

Appendix A. Sample Design

The major features of sample design are described in this appendix. Sample design features include target sample size, sample allocation, sample frame and listing, choice of domains, sampling stages, stratification, and the calculation of sample weights.

The primary objective of the sample design for the Georgia Multiple Indicator Cluster Survey was to produce statistically reliable estimates of most indicators, at the national level, for urban and rural areas, and for the 11 regions (Tbilisi, Kakheti, Mtskheta – Mtianeti, Shida Kartli, Kvemo Kartli, Samtskhe–Javakheti, Racha-Lechkhumi and Kvemo Svaneti, Imereti, Guria, Samegrelo and Zemo Svaneti, Adjara) of the country, excluding the disputed areas of Abkhazia and South Ossetia.

A multi-stage, stratified cluster sampling approach was used for the selection of the survey sample.

The target sample size for the Georgia MICS was calculated as 13860 households. In order to obtain sub-national estimates for as many indicators as possible, the calculation of the sample size was based on children under five assuming a hypothetical prevalence/coverage of 50 percent and the sample size was calculated for each region separately. The following formula was used to estimate the required sample size for these indicators:

$$n = \frac{[4 (r) (1-r) (f) (1.05)]}{[(0.075r)^2 (p) (n_h)]}$$

where

- n is the required sample size, expressed as number of households
- 4 is a factor to achieve the 95 per cent level of confidence
- r is the predicted or anticipated prevalence (coverage rate) of the indicator
- 1.05 is the factor necessary to raise the sample size by 5 percent for non-response
- f is the shortened symbol for *deff* (design effect)
- $0.075r$ is the margin of error to be tolerated at the 95 per cent level of confidence, defined as 7.5 per cent of r (relative sampling error of r)
- p is the proportion of the total population upon which the indicator, r , is based
- n_h is the average household size.

For the calculation, r was assumed to be 50 percent. The value of *deff* (design effect) was taken as 1.25 based on estimates from previous surveys, p (percentage of children aged 0-4 years in the total population) was taken as 5 percent, and n_h (average household size) was taken as 3.7 households.

The resulting number of households from this exercise was 1260 households which is the sample size needed in each region – thus yielding about 13860 in total. The average cluster size in the Georgia MICS was determined as 30 households, based on a number of considerations, including the budget available, and the time that would be needed per team to complete one cluster. Dividing the total number of households by the number of households per cluster, it was calculated that the selection of a total number of 42 clusters would be needed in each region.

Equal allocation of the total sample size to the four regions was targeted. Therefore, 42 clusters were allocated to each region, with the final sample size calculated at 13860 households (42 clusters * 11 regions * 30 households per cluster). In each region, the clusters (primary sampling units) were distributed to urban and rural domains, proportional to the size of urban and rural populations in that region. The table below shows the allocation of clusters to the sampling domains.

Table SD.1: Allocation of Sample Clusters (Primary Sampling Units) to Sampling Domains

Regions	Estimated Population size	Estimated no. of HHs	Total No. of PSUs	No. of PSUs selected	PSU size (No. of HHs)		
					Avg.	Min	Max
Adjara	377172	87527	1392	42	62.9	20.0	147.0
Guria	146115	39743	573	42	69.4	18.0	132.0
Imereti	699294	201213	2604	42	77.3	11.0	188.0
Kakheti	403100	109632	1603	42	68.4	12.0	160.0
Mtskheta-Mtianeti	126057	34484	524	42	65.8	24.0	143.0
Racha-Lechkhumi & Kvemo Svaneti	51933	17229	215	42	80.1	15.0	186.0
Samegrelo-Zemo Svaneti	466271	119148	1860	42	64.1	13.0	159.0
Samtskhe-Javakheti	209334	51381	848	42	60.6	12.0	123.0
Kvemo Kartli	494661	124031	2036	42	60.9	16.0	162.0
Shida Kartli	316557	83391	1176	42	70.9	16.0	150.0
Tbilisi	1074148	305896	3751	42	81.6	11.0	185.0
Total	4364642	1173675	16582	462	70.8	11.0	188.0

The 2002 census frame was used for the selection of clusters. Census enumeration areas were defined as primary sampling units (PSUs), and were selected from each of the sampling domains by using systematic pps (probability proportional to size) sampling procedures, based on the estimated sizes of the enumeration areas from the 2002 Population Census. The minimum PSU size in Georgia is 11 households and the maximum PSU size is 188 households. The average PSU size is 70.8 households. While constructing the sampling frame the PSUs that are smaller than 30 households will be merged with the neighbouring PSUs to achieve the minimum size of PSU equalling to 30 households.

Although the original sample design for the Georgia MICS 2005 called for approximately 14000 households with an equal number of clusters (42) of households in each of the 11 regions, stratified into urban and rural areas, this sample design was changed to use a more complicated stratification design, with unequal numbers of clusters in each stratum (Table SD.2). The rationale for this was for the selection to more closely follow the population distribution of the population.

Table SD.2. Originally planned distribution of clusters versus implemented, Georgia 2005

Regions	Planned	Implemented
Tbilisi	42	65
Kakheti	42	45
Mtskheta-Mtianeti	42	31
Shida Kartli	42	44
Kvemo Kartli	42	51
Samtskhe-Javakheti	42	39
Racha-Lechkumi	42	20
Imereti	42	51
Guria	42	33
Samegrelo	42	51
Adjara	42	45
Total	462	475

The sample was selected in four stages and in the first two stages, sample design was stratified according to 11 regions, 3 settlement types¹ (Large town, Small town, and Village), and 4 geographic strata² (Valley, Foothills, Mountain, and High mountain). This stratification was applied in all regions, except the city of Tbilisi where the region is stratified according to 10 districts. In total 49 separate strata were identified. The last two stages of the sample design were for the selection of clusters and households.

