number of households was selected in each sample EA is that a resident enumerator was assigned to each sample EA for conducting the interviews spread over the 12 months (that is, two sample households were interviewed per month).

The much higher design effects for the rural areas are unusual, since in many countries a higher clustering effect is found in urban areas, where there is greater variability in the socioeconomic characteristics between EAs. The very large design effect for the poverty rate in rural areas indicates a high clustering effect for this variable. This could be related to the variability between some sample rural EAs where most of the households are poor, and other rural EAs where most of the households are not poor. This could partly result from non sampling errors, due to variability in the quality of the survey data for auto-consumption between different sample rural EAs, for example. If the auto-consumption is under-reported in some EAs, most of the households in those EAs may be classified as poor. This is consistent with the high design effect for expenditures in rural areas. Therefore it was important to improve the quality control procedures for the 2011/12 HBS to ensure that all of the consumption data are captured for sample households in rural areas during the data collection.

The results in Tables 6 and 7 were used to simulate the sampling errors and design effects that could be expected based on alternative sample designs for the 2011/12 HBS, as described in the next section.

2.4. Sample Size and Allocation for 2011/12 HBS

The sample size for a particular survey is determined by the accuracy required for the survey estimates for each domain, as well as by the resource and operational constraints. The accuracy of the survey results depends on both the sampling error, which can be measured through variance estimation, and the non sampling error, which can only partially be measured through re-interview or validation studies. The sampling error is inversely proportional to the square root of the sample size. On the other hand, the non sampling error may increase with the sample size, since it is more difficult to control the quality of a larger operation. It is therefore important that the overall sample size be manageable for quality and operational control purposes. The sample size also depends on cost considerations and logistical issues related to the organization of the teams of enumerators and the workload for the data collection each month.

An important aspect of the sample design that will affect the sample size and level of precision is the number of households to be selected in each sample EA at the second sampling stage, since this determines the statistical efficiency of the sample design, measured by the design effect. The design effect depends on the intraclass correlation coefficient, a measure of the similarity between the households within a cluster for a particular variable, as well as the average number of households selected in each cluster. The design effect is generally higher for estimates of socioeconomic characteristics such as household income and expenditures. For this reason in most countries the number of sample households per cluster for socioeconomic household surveys such as the HBS is generally between 8 and 15. In the case of the 2007 Tanzania HBS, the number of households selected per EA was 24, given that resident enumerators were responsible for interviewing 2 sample households each month. As a result, the design effects for the estimates of the poverty rate and the mean expenditure per adult equivalent are relatively high, as shown in Tables 6 and 7.

In order to ensure consistency in the methodology for the 2011/12 HBS, the NBS decided to use the same data collection procedures using resident enumerators. It was recommended to use the same methodology of selecting 24 sample households per sample EA, since it would not be cost-effective for the resident enumerator to travel between two distant sample EAs. However an important methodological change has been implemented into HBS 2011/12, that was not present in previous HBS's, which has lead to an increase in the number of sample households per EA from 24 to 26 (see chapter 3).

The relatively high design effects noted above due to having a higher than usual number of households per EA had to be taken into account when determining the required sample size.

For the sample design of 2011/12 HBS based on the three major geographic domains, it was recommended to modify the sample allocation used for the 2007 HBS shown in Table 5. It can be seen in that table that only 30.6 percent of the sample was allocated to the rural domain even though 77.8 percent of the population is rural. One reason for this large oversampling in the urban strata is because there are two urban domains. However, this disproportionate sampling also makes the sample less efficient for national-level estimates. The large difference between the weights for the rural and urban sample households results in high design effects for the national-level estimates, as seen in Tables 6 and 7. It can also be seen in these tables that the design effects are higher for the rural stratum compared to Dar es Salaam and the other urban stratum.

Table 8 presents the proposed sample size and allocation for the 2011/12 HBS. The total of 400 sample EAs and 10,400 households is slightly smaller than the corresponding sample size for the 2007 HBS, but the distribution of the sample should be more effective. This sample allocation takes into account the sampling errors and design effects by domain from the 2007 HBS shown in Tables 6 and 7, as well as the distribution of the sampling frame. The approximate level of precision that can be expected from this proposed sample size and allocation is discussed later in this section.

Table 8. *Proposed Number of Sample EAs and Households for the 2011/12 HBS by Domain*

Domain	No. Sample EAs	No. Sample Households
Dar es Salaam	120	3,120
Other Urban	120	3,120
Rural	160	4,160
Tanzania	400	10,400

Although the proposed sample allocation in Table 8 still implies a higher sampling rate for Dar es Salaam and the other urban stratum compared to the rural stratum, this sample allocation is less disproportionate than that for the 2007 HBS.

The standard errors and design effects from the 2007 HBS data shown in Tables 6 and 7 were used to simulate the level of precision that can be expected from the two sampling alternatives. The ratio of the variance (square of the standard error) for a survey estimate based on the proposed sample design for the 2011/12 HBS to the corresponding variance based on the 2007 HBS data can be expressed as follows:

$$\frac{Var_{HBS11}(\bar{x})}{Var_{HBS07}(\bar{x})} = \frac{\frac{\sigma_x^2}{n_{HBS11}} \times DEFF_{HBS11}}{\frac{\sigma_x^2}{n_{HBS07}} \times DEFF_{HBS07}} \approx \frac{n_{HBS07}}{n_{HBS11}} \times \frac{DEFF_{HBS11}}{DEFF_{HBS07}},$$

where:

$Var_{HBS11}(\bar{x})$ = approximate variance (square of standard error) for estimate of a mean based on the proposed sample design for 2011/12 HBS

$Var_{HBS07}(\bar{x})$ = variance for estimate of a mean from 2007 HBS data based on the actual sample design

σ_x^2 = population variance (square of standard deviation) of variable x (such as household expenditure)

n_{HBS11} = proposed number of sample households (for domain of estimation) in 2011/12 HBS

n_{HBS07} = actual number of sample households (for domain of estimation) in 2007 HBS

From this ratio of the variances for the two surveys, we can calculate the approximate standard error of the estimate that would result from each alternative sample design for the 2011/12 HBS as follows:

$$se_{HBS11}(\bar{x}) = \sqrt{Var_{HBS11}(\bar{x})} \approx \sqrt{Var_{HBS07}(\bar{x}) \times \frac{n_{HBS07}}{n_{HBS11}} \times \frac{DEFF_{HBS11}}{DEFF_{HBS07}}}$$

For the sampling design based on the three major geographic domains, the number of households selected in each sample EA for the 2011/12 HBS will be 26 households and the design effects for the 2011/12 HBS should be similar to those of the 2007 HBS when 24 households were selected. In this case the formula for the approximate standard error that can be expected for the survey estimate for a particular domain can be simplified as follows:

$$se_{HBS11}(\bar{x}) \approx se_{HBS11}(\bar{x}) \times \sqrt{\frac{n_{HBS07}}{n_{HBS11}}}$$

Therefore the expected standard error will depend on the square root of the ratio between the number of sample households in the domain from the 2007 HBS and that proposed for the 2011/12 HBS for each domain under the first sampling alternative (shown in Table 8). This formula was used to estimate the standard errors and 95% confidence intervals for the estimates of the poverty rate and the mean expenditure per adult equivalent. These results are shown in Tables 9 and 10, respectively.

