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 Serbia Multiple Indicator Cluster Survey was to produce statistically reliable estimates of most indicators, at national level, for urban and rural areas, and for the six regions of the country: Vojvodina, Belgrade, West, Central, East and South-East Serbia. In order to look more deeply into ethnic disparities and to provide national estimates, a separate sample was designed for Roma living in Roma settlements.

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

Sample Size and Sample Allocation

The Serbia MICS3 sample was created by merging two samples: Serbia without Roma from Roma settlements and Roma living in Roma settlements. The average household size and the percentage of children under five in the total population were the factors which caused slightly different sample design for each of the samples.

Serbia is characterised by a very low fertility rate and a small number of household members. For example, one generation of children born makes up less than 1 percent of the population, and the average number of household members is around 3. Owing to these facts, the modification of the recommended sample plan had to be made, and that was the stratification of households in selected census block units into two categories: households with children and households without children under 5. The allocation of the sample in the category of households with children was significantly bigger than the allocation of the sample in the category of households without children.

In the case of the Roma population, the universe could be defined only for Roma who live in separate settlements. The birth rate and household size among Roma living in Roma settlements is higher than in other population groups, so key determinants were different to those used in calculations for Serbia.

The target sample size for the Serbia MICS was calculated as 7200 households in Serbia excluding Roma settlements and 1900 Roma households from Roma settlements.

For the calculation of the sample size, for both Serbia without Roma and Roma samples, the key indicator used was the percentage of children aged 0-4 years who had had Acute Respiratory infections. The following formula was used to estimate the required sample size for these indicators:

$$n = \frac{[\,4\,(r)\,(1\text{-}r)\,(f)\,(n_r)\,]}{[\,(me)^2\,(r)^2\,(p)\,(n_h)\,]} \qquad (1)$$

where

- *n* is the required sample size, expressed as the 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

- n_r is the factor necessary to raise the sample size by 100(n_r -1) percent for non-response
- *f* is the shortened symbol for *deff* (design effect)
- *me*r* is the margin of error to be tolerated at the 95 percent level of confidence, defined as *me* percent 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.

As far as the sample size for Serbia excluding Roma settlements is concerned, the following levels of parameters were included: *r* (percentage of children aged 0-4 years who had had Acute Respiratory infections) was assumed to be 12 percent. The expected non-response rate n_r, was determined at 15 percent. The value of *deff* (design effect) was taken as 1.5 based on estimates from previous surveys. The maximum relative error allowed (*me*) was 12 percent, *p* (percentage of children aged 0-4 years in the total population) was taken as 4.5 percent and n_h (average household size) was taken as 3.

For the Roma sample: *r* (percentage of children aged 0-4 years who had had Acute Respiratory infections) and the expected non-response rate (n_r) were assumed to be 12 and 15 percent, respectively, like the Serbia sample. Also the predicted value for design effect was the same, 1.5. The relative margin of error wanted *(me)* was 20 percent. The percentage of children 0-4 years in the total population, *p* was taken as 9.5, and n_h (average household size) as 4.7.

<u>The resulting number of households from these exercises was as follows:</u>

In the case of Serbia, excluding Roma settlements, the calculated sample size was 26000 households. Only a sample of that size would provide a significant number of children under 5 for drawing reliable conclusions. Therefore, in order to cut down the number of households in the sample, but not to lose estimation reliability, the stratification of the sample into categories with and without children aged 0-4 years was needed. For calculation of the necessary number of households in each category, the following formula was used:

$$n = (n_s)\ (n_c)\ (p_s) \qquad (2)$$

where

- *n* is the required sample size, expressed as the number of households
- n_s is the expected number of households with, or the number of households without children under 5 in a cluster, depending on what category the calculation is used
- n_c is the number of clusters in the sample, and
- p_s is the probability of selection of the household in each category.

Taking into account that the proportion of children under 5 in the total population, *p* was 4.5 percent, and if the average household size is 3, the estimated number of households with children was 13.5 per 100 households (the average number of households in each cluster). So the n_s was assumed to be 13.5 for the category with children, and 86.5 for the category without children. The probability of selection of a household (p_s) with at least one child out of all households with children was assumed to be 0.67, and the probability of selecting a household without children from all households with children in each cluster was 0.1. Supposing that 400 clusters were about to be selected, the total number of households was calculated at 3600 households with, and 3400 of households without children under 5, which makes a total of 7000 households.

For the Roma sample, the stratification of primary units was not needed. Using formula (1), the calculated sample size was 1800 households.

The average cluster size in the Serbia MICS was determined as 18 households, plus 3 backup households for both Serbia and Roma samples. Back-up households were to be interviewed only if some of the first 18 households were not found. In cases where a household refused to be interviewed, the substitution with a back-up household was not possible. The calculation was 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 400 clusters in Serbia without Roma from Roma settlements and 106 clusters in Roma settlements would be needed in all regions.