First stage of sampling: The number of clusters based on sample size calculations was 467 and these were allocated to regions based on the cube root of the number of households in the region. Because the number of clusters for the Racha-Lechkumi region was small (12 clusters), it was decided to increase the number of clusters in that region by 8 for a total of 20 clusters in that region for a total of 475 clusters nationwide.

Second stage of sampling: Within each region, another level of stratification was on a combination of the following: size of settlement (large town, small town, and village) and topography (valley, foothills, mountain, and mountain). The allocation of the number of clusters for a settlement/topography stratum was based on the square root of the number of households in each stratum. Some regions did not have each of the different size settlements or topography. Also, in Tbilisi, the *Rayons* (districts) were used for stratification.

Third stage of sampling: Within each stratum, clusters were selected with probability proportional to population size (PPS).

Fourth stage of sampling: Within each cluster, 30 households were systematically selected, resulting with 14,250 households.

¹ It should be noted that there is only large town settlement type in Tbilisi, whereas in Kakheti, Samtskhe-Javakheti, Guria, Mtskheta-Mtianeti and Racha-Svaneti there is no settlement of a large town. Other regions have all three types of settlement.

² There is valley stratum in all regions; the foothill stratum is in all regions, except Tbilisi. There is no high mountain stratum in the following regions: Kakheti, Tbilisi, Shida Kartli, Guria. It should be noted that there was a high mountain zone in Kakheti-Tusheti, however due to a small number of populations it was not selected as a separated stratum.

In preparing the address list of households, the record books of overall registration interviewers from 2002 Georgian population census were used. In all clusters selected, full address lists of households from these record books were entered into a database. The database included the following information: cluster, household address, and the number of residents (women and men). Following the entry process, among the total records (40,944 households), 14,000 households were selected. There was no updating of household listing prior to the survey. The similar procedure was applied for the 8 additional clusters from Racha-Lechkumi region.

Pre-testing of the questionnaires took place in 12 clusters, 6 in Tbilisi and 6 in Mtskheta–Mtianeti. Contrary to usual practice where pre-testing clusters are not taken from the survey sample, these 12 clusters were part of the full sample. Although it is not recommended to use the pre-test interviews as part of the analysis data files, because the 12 clusters included in the pre-test do not show any significant interviewing problems, and as excluding them would seriously bias and compromise the samples from the two regions it was decided not to exclude these clusters.

Calculation of Sample Weights

The Georgia Multiple Indicator Cluster Survey sample is not self-weighted. The basic weighting of the data has been done using the inverse of the probability of selection of each household.

The major component of the weight is the reciprocal of the sampling fraction employed in selecting the number of sample households in that particular sampling domain:

$$W_h = 1 / f_h$$

The term f_h , the sampling fraction at the h -th stratum, is the product of probabilities of selection at every stage in each sampling domain:

$$f_h = P_{1h} * P_{2h}$$

where P_{ih} is the probability of selection of the sampling unit in the i -th stage for the h -th sampling domain.

Different sampling fractions were used in each stratum. Therefore, sample weights were calculated at the stratum level and were used in the subsequent analyses of the survey data.

Based on the description of the sample plan, the basic probability of selection for a given household in a particular stratum is calculated as

$$P = [(a) (m_i)/\Sigma m_i] \times (30/m_i), \text{ where}$$

P is the probability of selection for each household in the stratum,
 a is the number of clusters selected in the stratum,
 m_i is the size of the i^{th} cluster in terms of number of households and
 30 represents the number of households selected in each cluster.

Note that Σm_i is equal to the total number of households in the stratum, that is, the frame total.

The above expression reduces to $P = [(30a)/\Sigma m_i]$, or 30 times the number of sample clusters divided by the stratum size.

The so-called design weight, W , is the inverse of P , or $\Sigma m_i/30a$.

A second component which has to be taken into account in the calculation of sample weights is the level of non-response for the household and individual interviews. The adjustment for household non-response is equal to the inverse value of:

$$RR = \text{Number of interviewed households} / \text{Number of occupied households listed}$$

After the completion of fieldwork, response rates were calculated for each sampling stratum. These were used to adjust the sample weights calculated for each cluster. Response rates in the Georgia Multiple Indicator Cluster Survey are shown in Table HH.1 in this report.

Similarly, the adjustment for non-response at the individual level (women and under-5 children) is equal to the inverse value of:

$$RR = \text{Completed women's (or under-5's) questionnaires} / \text{Eligible women (or under-5s)}$$

Numbers of eligible women and under-5 children were obtained from the household listing in the Household Questionnaire in households where interviews were completed.

The unadjusted weights for the households were calculated by multiplying the above factors for each enumeration area. These weights were then standardized (or normalized), one purpose of which is to make the sum of the interviewed sample units equal the total sample size at the national level. Normalization is performed by multiplying the aforementioned unadjusted weights by the ratio of the number of completed households to the total unadjusted weighted number of households. A similar standardization procedure was followed in obtaining standardized weights for the women's and under-5's questionnaires. Adjusted (normalized) weights varied between 0.145431 and 3.290674 in the 475 enumeration areas (clusters).

Sample weights were appended to all data sets and analyses were performed by weighting each household, woman or under-5 with these sample weights.