

DESIGNING AND SELECTING THE SAMPLE

This technical chapter is intended mainly for sample specialists, but also for the survey coordinator and other technical resource persons. It will help you:

- Determine the sample size
- Assess whether an existing sample can be used, or decide on an appropriate sampling frame for a new sample
- Decide on the design of the new sample
- Gain awareness of issues of sample size, reliability and cost related to subnational indicators and water and sanitation estimates
- Become better informed about weighting, estimation and sampling errors
- Learn about *pps* sampling and implicit stratification
- Find out about sample designs used in two countries

Conducting a Multiple Indicator Cluster Survey in your country will be carried out on a sample basis, as opposed to collecting data for an entire target population. There are different target populations in the survey – households, women aged 15 to 49 years, and children under five and in other age groups. The *respondents*, however, will usually be the mothers or caretakers of children in each household visited.¹ It is important to recognize that MICS4 is a national-level survey and a sample from all households in the country will be chosen, not just those households that have young children or women of childbearing age.

DETERMINING THE SAMPLE SIZE

The size of the sample is perhaps the most important parameter of the sample design, because it affects the precision, cost and duration of the survey more than any other factor. Sample size must be considered both in terms of the available budget for the survey and its precision requirements. The latter must be further considered in terms of the requirements for national

¹ A *household* is defined in the Multiple Indicator Cluster Survey as a group of persons who live and eat together. Any knowledgeable adult (defined, for the purposes of MICS4, as a person 15 years of age or older) is eligible to be the main respondent to the Household Questionnaire. However, in many cases, the respondent will be a mother or a primary caretaker for the practical reason that these individuals are more likely to be at home at the time of interview.

versus subnational estimates. Moreover, the overall sample size cannot be considered independently of the number of sample areas – primary sampling units (PSUs) – and the size of the ultimate clusters. So, while there are mathematical formulas to calculate the sample size, it will be necessary to take account of all these factors in making your final decision.

Getting Help

This chapter of the manual, though fairly detailed, is not intended to make expert sampling statisticians of its readers. Many aspects of sample design will likely require assistance from a specialist, either from within the national statistics office or from outside. These aspects may include calculation of the sample size, construction of frame(s), evaluation of the sample design options, applying the *pps* sampling scheme, computation of the weights, and preparation of the sampling error estimates. In any case, it is strongly recommended that the national statistics office in your country be consulted on the design.

Two general rules of thumb govern the choice on the number of primary sampling units and the cluster sizes: the more PSUs you select the better, as both geographic representation, or spread, and overall reliability will be improved; the smaller the cluster size, the more reliable the estimates will be.

EXAMPLE:

In a national survey, 600 PSUs with cluster sizes of 10 households each will yield a more reliable survey result than 400 PSUs with clusters of 15 households each, even though they both contain the same overall sample size of 6,000 households. Moreover, a cluster size of 10 is better than 15, because the survey reliability is improved with the smaller cluster size. In summary, it is better to strive for more rather than fewer PSUs, and smaller rather than larger clusters, provided other factors are the same.

In general, the more PSUs you select, the better your survey will be. However, the number of PSUs in your survey will be affected to a great extent by cost considerations and whether subnational estimates are needed (subnational estimates are discussed later in this chapter). Travel cost is a key factor. If the distances between PSUs are great and the same interviewing teams will be travelling from place to place (as opposed to using resident interviewers in each primary sampling unit), then decreasing the number of PSUs selected will significantly decrease overall survey costs. In contrast, if your survey requirements call for subnational estimates, there will be pressure on selecting more rather than fewer PSUs.

The choice of the cluster size is another parameter that has to be taken into account in

determining sample size. Its effect can be assessed by the so-called sample design effect, or *deff*. The *deff* is a measure that compares the ratios of sampling variance from the actual stratified cluster survey sample (MICS4 in the present case) to a simple random sample² of the same overall sample size. If, for example, the calculated value of the *deff* from the indicator survey were to be 2.0, this would tell you that the survey estimate has twice as much sampling variance as a simple random sample of the same size. The reason that smaller clusters are more reliable than larger ones is that *deff* increases with the number of sample households per cluster.

Several specific examples of choosing the number of PSUs and deciding on the cluster size are given at the end of this section on sample size.

The costs of simple random sampling preclude it from being a feasible option for MICS4 and for household surveys in general, which is why *cluster* sampling is used. The factors that contribute to sample design effects are stratification, the cluster size and the cluster homogeneity – the degree to which two persons (or households) in the cluster have the same characteristic. For example, the increased likelihood of two children living in close proximity both having received a given vaccination, compared to two children living at random locations in the population, is an example of cluster homogeneity.

Stratification generally decreases sampling variance, while the within-cluster homogeneity measure and the cluster size increase it. Hence, an objective in your sample design is to choose your cluster size so as to balance homogeneity, for which a smaller size is statistically more efficient, with cost, for which a larger cluster size is usually better.

To calculate the sample size for the survey, the design effect must be taken into account in the calculation formula. There are two problems, however. First, while the value of *deff* can be easily calculated *after the survey*, it is not often known prior to the survey, unless previous surveys have been conducted on the same variables. Second, the value of *deff* is different for every indicator and, in fact, every target group, because the cluster homogeneity varies by characteristic. It is not practical, of course, to conduct a survey with *different* sample sizes for each characteristic based on their variable *deffs*, even if we knew what they were.

The *deffs* will not generally be known for indicators prior to the survey, but it is expected that they will be quite small for many indicators, that is, those based on *rare subclasses* (for example, children aged 12 to 23 months).³ If there has been a previous household survey that collected similar data to that of the MICS, and used a very similar sample design, you may be able to use the *deffs* from this previous survey to assess the likely design effects for MICS4. Few household surveys calculate design effects, but the Demographic and Health Surveys (DHS) project is one

² A type of probability sampling in which n sample units are selected with equal probability from a population of N units, usually without replacement and by using a table of random numbers.

³ The mathematical expression for *deff* is a function of the *product* of the cluster homogeneity and the cluster size. Even if the cluster size is large in terms of total households, it will be small in terms of this particular target population (1-year-old children), and so the *deff* is likely to be small also.

good source of such information. It should be pointed out, however, that the design effect appearing in the sampling error tables of the DHS reports is the *deft*, which is the ratio of the standard error from the actual sample to that from a simple random sample of the same size; it is therefore the square root of the corresponding *deff*. When using the DHS report for the country as a source of information, it is important to remember to square the *deft* value appearing in the DHS table to obtain the corresponding value of *deff*.

In the calculation formula and table for sample size in the following sections, we have assumed the design effect to be 1.5 (which may be somewhat high for some indicators and will therefore provide a conservative or higher estimate of the sample size). In choosing a conservative *deff*, we want to ensure that the sample size will be big enough to measure all the main indicators. Nevertheless, a rule of thumb in choosing the cluster size and, by implication, the number of clusters *is to make sure that the cluster size is as small as can be efficiently accommodated in the field*, taking into account related considerations such as the number of PSUs and field costs (discussed above) and achieving interviewer workloads of a convenient size. If a larger cluster size (for example, more than 20 sample households per cluster) is used, it may be necessary to assume a larger *deff* for estimating the sample size.

CALCULATING THE SAMPLE SIZE

To calculate the sample size, using the appropriate mathematical formula, requires that several factors be specified and values for others be assumed or taken from previous or similar surveys. These factors are:

- The level of precision required, measured by the sampling error,
- The level of confidence desired
- The estimated (or known) proportion of the population in the specified target group
- The predicted or estimated rate, or prevalence, for the specified indicator
- The sample *deff*
- The average household size
- An adjustment for potential loss of sample households due to non-response.

The calculation of sample size is complicated by the fact that some of these factors vary by indicator. We have already mentioned that *deffs* differ by the type of characteristic. Even the margin of error wanted is not likely to be the same for every indicator (and, in practice, it cannot be). This implies that different sample sizes would be needed for different indicators to achieve the necessary precision. Obviously, we must settle upon *one* sample size for the survey.

The sample size calculation applies only to person-variables, even though it is expressed in terms of the number of households you must visit in order to interview individuals. This is because most of the important indicators for the MICS4 assessment are person-based. Household variables should not be used in the sample-size calculations because they require a different formula, as well as very different design effect (*deff*) values, in some cases as high as 10 or more,

for example for source of drinking water and type of sanitation.

The calculating formula is given by

$$n = \frac{[4 (r) (1-r) (deff) (1.1)]}{[(0.12r)^2 (p) (\tilde{n})]}$$

where

- **n** is the required sample size, expressed as number of households, for the KEY indicator (see following section on determining the key indicator)
- **4** is a factor to achieve the 95 percent level of confidence
- **r** is the predicted or estimated value of the indicator
- **1.1** is the factor necessary to raise the sample size by 10 percent for non-response
- **deff** is the design effect
- **0.12r** is the margin of error to be tolerated at the 95 percent level of confidence, defined as 12 percent of r (in this case the relative standard error of r would be about 6 percent)
- **p** is the proportion of the total population upon which the indicator, r, is based (that is, the proportion of the target subpopulation group), and
- **\tilde{n}** is the average household size (that is, average number of persons per household).⁴

If the sample size for the survey is calculated using a key indicator based on the smallest target group, in terms of its proportion of the total population, then the precision for survey estimates of most of the other main indicators will be better.

Observant users of an earlier Manual used for MICS2 will notice that this formula differs in that a target *relative margin of error* (value of 0.12) has been used instead of a target *margin of error* (**e** in the previous edition, with a value of either .05 or .03 for indicators with high and low proportion estimates, respectively); in this case, **0.12r = e**. In the MICS2 manual, a reliable estimate for the survey was defined differently, depending upon whether it represents high or low coverage. For the indicator estimates, it was recommended that the margin of error, or precision, be set at 5 percentage points for proportion estimates (for example, immunizations) that are comparatively high, greater than 25 percent, and at 3 percentage points for proportion estimates that are low, 25 percent or less. While a plausible rationale was given for presenting two margins of error defined in this way, users were nevertheless left with the often-difficult choice of which one to use for their survey, especially if the calculated sample sizes were widely

⁴ An alternative but equivalent formula is also provided further along in the text and also in the sample size template that is in the documentation on the MICS website.

different. By using the relative margin of error, this problem is avoided altogether because it scales the margin of error irrespective of whether an indicator with a high or low proportion estimate is chosen as the key one for sample size determination. Note, however, that the sample size is nevertheless larger for indicators with low proportion estimates, which is why it is important to choose carefully which indicator is *most key* for the survey (see the following section).