Table 9. Simulation of Expected Level of Precision for Poverty Rate by Domain for the 2011/12 HBS, Using 2007 Tanzania HBS Data

Domain	Estimate	SE	CV (%)	DEFF	95% Confidence Interval		No. Sample Households
					Lower	Upper	
Dar es Salaam	0.161	0.015	9.5%	1.33	0.131	0.191	3,120
Other Urban	0.242	0.020	8.1%	3.27	0.203	0.280	3,120
Rural	0.374	0.022	5.8%	19.60	0.331	0.416	4,160
Tanzania	0.333	0.019	5.6%	16.44	0.296	0.370	10,400

Table 10. Simulation of Expected Level of Precision of Mean Expenditure per Adult Equivalent Estimated for the 2011/12 HBS, using 2007 Tanzania HBS Data

Domain	Estimate	SE	CV (%)	DEFF	95% Confidence Interval		No. Sample Households
					Lower	Upper	
Dar es Salaam	34,196	1081.0	3.2%	5.64	32,078	36,315	3,120
Other Urban	30,690	1186.9	3.9%	3.56	28,364	33,016	3,120
Rural	23,373	730.7	3.1%	13.57	21,941	24,805	4,160
Tanzania	25,917	633.0	2.4%	8.83	24,677	27,158	10,400

It can be seen in Table 9 that the sampling error for the poverty rate is fairly reasonable for each domain, and the confidence intervals for the different domains do not overlap. As a result, this sample size is sufficient to determine a statistically significant difference between the poverty rates for these domains. In the case of the estimates of the mean expenditure per adult equivalent, the coefficient of variation for each domain is within 5%, which is considered good.

2.5. Sample Selection Procedures

The sample selection methodology for the 2011/12 HBS was based on a stratified two-stage sample design. The procedures used for each sampling stage are described separately here.

2.5.1. First Stage Selection of Sample EAs

At the first sampling stage the sample EAs for the 2011/12 HBS was selected within each stratum systematically with PPS from the ordered list of EAs in the sampling frame. The measure of size for each EA is based on the total number of households identified in the sampling frame of EAs from the 2002 Tanzania Census. The EAs within each stratum were ordered geographically by region, district, ward and EA codes, providing implicit geographic stratification of the sampling frame. This procedure ensures a proportional allocation of the sample by region within the other urban and rural strata. Within each stratum the following first stage sample selection procedures were used:

- (1) Cumulate the measures of size (number of households) down the ordered list of EAs within the stratum. The final cumulated measure of size will be the total number of households in the frame for the stratum (M_h).
- (2) To obtain the sampling interval for stratum h (I_h), divide M_h by the total number of EAs to be selected in stratum h (n_h): $I_h = M_h/n_h$.
- (3) Select a random number (R_h) between 0 and I_h . The sample EAs in stratum h will be identified by the following selection numbers:

$$S_{hi} = R_h + [I_h \times (i - 1)], \text{ rounded up,}$$

where $i = 1, 2, \dots, n_h$

The i -th selected EA is the one with a cumulated measure of size closest to S_{hi} but not less than S_{hi} .

An Excel file was used for selecting the sample EAs in each stratum for the 2011/12 HBS following these procedures, based on the final allocation of the sample EAs. The Excel file had a separate spreadsheet for each stratum, showing the ordered list of EAs with the corresponding information from the sampling frame of EAs. Each spreadsheet documents the first stage systematic selection of sample EAs with PPS for a stratum. The file has a summary spreadsheet with the frame information for all sample EAs, and formulas for calculating the probabilities and weights based on the information in the frame.

Once the 400 EAs had been selected, the next task was to find the relevant map from the Cartography Department. Of the original 400 selected EAs 13 maps could not be found (3.3%, see Table 12) and it was necessary to select replacement EAs. The missing maps were concentrated in Dar es Salaam (9 out of the 13).

A random process for selecting the replacement EAs, using a PPS sampling procedure was used. Each replacement EA was selected with PPS from the sampling frame within the same ward as the original sample EA with a missing map. The weight was calculated using the measure of size of the replacement EA.

A listing of households was conducted in each sample EA prior to the 2011/12 HBS data collection in order to select the sample households. Supervisors verified the boundaries of the sample EA in order to ensure good coverage of the listed households.

2.5.2. Second Stage Selection of Sample Households within a Sample EA

A systematic sample of m_{hi} 26 households was selected from the listing for each sample EA. The systematic sample of households was selected from the household listing for each sample EA using the following procedures:

- (1) All the households in valid (occupied) housing units in the EA were assigned a serial number from 1 to M'_{hi} , the total number of households listed in the EA.

(2) To obtain the sampling interval for the selection of households within the sample EA (I_{hi}), divide M'_{hi} by m_{hi} , and maintain 2 decimal places.

(3) Select a random number (R_{hi}) with 2 decimal places, between 0.01 and I_{hi} . The sample households within the sample EA will be identified by the following selection numbers:

$$S_{hij} = R_{hi} + [I_{hi} \times (j-1)], \text{ rounded up,}$$

where $j = 1, 2, 3, \dots, m_{hi}$

The j -th selected household is the one with a serial number equal to S_{hij} .

A household selection table for selecting 26 sample households and a reserve of 4 random households for possible replacement in each sample EA was developed for this purpose. An Excel spreadsheet was used for generating this table, using the random number function and the formulas specified above.

2.6. Weighting Procedures

In order for the sample estimates from the 2011/12 HBS to be representative of the population, it was necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample household was equal to the inverse of its probability of selection (calculated by multiplying the probabilities at each sampling stage). Since all survey data was processed by computer, it was easy to attach a weight to each sample household record in the data files, and the tabulation programs weighted the data automatically. The sampling probabilities at each stage of selection were maintained in an Excel spreadsheet with information from the sampling frame for each sample EA so that the corresponding overall probability and corresponding weight could be calculated.

The overall probability of selection for sample households in the 2011/12 HBS can be expressed as follows:

$$p_{hi} = \frac{n_h \times M_{hi}}{M_h} \times \frac{m_{hi}}{M'_{hi}},$$

where:

p_{hi} = probability of selection for the sample households in the i -th sample EA in stratum h

n_h = number of sample EAs selected in stratum h for the 2011/12 HBS

M_h = total number of households in the sampling frame of EAs for stratum h

M_{hi} = total number of households in the frame for the i -th sample EA in stratum h

m_{hi} = number of sample households selected in the i -th sample EA in stratum h

M'_{hi} = total number of households listed in the i -th sample EA in stratum h

The two components of this probability of selection correspond to the individual sampling stages.

The basic sampling weight, or expansion factor, is calculated as the inverse of this probability of selection. Based on the previous expression for the probability, the weight can be simplified as follows:

$$W_{hi} = \frac{M_h \times M'_{hi}}{n_h \times M_{hi} \times m_{hi}},$$

where:

W_{hi} = basic weight for the sample households in the i-th sample EA in stratum h

If m_{hi} is constant for each stratum (26, for example), the sample will be approximately self-weighting within each stratum. These weights vary based on the difference between the number of households listed in each sample EA and the corresponding number from the sampling frame.

