

SURVEY DESIGN FOR THE 2005 TAJIKISTAN MULTIPLE INDICATOR CLUSTER BASED SURVEY)

Oleg Benes,
June 30, 2005

I. SAMPLE DESIGN

The purpose of this document is to provide recommendations for the sample design of the 2005 MICS survey, and the corresponding selection procedures to be done prior to the survey implementation.

1. OBJECTIVES OF THE SAMPLE DESIGN

- (1) The 2005 MICS survey is designed to allow reliable estimation of most variables for a variety of health indicators at the various domains of interest based on a representative probability sample.
- (2) The major domains to be distinguished in the tabulation of important characteristics for the eligible women and children up to five population are:
 - Tajikistan as a whole
 - Total urban and total rural areas
 - The 5 main administrative regions of the country (Dushanbe, RRP, Sogd, Hatlon, GBAO)
- (3) The primary objective of the 2005 MICS is to provide estimates with acceptable precision for important population characteristics and maternal and child health indicators as vaccination coverage, diarrhea prevalence, infant mortality for children under five etc..
For instance, with any prevalence estimate for any population group (women or child under five) being equals to 15% or more, *with a denominator based on 1200 cases or more*, the relative error (absolute error divided by prevalence value) for the prevalence estimate is less than 10%.
- (4) The population to be covered by the MICS2005 is defined as all women age 15-49 and children under five years of age in Tajikistan.

2. SAMPLE SIZE

For sample size calculation there was used the formula recommended for MICS2005, that emphasizes *relative margin of error** instead of % of absolute error¹. Calculation was based on estimates for one region-based domain. Due stratification included 5 main administrative regions of the country additionally to urban/rural domains, the overall sample size was obtained by increasing the domain-based estimate by a factor of 5.

$$n = [4 (r) (1 - r) (deff) (1.1)] / [(.13r)^2(p)(ave-size)]$$

¹ www.childinfo.org/mics3/workshop/presentations/sampling

where

- n is the required sample size, expressed as number of households, for the KEY indicator
- 4 is factor to achieve 95 percent level of confidence,
- r is anticipated prevalence (coverage) rate for key indicator ($=0.65$),
- 1.1 is factor to raise sample size by 10 percent for potential nonresponse,
- $deff$ is shortened symbol for design effect ($=1.5$),
- $0.13r$ is margin of error to be tolerated at region-base domain, defined as 13 percent of r (13 percent thus represents the relative sampling error of r),
- p is proportion of total population that one birth cohort comprises ($=0.026$) and
- $ave-size$ is average household size ($=5.85$)².

Based on the estimated prevalence of the measles vaccination coverage of 65% for children < 1 year of age, it is expected that the relative error for the national estimates is less than 5% and for urban/rural estimates – less than 6%. For region-based estimates a higher relative error will be accepted in order to achieve a reasonable compromise between the need for such estimates and budgetary constraints.

3. SAMPLE FRAME

The last census was conducted in Tajikistan by the State Statistical Committee in 2000 and covered all five major administrative areas of the Country (Dushanbe, RRP, Sogd, Hatlon, GBAO). Each of such division, for last census purposes, was further subdivided in convenient areas called “census section”, “instructor’s sector” and “enumeration sector”(ESs). Each ES is either totally urban or rural. The list of ESs has census information on population information for each ES, and the ESs list are grouped by administrative unit. The size of the rural ES was 342 persons and the size of the urban ES was 378 persons. There were established 17923 ES in Tajikistan. The available demarcated cartographic material for each ES in localities with more than 5000 population made last census material as the most appropriate sample frame for the MICS2005.

4. STRATIFICATION

In the census frame, ESs are grouped by location in the major five administrative areas. Each ES is classified as urban or rural. Therefore such stratification by administrative regions and urban and rural areas is also reflected in the MICS sample. The following is the list of strata used for the sample selection:

Sampling strata	
1. Dushanbe	Urban
2. RRP	Urban
3. RRP	Rural
4. Sogdyiskaya oblast	Urban
5. Sogdyiskaya oblast	Rural
6. Hatlonskaya oblast	Urban
7. Hatlonskaya oblast	Rural
8. GBAO	Urban
9. GBAO	Rural

2 Basic methodological and organizational issues of the 2000 population census in the Republic of Tajikistan, State Statistics Committee, Dushanbe 2005

5. CLUSTER SIZE

The size of the primary sampling unit (cluster) was defined based on estimated interview time for one household to be equal to 60 minutes. Considering the 8 hours working day, one interviewer is supposed to complete interview in 6 households. One team of four interviewers is expected to complete interview in 24 households, allowing also some time for movement and call back visits.