DEFINING AND CHOOSING THE KEY INDICATOR TO CALCULATE SAMPLE SIZE

The recommended strategy for calculating the sample size is to choose an important indicator that will yield the largest sample size. This will mean first choosing a target population that comprises a small proportion of the total population (p in the above formula). This is generally a *target population of a single-year age group*.⁵ In MICS4, this is children aged 12 to 23 months, which in many MICS4 countries comprise about 2.5 percent of the total population. *We recommend using 2.5 percent unless you have better estimates available for your country.* If, for example, your figure is higher (3.5, 4 or 5 percent) your sample sizes will be considerably less than those in Table 3, so it is very important to use your best estimate of p for this target population. Second, the particular indicator must be chosen for this same target population. We will label it the *key indicator* (but only for purposes of calculating the sample size).

Table 1
Criteria for Choosing Key Indicator Based on Level of Proportion Estimate

Low proportion estimate is *undesirable*

- Use of improved water sources or sanitation facilities
- School attendance
- Antenatal care and institutional deliveries
- Breastfeeding rates
- Immunization rates

Low proportion estimate is *desirable*

- Mortality rates
- Underweight, stunting or wasting prevalence
- Child labour

In making your decision on the key indicator, you will need to choose one with a low proportion estimate. Some indicators with a low proportion estimate should be excluded from consideration,

⁵ In making your choice of the lowest-percentage population groups, it is strongly recommended that you exclude from consideration the four-month age groupings of children that form the basis for the breastfeeding indicators, because the necessary sample sizes would likely be impractically large.

however. This can be explained by reviewing the indicators in Table 1, where examples are given of indicators for which a low proportion estimate is *undesirable* and the associated goal is focused on *raising* the rate (for example, the immunization rate for DPT, that is, diphtheria, pertussis and tetanus). The second set of indicators in Table 1 provides examples for which the opposite is true – a low proportion estimate is desirable and the goal is to lower it further (an example is stunting prevalence). *It would not make sense to base your sample size on indicators for which low coverage is desired and the proportion estimate is already very low; such indicators should be excluded when choosing the key indicator.*

Table 2 provides suggestions for choosing the target group and key indicator for purposes of calculating the sample size directly or finding the sample size in Table 3. Note that the infant mortality rate (IMR) or the maternal mortality ratio (MMR)⁶ are not mentioned as candidates for the key indicator. This is because the sample sizes that would be necessary to measure these indicators are much too large – in the tens of thousands – and it would be impractical to consider them. This does not necessarily mean that such indicators should not be measured in the survey, but rather that the sample size for the survey should not be based on them. The survey results for these indicators will have larger sampling errors and, hence, wider confidence intervals than the other indicators.

⁶ Regarding sample size for measuring the maternal mortality ratio: A 1997 guide by WHO and UNICEF entitled ‘The Sisterhood Method for Estimating Maternal Mortality’ recommends that if the maternal mortality ratio is 300 (per 100,000 live births), it can be estimated with a sample size of about 4,000 respondents with a margin of error of about 60, utilizing the indirect sisterhood method.

Table 2
Checklist for Target Group and Indicator

To decide on the appropriate target group and indicator that you need to determine your sample size:

1. Select two or three target populations that comprise small percentages of the total population. Normally, these target groups should not be narrower than 1-year age groups, or wider than 5-year age groups. In MICS4, these will typically be children aged 12-23 months, or children under 5 years, which would, in many countries, comprise 2-4 percent and 10-20 percent of the total population, respectively.
2. Review important indicators based on these groups, ignoring indicators that have very low (less than 5 percent) or very high (more than 50 percent) prevalence; do not consider an indicator with prevalence above 50 percent. Begin calculations with your smaller group. If the indicators based on this group are high coverage, perform calculations for the wider age group, for which indicators might have lower coverage.
3. In general, pick an indicator that has a fairly low coverage rate, on the order of 15 to 20 percent, for target populations comprising 10 to 15 percent of the population. For target populations comprising less than 5 percent of the population, pick an indicator that has slightly higher coverage, above 20 percent, but below 50 percent.
4. Do not pick from the desirably low coverage indicators an indicator that is already acceptably low.

In making your choice, you must also consider the *relative* importance of the various indicators in your country. For example, you would not want to use an indicator that requires a very large sample size if that indicator is of comparatively small importance in your country.

USING THE SAMPLE SIZE TABLE

Table 3 shows sample sizes already calculated on the basis of MICS4 requirements, plus certain assumptions. You may use the table values, if they fit your situation, to get your sample size. Otherwise, you or your sampling specialist may calculate the sample size directly, using the formula given below.

If the parameters in Table 3 fit the situation in your country, you can find the sample size without having to calculate it using the formula above. In Table 3, the level of confidence for the precision of the estimates is pre-specified at 95 percent. Varying values of the average household size and coverage rate, r , are used – from 4.0 to 6.0 and from 0.25 to 0.40, respectively. The *deff* is assumed to be 1.5 and the precision (margin of error) level is set at 12 percent of r , that is, a 12 percent relative sampling error. The table reflects a 10 percent upward adjustment in sample size to allow for potential non-response in the survey.

It is crucial to note that the table also assumes that the target population for your key indicator comprises 2.5 percent of the total population. If it is a different value, you cannot use the table to find the required sample size. In general, the table cannot be used when any of the assumed values for the parameters of the formula do not fit your situation. More about what to do in this case is provided later in this section.

Table 3
Sample Size (Households) to Estimate Indicators for Smallest Target Population
(with Relative Sampling Error of 12 Percent of Estimated Rate at 95 Percent Confidence Level)

Average Household Size (number of persons)	Value of Indicator (r)			
	<i>r</i> = 0.25	<i>r</i> = 0.30	<i>r</i> = 0.35	<i>r</i> = 0.40
4.0	13,750	10,694	8,512	6,875
4.5	12,222	9,506	7,566	6,111
5.0	11,000	8,556	6,810	5,500
5.5	10,000	7,778	6,191	5,000
6.0	9,167	7,130	5,675	4,583

Use this table when your

- Target population is 2.5 percent of the total population; this is generally children aged 12-23 months
- Sample design effect, *deff*, is assumed to be 1.5 and non-response is expected to be 10 percent
- Relative sampling error is set at 12 percent of estimate of coverage rate, *r*

If all the assumptions for the parameter values of the formula pertain in your country, then one of the sample sizes in Table 3 should fit your situation. In some cases, the parameters may apply, but the coverage rate you choose has to be interpolated. For example, if your coverage rate is between 30 percent and 35 percent you may figure the sample size by interpolating a value between the third and fourth columns of the table. To illustrate: in the last row for a 32.5 percent coverage rate, your sample size would be halfway between 7,130 and 5,675, or about 6,403 households.

A stepwise illustration of the use of Table 3 can be given as follows:

- First, satisfy yourself that all the parameter values used in Table 3 apply in your situation.
- Next, from Table 2, pick the indicator that has the lowest coverage, excluding any indicator that is already acceptably low. Suppose it is measles immunization at 35 percent.
- Next, find the average household size in Table 3 that is closest to that in your country (assuming it is within the ranges shown). Suppose it is 5.5 persons.
- Finally, find the value in Table 3 that corresponds to a household size of 5.5 persons and 35 percent coverage rate. That value is 6,191.

The figures should not, however, be taken as exact, but only as approximate sample sizes; remember that several assumptions were made in calculating the sample sizes. It would make sense to round the sample sizes up or down depending upon budget constraints. In this example, you might decide that 6,100 or 6,200 would be appropriate after you consider travel costs between primary sampling units, cluster sizes and interviewer workloads.

USING THE SAMPLE SIZE FORMULA⁷

What happens to the sample size calculations if all the assumptions of the parameter values pertain, except that the proportion of children aged 12-23 months in your country is not 2.5 percent, but instead is closer to 2.0 percent? In that case, simply multiply all of the numbers in Table 3 by 2.5/2, or 1.25, to come up with the sample sizes. This is important since the sample sizes are significantly larger, an increase of 25 percent.

There are situations, however, when it is best to ignore Table 3 and calculate the sample size directly using the formula below. The formula must be used when any of the parameter values in your country differ from the assumptions used in Table 3. Table 4 outlines the conditions under which the formula should be used. If we assume, as a rule of thumb, that a *deff* value of 1.5 should be used in the formula for determining the sample size when your cluster size is 20 households, then you would need to adjust the *deff* value lower or higher if you use a cluster of, say, size 10 or size 30.

We have already discussed why the sample size must be larger if *p* is smaller than 0.025. Unless the illustrated example above conforms to the situation in your country, the formula below should be used for calculating the sample size. To repeat, the formula below must be used if any of the other parameter values fit the criteria outlined in Table 4.

⁷ An Excel template for calculating sample size can be found at www.childinfo.org

Table 4
Checklist for Use of Sample Size Formula

The formula to determine your sample size is given by:

$$n = \frac{[4 (r) (1-r) (deff) (1.1)]}{[(0.12r)^2 (p) (\bar{n})]}$$

Use it if any (one or more) of the following applies in your country:

- The proportion of 1-year-old children (p) is other than 0.025
- The average household size (\bar{n}) is less than 4.0 persons or greater than 6.0
- The coverage rate of your key indicator (r) is under 25 percent
- The sample design effect (*deff*) for your key indicator is different from 1.5, according to accepted estimates from other surveys in your country
- Your anticipated non-response rate is more or less than 10 percent (adjust the factor 1.1 accordingly).

Do not change the confidence level in the formula, but keep its value at 4.

Using the formula is quite easy, since it is basic arithmetic once the parameter values are inserted. For example, for $r = 0.25$, $deff = 1.6$, non-response adjustment = 1.05, $p = 0.035$ and $\bar{n} = 6$, we have

$$n = \frac{[4 (0.25) (1-0.25) (1.6) (1.05)]}{[(0.12 \times 0.25)^2 (0.035) (6)]} = \frac{1.26}{0.000189} = 6,667$$

In previous MICS surveys, the typical sample size has ranged between 4,000 and 8,000 households. That range is a plausible target for you to strive for when doing your calculations on sample size, considering both reliability requirements and budgetary constraints. As we mentioned already, MICS4 will produce estimates of many indicators, each of which will have its own level of precision. It is therefore useful to examine the approximate levels of reliability – standard errors and confidence limits – on your indicators for a particular sample size.