2.7 Weighting to take account of non response

It was also important to adjust the weights to take into account the non-interviews in each sample EA. The final weight (W'_{hi}) for the sample households in the i-th sample EA in stratum h can be expressed as follows:

$$W'_{hi} = W_{hi} \times \frac{m'_{hi}}{m''_{hi}},$$

where:

m'_{hi} = total number of valid (occupied) sample households selected in the i-th sample EA in stratum h

m''_{hi} = total number of sample households with completed interviews in the i-th sample EA in stratum h

The sampling probabilities at each stage of selection were maintained in an Excel spreadsheet with information from the sampling frame for each sample EA so that the corresponding weights could be calculated. Following the listing and data collection for the HBS the total number of households listed in each sample EA and the final number of household interviews completed, including replacements were added to this file. Original sample households which could not be interviewed were replaced from the reserve sample of households for each EA (four replacement households provided in each EA). By the end of the data collection 10,186 households were successfully interviewed for the survey out of the sample size of 10,400 households. The completed interviews included 398 replacement households, for a replacement rate of 3.9%. In relation to when households should be replaced with a reserve there was no reason to treat refusals differently from the category respondent not at home, for example because if there was a concern about the refusal rate in some urban areas, the effective sample size would decrease if the refusals are not replaced, and the interviewer workloads would vary. Therefore there was a standard approach to replace all non-interviews, after all attempts are made to interview the original sample household. The supervisor made the decision to replace a household after making a final attempt to interview the original sample household.

The distribution of the final sample EAs enumerated and sample households interviewed by replacement status is presented in Table 11.

Table 11. Final Distribution of Sample EAs for the 2011-2012 HBS by Geographic Domain

Domain (Stratum)	Number and percentage of Sample EAs Enumerated				
	Total	Original	%	Replaced	%
Dar-es-Salaam	120	111	92.5	9	7.5
Other Urban	120	117	97.5	3	2.5
Rural	160	159	99.3	1	0.7
Tanzania	400	387	96.7	13	3.3

Table 12. *Final Distribution of Completed Interviews for the 2011-2012 HBS by Geographic Domain*

Domain (Stratum)	Number and percentage of Completed Household Interviews				
	Total	Original	%	Replaced	%
Dar-es-Salaam	3,016	2,890	95.8	126	4.2
Other Urban	3,040	2,894	95.2	146	4.8
Rural	4,130	4,004	96.9	126	3.1
Tanzania	10,186	9,788	96.1	398	3.9

2.8 Final Response Rate

The final household response rate is in reference to the original households selected. A total of 9,788 original sample households (out of 10,400) were interviewed, so the final household response rate was 94.1%.

The Final Household Outcome for all contacted households (including replacements) is shown in Table 13.

Table 13. *Final Household Outcome by strata, HBS 2011-12*

Final Household Outcome	Other Urban		Rural		Dar e Salaam		Mainland Tanzania	
	N	%	N	%	N	%	N	%
Household interviewed	3040	92.9	4130	96.2	3016	92.8	10186	94.2
Non contact after 3 calls	2	0.1	1	0	0	0	3	0
Address temporarily empty	1	0	0	0	0	0	1	0
Household Refused	213	6.5	149	3.5	193	5.9	555	5.1
Household refused diary ³	12	0.4	11	0.3	38	1.2	61	0.6
Address empty/derelict	4	0.1	1	0	3	0.1	8	0.1
<i>Total</i>	<i>3272</i>	<i>100</i>	<i>4292</i>	<i>100</i>	<i>3250</i>	<i>100</i>	<i>10814</i>	<i>100</i>

The nonresponse adjustment factor for the weights of most EAs is equal to 1 because of the replacements. The numerator of the non-interview adjustment factor excludes the eight non-interview households in the category "Address empty/derelict" as they are not considered as valid households⁴. The weight adjustment factors vary from 1 to 1.5294.

2.9. Adjusting the weight by the Population and Housing Census (PHC) 2012

Following the calculation of the weights using the specified weighting procedures it was important to compare the weighted distribution of households and population from the HBS to the original 2002 frame. Table 14 shows the distribution of the total number of households in the 2002 Tanzania Census frame by stratum and the corresponding final weighted distribution of households and population from the 2011/2012 HBS data.

Table 14. *Distribution of Total Number of Households in 2002 Tanzania Census Frame by*

³ These 61 households did complete questionnaires 1- 4 but refused to complete the diary. As the poverty variables are based on the 10,186 households who completed all sections of the questionnaires and the diary these 61 households have been classified as non-responders. If they were included in analysis an additional set of weights would need to be created and this would add complexity and potential confusion during data analysis.

⁴ As the listing exercise took place just prior to data collection and the enumerators did not include empty or derelict buildings onto the list, the numbers recorded during data collection were very low.

*Geographic Domain and Corresponding Weighted Distribution of Households
and Population from Final 2011/2012 HBS Data*

Domain (Stratum)	2002 Census Frame		Weighted HBS Estimates		
	Total Households	%	Total Households	%	Total Population
Dar-es-Salaam	602,549	8.8%	755,935	9.4%	3,009,259
Other Urban	1,228,478	18.0%	1,644,963	20.4%	7,657,942
Rural	4,995,513	73.2%	5,680,451	70.3%	30,132,726
Tanzania	6,826,540	100.0%	8,081,348	100.0%	40,799,927

The timing of the HBS 2011-12 was fortunate in that it coincided with the new 2012 Population and Housing Census (data collection took place in August 2012). As a validation exercise to check the weights and the estimates coming from HBS 2011-12 it was extremely useful to compare HBS 2011-12 with the new Census data. In doing this two important tasks had to take place to compare the databases:

1. A one-to-one mapping between the 2002 strata and the corresponding areas in the 2012 Census had to be undertaken. This time-consuming activity was compiled by NBS.
2. Limit the comparison to the household-based population from the 2012 Census (i.e. exclude the institutional population).

In comparing the two results (PHC 2012 and HBS 2011-12) we can see there was an apparent under-estimation of about 1.5 million people from the HBS.

Table 15. Distribution of Population by stratum in 2012 Tanzania PHC and 2011-2012 HBS

Domain	PHC 2012	HBS 2012	Difference between PHC 2012 and HBS 2011-12	Adjustment Factor
Dar es Salaam	4,262,497	3,009,259	1253238	1.416
Other Urban	7,922,947	7,657,942	265005	1.034
Rural	30,084,707	30,132,726	-48019	0.998
Total	42,270,151	40,799,927	1,470,224	1.036

We can assume some level of measurement error in both data collection exercises. As a Census of 45 million people is a massive exercise we might fairly assume that measurement error is higher in the Census. However we assume the PHC 2012 is correct and the HBS 2011-12 weights were adjusted further with an adjustment factor (see Table 15) to bring all strata population numbers to match those from PHC 2012.

The large discrepancy between the Dar es Salaam estimates of population size is disconcerting and could suggest that the:

1. The listing in Dar es Salaam was not of good quality.
2. The population of Dar es Salaam grew considerably between October 2011 when listing was undertaken and August 2012 when Census 2012 was undertaken. There is some anecdotal evidence that the population of Dar es Salaam swelled during Census 2012 due to several institutions closing and people returning home.

In terms of quality of the listing there is some useful evidence that can be gathered by examining the weights and their links to the high design effects of the poverty estimates (discussed in Chapter 6).

A factor that can contribute to high design effects is the variability in the weights. Table 16 shows the statistical distribution of the household weights by stratum, including the mean weight, the minimum and maximum values of the weight and the standard deviation of the weights.

Table 16. Statistical Distribution of the Adjusted Household Weights by Stratum for the HBS Data

Stratum	Number of Sample Households	Minimum Weight	Maximum Weight	Mean Weight	Standard Deviation of Weights
Other Urban	3040	107.47	3518.18	559.83	512.50
Rural	4130	255.44	3954.37	1373.22	543.14
Dar-es-Salaam	3016	109.42	1020.99	355.02	179.32

It can be seen in Table 16 that the household weights are much more variable for the rural and other urban strata, where the standard deviation of the weights is about 3 times higher than that for Dar-es-Salaam. The lower variability of the weights for Dar-es-Salaam also resulted in lower design effects for this stratum.

