The Gambia MICS 3 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 Gambia 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 eight regions: Banjul, Kanifing, Brikama, Mansakonko, Kerewan, Kuntaur, Janjanbureh and Basse. Regions were identified as the main sampling domains and the sample was selected in two stages. Within each region, at least 14 and at most 99 census enumeration areas were selected with probability proportional to size. After a household listing was carried out within the selected enumeration areas, a systematic sample of 6175 households was drawn. All enumeration areas were accessible and were therefore visited. The sample was stratified by region and is not self-weighting. For reporting national level results, sample weights are used.

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

Sample Size and Sample Allocation

The target sample size for the Gambia MICS was calculated as 6175 households. For the calculation of the sample size, the key indicator used was the proportion of children under five years of age reported ill during the last 2 weeks who received increased fluids and continued feeding during the MICS2 survey, 2000. The following formula was used to estimate the required sample size for these indicators:

$$n = \frac{4\big((r)(1-r)(f)(1.06)\big)}{(0.06r)^2(p)(n_h)}$$

Where:

- n approximately is the required sample size, expressed as number of households, for the KEY indicator.

- 4 is the factor to achieve 95 per cent level of confidence.

- r = 0.23 is the anticipated level (coverage) of the key indicator – proportion of children under five years of age reported ill during the last 2 weeks who received increased fluids and continued feeding during the MICS2 survey, 2000.

- 1.06 is the factor to raise sample size by 6 per cent for 94 per cent response rate for children under five.
- f = 2.08 is the shortened symbol for design effect, *deff*,
- 0.06r is the margin of error to be tolerated, defined as 6 per cent of r (6 per cent thus represents the relative sampling error of r),
- p, is the proportion of the smallest group in the total population. Children less than one year or children 12-23 months are among the smallest group of the study population. However, since indicators for these groups are either very low or very high a fairly larger group, ie., children under five was considered, which gives p = 14 per cent.
- n_h = 9 is the average household size.

Formula 1 above gives about 6253 households. However, 19 households per EA, was the sample take that would give number of households (6175) nearest to 6253 households. Hence, the actual sample size chosen was 6175 households.
The average cluster size in the Gambia MICS was determined as 19 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 325 clusters would be needed for the entire country.

The clusters or EAs were allocated to the eight regions in proportion to their population size. The table below shows the allocation of clusters to the sampling domains.

Table SD1: Sample allocations MICS3, 2005, provisional census population and households by LGA, 2003

LGA	Census number of households, 2003	Census population, 2003	Census EAs, 2003	Sampled EAs, 2005	Households in EAs selected, 2003	(Sample size) Households to be selected for interviews, 2005
Banjul	6903	35061	92	14	1313	266
Kanifing	49227	322735	634	99	8201	1881
Brikama	45219	389594	724	89	5728	1691
Mansakonko	8469	72167	155	19	1244	361
Kerewan	18298	172835	322	40	2534	760
Kuntaur	7140	78491	124	14	961	266
Janjanbureh	10138	107212	179	22	1274	418
Basse	12638	182586	247	28	1512	532
Total	158032	1360681	2477	325	22767	6175

Sampling Frame and Selection of Clusters

The soft copy of the 2003 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 2003 Population Census. The first stage of sampling was thus completed by selecting the required number of enumeration areas from each of the 8 regions by urban and rural areas separately.

The standard clusters were cumulated along the EAs. The cumulative total T_i for the i-th EA is $T_{i-1} + X_i$, where i = 1, 2, 3,.....N(N = 1453 EAs for rural areas); and X_i is the number of standard clusters in the i-th EA. One can define a range or interval for each EA as follows (T_{i-1} to T_i). T_{i-1} is the lower limit of the range and T_i is the upper limit of the range.

The range defined, associates each EA with a range of numbers which is proportional to the size of the EA. Any selection of EAs that make use of the range can be described as PPS sampling, size being standard clusters in each EA.

In implementing PPS systematic sampling, two separate datasets were used - one corresponds to the urban sampling frame and the other to the rural sampling frame.
Using the urban sampling frame T_N was 995.28 standard clusters for 1024 EAs. With a sample size, n = 2945 households for the urban areas, in 155 EAs, the sampling interval, k, becomes 6.4 and the random start, r, which is 6 was randomly selected from 1 to 6.
By using an SPSS programming syntax, 155 urban EAs were selected by PPS systematic procedure.

The i-th EA, was selected if ($T_{i-1} < r + jk \leq T_i$), where j = 1, 2, ...,n-1, n = 155 EAs, k = T_N/n. Thus, the probability of selecting the i-th EA, p_i, is X_i/k.

By using the rural dataset, the above process was repeated with a cumulative total of 1481.69 rural standard clusters, a sample size of 3230 households in 170 EAs, a sampling interval of 8.7 and a random start of 3.

Listing Activities

Since the sample frame (the 2003 Population Census) was not up to date, household lists in all selected enumeration areas were updated prior to the selection of households. For this purpose, listing teams were formed, who visited each enumeration area, and listed the occupied households. Compound and Household Listing Forms were completed for this purpose.

Selection of Households

Lists of households were prepared by the listing teams in the field for each enumeration area. The households were then sequentially numbered from 1 to n (the total number of households in each enumeration area) by the field supervisor in the field, where selection of 19 households in each enumeration area was carried out using circular systematic selection procedures.

Calculation of Sample Weights

The Gambia Multiple Indicator Cluster Survey sample is not self-weighted. The method of proportional allocation of households to each of the regions, results in different sampling fractions for the 8 regions. For this reason, sample weights were calculated and these were used in the subsequent analyses of the survey data.

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

$f_h = P_{1h} * P_{2h} * P_{3h}$

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

Since the estimated numbers of households per enumeration area prior to the first stage selection (selection of primary sampling units) and the updated number of households per enumeration area were different, individual sampling fractions for households in each enumeration area (cluster) were calculated. The sampling fractions for households in each enumeration area (cluster) therefore included the probability of selection of the enumeration area in that particular sampling domain and the probability of selection of a household in the sample enumeration area (cluster).

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 = Number of interviewed households / Number of occupied households listed

After the completion of fieldwork, response rates were calculated for each sampling domain. These were used to adjust the sample weights calculated for each cluster. Response rates in the Gambia 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-five children) is equal to the inverse value of:

RR = Completed women's (or under-5's) questionnaires / Eligible women (or under-5s)
Numbers of eligible women and under-five 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-five's questionnaires.

Table: MICS III weights

LGA	Residence	hhweight	wmweight	Chweight
1	1	1.159443	1.154301	1.222071
2	1	1.021936	1.046005	1.058057
3	1	.756362	.764721	.767337
3	2	1.063019	1.078038	1.085504
4	1	.918385	.958718	.934663
4	2	1.030991	1.047510	1.056498
5	1	1.000438	.993780	1.017013
5	2	.936575	.933412	.953825
6	1	1.738290	1.781459	1.755386
6	2	1.112196	1.162675	1.149876
7	1	.782617	.774395	.790314
7	2	.903480	.900175	.913660
8	1	1.613439	1.596489	1.629307
8	2	.807293	.831640	.828832

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