Table 5 illustrates those levels of reliability for a sample of 6,000 households, which may be regarded as a typical sample size for producing comparatively reliable estimates for most indicators of interest in MICS4.

Table 5
Expected Reliability Measures (Standard Error and Confidence Interval) for Sample of
6,000 Households under Various Demographic Alternatives

Average household size	Size of indicator r	Size of subpopulation p	Number of sample persons in Subpopulation	Number of persons with indicator	Standard error	Confidence interval (95% level)	
						Lower	Upper
4	0.10	.025	540	54	.016	.068	.132
		.05	1,080	108	.011	.078	.122
		.125	2,700	270	.007	.086	.114
		.20	4,320	432	.006	.089	.111
	0.20	.025	540	108	.021	.158	.242
		.05	1,080	216	.015	.170	.230
		.125	2,700	540	.009	.181	.219
		.20	4,320	864	.007	.185	.215
	0.30	.025	540	162	.024	.252	.348
		.05	1,080	324	.017	.266	.334
		.125	2,700	810	.011	.278	.322
		.20	4,320	1,296	.009	.283	.317
	0.50	.025	540	270	.026	.447	.553
		.05	1,080	540	.019	.463	.537
		.125	2,700	1,350	.012	.476	.524
		.20	4,320	2,160	.009	.481	.519
5	0.10	.025	675	68	.014	.072	.128
		.05	1,350	135	.010	.080	.120
		.125	3,375	338	.006	.087	.113
		.20	5,400	540	.005	.090	.110
	0.20	.025	675	135	.019	.162	.238
		.05	1,350	270	.013	.173	.227
		.125	3,375	675	.008	.183	.217
		.20	5,400	1,080	.007	.187	.213
	0.30	.025	675	203	.022	.257	.343
		.05	1,350	405	.015	.269	.331
		.125	3,375	1,013	.010	.281	.319
		.20	5,400	1,620	.008	.285	.315
	0.50	.025	675	338	.024	.453	.547
		.05	1,350	675	.017	.467	.533
		.125	3,375	1,688	.011	.479	.521
		.20	5,400	2,700	.008	.483	.517
6	0.10	.025	810	81	.013	.074	.126
		.05	1,620	162	.009	.082	.118
		.125	4,050	405	.006	.088	.112
		.20	6,480	648	.005	.091	.109
	0.20	.025	810	162	.017	.166	.234
		.05	1,620	324	.012	.176	.224
		.125	4,050	810	.008	.185	.215
		.20	6,480	1,296	.006	.188	.212
	0.30	.025	810	243	.020	.261	.339
		.05	1,620	486	.014	.272	.328
		.125	4,050	1,215	.009	.282	.318
		.20	6,480	1,944	.007	.286	.314
	0.50	.025	810	405	.022	.457	.543
		.05	1,620	810	.015	.470	.530
		.125	4,050	2,025	.010	.481	.519
		.20	6,480	3,240	.008	.485	.515

Column 4 in Table 5 shows the expected number of persons to be interviewed in a 6,000-household sample, assuming a non-response rate of 10 percent. For example, in a country where the average household size is four persons, the number of sample persons in a subpopulation comprising 2.5 percent of the total population (say, children aged 12-23 months) would be about 540, instead of 600, after allowing for non-response. Of those, column 5 shows the expected number of sample persons that would have the characteristic, r . The expected number of sample persons is 54 if r is 10 percent, 108 if r is 20 percent, 162 if r is 30 percent and 270 if r is 50 percent.

Observe that the expected standard error varies considerably depending on the size of the subpopulation and the size of the indicator. An important reliability measure for evaluating your results is the confidence interval, the last column in Table 5. The confidence interval, or CI, shows the range around which your estimate can be expected to vary from the true value in the population, taking account of the standard error. It is computed by adding and subtracting twice the standard error (for a 95 percent level of confidence) to the indicator estimate. Looking at the very last row of Table 5, the confidence interval shown is $|0.485 - 0.515|$ for an indicator estimated at 0.50. This means that if you estimate the indicator coverage to be 50 percent, then you can be confident with 95 percent assurance that the true value of the indicator in the population is between 48.5 percent and 51.5 percent.

USING AN ALTERNATIVE SAMPLE SIZE FORMULA

It was mentioned in an earlier footnote there is an alternative formula that may be used instead; that formula also appears in the sample size template of the MICS documentation. The formula differs from the one above but provides the same results. The alternate formula makes use of the identity that the average number of target (base) population persons per household (labelled as k_h) is the same as the quantity, $(p \text{ times } \tilde{n})$, that is, the target population proportion times the average household size. Use of k_h means one less parameter to have to deal with. The formula is thus

$$n = \frac{[4 (r) (1-r) (deff) (1.1)]}{[(0.12r)^2 (k_h)]}$$

The value of k_h may be determined in either of two ways, depending upon what figures are most readily available from your latest census:

- a. Multiply the average household size by the proportion that the target population (upon which r is based) is of the total population. For example if the target group is children under 5, it is typically about 15 percent of the total population; if the average household size is 5, then k_h is .15 times 5, or 0.75. Both the target group proportion and the average household size can be found, or derived, from the most recent census figures.

- b. Divide the number of persons in the target group, according to the most recent census, by the number of households, again according to the census. For example, suppose the target group is again children under 5 and there are 3,550,200 of them. Suppose, further, there are 4,733,600 households in the country whose population is 23,668,200. Then k_h equals $3550200/4733600$, or 0.75.

It was mentioned in an earlier footnote that an Excel spreadsheet is available in the MICS documentation that may be used as a template to help you calculate sample sizes for candidate indicators in order to choose the best sample size for your MICS. It is only a matter of substituting the appropriate values of the variables for your country in the template table and the sample size will be calculated automatically. Hence, you can avoid altogether the need to carry out the arithmetic manually and it is a convenient way of examining sample sizes quickly for several indicators. As noted earlier, *deff* values can be obtained from previous MICS or DHS reports in many countries for some of the important indicators.⁸

DECIDING ON THE NUMBER OF PRIMARY SAMPLING UNITS AND CLUSTER SIZES – ILLUSTRATIONS

At the beginning of the section on sample size, we discussed how the number of PSUs and the size of the clusters (number of households selected in each PSU) play a role in sample size. We emphasized that sampling reliability is improved with more PSUs and smaller cluster sizes. We conclude the section with three examples using different scenarios to illustrate the interrelationship of sample size, number of PSUs and cluster size.

EXAMPLE 1:

Target group:	Children aged 12 to 23 months
Percent of population:	2.6 percent
Key indicator:	DPT immunization coverage
Value of indicator:	40 percent
<i>Deff</i> :	No information
Average household size:	6

Under this scenario, use Table 3 because the coverage rate of the key indicator and the household size can be found in the table. The target population, comprising 2.6 percent, is also very close to the 3 percent figure that Table 3 is based upon. With no information on the design effect, it is assumed to have a value of 1.5, and the non-response adjustment factor is assumed to be 1.1, corresponding to an expected 10 percent non-response rate. The sample size for an average household size of 6.0 persons for a 40 percent coverage rate is then found to be 4,583 households.

Suppose that your country is relatively large in geographic size and, further, that there are a large number of provinces, say, 15. You and your sampling staff have concluded, therefore, that you need to have a minimum of 300 PSUs in order to achieve good

⁸ As already mentioned, the DHS tables, however, show the *deft* value, instead of *deff*, and that number must be squared for use in the MICS template. For example, a DHS *deft* of 1.43 should be multiplied times itself to give a *deff* of 2.04 to use in the template.

geographic spread and sufficient representation in each province. Moreover, you have decided that the budget for the survey would support that number of PSUs. The cluster size would then be calculated as 4,583 divided by 300, or about 15-16 households.

Instead of targeting 300 PSUs as your number, you and the survey and sampling staff may have decided, alternatively, that you wanted clusters of a certain size, say, 10, in order to meet operational requirements such as interviewer workload distribution. In this case, you would divide 4,583 by 10 to give you the number of PSUs – about 458. You would then review this number in terms of cost and other considerations, and either accept it or adjust your cluster size. You might conclude that 425 is the maximum number of PSUs you can field because of travel costs, in which case you would have to adjust the cluster size to 11 (that is, $4,583/425$).

EXAMPLE 2:

Target group:	Children aged 12 to 23 months
Percent of population:	2.5 percent
Key indicator:	Polio immunization coverage
Value of indicator:	26 percent
<i>Deff</i> :	No information
Average household size:	6

Under this scenario, you may still use Table 3 since, except for the coverage rate of the key indicator, all the parameters of the table pertain, given that we can again assume the design effect is 1.5 and the non-response adjustment factor is 1.1. For the estimated indicator value, r , we can use the column for 25 percent since the estimated value of 26 percent is so close. The sample size for an average household size of 6.0 persons is found in the table to be 9,167 households.

Because of cost considerations and field workloads, suppose the survey team decides it wants cluster sizes of 30 households, if possible. Here, dividing 9,167 by 30 gives 306 PSUs, and you may decide that this is an acceptable number for the fieldwork. If, on the other hand, you decided that you would like to have about 400 PSUs for geographic spread and also to have enough PSUs to enable subnational estimates to be made for five regions, you would divide 9,167 by 400, which gives you 23 as your cluster size. Recall that the smaller the cluster size the more reliable the indicator estimates will be (for all indicators, not just the key indicator). You may decide, therefore, to use the 400-PSU design with its average cluster size of 23 households, bearing in mind that it will be more expensive than 306 PSUs due to travel costs.

It should also be pointed out that increasing the cluster size will also increase the *deff*, so this should be taken into account in determining the sample size. For example, the *deff* of 1.5 may be based on a cluster size of 20 households. Therefore if the cluster size is increased to 30, a slightly higher value of *deff* (for example, 1.7) could be used in the formula for calculating the sample size.