One cause of the variability of the weights within a stratum is the difference between the number of households listed in each sample EA and in the corresponding number of households in the sampling frame. As the sampling frame was based on the 2002 Tanzania Census there were many changes in the distribution of the households during the approximately 9 years between the Census and the HBS. It is interesting that these differences appear to be greater in the rural and other urban strata compared to Dar-es-Salaam (see Table 17). The mean average overall growth in the number of households per EA is 38 households in urban Eas, 15 in rural Eas and 26 households in Dar es Salaam.

Table 17. Average difference in number of households from the sample frame (PHC 2002 and the HBS 2011-12 listing

	Region	Average difference in number of households from the Sample frame (PHC 2002) and the HBS 2011-12 listing		Average annual growth in population (percent) from PHC 2012 ⁵
		Urban EAs	Rural EAs	
1	Dodoma	-47	13	2.1
2	Arusha	-21	15	2.7
3	Kilimanjaro	30	2	1.8
4	Tanga	63	63	2.2
5	Morogoro	105	2	2.4
6	Pwani	22	-26	2.2
7	Dar es Salaam	26	-	5.6
8	Lindi	-23	49	0.9
9	Mtwara	69	43	1.2
10	Ruvuma	220	125	2.1
11	Iringa	49	1	1.1
12	Mbeya	22	6	2.7
13	Singida	11	28	2.3
14	Tabora	-17	-15	2.9
15	Rukwa	16	-8	3.2
16	Kigoma	1	49	2.4

⁵ Data from p2 of 2012 Population and Housing Census, Population distribution by administrative areas. March 2013.

17	Shinyanga	34	-19	2.1
18	Kagera	15	21	3.2
19	Mwanza	68	39	3.0
20	Mara	-10	58	2.5
21	Manyara	-2	-23	3.2

3. The 28 day diary of household expenditure and consumption.

3.1. Background

The NBS is committed to producing high quality data and reviews previous surveys to learn useful lessons from past experiences and improve survey quality over time. Prior to HBS 2011-12 data collection a debate took place among the HBS TWG on whether Tanzania HBS should continue to use the 28 day diary, especially as for HBS 2011-12 Form II had been introduced with the aim of capturing more non-food expenditure.

It was agreed to keep the 28 day diary, but to administer it in a way that would facilitate clearer evaluation. To enable this, a “staggering scheme” was introduced (developed by Mr Ahmed Makbel). This scheme aimed to remove the potential bias of beginning the reporting period of the diary on the 1st of every month.

3.2. How the diary was implemented in 2011-12

HBS 2011-12 data collection began on October 1st 2011 and the first of the twenty-six households in each of the 400 EAs did their dairy keeping for four weeks from 1–28 October. Two weeks after the first interview the interviewer contacted the second household on their list and began the diary with that household (from Oct 15 – Nov 11). Then two weeks later they contacted the third household and the diary period was 29 Oct to Nov 25 and so on, until 26 households had been interviewed over the 12 month⁶ period (all 26 recording periods are shown in Table 1 below).

3.3. Analysis undertaken to examine the diary data.

In the sections below a comparison is made between results of research conducted on HBS 2007 by Waly Wane (World Bank) in 2011 and the results coming from HBS 2011-12 data. The basis for the HBS 2011-12 analysis is from file *SECTBI_allcorrected_2.sav* of the diary (food consumption). This file consists of 1,736,572 cases. The data was aggregated to create total food expenditure per week and total number of transactions per week. The data was weighted using the household weight (weight). Examination of the data revealed 1.6% of errors (dates and recording period). These 27,030 records were removed from the analysis.

Table 1: The 26 recording periods (sequences) for the 28 day diary in the 2011-12 HBS

1	1 – 28 Oct	10	4 Feb – 2 March	19	9 June – 6 July
2	15 Oct – 11 Nov	11	18 Feb – 16 March	20	23 June – 20 July
3	29 Oct – 25 Nov	12	3 March – 30 March	21	7 July – 3 Aug
4	12 Nov – 9 Dec	13	17 March – 13 April	22	21 July – 17 Aug
5	26 Nov – 23 Dec	14	31 March – 27 April	23	4 Aug – 31 Aug

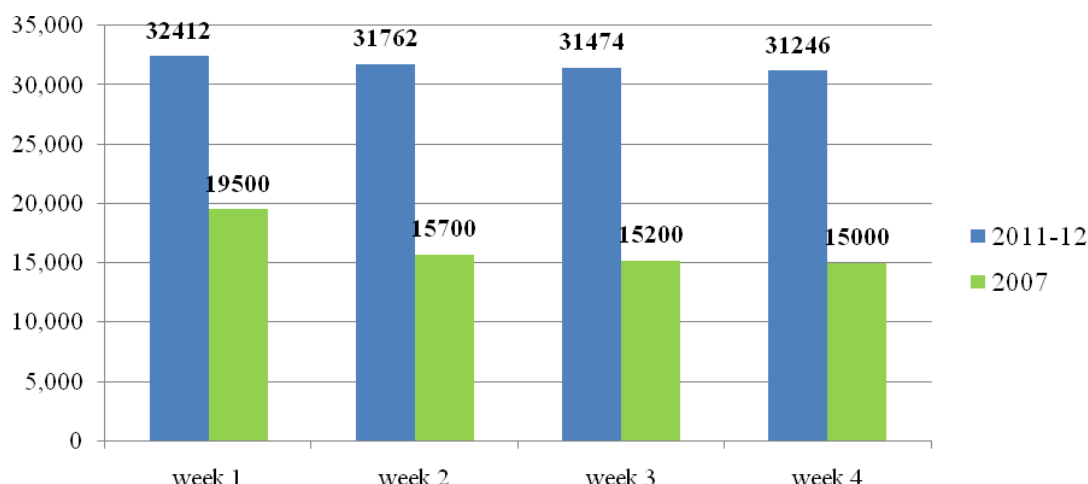
⁶ 54 weeks of fieldwork were used to implement the full staggering scheme.

6	10 Dec – 6 Jan	15	14 April – 11 May	24	18 Aug – 14 Sept
7	24 Dec – 20 Jan	16	28 April – 25 May	25	1 Sept – 28 Sept
8	7 Jan - 3 Feb	17	12 May – 8 June	26	15 Sept – 12 Oct
9	21 Jan – 17 Feb	18	26 May – 22 June		

3.4. Average household consumption

Figure 1 show that in HBS 2007 the average household weekly food consumption experienced a significant drop after the 1st week but then remained fairly constant. In HBS 2011-12 this issue persisted with a small decrease in reporting household consumption over whole four week period. **Household weekly consumption was, on average, 1,166 Tshs less in the last week than in the first week.** One should not conclude that this is because households had less funds left at the end of month (as the weeks now start at all times during the month) this is a measure of the “cost” of respondent and/or enumerator fatigue.