6. SAMPLE ALLOCATION

The sample for the MICS2005 is designed to be nationally representative, stratified and selected into two stages. Each of the five administrative regions of the country was stratified into two groups: 1) Urban areas and 2) Rural areas. As Dushanbe municipality does not include any rural area, the stratification scheme resulted in 9 strata. The total sample for the MICS 2005 survey is to select 6960 selected households, and based on the level of non-response found in the Tajikistan MICS2000 approximately 11,000 women age 15-49 and 6300 children under five will be interviewed.

Table 1. Expected ratios of completed interviews of Women and Children under five in selected households by Urban/Rural areas

Domain	Census 2000		MICS2000					
	Estimated # of HH	Distribution of HH	Selected No. of HHs*	Sample HH distribution	Completed eligible women	Completed eligible Kids	Completed Women / selected HH	Completed Kids / selected HH
Tajikistan: Urban	359807	0.34	984	0.265	1446	738	1.47	0.75
Tajikistan: Rural	687213	0.66	2736	0.735	4760	2797	1.74	1.02
Total	1047020	1	3720	1	6206	3535	1.67	0.95

* Of the 3720 households selected for MICS2000, 3720 were successfully interviewed. This is explained by the use of replacement: for each cluster 28 HH were selected, with the first 24 being approached. If any of the first 24 HH were not available, a replacement HH was taken from the last four of the 28 selected.³

In the following table the assumption was made that the expected ratios of completed interviews of Women and Children under five in selected households in each of the five urban strata and each of the four rural strata will follow the total Urban/Rural patterns. Based on MICS2000 urban/rural response rates and proportional distribution of rural/urban population in each of the five administrative regions, there were estimated response rates for each of the administrative region.

3 The status of women and children: Tajikistan, 2000, Multiple indicator Cluster Survey

Table 2. Expected number of selected households to reach the target of completed interviews by administrative regions

Domain	Census 2000					MICS2000	
	Estimated # of HH			Distribution of HH		Compl. Women / selected HH	Compl. Kids / selected HH
	Total	Urban	Rural	Urban	Rural	W/HH	CH/HH
1. Dushanbe	139,239	139,239		1.00	-	1.47	0.75
2. RRP	204,390	32,407	171,983	0.16	0.84	1.70	0.98
3. Sogdyiskaya oblast	365,136	117,955	247,181	0.32	0.68	1.65	0.93
4. Hatlonskaya oblast	307,322	65,576	241,746	0.21	0.79	1.68	0.96
5. GBAO	30,933	4,630	26,303	0.15	0.85	1.70	0.98
Tajikistan	1,047,020	359,807	687,213	0.34	0.66	1.65	0.93

The primary sampling unit (PSU) – or cluster - for the MICS2005 is defined on basis of ES from the 2000 census frame, as having one (or more) ES per PSU.

In rural places the selection of PSU will be done systematically in one-stage, and carried out independently for each of the four rural strata. In each rural stratum implicit geographical stratification will be introduced by ordering rayons from North to South in a serpentine manner. Within each rayon, ES will be ordered sequentially by djamaat, village and census code.

The selection procedure of PSU in urban areas will be the same as for the rural areas, again independent for each of the five urban strata

The following table shows the distribution of about 6,960 households for selection in each domain.

The number of PSUs were calculated by major domain and includes proportional and disproportional sample allocation.