EXAMPLE 3:

Target group:	Children aged 0 to 11 months
Percent of population:	3.5 percent
Key indicator:	Adequately fed infants
Value of indicator:	24 percent
<i>Deff</i> :	1.4 (from a previous survey)
Average household size:	4

Expected non-response rate 10 percent

Under this scenario, you would have to calculate the sample size by using the formula provided in this section, since several of the parameters differ from those used or assumed in Table 3. They include the values for p , d_{eff} and the non-response adjustment factor, the latter of which is based on an expected non-response rate of 5 percent instead of 10 percent, judging from similar surveys in your country. The formula yields a figure of 10,303 households.

Suppose that the survey staff has concluded that the survey can handle a maximum of 300 PSUs because of cost considerations. In this case, you would set 300 as fixed and figure the cluster size by dividing 10,303 by 300, which gives 34 households as the cluster size. Here you would have to evaluate whether a cluster size that big will give sufficiently reliable estimates for indicators other than the key ones.⁹ If we assume that the maximum cluster size should not be greater than 30 households, the number of PSUs that would be needed for 10,303 households is 343. Thus, the choice would have to be made whether to accept the lower reliability of a 300-PSU design or the higher cost of a 343-PSU design.

DETERMINING WHAT SAMPLE TO USE

Once you have decided on the sample size and made initial determinations about the number of PSUs, the next task is to decide what sample to use for the survey. Designing, selecting and implementing a proper probability sample from beginning to end is a time-consuming and expensive process (probability sampling is discussed in the next section). For MICS4, there is the need to produce the indicator estimates in a comparatively short time frame, and you may not have sufficient time to design a new sample for the survey. Hence, there are two major steps to be followed in determining what sample to use for your survey:

Step 1: Determine if an existing sample can be used.

Step 2: If no suitable existing sample can be found, develop a sample specific to MICS4.

In this section we discuss step 1. If there is a suitable existing sample for MICS4, you need not review the optional sample designs presented for step 2, which is discussed in the next section. It is nevertheless useful to review the next section in order to assure yourself that the existing sample you plan to use is a proper probability sample with a reasonably current sampling frame.

⁹ While the design effect is very low for the key indicator of this example, and hence the reliability of that estimate would be expected to meet the precision requirements set, other indicators that have a much higher intra-cluster correlation than that for children under age one would be expected to have considerably higher sampling errors with a cluster size over 30 compared to, say, 20 or 25.

USE OF AN EXISTING SAMPLE – OPTION 1

Fortunately, most countries have well-developed survey programmes through their national statistical offices or health ministries. It may be possible in your country, therefore, to use an already existing sample, one that has been designed for other purposes. *This is the recommended option for your survey if the existing sample is a valid probability sample and is available.* The existing sample must be evaluated to see if it meets the requirements of probability sampling (which is discussed in a subsequent section). Many countries use master samples developed from master sampling frames from which subsets are selected for use in particular surveys. That possibility should obviously be explored for your MICS.

There are various ways in which an existing sample may be used:

- Attaching MICS4 questionnaire modules to the questionnaires to be used in another survey
- Using the sample, or a subset, from a previous survey
- Using the household listings in the sample enumeration areas (or clusters) of another survey
- Using the enumeration areas or clusters from a previous survey with a fresh listing of households.

Of these choices, there are advantages and limitations to each. Timing considerations are also a key factor. For example, the first choice is only an option if another survey is going to be carried out within the prescribed time frame for the MICS. This choice – attaching the questionnaire modules to another survey, sometimes called ‘piggy-backing’ because the data for both surveys are collected simultaneously – has obvious appeal *since the sampling will have already been done, thus saving the sampling costs as well as sharing the costs of the fieldwork for MICS4.* A major limitation, however, can be the burden it places on the respondent, since MICS4 questionnaires are quite long and the parent survey may have its own lengthy questionnaire. These aspects must be carefully evaluated and discussed with the host survey sponsors and management team.

The second choice, using the sample from a previous survey, also has the advantage that the sample design is already in place, again saving sampling costs. If the sample size for the previous survey was too large, it would be a simple matter for the sampling statistician to sub-sample the original sample to bring the size into compliance with MICS4 requirements. By contrast, however, if the sample size is too small, expanding it is more problematic. There is also the limitation of revisiting the same households from the previous survey, again because of potential problems this could pose in terms of respondent burden and/or conditioning. Finally, the previous survey must be very recent for this to be a viable choice.

The third choice, using the household listings in sample enumeration areas from a previous survey as a frame for selecting the MICS4 sample, has a dual advantage: (1) the first-stage units are already sampled and (2) household listings are already available. Therefore, again, most of the sampling

operations and costs will already have been achieved. An advantage is that different households would be selected for MICS4, thus eliminating the problems of respondent burden, fatigue or conditioning. A limitation is that the household listings would be out of date if the previous survey is more than a year or two old, in which case this choice would not be viable. In fact, when the household listings are out of date, then the fourth choice above can be considered. This choice requires making a fresh listing of households in the sample enumeration areas before sample selection. While this has the limitation of having to carry out a new household listing operation, with its associated expense, the advantage is that first-stage units would have already been selected and the sample plan itself is basically in place without further design work.

Table 6
Option 1 – Existing Sample

Pros

- Saves time and cost
- Likely to be properly designed with probability methods
- Adjustments to fit MICS4 can be simple

Cons

- Requires updating if old
- Respondents may be overburdened
- Indicator questionnaire may be too long if 'piggy-backed'
- Adjustments to fit MICS4 can be complex

Each of these points should be carefully evaluated and a determination made about the feasibility of implementing the necessary modifications before you decide to use an existing sample.

An existing sample that may be an excellent candidate is the Demographic and Health Survey (DHS).¹⁰ Many countries have conducted these surveys recently and others plan to in the coming months.¹¹ The measurement objectives of DHS are quite similar to the MICS. For that reason, the sample design that is used in DHS is likely to be perfectly appropriate for your use.

Under what circumstances is it appropriate to use the DHS sample? You must evaluate its availability, timeliness and suitability in terms of your requirements. *Either a recent but pre-2007 DHS sample could be used to field the MICS, or an upcoming DHS could be used with the MICS4 as a supplement.* The DHS will undoubtedly be designed as a probability sample.

¹⁰ Sampling matters are described in *Demographic and Health Surveys: Sampling Manual, Basic Documentation* – 8. Calverton, Maryland: Macro International Inc., 1987.

¹¹ It should be noted, however, that conducting MICS4 is not recommended if a DHS has been done since 2007, or will be done in 2009 or early 2010.

Therefore, you need only evaluate whether (1) its sample size is large enough for MICS and (2) the number of PSUs and cluster sizes are within the ranges that are discussed in this manual. Finally, it would require agreement and cooperation with the DHS sponsoring or implementing agency in your country, noting the constraints mentioned above about overburdening respondents.

Another survey that many countries have implemented and whose sample may be appropriate for your use is a labour force survey (LFS). While the measurement objectives of labour force surveys are quite different from the objectives of MICS4, labour force surveys are frequently designed in a very similar fashion to Multiple Indicator Cluster Surveys in terms of stratification, sample size and other sampling criteria. In some countries the LFS may have a larger number of sample PSUs than needed for the MICS, but a subsample of these PSUs can be selected as mentioned previously. Also, the MICS generally uses a larger cluster size than the LFS, but a new sample of households can be selected from the same or new listing, as described in the different options.

DEVELOPING A SAMPLING FRAME FOR A NEW SAMPLE

When an existing sample cannot be used, it will be necessary to use and/or develop a sampling frame of households from which to select a new sample for MICS4. The frame should be constructed in accordance with the tenets of probability sampling.

PROPER PROBABILITY SAMPLE DESIGN AND SAMPLING FRAME

Design of an appropriate probability sample for the survey is just as important as development of the various questionnaire modules in terms of producing results that will be valid and, as far as possible, free from bias. There are a number of ways you can design a probability sample, and each country will undoubtedly have its own conditions and data needs that dictate the particular sample plan it adopts. There are certain features that should be observed by all countries, however, to meet the requirements of a scientific probability sample:

- Use accepted probability sampling methods at every stage of sample selection
- Ensure that the sampling frame is complete and does not have gaps in covering the target population
- Select a nationally representative sample
- Ensure that the field implementation is faithful to the sample design
- Ensure that the sample size is sufficient to achieve reliability requirements.

In addition to these five requirements, there are other features of sample design that you are strongly recommended to adopt, although each may be modified in certain ways depending upon individual country situations and needs. They include:

- Simple, as opposed to complex, sampling procedures
- Use of the most recent population census as the sampling frame
- A self-weighting sample within domain, if possible.

Scientifically grounded probability sampling methods for surveys have been practised in most countries of the world for decades. If a sample is not accurately drawn from the whole population of interest by using well-known probability techniques, the survey estimates will be biased. Moreover, the magnitude of these biases will be unknown. It is crucial to ensure that the sampling methodology employs probability selection techniques at every stage of the selection process.

To avoid sample bias, you should use *probability sampling* to select the respondents. Sample bias depends on the selection *techniques*, not the sample size. Increasing the sample size will not eliminate sample bias if the selection techniques are wrong.

Probability sampling is a means of ensuring that all individuals in the target population¹² have a known chance of being selected into the sample. Further, that chance must be non-zero and calculable. A sure sign of not having a probability sample is when the sampling statistician cannot calculate the selection probabilities of the sample plan being used.

Examples of sampling methods that are not based on probability techniques are judgement samples, purposive samples and quota samples. The random walk method of selecting children is a quota sample procedure. It is important that you not use such procedures for MICS4.

The best way to control sampling bias is to insist on strict probability sampling. There are other biases, non-sampling in origin, including non-response, erroneous response, and interviewer errors, but these will occur in varying degrees anyway, no matter what kind of sampling methods are used. Appropriate steps must be taken to control these non-sampling biases as well, including such measures as pre-testing, careful interviewer training and quality control of fieldwork.

In probability samples, every person in the target population has a chance of being selected, the selection chance is non-zero and is calculable mathematically, and probability techniques are used in every stage of selection.