Figure 1: Average weekly household food consumption (Tshs) by week of diary keeping (HBS 2007⁷ and HBS 2011/12)

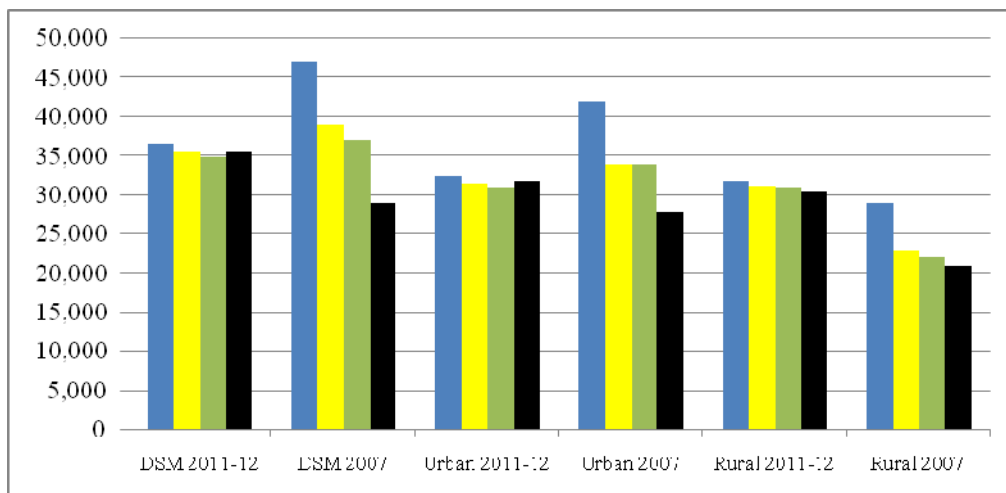


In 2007 a larger reduction in consumption over time was noticed in relation to non food items. This was the trigger to create Form II to more efficiently gather this data.

Patterns of consumption were also examined by strata. As can be seen in Figure 2, in 2007 the reduction in consumption over time was most visible in Dar es Salaam. In 2011-12 this issue did not occur so severely, but was again still visible. Note that in Figure 2 the 2011-12 data is for food only while in HBS 2007 it is total expenditure, this is why the amounts are higher.

⁷ The figures for HBS 2007 are taken from graphs produced by Waly Wane, World Bank, not tables hence the use of numbers rounded to the nearest hundred.

Figure 2: Average weekly household consumption (Tshs) by week of diary keeping period by strata (weeks 1 to 4) HBS 2007⁸ and HBS 2011/12

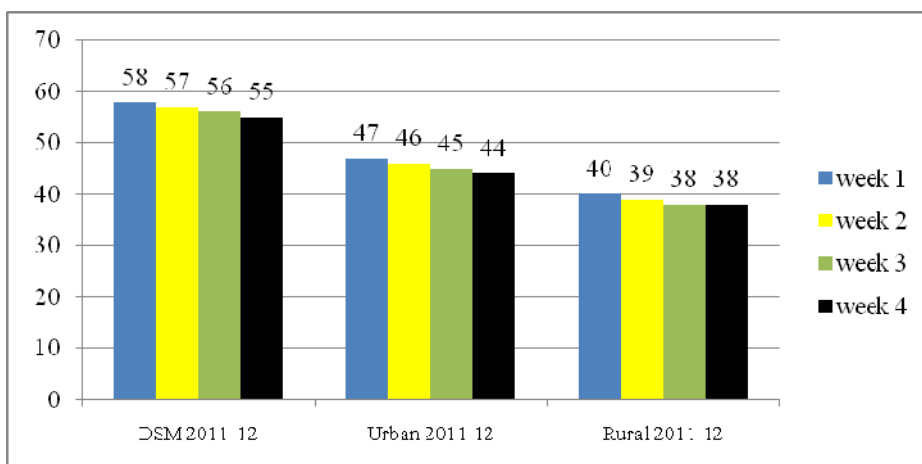


3.5. Average number of transactions

The data was also examined to look at the number of transactions. In HBS 2007 the number of transactions fell during the diary reporting period as respondents became increasingly fatigued. In HBS 2011-12 this continued to occur, but not so severely. The average number of transactions in week 1 was 43, falling to 42 in week 2 with a further slight fall to 41 average transactions in weeks 3 and 4. This demonstrates a quite a lot of dedication on behalf of respondents and enumerators when the urge to lump together expenditures is a natural inclination.

The differing numbers between the strata would not seem to be a quality concern, rather a reflection of the lower levels of consumption (and therefore transactions) outside Dar es Salaam.

Figure 3: Average number of transactions per household by diary keeping week and strata (HBS 2011-12)



⁸ HBS 2007 data is total consumption (food and non food) HBS 2011-12 is food only

3.6. Number of transactions throughout the 12 months of the survey

Figure 4, produced by Waly Wane from the 2007 HBS data, illustrates that as the survey went on the number of transactions became less. The same trend can be seen in Figure 5, coming from the HBS 2007 Final Report. An issue to note from Figure 5 is that the monthly average number of transactions per household is around 200 a month (in HBS 2000-01 it was around 300 transactions a month). The HBS 2011-12 diary is averaging **40 transactions a week (for food only)**. This is a sign of the similar quality between HBS 2007 and HBS 2011-12 in this perspective.

Figure 4. Daily number of transactions throughout 2007 (HBS 2007)

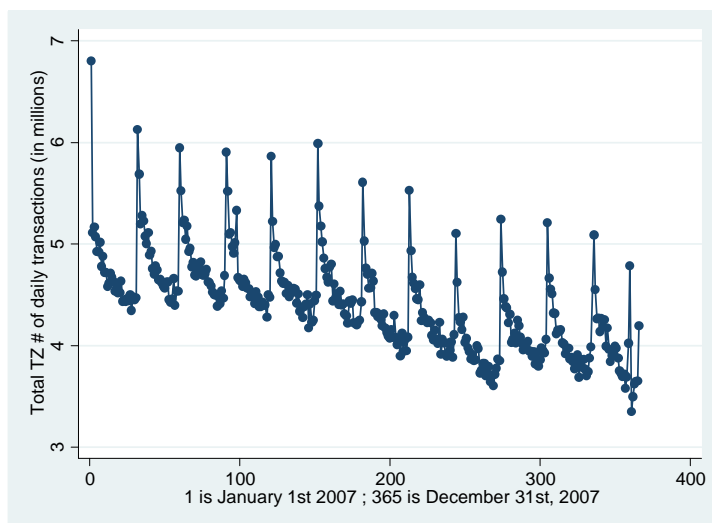
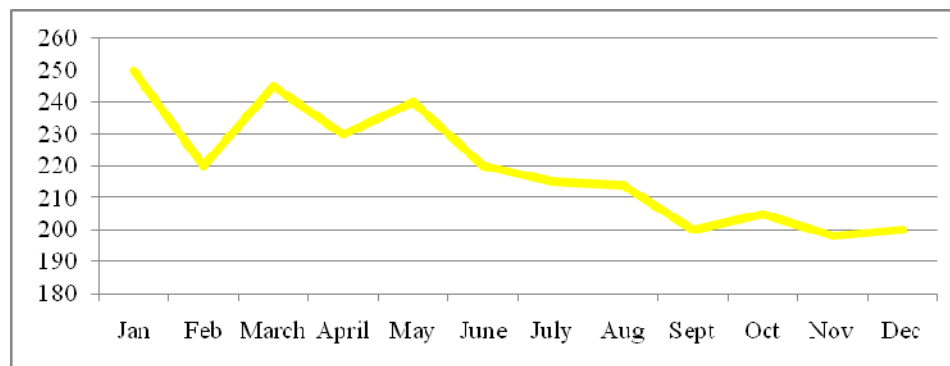


Figure 5: Average monthly transactions per household by month (HBS 2007)



Figures 6 and 7 show that HBS 2011/12 was not immune to reductions in both consumption and number of transactions over time, but the decline was not so steep. The main outlier in the HBS 2011-12 data relates to the final recording period (15 September – 12 October 2012) when enumerators probably felt that they had done the 12 months of data collection and didn't put in much effort in the last two "extra" weeks (1-12 Oct 2012).