Allocation of the total sample size to the six regions was targeted with probability proportional to the regions' size. Therefore, 400 Serbia excluding Roma, plus 106 Roma sample clusters were allocated across the regions, with the final sample size calculated at 9108 households ((400+106) clusters * 18 households per cluster). In each region, the clusters (primary sampling units) were distributed to urban and rural domains, proportional to the size of the 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

Region	Population (Census 2002)			Roma Population (Census 2002)			Number of Clusters Serbia without Roma			Number of Clusters Roma in Roma settlements		
	Total	Urban	Rural	Total	Urban	Rural	Urban	Rural	Total	Urban	Rural	Total
Vojvodina	2031992	1152295	879697	29057	12593	16464	62	45	107	13	13	26
Belgrade	1576124	1281801	294323	19191	16040	3151	67	17	84	17	5	22
West Serbia	835225	322919	512306	6294	1345	4949	18	26	44	1	3	4
Central Serbia	1301656	636412	665244	7320	5446	1874	36	37	73	4	2	6
East Serbia	694905	326326	368579	8452	6924	1528	17	20	37	7	1	8
South - East Serbia	1058099	506143	551956	37879	24415	13464	29	26	55	27	13	40
Total	**7498001**	**4225896**	**3272105**	**108193**	**66763**	**41430**	**229**	**171**	**400**	**69**	**37**	**106**

Sampling Frame and Selection of Clusters

The 2002 Serbian Population Census framework was used for the selection of clusters. Census enumeration areas (app. 100 households) 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 first stage of sampling was thus completed by selecting the required number of enumeration areas from each of the 6 regions by urban and rural areas separately.

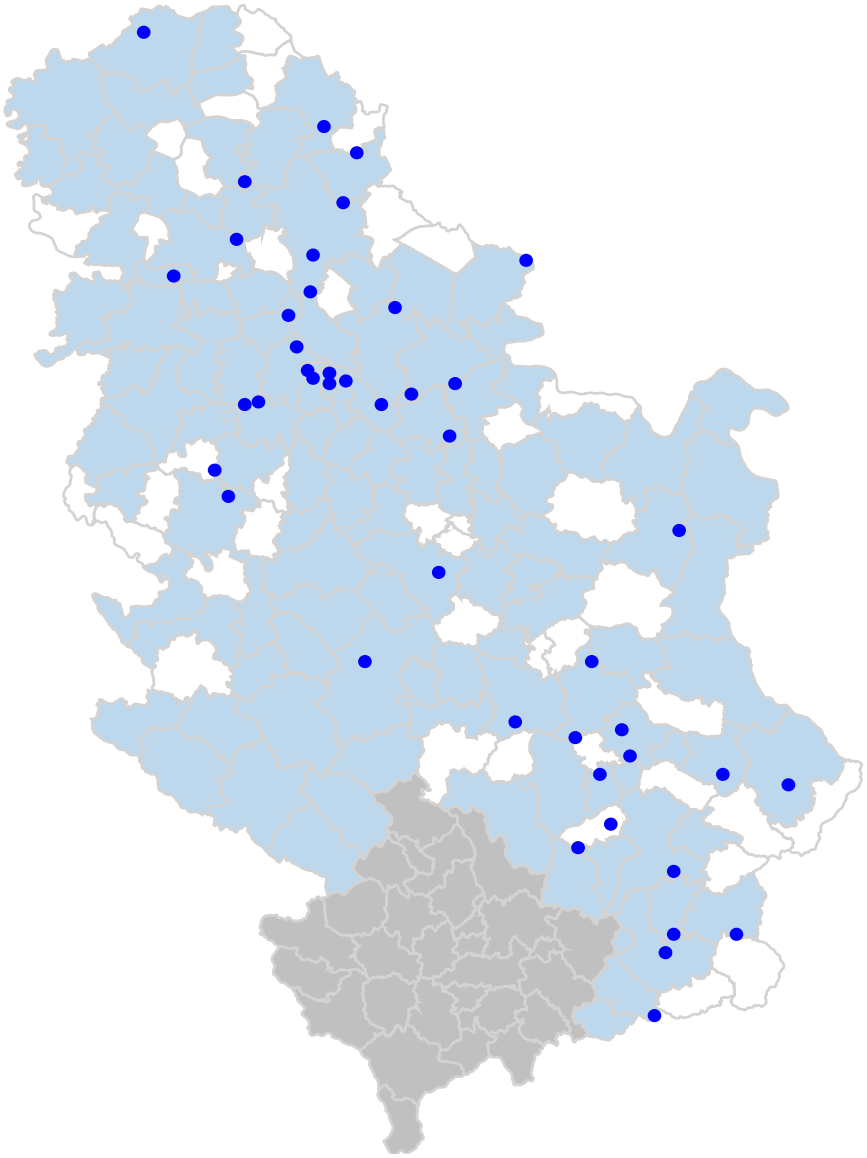
Listing Activities

Since the sample frame (the 2002 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. The Statistical Office of the Republic of Serbia and The Strategic Marketing Research Agency were responsible for updating household lists. The listing exercise was performed by teams which were the direct implementers of the field work during the course of data collection that came later. The whole territory of Serbia was divided into 18 districts according to the regional network of institutions responsible for listing and fieldwork. In each district a team of people was selected – one supervisor for the district and the interviewers (whose number depended on the number of clusters in the region). Criteria for the selection of the interviewers and supervisors were their qualifications, communication skills, experience in fieldwork and knowledge of the region where research was to be conducted. A total of 47 teams were formed. For each team, the list of all households in the selected cluster from the last census was provided. The interviewers' task was to go to the addresses listed and to mark any change that had happened, e.g. the dwelling didn't exist any more, the household had moved away from the dwelling and another household was living there, and to note the number of children under five living in the household. The listing process was performed during September 2005. Besides providing updated information on households, updating household lists made interviewers more acquainted with the field.

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) at the Strategic Marketing Research Agency. Selection of 18 plus 3 back-up households with equal probability in each enumeration area was carried out using the method of random start and equal random walk (simulation of the SRSWoR scheme). In the case of the Serbia without Roma from Roma settlements sample, before the selection of households, updated census block units were put into two categories: households with children and households without children under 5.

Map SD.1: Allocation of Sample Clusters



Municipalities in which the Serbia sample was selected

Roma settlements from the Roma sample

Note: a grey zone represents Kosovo, which is officially (according to UN Resolution 1244) territory under jurisdiction of UMNIK and the international community