Table 3. Proportional and disproportional sample allocation

Domain	Proportional allocation (HHs)	Sqrt (HHs)	Distribution Sqrt ()	Sqrt allocation (HHs)	Adjusted sample (HHs)	Expected Completed Women	Expected Completed Kids
1. Dushanbe	920	373.15	0.17	1,194.73	1,190	1749	893
2. RRP	1350	452.10	0.21	1,447.50	1,450	2461	1420
3. Sogdyiskaya oblast	2412	604.26	0.28	1,934.71	1,930	3189	1803
4. Hatlonskaya oblast	2030	554.37	0.26	1,774.95	1,770	2977	1707
5. GBAO	204	175.88	0.08	563.12	560	952	550
Tajikistan	6915	2159.75	1	6915	6900	11327	6372

While the samples for the five largest domains would be sufficiently large for providing reliable estimates, it is not the case for GBAO. For this reason it was necessary to increase the sampling rate for the above mentioned domain relative to other domains. Based from the experience of other demographic and health surveys a minimum sample of 1000 women is required to obtain estimates of fertility rates at an acceptable level of sampling errors.⁴

A suggested final adjusted sample is provided to have at least 1000 eligible children up to 5 completed

in each domain that will yield not less than 200 children of one birth cohort.

The selected households are distributed in 290 clusters in Tajikistan.

The following table shows the distribution of selected households for the MICS2005, and the number of ESs by each of the nine domains in Tajikistan

Table 4. Final recommended sample size

Domain	Expected Completed Kids	Expected Completed Women	Household sample for selection	No. of PSU for selection
Total				
Dushanbe	1,278	2,504	1,704	71
RRP	1,269	2,199	1,296	54
Sogdyiskaya oblast	1,256	2,221	1,344	56
Hatlonskaya oblast	1,273	2,220	1,320	55
GBAO	1,272	2,202	1,296	54
Tajikistan	6,347	11,347	6,960	290
Urban				
Dushanbe	1,278	2,504	1,704	71
RRP	162	317	216	9
Sogdyiskaya oblast	324	635	432	18
Hatlonskaya oblast	216	423	288	12
GBAO	144	282	192	8
Tajikistan	2,124	4,162	2,832	118
Rural				
Dushanbe	-	-	-	-
RRP	1,104	1,879	1,080	45
Sogdyiskaya oblast	932	1,587	912	38
Hatlonskaya oblast	1,055	1,795	1,032	43
GBAO	1,129	1,921	1,104	46
Tajikistan	4,220	7,182	4,128	172

Neither the 290 ESs distribution between domains areas nor the households sample distribution is proportional to the last census distribution. That is due to the disproportional number of ESs and therefore the household sample for the MICS2005 is not a self-weighted household sample.

7. SAMPLE SELECTION

The MICS2005 sample is selected using a stratified two-stage cluster design, following the “**MICS Sampling Option 2**”. Taking into consideration the small size of the ES (less than 100 households), there is no need to create standard segments.

In the first stage, 290 PSU will be selected in each stratum with equal probabilities.

The numbers of clusters in each domain area was calculated dividing its total allocated number of households by the sample take of 24 (number of households for selection per ES).

In each domain area, clusters are selected systematically with probability proportional to size. The selection is done using the following formula:

$$P_{li} = (b m_i / \Sigma m_i)$$

where

b: number of ESs in the MICS2005 in a given domain area,

m_i : measure of size of i-th ES

Σm_i : measure of size for the corresponding domain area

In each selected ES, a household listing operation will be carried out during the next 45 days. The updated list of households obtained will be used as the frame for the second-stage sampling. Households will be selected to achieve a fixed sample take per cluster. However, since the MICS2005 sample is unbalanced among domain areas, it will require a final weighing adjustment procedure to provide estimates at the national domain of study.

In a given domain for the i-th cluster, if (c) is the fixed number of households selected out of the total households (L_i) -found in the 2005 listing process- then the household probability in the selected i-th cluster can be expressed as

$$P_{2i} = (c / L_i)$$

The final households overall probability in the i-th cluster could be calculated as

$$f_i = P_{1i} * P_{2i}$$

and the sampling design weight for the i-th cluster is given as

$$1/f_i = 1 / (P_{1i} * P_{2i})$$

II. UPDATING LISTS OF HOUSEHOLDS

Before the start of the fieldwork, the selected PSUs will be updated (i.e. mapping and household enumeration) so that complete household lists are available for the final selection of households in the sample. The updating will be carried out by 13 enumeration teams for a period of five weeks. Each team is composed of two trained enumerators. Two supervisors should be assigned to oversee the listing/mapping activities. The implementing agency (IA) will organize a 3-day training session for enumeration fieldwork within the 7 days prior to the commencement of fieldwork. Enumeration activities will include taking georeference points with GPS units. The Household Listing Manual of the DHS program will be adapted and utilized during this training.