Peaks around the sixth recording period, incorporating the approach to Christmas is clear. Easter Monday was on April 9th in 2012 (recording period 13) and there is a peak in this period. [Eid ul-Fitr](#) was on August 21st 2012 (recording period 23) and a small peak is visible during that time also.

Figure 6: Average weekly household consumption (TShs) by recording period (HBS 2011-12)

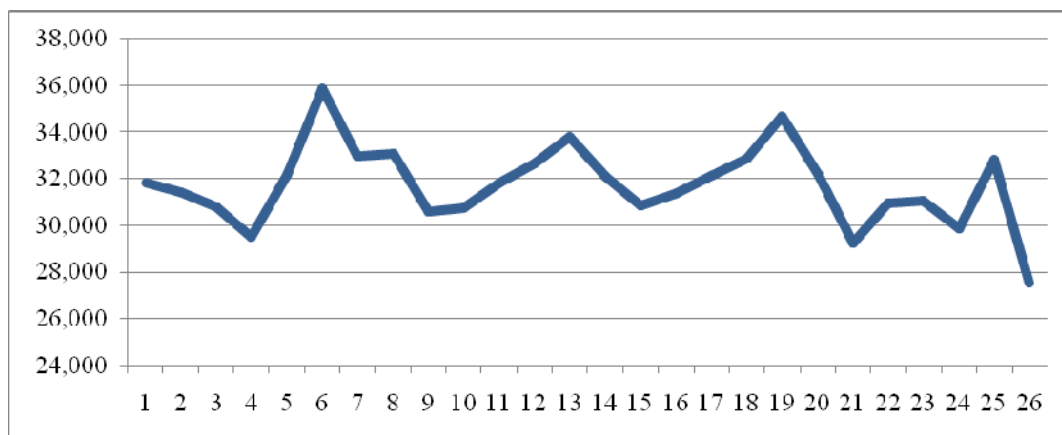
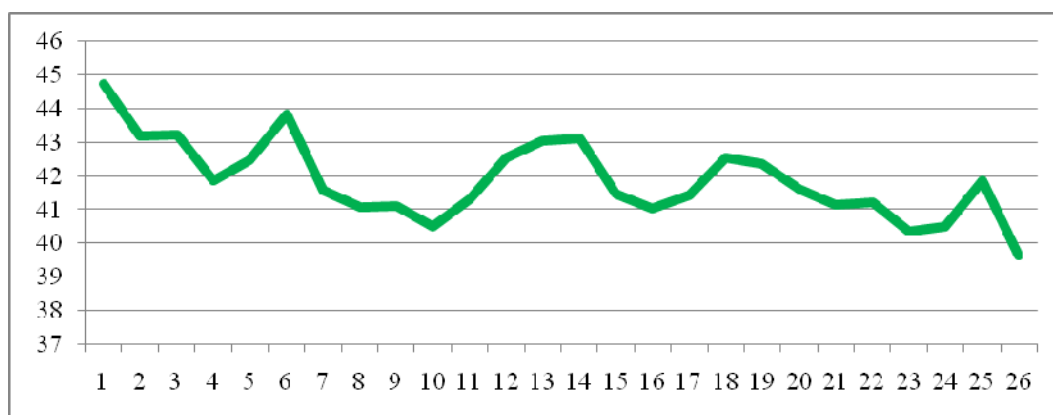


Figure 7: Average number of transactions a week per household by recording period (HBS 2011-12)



3.7. Overall non-completion of the diary

In 2007 the number of households that filled out their diary steadily decreased over time again. The downward slope accelerated week by week. By the end of the 28 days **15%** of households had dropped out of filling their diary. Overall in HBS 2011-12 non response on the diary was **7.8%**. From the 10,186 interviewed households 9,394 fully completed the diary.

4. Quality control during fieldwork and editing

4.1 Quality control during fieldwork

As part of the entire process of HBS data collection, a quality control team was formed as an integral part of the process. The team is required to make sure that the data being collected meet the expected quality and standards. The team, composed mainly of three staff from the NBS main office (Ahmed Makbel, Ibrahim Masanja, and Emilian Karugendo) visited the following regions;

- Tanga
- Kilimanjaro

- Mwanza
- Mara
- Tabora
- Shinyanga
- Rukwa
- Mbeya
- Ruvuma, and
- Iringa

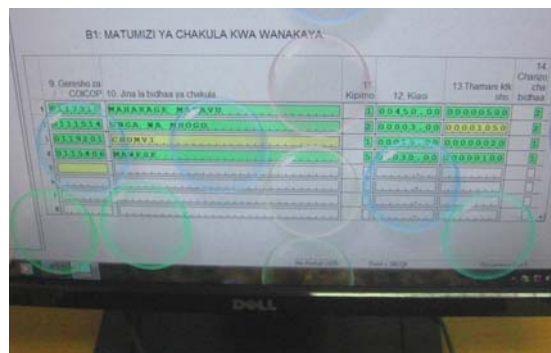
In addition, the HBS Quality Assurance consultant, Ms. Rachel Smith-Govoni, visited Tanga, Kilimanjaro, Arusha, Mwanza, Pwani (Kilindoni) and Kigoma regions during October and January. In all of these regions, enumerators were visited and accompanied by the quality control supervisor to the households that are being interviewed. From Rukwa, Mbeya, Ruvuma, and Iringa regions, **98 households that were being interviewed between November 2011 and February 2012 were visited** by quality control supervisors.



However, general supervision from the main office had already been conducted to all regions as of February 2012. During the visits to the households, the supervisor who at times was accompanied by the respective supervisor and the regional statistical manager conducted some re-interview of the HBS questionnaire forms 1 to 4 and checks the completeness of the individual diaries of the household members. A report of these visits is available on request and the Guide of checks that were undertaken is shown in Appendix 1.

4.2 Quality control during editing

A team of **of ??** editors were used in HBS 2011/12. They were all trained and given the instructions in Appendix 2 as to what items to check. Editors work was also monitored by several NBS staff members. Once editing had taken place the forms were put into the data entry system. The CSPro program included all skips and a number of checks. The program is available on request.



Appendix 1. HBS 2011/12 Guide for staff members and DPs who do field checks

Introduction

The 2011/12 Household Budget Survey data collection exercise has taken-off. It is good practice to undertake visits by supervisors from the main office and this guidance should help field visit staff focus on the priority areas of the completed HBS questionnaires that should be examined. HBS 2011/12 has a total of six questionnaires, Form I through to Form VI and all are administered in every selected household. These forms must be checked on the following sections to ensure of their completeness.

HBS Form I

Section 1: All listed members of the household must have their relationship to the Head of Household established. From experience of the pilot and field practical during the training of enumerators and supervisors, some enumerators were confusing the sex of the household members with the codes of the relationship to the head. This might lead to a finding of some households with multiple household heads, multiple spouses, etc.

