THE THIRD SOMALI MULTIPLE INDICATOR CLUSTER SURVEY 2006

SAMPLE AND SURVEY METHODOLOGY

The major features of sample design for the Somalia Multiple Indicator cluster Survey are described below. 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 Somali 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 three zones: North West, North East and Central South, of the country. Urban and rural areas in each of the three zones were defined as the sampling domains.

A multi-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 Somali MICS was calculated as 6000 households. For the calculation of the sample size, the key indicator used was the polio coverage among children aged 12 – 23 months. The following formula was used to estimate the required sample size for these indicators:

$$n \quad = \quad \frac{[\,4\,(r)\,(1\text{-}r)\,(f)\,(1.1)\,]}{[\,(0.12r)^2\,(p)\,(n_h)\,]}$$

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
- *1.1* is the factor necessary to raise the sample size by 10 per cent for non-response
- f is the shortened symbol for *deff* (design effect)
- *0.12r* is the margin of error to be tolerated at the 95 per cent level of confidence, defined as 12 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* (polio coverage) was assumed to be 33 percent. The value of *deff* (design effect) was taken as a default of 1.75, *p* (percentage of children aged 0-4 years in the total population) was taken as 3.6 percent, and n_h (average household size) was taken as 6 members per household.

The resulting number of households from this exercise was 5776 households. The average cluster size in the Somali MICS was determined as 24 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 250 clusters would be needed. Of these 130 clusters would be required in the Central South Zone and 60 clusters in the North West and North East Zone respectively.

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

Zone	Population (2005 UNDP Estimates)				Number of Clusters		
	Total	Urban	Rural		Urban	Rural	Total
North West	1,828,739	773,463	1,055,276		25	35	60
North East	863,077	307,307	555,770		21	39	60
Central South	4,810,858	1,574,833	3,236,025		42	88	130
Total	7,502,674	2,655,603	4,847,071		89	161	250

Sampling Frame and Selection of Clusters

Whilst perfect sampling frames are rarely encountered in practice anywhere, Somalia presents a particular challenge. Without a published census for more than 40 years and without a central statistical authority collecting data on a regular basis, Somalia lacks the basics tenets that a conventional cluster sampling approach is built on. However, in recent years, UN agencies combined have made progress towards compiling population estimates and collating settlement data. From drawing on these sources and engaging with local partners it was possible for UNICEF to create a sample design which made efforts to achieve a nationally representative sample giving Somalia the opportunity to partake in the MICS3.

The design follows a 4 stage-sample approach. The first stage is the selection of the districts in each of the 18 regions of the country selected using probability proportional to size (pps)[1]. The second stage is the selection of the secondary sampling units which

---

[1] It should be noted that the decision to distribute an equal number of clusters to the two northern zones of the country was to avoid political disputes over population figures. From thereon however the pps method was adhered to.

are defined as permanent and temporary settlements. The third stage is the selection of the cluster(s) within the settlement and the fourth stage is the selection of the households to be interviewed.

Once the districts had been selected great efforts went into compiling a complete list of permanent and temporary settlements within these districts. The main source was the WHO immunisation campaign data, this data was later backed up by the UNDP settlement survey for at least two out of the three zones. Other sources also contributed such as FAO data on water points which could act as proxy for surrounding nomadic areas and temporary settlements. Finally lists were shown to the NGO partners implementing the survey and UNICEF staff on the ground for additional contributions to recent movement of internally displaced persons and nomads. The settlement lists were then sorted into urban and non urban. The first two stages of sampling were thus completed by selecting the required number of clusters from each of the 3 zones by urban and rural areas separately.

Mapping and Listing Activities

For settlements over the estimated size of 150 households some form of segmentation through sketch mapping was necessary. For several district capitals it was possible to use maps from UN Habitat to assist the personnel deployed in sketch mapping. However for most of the larger non-urban settlements there were no maps available. The most important aspect of the sketch mapping was to divide the settlements into roughly equal sizes by estimating the number of households and to clearly delineate the segments using identifiable boundaries.

Once sketch maps were prepared survey coordinators were then in a position to randomly select the cluster(s) where household would be selected. It must be added at this point that finding people trained in cartographic techniques is rare in Somalia. Thus the quality of the maps varied significantly across the country and resources and time also did not allow for a full household count.

Selection of Households

For the final stage of sampling, the Somali MICS had no other option than to use the method used in MICS2 of the Expanded Program for Immunization (EPI) random walk method; the expense of household/dwelling listing would simply be too considerable.

Whilst the EPI method is quick and approximately self-weighting, it is recognised that this is not a probability sample, and so cannot ensure objectivity of household selection. However with an experienced team it is argued that the EPI method can still produce accurate results. In order to try and avoid the subjectivity involved in selecting households some measures were put in place. For example instead of relying on an arbitrary decision regarding the central point of a cluster, supervisors selected at least

three or four possible starting points and then randomly choose one of them. Moreover only supervisors were able to select and number the households, not interviewers. Significant time was spent training supervisors on how to select households in order to avoid some of the criticisms typically directed towards this method.

For clusters falling in nomadic areas (the temporary settlements) the survey teams were instructed to interview the first 24 households that they came across. Typically nomads do not move in large numbers, therefore in order to ensure representation of nomads in the sample it was necessary to assume a more purposive method of sampling for this group.

Calculation of Sample Weights

The Somalia Multiple Indicator Cluster Survey sample is not self-weighted. Essentially, by allocating equal numbers of households to each of the zones in the North, different sampling fractions were used in each zone since the size of the zones varied. For this reason, sample weights were calculated at the regional level 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 f_h, the sampling fraction at the *h-th* stratum, is the product of probabilities of selection at 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.

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 Somali 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 = Completed women's (or under-5's) questionnaires / 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 0.3 and 2.5 in the 250 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.