A second required feature of sample design for MICS4 is that the sample should be national in scope and coverage. This is necessary because the indicator estimates must reflect the situation of the nation as a whole. It is important to include, to the extent practicable, difficult-to-

¹² MICS4 has different target populations depending upon the indicator. Examples include children aged 0 to 11 months, 12 to 23 months, under 5 years, children under 5 years with diarrhoea, women aged 15 to 49 years, and the total population.

enumerate groups to ensure complete national coverage. Such groups might be nomads, homeless or transient persons, or those living in refugee camps, military quarters, as well as settlements in isolated areas that are difficult to access. It is quite likely that children in particular, living in such situations, have different health conditions from those found in more stable or traditional living environments, and excluding them would result in biased indicator estimates. Given that MICS4 is a household survey, it may exclude homeless persons and those living in institutions or group quarters, which may represent 1 percent or less of the total population. However, this should be made very clear in describing the target population in the survey documentation.

One of the crucial ways in which the sample can be truly national in scope and therefore consistent with proper probability sampling is by ensuring that the frame used covers the entire population of the country. The sampling frame is discussed in more detail below.

For probability sampling to be effective, it is essential that the field implementation of the sample selection plan, including the interviewing procedures, be faithful to the design. There have been numerous occasions where lax fieldwork has ruined an otherwise perfectly acceptable sample design. The field supervisors must make certain that the sample selection procedures are followed strictly.

A crucial feature of valid probability sampling is the specification of precision requirements in order to calculate the sample size. This topic was discussed in the previous section on determining sample size. We have recommended that the precision for the key indicator be set at a relative margin of error of 12 percent at the 95 percent level of confidence, and those are the criteria under which the calculation formula for sample size is based. If your key indicator is, for example, one with a 20 percent prevalence rate, then the 12 percent relative error translates into a margin of error of 2.4 percentage points, and the confidence interval on your survey estimate of 20 percent would be $| 17.6 - 22.4 |$.

Your sample should be designed as simply as possible. It is well known that the more complex the sample plan is, the more likely it is that its implementation will go wrong. This can be especially troublesome at the field level if complicated sampling procedures have to be carried out. Moreover, the operational objective to produce the survey results in a timely manner may not be met.

A sample plan is said to be *self-weighting* when every sample member of the target population is selected with the same overall probability. The overall probability is the product of the probabilities at each of the stages of selection. A self-weighting sample is desirable because various estimates can be prepared, for example, percentage distributions, from the sample figures without weighting, or inflating, them. In keeping with the desire for simplicity in sample design, it is better to have a self-weighting design than a more complicated, non-self-weighting one. Still, *self-weighting should not be considered a strict criterion*, because weighting the sample results to prepare the estimates can be easily handled by today's computers. Moreover, there are

some situations where the sample design cannot be self-weighting, for example when reliable estimates are required for different geographic domains such as regions or provinces. Weighting is discussed in more detail further along in this chapter.

EXAMPLE:

Suppose that in your country you will need separate urban and rural indicator estimates, and suppose further that you want the estimates to be equally reliable. This would necessitate selecting a sample of equal size in the urban and rural sectors. Unless the urban and rural populations are equal, the sampling rates in each would be different. Hence, the overall national sample would require weighting for correct results and, therefore, the survey sample would not be self-weighting

CENSUS SAMPLING FRAME AND WHEN UPDATING BECOMES NECESSARY

It is strongly recommended that the most recent population census be used as the basis for the sampling frame, updated if necessary. Nearly all countries of the world now have a recent population census, that is, one conducted within the last 10 years. The frame is essentially the set of data and materials from which the survey sample is selected. A good sampling frame is one that is complete, accurate and up to date, and while no frame is 100 percent perfect, the population census comes closest in most countries. *The prime use of the census for our survey is to provide a complete list of enumeration areas with measures of size, such as population or household counts, for selection of the first-stage sampling units.* Maps are usually part of the census of population in most countries, and these might include sketch maps for the enumeration areas. The maps are a useful resource because the selected enumeration areas will likely have to be updated with a new listing in terms of the *current* households residing therein, especially if the census is more than a year or two old. It is recommended, however, that if the census is no more than two years old, the household listings may be used at the second or third stage of sampling without re-listing unless there is suspicion about bias in the original listings.

Many countries are in the process of preparing for the year 2010 round of censuses. Very few of those are likely to have been completed before your MICS must be designed. It is likely you may have to depend on the 2000 round of censuses. This brings us to the very important issue of whether the census frame will need to be updated for MICS. *It is recommended, in general, that updating not be undertaken if the census frame was created in 2006 or later, with one exception.* In countries where there have been dramatic shifts in population since 2006, especially in highly urbanized areas that have expanded in specific zones due to massive new construction of residential units, an updating operation should be undertaken in such zones. You may decide, however, that this is not necessary if your population census is so recent that it precedes your survey by 12 months or less.

If the census frame in your country was prepared before 2006, updating is recommended.

The reason for updating should be apparent. It is necessary to ensure that coverage of the total population is as accurate and complete as possible. The recommended steps for updating the census frame are the same under either scenario, that is, whether in large-scale urban developments since 2006 or for general updating of an old census frame prepared prior to 2006. The difference is in the scope and scale of the updating operation. Updating an old, pre-2006, census frame is considerably more demanding and expensive than updating the more recent frames. In either case, however, the operation must take place for the entire sampling frame and not just those enumeration areas – PSUs – that happen to be selected in the sample. In fact, the information gathered in updating is used to select the sample.

It is important to be aware that updating a frame is a major statistical operation. If updating becomes necessary, it cannot be ignored in your costing algorithm when you are preparing your budget. Moreover, you are strongly urged to engage the services of your national statistical office when updating is deemed necessary. The specific steps are as follows:

1. Identify the zones, especially in large cities, where there has been massive residential construction since the census was conducted, irrespective of whether your population census is pre- or post-2006.
2. Identify new zones, such as squatter communities that have become highly populated since the census. These may include zones that were ‘empty’ or very sparsely populated at the time of the census.
3. Ignore old, stable residential zones where little change occurs over time.
4. Match the zones identified in steps 1 and 2 with their census enumeration areas, taking into account overlapping boundaries.
5. In the affected enumeration areas, conduct a canvass of each one and make a quick count of dwelling units. Note that quick counting only entails making a rough count of dwellings without actually enumerating the occupants. The quick count should not entail knocking on doors at all, except in the case of multi-unit buildings where it is not obvious from the street how many flats or apartments there are.

Use the new quick count of dwellings¹³ to replace the original count of households in the census frame. This is its new ‘measure of size’, a count necessary to establish the probabilities of selecting the sample enumeration areas.

It is obvious that updating the frame prior to sample selection is not a trivial operation, but rather time-consuming and costly. That is one reason why it is recommended to use an existing sample, whenever possible.

Finally, it should be noted that some countries have no recent census nor acceptable cartography

¹³ It is recognized that the number of dwelling units may not equal the number of households. However, it is only important to obtain a rough estimate in order to establish the measure of size. For example, if 120 dwellings were ‘quick counted’ in an enumeration area that was selected for the sample, and it was later found that 132 households occupied these dwellings, the validity and reliability of the sample results would not be seriously affected.

from which a frame can be constructed. In such situations updating the census frame becomes irrelevant. What to do in such situations in order to construct a sampling frame for MICS becomes very country-specific and pertinent guidelines cannot be provided. They depend on what materials, geographical structure and hierarchy and other resources exist. In such cases it is imperative that an expert sampling practitioner with much experience in developing countries be engaged to assist in designing the frame and the sample plan for MICS.

USING A NEW SAMPLE FOR MICS AND DECIDING ON ITS DESIGN

When a suitable existing sample is not available for use in MICS4, either for a stand-alone survey or a supplement to another survey, a new sample will have to be designed and selected, starting with the preparation of the sampling frame (discussed above).

In this section, we outline the main properties that the design of the MICS4 sample should possess. Two options are presented below, preceded by a summary of general features.

In the most general terms, your survey sample should be a probability sample in all stages of selection, national in coverage, and designed in as simple a way as possible so that its field implementation can be easily and faithfully carried out with minimum opportunity for deviation from the design. In keeping with the aim of simplicity, both stratification and the number of stages of selection should be minimal. Regarding stratification: its prime purpose is to increase the precision of the survey estimates and to permit oversampling for subnational areas when those areas are of particular interest. A type of stratification that is simple to implement and highly efficient when national level estimates are the main focus is implicit stratification. This is a form of geographic stratification that, when used together with systematic *pps*¹⁴ sampling (see illustrations near the end of this chapter), automatically distributes the sample proportionately into each of the nation's administrative subdivisions, as well as the urban and rural sectors. Implicit stratification is carried out by geographically ordering the sampling frame of EAs in a serpentine fashion, separately by urban and rural, before applying systematic *pps* sampling.

Further, the design should be a two-stage or three-stage sample. The first-stage, or primary sampling units (PSUs), should be defined, if possible, as census enumeration areas, and they should be selected with *pps*. The enumeration area is recommended because the PSU should be an area around which fieldwork can be conveniently organized; it should be small enough for mapping, segmentation, or listing of households, but large enough to be easily identifiable in the field.

The second stage would be the selection of segments (clusters), and the third stage the selection of the particular households within each segment that are to be interviewed in the survey. These households could be selected in a variety of ways – through sub-sampling from an existing list of

¹⁴ This is probability proportionate to size (*pps*) and it refers to the technique of selecting sample areas proportional to their population sizes. Thus, an area containing 600 persons would be twice as likely to be selected as one containing 300 persons at the first stage of selection.

households in each segment or a newly created listing. Often, the primary sampling units are themselves the clusters to be listed; in which case the second stage mentioned above is eliminated. The selection of households directly from the PSU listing then becomes the second stage in a two-stage design.

There is, of course, room for flexibility in this design, depending on country conditions and needs. The design is likely to vary a good deal from one country to another with respect to the number of sample PSUs, the number of segments or clusters per PSU, and the number of households per segment, and, hence, the overall sample size.

As a very general rule of thumb:

- The number of PSUs should be in the range of 250 to 350 at the national level, unless there are subnational geographic domains of analysis.
- The cluster sizes (that is, the number of households to interview in each segment) should be in the range of 10 to 35, depending upon which of two options described below is followed.
- The overall sample size should be in the range of 4,000 to 14,000 households.

A country may decide, for its own purposes, that it wants indicator estimates for a few regions in addition to the national level. In that case, its sample design would undoubtedly include a different stratification scheme and a greater number of PSUs, so as to ensure adequate geographic representation of the sample areas in each region. In addition, the sample size for the survey would have to be increased substantially in order to provide reliable estimates for regions, or for other subnational domains (discussed in more detail later in this chapter).