Age is an extremely important demographic variable. There are two questions under the age variable; one is recording the month and year of birth and the second is age in completed years. Consistency of these two must be checked. Specifically, the following must be checked;

1. Ensure the Year of Birth (Q3) and the Age (Q4) match.
2. Q5. Ensure the Head of Household is listed on Row 1.
3. Q5. If person 01 is married to person 02 then person 02 must be married to person 01. People cannot be married to themselves. Ensure the mother or father than anyone is older than the child by say 12 years.
4. Question 5 cross checked by Questions 10 and/or 11; and
5. Relationship is 3 and the head is a female age diff between. head and child should be at least 12yrs and if relationship is 6 and same female head, then age diff should be at least 25
6. Relationship is 3 and the head is a male age diff between. head and child should be at least 15yrs and if relationship is 6 and same male head, then age diff should be at least 30
7. Question 6 Cross checked by Question 15
8. Q10 and Q11. Father or mother number can be a person number not in the household (96).
9. This should only have a value if respondents age is between 0 and 17, otherwise it should be blank.
10. Q10 check against age in Q4
11. Q14 check against age in Q4
12. Q37 check against age in Q4
13. Q44 check against age in Q4

HBS Form II:

This form covers information on housing and purchase of durable and semi-durable household items for different reference periods. General completeness of all sections in this form must be checked. For example it is anticipated that a household will have some expenditure on clothing and/or footwear for some of the members in the 12 month period that precedes the survey month and so the check should check these items are recorded in Form II and verify with a household member if they do not appear in the questionnaire.

In this form many amounts will be recorded. Ensure the writing is legible so all amounts can be clearly understood (for example between 1's and 7's). In addition when there is no expenditure ensure only "0" is placed in the right hand box so that it is clearly visible to data entry staff that there is no amount for this answer.

HBS Form III:

This form contains sections 12 through 14. Section 12 starts with transferring names and member numbers from Form I of all household members with at least five (5) years of age. **The transferring must be done in**

accordance with the instructions that require all members of the household that qualify for this section to be copied here with their respective member number from Form I and not re-serializing them afresh. This transferring of member number that do not necessarily run in sequence, i.e. the numbers CAN be 01, 03, 05, 06 for example is a common error – please check this very carefully.

All skips are important not only in this section but throughout all the four questionnaires. Section 12 covers economic activities that must carefully be probed. It should be remembered that most enumerators will confuse from time to time full time students as to where or which codes are relevant for them. Being a student is not an economic activity, what should be recorded in this section is the non-studying economic activity even if the person performed any work ⁱfor an hour in the reference period of one week ago. All questions in this section are equally important but special attention must be paid to question number 5 as this determines the relevance of section 13 in this same form III. If no single household member has a **YES** response in this question, then the entire section 13 is not applicable. Given scarcity of formal employment is both public and private sectors, most Tanzanians especially women are engaging themselves in some sort buying and selling some items directly or after making some transformation on them. A **NO** response (at Q5) might be a result of reluctance to probe on the side of an enumerator

Business or Self Employment

The word commerce generally means the exchange of goods and services, but commerce has its own branches, one of its branches is trade which means the direct buying and selling of goods. There are two branches under trade which is Whole sale trade and Retail trade sale. In this form, it has been realized that there is confusion to some of our enumerators especially in this part of employment. Here are categories of employments; that is

Self employment with employees (workers) and without workers

Let's take the example of one member of a certain household who owned the a garage, and in that premises has shop for spare parts which he/she sells to customers who bring their motor vehicles for repair. In this scenario, we should ask the main activity being undertaken in that workshop. If it is selling of spare parts then we consider him as doing business of direct buying and selling of spare parts.

If the main activity is motor vehicle repairs then we consider him as rendering services, even though whether one is rendering service or is directly engaged to trade, they all mean to make profits. Ref: to the ISIC and TASCO Codes.

Therefore, an example of activities that are not considered as business under **section 12 question 5** include the following:

- Owning and running motor vehicles repair workshop/garage,
- Owning and running a school,
- Owning and running a health facility,
- Owning and running public transportation (passengers) vehicles, etc.

Cross-checking questions under section 13

1. Question 8 item 25 cross-checked by question 8a; the sum of amounts in 8a must be less or equal to that in 8(25).
2. Question 18 Cross checked by Question 21

At section 13, Q19 – Q25, at the top of the page, put two boxes and fill-in number of the household member who own those employees. The box is similar to that found at Section 13, Q1.

Section 14

There must be a box completed for every household member aged five years and above even if all that is entered is the Individual ID and Code 2 “no” for every type of income. Do not leave the box blank even if the individual does not receive any of these incomes.

The questionnaire has space to record the details of five individuals aged 5 years and above. If more than five people aged 5 years and above are in the HH use another Form III and mark the box on the Cover Page with an X to indicate a further form was used.

Form IV:

Check of proper identification transferring from Form I's cover page to forms IV's. Also checks of the skips and proper coding are the only issues to concentrate when checking completeness of this questionnaire. And this is a general rule for all of the questionnaires.

Check question 2, section 16 the size of the farm planted must be less or equal to the size reported in question 2, section 15 for the particular (same) farm

Form V:

This form contains data mainly from Form VI (below) and all records in Form V are transferred/copied from Form VI. Therefore it is required to check that all items from all individual diaries are moved into respective sections of this form.

Specifically, check the following;

Ensure unique HH ID (Region, District, Ward, EA, Household Number) is transferred correctly onto the Cover Page.

On the Cover page at question 1 check that the list of people and their Individual IDs are the same as those on page 1 of Form I.

Part A – ensure every food and non food product written onto the form in question 3 has questions 2, 4, 5, 6, 6 and 7 coded

Part B1 – ensure every food product written onto the form in question 10 has questions 9, 11, 12, 13 and 14 coded

Part B2 – ensure for every day of the diary questions 15, 16 and 17 contain a number for HH members and non household members. None should be “0” not blank. Ensure the number of HH members in each of the questions 15,16, 17 is not higher than the number of HH members listed on the cover page.

Part B3 – ensure every food product eaten outside the home written onto the form in question 19 has questions 18, 20 and 21 coded. Ensure the number of HH members in Q20 is not higher than the number of HH members listed on the cover page.

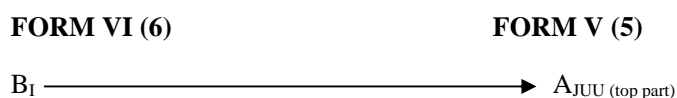
Part B4 – ensure that for every day of the diary that if food is eaten outside the home by a HH member, i.e. there is at least one row completed in Part B3 that 22, 23 and 24 contain a number for HH members. None should be “0” not blank. Ensure the number of HH members in each of the questions 22, 23 and 24 is not higher than the number of HH members listed on the cover page.

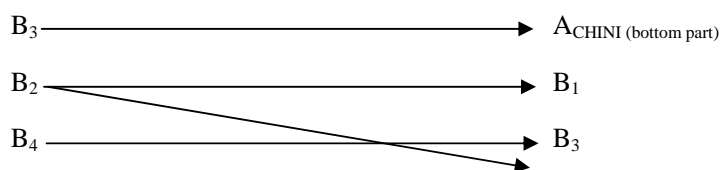
*Note: Form V – the household diary. It is assumed that the transfer of daily records from individual members' diaries (form VI) is done accordingly by the enumerator. **Please verify if the transfer is done accordingly.***

Form VI:

This an individual form for all household members age 5 years and above. A supervisor must firstly make sure that all members of the household that qualify for this form are actually having it and that they understand what is required to be recorded therein. A supervisor must visit households to insist on the importance of individuals to complete the diary as it takes a major importance in the measuring of poverty. It should be insisted that all transactions must be recorded even those that are own produce, gathered, received as gifts or payment in kind, etc. These take a major part of rural consumption but very unfortunately individuals tend to marginalise them and in most cases do not record/regard them as important.