The IA will provide supplies, office space and transportation arrangements for the enumeration fieldwork. At least 13 vehicles will be needed for five weeks while the teams update selected PSU maps and carry out the listing of households.

III. PREPARATION OF THE QUESTIONNAIRES

Three questionnaires will be used during the MICS2005: 1) a household questionnaire; 2) an individual questionnaire for women age 15-49 years; and 3) an individual questionnaire for children up to five years. These three instruments will be based on the questionnaires developed within the framework of the MICS2005 program, and will be adapted to the specific conditions of Tajikistan in consultation with the steering committee.

MICS questionnaires used in Tajikistan:

Household modules	Women's modules	Under-5 modules
<ul style="list-style-type: none"> • Household Information Panel • Extended Household Listing • Education • Water and Sanitation • <i>Additional Household Characteristics</i> • <i>Insecticide Treated Nets + Source and Cost of ITN</i> • Child Labor • <i>Child discipline</i> • <i>Maternal Mortality</i> • Salt Iodization 	<ul style="list-style-type: none"> • Women's Information Panel • Child Mortality • Tetanus Toxoid • Maternal and Newborn Health + Malaria Prevention • Marriage/Union • Contraception + Unmet Need • <i>Attitudes Toward Domestic Violence</i> • Sexual Behavior • HIV/AIDS • Tuberculosis 	<ul style="list-style-type: none"> • Under-5 Child Information Panel • Birth Registration and Early Learning • <i>Child Development</i> • Vitamin A • Breastfeeding • <i>Care of Illness + Source and Cost of Supply of ORS and Antibiotics</i> • <i>Malaria + Source and Cost of Supply of Antimalarials</i> • Immunization • Anthropometry

* **MICS Core modules** / MICS Additional modules / *MICS Optional modules*

After the definitive questionnaires have been agreed upon, they will be translated into Russian and Tajik. There should be taken into consideration the high proportion of Uzbek population (15%) in the country that may require translation of questionnaires in Uzbek too. These translated questionnaires will be used during training and practice in the field. Translation of documents must be completed before the pretest. In addition to the questionnaires, other technical documents will be needed and produced prior to training: the manual/guidelines/worksheets for interviewers, team leaders/supervisors/editors.

Anthropometric measurements:

All children under age 5 years will be weighed and measured in order to determine their nutritional status (wasting, insufficient weight and stunting).

Tajikistan has conducted with the international support a series of Nutritional surveys during 1999-2004 that revealed an improving over the time nutritional status of children.. All above mentioned surveys covered the four country regions except Dushanbe, thus not being able to produce countrywide estimates.

Taking into consideration that the anthropometry module belongs to the core questionnaires of MICS, Tajikistan should be encouraged to use the MICS opportunity to conduct anthropometric measurements for children under five and obtain national-wide indicators related to nutritional status of children.

Salt test:

In all the households surveyed, salt used for cooking will be tested to determine its iodine content. Results of the test will determine the proportion of women and children living in households using sufficiently iodized salt.

IV. PRE-TEST

The IA will make the necessary arrangements for organizing a pretest of the MICS2005 as soon as the questionnaires have been finalized and translated. The objective of the pretest is to detect problems in the questionnaires and their translations, as well as to evaluate the time necessary for conducting the interviews.

UNICEF will provide external consultancy support to the process. It also will pay attention that selected candidates have good learning capacity in order to allow local capacity building. Participants of the pretest are intended to act afterwards as team leaders/editors and national coordinators.

There will be recruited 28 candidates for the pretest. Questionnaires will be pretested in both Tajik and Russian language. Training for the pretest will last about one week and will include theoretical and classroom practice followed by 1-2 days fieldwork, during which around 250 women and about 140 children will be interviewed: interviews will be conducted in Tajik and in Russian so that eventual problems in translation can be identified. During training, health experts will present basic information on the various areas covered by the survey.

Results from the pretest will be used to modify the survey instruments and field procedures, if necessary.

V. TRAINING OF DATA COLLECTION STAFF

Thirteen teams of field staff will be recruited by the IA for the main survey. Each team will be composed of a fieldwork supervisor (male), a field editor (female) and four female interviewers. All interviewers will receive training to take height and weight measurements. All candidates for the field staff positions will be selected on the basis of their maturity, their skill in communicating with respondents, their level of education, their knowledge of Tagik and Russian languages and availability to work away from home for an extended period. Those participants demonstrating the strongest qualifications will be selected for fieldwork.