STANDARD SEGMENT DESIGN – OPTION 2

It was mentioned above that the Demographic and Health Surveys project might provide a suitable existing sample for use in MICS4 (recall that we referred to use of an existing sample as Option 1). The standard DHS sample design is, in fact, a good model for MICS4, if you decide that a new sample has to be designed. The DHS sample model has also been used in other health-related survey programmes such as the PAPCHILD surveys in the Arab countries.¹⁵

¹⁵ See *The Arab Maternal and Child Health Survey, Basic Documentation 5: Sampling Manual*. Cairo: League of Arab States, 1990.

The DHS and PAPCHILD sample models are based on the so-called *standard segment design*, which has the benefits of probability sampling methodology, simplicity and close relevance to the MICS4 objectives, both substantive and statistical. The sampling manuals for DHS and PAPCHILD note that most countries have convenient area sampling frames in the form of enumeration areas of the most recent population census. Sketch maps are normally available for the enumeration areas, as are counts of population and/or households. The census enumeration areas are usually fairly uniform in size. In many countries, there are no satisfactory lists of living quarters or households, nor is there an adequate address system, especially in many rural areas. Consequently, it is necessary to prepare new listings of households to bring the frame up to date.

To apply the standard segment design to the MICS4, first arrange the census frame of enumeration areas in geographic sequence to achieve implicit stratification. Some enumeration areas are so large that it is not economically feasible to carry out a new listing of all households if they are selected. Instead, it is more efficient to use segments. This is done by assigning each enumeration area a measure of size equal to the desired number of ‘standard segments’ it contains. In the DHS and PAPCHILD sampling manuals, it is recommended that the number of standard segments be defined (and computed) by dividing the census population of the enumeration area by 500 and rounding to the nearest whole number. Note that in cases where you are updating your census frame, the *count of dwellings (multiplied by 5)* you obtained in the last step of the updating operation (described in the preceding section on frames) should be used instead of the census population figure. The multiplication factor of 5 is necessary to approximate the current population count in the updated enumeration areas so that its measure of size is defined the same as those enumeration areas that are not updated. *This size for the standard segment is recommended for MICS4, if you decide to use Option 2.*

The next step is to select sample enumeration areas using probability proportionate to this measure of size. Note that the measure of size is also the number of segments. In many cases, you may find that the average size of an enumeration area is about 500 persons (equivalent to 100 households when the average household size is five); therefore, the typical measure of size will be one.

Segmentation, using the available maps, is the next phase of operation. *When the number of segments in a sample enumeration area is equal to one, no segmentation is necessary*, because the segment and the enumeration area are one and the same. If the number of segments is greater than one, then segmentation will be necessary. This entails subdividing the sampled enumeration area into parts (equal to the number of segments), with each part containing roughly the same number of households. Segmentation may be done as an office operation if the maps are accurate enough. Otherwise, a field visit would be necessary, especially in cases where identifiable internal boundaries within the enumeration area are not clearly delineated (see next subsection below, “How to Map, Segment and List Households,” for more details).

Table 7
Option 2 – Summary of Standard Segment Design

Features

- Three-stage sampling with implicit stratification
- Selection of enumeration areas by *pps*
- Mapping and segmentation in enumeration areas with more than one standard segment
- Selection of one segment at random in each enumeration area
- Listing of households in sample segments
- Systematic selection of sample households in segments

Parameters

- Usually 250 to 400 sample enumeration areas (PSUs)
- Standard segments of 500 population (about 100 households)
- Systematic sample of 10 to 35 households selected from listing in each sample segment (differs from Option 3 below)
- Sample size of usually 4,000 to 14,000 households*

**Note that, in general, we would not recommend the minimum number of PSUs times the minimum cluster size (250 times 10), since 2,500 is most likely too small a sample size to measure the important indicators reliably in most countries.*

After segmentation, one segment is selected at random in each sample enumeration area. In all selected segments, *a new household listing is undertaken*. Again, this will typically be about 100 households. Then, from the listings, using a fixed fraction, choose a systematic sample of households in each sample segment for interview.

Table 8
Option 2 – Standard Segment Design

Pros

- Probability sample
- Minimal mapping and segmentation
- Amount of listing is reduced
- Somewhat more reliable than Option 3 (below)
- Partially updates old sampling frame
- Self-weighting design

Cons

- Listing, though minimal, is necessary in every sample segment
- May give widely variable segment sizes, especially if frame is old and not updated
- If a constant sampling fraction is used to select the households within each segment, the number of sample households will vary by segment, and there is less control over the final sample size.

EXAMPLE:

It might be decided to select one fifth of the newly listed households in each sample segment. Thus, if there are, say, 300 segments, then the number of households selected in each segment would be approximately 20 (though it would vary by PSU) and the overall sample size would be approximately 6,000 households.

The standard segment design is convenient and practical. In a typical country, that is, one where the enumeration area averages about 100 households, very little actual segmentation would have to be done. Moreover, the amount of household listing is also limited.

The sample households under Option 2 are contained within non-compact clusters,¹⁶ and the sample is self-weighting if the same sampling fraction is used for selecting the households within each segment. The number of households selected in each sample PSU will vary somewhat because the PSUs are selected based on their census sizes (except for those that have been updated), which will likely be different from the actual sizes when the new household listing is made.

EXAMPLE:

Suppose the within-segment selection rate is calculated to be 1 in 5 of the listed households. If a segment is selected on the expectation of 98 households based on the census, but the listing shows there are now 112 households, then a one-fifth sample of the households will yield 22 or 23 households (the correct number), instead of the expected 19 or 20. The procedure not only reflects population change correctly, but it also retains the self-weighting nature of the sample. The deviation in the average segment size should not be great, unless an old census frame that has not been updated is used.¹⁷

One issue under both Option 2 and Option 3, discussed next, is how to control the overall sample size when designing a self-weighting sample. This requires selecting the households in sample EAs with a probability that is determined prior to the listing. The differences between the number of households in the frame and the listing in some sample EAs can lead to an overall effective sample size that is at least 10 percent lower or higher than the target size. This will make it more difficult to control the cost and timing of the fieldwork. There is also the issue of variable workloads by EA; there are cases in DHS where they have upper and lower limits on the number of households to be selected, which means that the sample is no longer self-weighting.

¹⁶ A non-compact cluster is one in which the households selected for the sample are spread systematically throughout the entire sample area. A compact cluster is one in which each sample household in a given segment is contiguous to its next-door neighbour. Non-compact clusters give more reliable results than compact clusters, because of their smaller design effects.

¹⁷ There is an alternative procedure when the population is thought to have changed significantly, so that the average segment size might be too variable for efficient field assignments. The number of households selected in each segment may be fixed rather than the fraction of households to be selected, in which case a different sampling interval would have to be calculated and applied in each sample segment. Each segment would then have a different weight and this would have to be accounted for in the preparation of the indicator estimates.

Most countries now seem to prefer a fixed number of sample households per EA and live with slightly variable weights. With the compact cluster in Option 3 (following subsection) it becomes even more difficult to control the overall sample size, given the variability in the size of the compact clusters.

MODIFIED SEGMENT DESIGN – OPTION 3

We have discussed the use of an existing sample as the preferred option for MICS4, whenever a well-designed existing sample is available and relevant. We have also discussed using the DHS and PAPCHILD model sample plan, the standard segment design, as the next best option whenever your country has to design the indicator survey sample from scratch.

Option 3 uses a modification of the standard segment design. The *modified segment design* is similar to the standard segment design, but there are important differences.¹⁸ Rather than creating standard segments with a population size of 500 in each sample enumeration area, the latter is subdivided into a predetermined number of smaller segments. This predetermined number is equal to the number of census households (or the updated dwelling count) in the enumeration area divided by the desired cluster size and rounded to the nearest whole number. Note here we are using households (or dwellings for updated frame areas) rather than population, which was used for Option 2. Hence it is not necessary to multiply the dwelling count in updated areas by 5.

EXAMPLE:

If the desired cluster size is 20 households, and there are 155 households in the enumeration area, then 8 segments would be created.

As with Option 2, enumeration areas are sampled with probability proportionate to the number of segments they contain. Each selected enumeration area is then segmented into the predetermined number of segments using sketch maps together with a quick count of current dwellings. Carefully delineated boundaries must be formulated in the segmentation, and the number of dwellings in each segment should be roughly equal, although it need not be exact. Note that the quick count can, again, be based on dwellings rather than households, just as is done for frame updating (refer to that section for details).

After segmentation, one (and only one) segment is selected at random within each sample enumeration area. *All the households contained within the boundaries of the sample segment are then interviewed for the survey*, the segment thus forming a compact cluster of households.

The other features of the modified segment design are essentially the same as the standard segment design – three-stage sampling, implicit stratification, *pps* selection of enumeration areas.

¹⁸ See a complete description of the modified segment (or cluster) design in: Turner, A., R. Magnani, and M. Shuaib. 1996. 'A Not Quite as Quick but Much Cleaner Alternative to the Expanded Programme on Immunization (EPI) Cluster Survey Design'. *International Journal of Epidemiology* 25(1).

The modified segment methodology has an advantage over the standard segment design in that no household listings need be undertaken, thus eliminating a major survey expense. The quick-count operation and sketch mapping do, however, bear an additional expense, but the cost of the quick count is minimized since it can be done by visual inspection rather than actually knocking on doors to speak to respondents. In addition, the procedure compensates somewhat for using a sampling frame that may be outdated by interviewing all the current households in a sample segment, no matter how many there were at the time of the census.