Figure 1: Illustration on how to transfer information from Form VI to Form V





On an ordinary sheet please ask the household member(s) to remember and list all food items consumed in the preceding day. Cross check this list with the diary (form VI) items. This should be done with probing question such as “..... is that all? Can’t you remember any other that you may have forgotten to record?”, etc. A supervisor must do this while temporary keeping the individual diary (form VI) of the member s/he is interviewing. This will help insist on recording all items such as own produced ones, gathered, etc that a member may ignore and not include them in the individual diary.

This brings into one’s mind that a supervisor must randomly choose some households whose interviews have been on-going or recently completed to verify some information s/he may be doubting and as a matter of questionnaire completeness check, must ask quick questions to the household visited; such as listing all household members with their relationship to the household head, their ages, sex, and survival-ship of biological parents of all under 18 years old members of the household. This should be done on the loose form I questionnaire sheets to be collected from and submitted back to the Desk Officer at the time of return.

It should be noted, however, that in case a visit is made to a household whose interviews have been completed, the interval between the supervisor’s visit and the period since interview completion, such interval is suggested to be within seven days.

Check All “other” response categories, in all questionnaires must bear specifications to facilitate post coding activities.

Conclusion

The supervisor must submit a written supervision report covering all of the above parts and others that are not enumerated as these are just the quick ones that are easily seen or that an eye can easily catch. A supervisor is expected to go beyond these and also include even administrative issues that s/he thinks they are of importance to the success of the survey. The report must be submitted to the survey Desk Officer (Stephen Maganda) within seven days of return so that the findings can be communicated.

Appendix 2: 2011/12 HBS Editing specifications

- Questionnaires for every household should contain the following: The listing form, Control form, Form 1, Form 2, Form 3, Form 4, Form 5 and Form 6.
- The Identity parts should be unique within a household, in all the forms; having exactly 13 digits. (ID = reg, dist, ward, ea, selected, hhno.)
- Handwriting in all questionnaires should be legible and clearly visible to the third party (data entrant).
- If extra forms are used to contain extra household members, these forms should be stapled together and ensure that this is clearly visible to the data entrant.

CONTROL FORM

- If the head of household was replaced, then both the names of the previously selected and the replaced heads should appear in the control form.
- The total number of household members recorded in the control form should exactly agree with not only the listed members in Form 1 but also in Form 5.
- If result code is 1 then all the forms should be completely filled in.

FORM 1

- Head of household should be listed first on line number one.
- There should be one and only one head of household.
- Age of head of household should be reasonable, (say not less than 10 years).
- Household members should be listed sequentially with no gaps in between.
- Ensure consistency between the reported date/year of birth (Q3) and age (Q4).
- If there is a child in the household, (Q5 = 3) ensure that the age difference with the head (the father/mother) is reasonable (say, not less than 12 years).
- If there is a grandchild in the household, (Q5 = 6) ensure that the age difference with the head (the grandfather/mother) is reasonable (say, not less than 25 years).
- If there is a parent in the household, (Q5 = 7) ensure that the age difference with the head is reasonable (say, not less than 12 years or not above 70 years).
- If there is a spouse in the household, ensure that the sex is opposite to that of the head.
- Age of spouse should be at least 12 years.
- If there is a spouse in the household, ensure that the marital status of that spouse and that of head is 'married' or living together (Q13=1,2,3,).
- Ensure consistency between citizenship (Q6) and immigration/place of birth (Q15, Q17).
- Ensure at Q20 that age at start schooling is reasonable.
- Ensure consistency between duration of stay (Q14) with age (Q4)
- Ensure consistency between parents' information (Q10, Q11) with existence of parents or relationship (Q5).
- Ensure consistency between education (Q23, Q24) with reported age (Q4).
- Ensure consistency between the reported reason (Q25) with sex, age, relationship and marital status (Q2, Q 4, Q 5 and Q13). If reason is 'too old', age should be 50+ years. If reason is 'too young', age should be under 10 years. If reason is 'pregnancy', sex should be female and age reasonable (12:49).
- Ensure consistency between Q37 and Q44 with reported age (Q4).

FORM 2

- Check consistency between the reported numbers of rooms (Q4) with household size.
- At section 2, Q31; ensure that the listed items are those listed in Q27.
- At section 6, if Q4 is filled in then Q1 should have a 'Yes' and Q2 should not be zero.
- At section 6, if Q5 is filled in then Q1 should have a 'Yes' and Q3 should not be zero.
- At section 6 Q1, if there is any new transport equipment, this should appear in Q4.
- At section 6 Q1, if there is any second hand transport equipment, this should appear in Q5.
- At section 6, Q7; ensure that the listed transport equipments, are those listed in Q5.
- At section 6, Q19; ensure that the listed item are those listed in Q15.
- At section 7, Q5; ensure that the listed items, are those listed in Q1.
- Check consistency of responses at section 7, Q9, Q11, Q13, Q15 with education status of household members (Form 1, Q21, Q23, Q24)
- At section 9 ensure consistency between existence of electricity and types of energy with payments made at section 2, Q10 and Q11.

FORM 3

- Ensure that this form contain household members aged 5 years and above only.
- Ensure that serial number of individual household members is transferred correctly from Form 1.
- If an individual household member has no secondary activity, then Q33 to Q50 should be blank.
- Main activity falls under Q11 to Q32, and secondary activity falls under Q33 to Q50.
- Ensure that if Q10a = 4,5 then Q20 to Q29 is filled-in (main activity component).
- Ensure that if Q10b = 4,5 then Q38 to Q47 is filled-in (secondary activity component).
- Ensure that if Q16 is filled-in, (existence of inactive household member) then Q17 to Q50 is blank.
- Ensure that Q30 to Q32 will be answered only for students (Q10a = 5) in main activity.
- Ensure that Q48 to Q50 will be answered only for students (Q10b = 5) in secondary activity.
- For all household members, section 13 should all be blank if there is no 'yes' at Section 12, Q5.
- At Section 13, Q19 to 25 should be filled-in only if there are employees at Q18.

- Section 13, If Q8 line 25 is 0, then Q8a should be blank.
- Section 13, ensure that there is consistency between Q18 and Q21.
- Ensure that section 14 is filled in for all household members aged 5 years and above.

FORM 4

- If a household is in Urban area then Form 4 might contain no data at all.
- If a household reports to have ever cultivated a crop (section 16) then a farm should exist (section 15).
- Total area of all cultivated crops (section 16, Q2) should not be greater than total area of all farms owned (section 15, Q2).
- If a household reports to have ever processed crops (section 17) then cultivated crops (section 16) should exist.
- If a household reports to have no land (section 15) and not cultivated a crop (section 16) then section 19B should have all responses 'no'.
- At section 16, ensure that amount in kgs sold in Q4 is not greater than that harvested at Q3.
- At section 16, ensure that amount in kgs consumed in Q6 is not greater than that harvested at Q3.
- Ensure that section 16, Q9 to Q11 is filled-in only if Section 15 Q3 is code 1,2 or 5.
- Ensure that all codes used at section 17, Q3 are appropriate. That is, if Q2 is code 1 - 'zao limesindikwa' then use codes 1:7; if Q2 is code 2 - 'zao ni mabaki' then use codes 8:17.

FORM 5 and 6

- Ensure that transactions from Form 6 have all been transferred into Form 5.
 - Ensure that there are transactions for the complete 28 days for every household.
 - The total number of household members recorded in form 5 should agree with that found in the control form and Form 1.
 - COICOP codes for items in section A1 and B1 should be for food items only.
 - COICOP codes for items in section A2 should be those of non food items only.
-