The candidates will participate in two weeks of training dealing with all aspects of the survey. A greater number of candidates than needed will be trained so that the best of these may be chosen and so that reserve staff is available in case of need. The training will take place in a venue of sufficient size to accommodate the candidates. The local trainers will be chosen from the candidates participated in pretest. National field work coordinators should play a key role in assisting the training process. UNICEF will provide consultancy support in order to facilitate it if necessary. Presentations on specific topics will be made by personnel from the Ministry of Health and other appropriate institutions.

The training program will include: a detailed description of the content of the questionnaires; a presentation of interview techniques and explanations on the way to fill out the questionnaires. The training will include presentations, practice interviews in the classroom, and practice interviews in the field. Each interviewer will carry out at least ten interviews throughout the training period.

Finally, at least one day will be reserved for team leaders to train them in the manner of observing interviews in the field, editing completed questionnaires and controlling the quality of the height and weight measurement .

After having completed the training session, every field staff member must have an in depth knowledge of their role in the collection of the data in order to achieve maximum effectiveness in the fieldwork.

VI. DATA COLLECTION WORK

Data collection will last about five weeks. Each team will complete interviews in five to six clusters per week. The IA will make the necessary central office space available to survey personnel as well as vehicles and communication equipment for field activities. About thirteen (13) vehicles (sufficiently large to hold 6 people) will be necessary for the collection teams, at least two (2) vehicles for technical supervision from the IA.

There must be close communication between the central office of the IA and field personnel during collection work. Details concerning supervision and communication will be discussed during training and will be included in the interviewer's and supervisor's manuals.

Quality control will be ensured through supervision and follow-up of teams during fieldwork. Team leaders will be responsible for the quality of work of their team: they will hold regular meetings with their team in order to reinforce their training and correct any errors committed during collection. In addition, team leaders must re-interview about 5% of households in order to control data quality. These re-interviews, limited to certain sections of the questionnaire, will be undertaken before leaving the cluster.

In order to strengthen quality control, the IA will designate two (2) coordinators who will be present in the field throughout the period of the survey. They will have to visit the field staff regularly, and will be responsible for quality control of interviews (assist in interviews and verification of questionnaires), as well as for follow up of the anthropometric measurements.

Finally, the IA will produce a set of quality control tables at least once every two weeks throughout data collection in order to verify the quality of the data. In the event that after examining these tables, it appears that a team and/or an interviewer is doing work of particularly mediocre quality, then this team and/or this individual will be dismissed and replaced in the shortest possible time. UNICEF HQ will provide guidance in the programming and design of these quality control tables.

VII. DATA PROCESSING

After editing and correction in the field the questionnaires from all regions will be sent to the IA on a weekly base. Two data entry coordinators will supervise the data processing activities overall and will be in charge of the management and archiving of the questionnaires. Once having arrived at the central office, the questionnaires will be registered and verified against the sample by 1-2 office editors. Entry and verification (double entry) of all questionnaires will be carried out by twelve data entry personnel.

All data processing activities will take place in Dushanbe where the IA will provide computers for the period of data processing. The data processing personnel will attend a portion of the training for male/female interviewers in order to familiarize themselves with the questionnaires and to understand their internal logic.

Data entry, editing and tabulation will be achieved by using CSPro, software developed by the MEASURE DHS project and the Bureau of the Census of the United States. UNICEF HQ will provide entry, editing, and tabulation programs using CSPro and will provide assistance to the IA for installing the programs and the entire data processing system for the MICS. The IA will provide a room sufficiently large to contain 12 data entry stations and furnished with shelves to allow for storage of the questionnaires. This workspace must be secure and electricity must be assured on a regular basis.

Data entry will be done using a program that monitors the range of data and the logic of skips in the questionnaire, as well as internal coherence. Editing of data will include the verification of data ranges and thorough monitoring of the internal consistency of the data. All errors detected during the editing process will be corrected. Once the verification and cleaning of data has been completed, imputed dates will be added to the data file, weighting factors will be calculated and added to the data file and all necessary tables for both the preliminary and final reports will be produced and verified. These tables will be based on the standard tabulation plan of the MICS program, adapted to the specific needs of Tajikistan data users and other stakeholders.