Table 9
Option 3 – Summary of Modified Segment Design

Features

- Three-stage sampling with implicit stratification
- Predetermination of number of segments by PSU
- Selection of census enumeration areas by *pps*
- Mapping and segmentation in all sample enumeration areas
- Selection of one segment at random in each enumeration area
- Interview of all sample households in selected segment

Parameters

- Usually, 250 to 400 sample enumeration areas (PSUs)
- Compact cluster of 20 to 30 households (minimum size 20)
- Usual sample size of 5,000 to 12,000 households*
- *Segment size* and *cluster size* are synonymous (unlike Option 2)

**Note that the range of sample sizes is different from those for Option 2 in Table 7 because the recommended compact cluster sizes are different.*

A limitation of the modified segment design is that the segments (the clusters) are compact. Therefore, with the same sample size, the sampling reliability for this design will be somewhat less than the standard segment design, where the clusters are non-compact. This could be compensated, however, by sampling more enumeration areas with smaller compact clusters within the enumeration areas. Another limitation is that the segmentation itself requires comparatively small segments to be delineated, which may not be practical in some countries. It can be very problematic in small areas where there are not enough natural boundaries such as roads, lanes, streams, etc. for the segmentation to be accurate or even adequate. *For this reason, it is recommended that the segment size under this option be at least 20 households; and to compensate for the decrease in reliability with the compact segment, it should not be greater than 30 households.* Boundary delineation is extremely important when forming segments, in terms of controlling sampling bias. Another limitation with using compact clusters is that it becomes more difficult to control the overall sample size, especially when there are differences between the measure of size of the EA from the census and the actual household count during the enumeration. Therefore some countries prefer a two-stage design with a fixed number of households selected at the second stage, although it is sometimes necessary to segment very

large EAs. Although the weights will vary somewhat by cluster, the computer software can easily handle the weighting of the data for the tabulations and analysis. The *pps* selection of the EAs at the first stage will limit the variability in the weights.

SHORTCUT DESIGNS – NOT RECOMMENDED

In the first round of MICS in 1995, considerable attention was devoted to a method called ‘random walk’, which is used in the Expanded Programme of Immunization. The chief objection to using this method for MICS4 is that household selection is not based on probability sampling methods, but rather on a procedure that effectively gives a quota sample.

Since MICS4 has large sample sizes, the random walk method is inappropriate. It is sometimes argued that the small-scale, Expanded Programme of Immunization surveys, with their correspondingly small sample sizes, are dominated more by

sampling variance than by bias, thus justifying somewhat the use of the random walk method. For MICS4, however, that same argument leads to the reverse conclusion – that bias is of greater concern than sampling variance, due to the much greater sample sizes, and so stricter probability methodologies should be used at each stage of selection. As an example of the potential bias, the random walk method will miss the less accessible households in a sample village, which may also have less access to health services and other facilities, so the resulting bias may overestimate immunization rates that are being measured.

Shortcut procedures – such as random walk – which depart from probability designs are not recommended for MICS4 and should not be used.

Table 10
Summary Checklist for Sample Size and Design

- Determine target group that is a small percentage of total population
- Determine estimated rate for same target group
- Choose sample size from Table 3 if your country situation fits the table assumptions and conditions; OR ELSE, calculate sample size with the formula provided in this chapter
- Decide on cluster size (number of households selected per segment), usually in a range of 10 to 35 households
- Divide sample size by cluster size to get the number of PSUs (sample areas)
- Review your choices of n, cluster size, and number of PSUs, IN ORDER TO choose among Options 1, 2 or 3 for sample design

HOW TO MAP, SEGMENT AND LIST HOUSEHOLDS

Two of the options in the preceding section involve segmentation in the field as a key operational step. Segmentation has to be carried out in many primary sampling units (PSUs), however they are defined (some may be census enumeration areas, for example), prior to the selection of the households for the survey. Recall that under Option 2, the households within each selected segment must be *listed and then sub-sampled* for interview, whereas in Option 3, *all* the households in each selected segment are included in the sample. We now give detailed guidelines on how to map and segment your PSUs.

In this section, the following definitions are used:

- A *household* is a group of persons who eat and live together.
- A *dwelling* is a building or residential unit. It may include one or more households, as in the case of compounds or apartment buildings.

Sample updating activities need to be implemented well in advance and before data collection by trained listing teams, each consisting of two enumerators. A field coordinator will monitor the entire operation.

The responsibilities of the *coordinator* are:

- Obtain the basic materials and identification for all the PSUs included in the MICS4 sample.
- Arrange for the reproduction of all listing materials (listing manual, mapping and listing forms).
- Assign teams to PSUs.
- Obtain travel allowances for the teams.
- Arrange for transportation of the teams to the field.
- Monitor the receipt of the completed listing forms at the central office.
- Verify that the quality of work is acceptable.

The responsibilities of the *enumerators* are:

- Contact local officials in each PSU/segment to inform them about the listing operation and to obtain their cooperation.
- Identify the boundaries of the PSU/segment.
- Draw a map showing the location of the PSU/segment.
- Draw a detailed sketch map of the segment.
- List all the households in the segment in a systematic manner.
- Communicate to the coordinator any problems encountered in the field and follow his/her instructions.

CREATING SEGMENTS – SAMPLE OPTIONS 2 AND 3

Follow the steps below for each PSU selected:

Step 1: Ascertain the number of segments to be created. The number of segments is pre-specified in the sample selection by the sampling staff, and in that regard the coordinator should be told how many segments to create by the sampler.

Under Sample Option 2, the number of segments was determined at the first stage of selection by dividing the total number of *persons* in the PSU by the number of people in a *standard* segment, usually 500, rounding to the nearest whole number. Note that for a large number, perhaps a majority of PSUs, the number of segments under this option will be equal to one – in such cases, *no segmentation will be necessary*. Only when the number of segments is two or greater must mapping and segmentation be carried out.

Under Sample Option 3, the number of segments, again, has already been predetermined by the sampling staff at the first stage of selection, though computed differently than under Option 2. In Option 3, the number of segments is calculated by dividing the total number of *households* in the PSU by the desired cluster size, and rounding the result to the nearest whole number. Table II gives an illustration of the number of segments that the sampling staff would create for Sample Option 3 in a given PSU when the desired cluster size is equal to 20 households.

EXAMPLE (SAMPLE OPTION 3):

Suppose the frame (for example, the last population census) recorded 115 households in the PSU and the average cluster size is to be 20 households. Dividing 115 by 20 gives 5.75. Rounding it to the nearest whole number gives 6. *Exactly* 6 segments are to be created in this PSU, even if the *current number* of households is quite different from the original 115. Using Table 11 would give the same result by looking up the number of households (115) in the table and reading off the number of segments (6).

Table 11
Determining the Number of Segments to Create in a Given PSU
(Cluster Size = 20 Households) – Sample Option 3

<i>Number of Households</i>	<i>Number of Segments</i>
1-29	1
30-49	2
50-69	3
70-89	4
90-109	5
110-129	6
130-149	7
150-169	8
170-189	9
190-209	10
<i>Etc.</i>	

In some cases, the PSU may be too large, in terms of population, for easy segmentation and mapping. You may divide the PSU first into, say, four quadrants with approximately the same area and then select one of the quadrants at random before carrying out the segmentation in the sampled quadrant. In such cases, the sampling staff should be consulted. In our example, the sampler would then divide the frame count of the number of households in the PSU by four and recalculate the number of segments to create in the chosen quadrant. A note must be made of these special cases so that the correct survey probabilities and weights can be calculated later at the estimation and analysis stage.

EXAMPLE (SAMPLE OPTION 3):

With a cluster size of 20 households, a PSU with 560 households would result in $560/20$, or 28, segments – too large a number for efficient segmentation. Instead, the PSU may be divided into four quadrants of about 140 households each. One quadrant would be chosen at random and, according to Table 11, seven segments would be created in the chosen quadrant.

NOTE: Steps 2 through 7 apply (a) only to those PSUs in Sample Option 2 that require segmentation and (b) to all PSUs in Sample Option 3. Go to Step 9 for those PSUs in Sample Option 2 that **do not require** segmentation.

Step 2: To prepare the sketch map, go to the field with the available boundary map and locate the outer boundaries of the PSU. If a map is not available, get help from a local person and draw a map marking these boundaries, identifying the names of roads, lanes and streets and showing physical boundaries such as streams, rivers and so forth (use standard symbols suggested for mapping).

Step 3: Draw internal markers that will help identify locations and aid in establishing a path of travel. These markers include internal streets, paths, streams and so forth (use standard symbols suggested for mapping).

Step 4: Draw a small square representing each dwelling unit in its appropriate location on the map. For help in later locating the households, it is also useful to mark other prominent buildings – schools, churches, mosques and so forth (use standard symbols suggested for mapping).

For purposes of making the segments, absolute precision in the count of households is not necessary. A quick count of dwellings can be used instead of counting households.

Note that the sketching should be a quick count operation for locating dwellings, which does not require knocking on doors to inquire about households or the names of occupants. An exception is multi-dwelling buildings that appear to include several households – for example, in the case of large compounds or apartment buildings, where you should ask about the number of households and record it on your map.

Step 5: Count the number of dwellings in the PSU and divide it by the number of segments that have been specified by the sampling staff (step 1, above), whether Sample Option 2 or 3. This will give you the approximate number of households that each segment should contain.

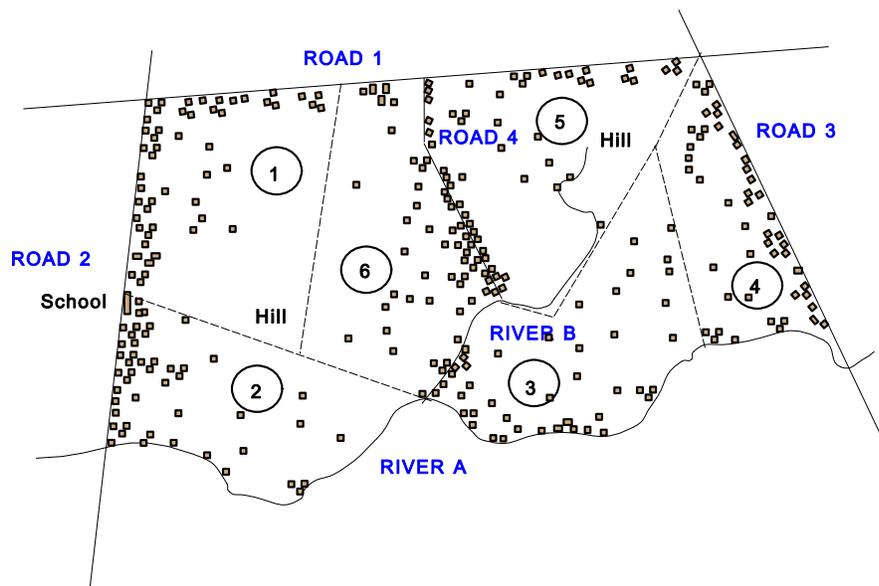
Step 6: Based on the sketch map, divide the PSU into the pre-specified number of segments with roughly the same number of dwellings. This means that the size of the segments in terms of geographic area may vary considerably – densely populated areas will have geographically small segments and low-density segments will be large. Assign each segment a number.

EXAMPLE:

The pre-specified number of segments is six and, after mapping, your PSU was found to contain 241 dwellings. You should then divide it into six segments of about 40 dwellings each.

It is essential to create segments in such a way that segment boundaries can be easily identified. *You may have to relax the criterion of equal-sized segments in order to achieve this.* The segment will be visited later, perhaps by some other field worker, so the boundaries must be readily identifiable on the sketch map and on the ground.

Table I.4.12
Sample PSU and Segments



Step 7: After segmentation has been completed, the next step is to select one and only one segment at random in each of the PSUs. Using a random number table, choose a random number between one and the number of segments created in the PSU to identify which segment is to be selected.

EXAMPLE:

If the PSU has been divided into six segments, using a random numbers table choose a random number between one and six to determine which is selected. Do this independently for each sample PSU.

If possible, the person selecting the segment should not be the same person who created the segmentation in the first place, in order to prevent the possibility of inadvertent bias.

Table 13
Using Segmentation in Urban Areas

Segmenting urban areas is easier than segmenting rural areas. Cities and towns are usually organized into blocks or some similar units. When using census enumeration areas, maps are usually available showing streets and blocks. If unavailable, these maps can be easily drawn. A quick drive through the area will let you find out whether the number of dwellings appears to be about the same in every block. If so, the segmentation may be based on the blocks or parts of the blocks.

EXAMPLE:

Let us suppose that your urban PSU, say a census tract, includes 18 very similar blocks and that you require 7 segments. Divide 18 by 7 to obtain 2.6 blocks per segment. Using the map, divide the area into 7 segments with 2.5 blocks each (the last will have to include 3 blocks to complete 18).

If the number of dwellings in the blocks varies, proceed as for the rural areas, counting the total number of dwellings in the area and later dividing them into the required number of segments.

When sketch mapping rural areas you do not have to worry about separating dwellings from households. However, in urban areas, you should ask how many households are present in tall buildings (for example, those with more than two floors).

IMPORTANT: *Step 8 applies to Sample Option 3 only.*

Step 8: Interview all households in the selected segment. Note that the number of households is likely to be different from the desired cluster size because of population changes since the frame was constructed. It is essential to interview all the households no matter what the original, desired cluster size is.

EXAMPLE

If the desired cluster size is 20 households and the selected segment in a given PSU contains 27 households, all 27 households must be included in the sample for interview. Likewise, if the selected segment is found to contain fewer than the desired 20, say 15 households, then those 15 would be interviewed.

IMPORTANT: *Steps 9 to 11 apply in Sample Option 2 to all selected segments, including those in segmented and non-segmented PSUs.*

Step 9: Make a new listing of the households in the sample segment. Include in the listing the address or other information that will pinpoint its location, the name of the head of the household, and the number of household members. This step may be carried out by the supervisor or an interviewer.

Step 10: From the listing, apply the sample interval and the random start that has been set already by the sampling staff. This will be done back in the office. Then select the sample households accordingly (that is, systematically) until the end of the list is reached. Note that the number of households selected into the sample, in a given segment, may differ from the *expected* number because of population changes that may have occurred since the frame was constructed and because sampled segments will differ somewhat in size (even without population change). *It is essential, however, to use the within-segment selection rate that was determined on the basis of the census measure of size rather than the current measure; this will maintain a self-weighting sample.* If it is alternatively decided to select a fixed number of sample households in each segment, a new sampling interval and random start will have to be determined; in this case the weights will vary by segment, so it is important to maintain a good record of the listing information so that the appropriate weights can be calculated later.

Step 11: Make a list of sample households to give to the survey interviewer. Include in the list the information obtained from the listing operation described in step 9. Mark the sampled households on the segment map to help the interviewer locate them.

SPECIAL TOPICS FOR THE MICS4 SAMPLE

In this section we discuss a few other important topics that should be taken into account in planning the sampling aspects of MICS4 in your country. Those topics include subnational estimates, estimating change and analytical subgroups, and water and sanitation indicators.

SUBNATIONAL ESTIMATES

Thus far we have been concerned with sample sizes necessary to generate national estimates of indicators. Many countries, however, will also want to use MICS4 to provide subnational figures – for example, at urban-rural, regional, state, provincial, or possibly district levels. Such data would be used for identifying areas where greater efforts are needed, as well as for programming and evaluation purposes.

A crucial limiting factor in providing reliable subnational estimates is sample size. For each reporting *domain* (that is, subnational area such as a region or urban-rural areas), the overall sample size must be increased substantially for the results to be acceptably reliable. If *equally*

reliable results are wanted for each domain with standard errors and confidence intervals comparable to what was required for the national level, the sample size must be increased. It is common practice to increase the *national* sample size, *n*, by a factor close to the number of estimation domains, thus selecting *n* cases *in each domain*. So, if equally reliable data are wanted for five regions of the country, the sample size that is calculated for the national estimates on the basis of Table 3 or directly from the sample size formula would have to be *multiplied by a factor of about 5* in order to obtain the regional estimates. Similarly, if domain data are wanted for 10 provinces, the national sample size would have to be increased about tenfold. This, of course, increases the overall sample size (and cost) substantially and is likely to be impractical for most countries.

An alternative formulation suggested by Vijay Verma is to increase the national sample size by a factor of $D^{0.65}$, where *D* is the number of domains wanted.¹⁹ The reliability of each domain's estimate is somewhat less than that of the national estimate under this approach.

EXAMPLE:

If there are 10 domains, the factor by which to multiply the national sample size is equal to 10 raised to the power of 0.65. That value is 4.47. Thus, instead of multiplying the national sample size by 10, it would be multiplied by 4.47.

Table 14
Overall Sample Sizes for Domains Under 2 Options and Reliability Implications

Number of sample households national	Number of estimation domains	Overall sample size for same reliability in domains as national	$D^{0.65}$ option multiplier	Overall sample size for decreased reliability in domains vs national	Ratio of sampling error ($D^{0.65}$ option), domains to national
		Col 1 x Col 2	(Col 2) ^{0.65}	Col 1 x Col 4	Square root (Col 2/Col 4)
3500	5	17500	2.85	9963	1.33
3500	10	35000	4.47	15634	1.50
3500	15	52500	5.81	20348	1.61
3500	20	70000	7.01	24532	1.69
5000	5	25000	2.85	14233	1.33
5000	10	50000	4.47	22334	1.50
5000	15	75000	5.81	29069	1.61
5000	20	100000	7.01	35046	1.69
6500	5	32500	2.85	18503	1.33
6500	10	65000	4.47	29034	1.50
6500	15	97500	5.81	37790	1.61
6500	20	130000	7.01	45560	1.69

¹⁹ See 'A Critical Review of MICS Sampling Methodology', a report by Vijay Verma to UNICEF, April 1995.

While the preceding example yields a considerably smaller sample size (less than half as big as multiplying by 10), there are of course consequences to the reliability of the estimates for each domain and to their confidence intervals. In that example, the standard error on any estimate at the domain level would be higher than that for the national level by a factor approximately equal to the square root of 10/4.47, or 1.5. Confidence intervals would be comparably wider. Table 14 shows the impact on sample size of domain requirements under two options.

Other compromises may have to be considered, especially if the number of domains is large. One is to restrict the separate reporting domains, such as provinces, to only those that exceed a certain population size. The remaining sub-areas could be combined into regional groupings. Another alternative is to allow the precision levels for the domain estimates to be less strict than that for the national estimate. For example, the margin of error for the key indicator for the national estimate is set at 12 percent of r (see section on sample size), but the separate reporting domains might have their margins of error relaxed considerably – even as high as 25 to 30 percent of r . Further, these two alternatives could be used in combination.

It is obvious from a quick look at Table 14 that the sample size increases dramatically when even a few domains are wanted and that is the case under either of the two options shown in the table. The second option has the added disadvantage of giving much larger sampling errors than you might want to accept for your domain indicators, even though the sample size is considerably less than option one (but still large). Clearly, few countries can entertain the prospect of supporting such huge sample sizes. Moreover, the logistical and technical problems with mounting household surveys that have huge samples only get worse as the sample size grows bigger. Non-sampling error is particularly troublesome to control and may be far worse on the overall results, in terms of survey bias, than sampling error when massive samples are used.

Still, the pressure on countries to produce small area estimates (regional, provincial, etc.) has never been more intense. That is why many countries in previous MICS programmes have insisted on producing and publishing domain estimates even when the sample size was not appropriately increased to deal with the issue of reliability. For example, consider the case where a sample of 5000 households is settled upon as a suitable size for the national indicators and there are insufficient resources to increase the sample size for 10 provinces for which the government wants to produce data. The survey team decides to tabulate the provincial data anyway. Note here that the average sample size in each of those domains is 500 households. Suppose, further, that the national estimate of characteristic p is 0.2 and the standard error is estimated to be 0.0085 based on 5000 sample households. The confidence interval around the national estimate is $p \pm$ twice the standard error, or $|.183$ to $.217|$. The standard error on the same characteristic for the average domain is larger by a factor equal to the square root of 10 (that is, $\sqrt{(5000/500)}$), or 3.16. Thus the domain standard error is 0.027 and the confidence interval is a horrific $|.146$ to $.254|$. It is no wonder that this result is thought to be useless. It cannot be emphasized too much that trying to get meaningful data from comparatively small national samples spread over too many domains is virtually impossible.

ESTIMATING CHANGES AND SUBGROUP ANALYSES

Some of the international goals are expressed as expected reductions, such as decreasing the prevalence of malnutrition by 20 percent in a 5-year period. You may also have a MICS or DHS survey from some years ago that provided estimates of indicators covered in MICS3, and you may want to make an assessment of changes that have taken place since then. This type of assessment would require two surveys, one at the beginning and the other at the end of the assessment period. The size of the sample necessary to measure the change between two time periods is highly dependent on the magnitude of the change, as well as the magnitude of the two estimates at each point. It is a somewhat complicated matter and impractical to provide short, general guidelines for estimating change over time. It is recommended that you seek the help of the national statistics office or specialized sampling assistance if your plans include change measurement. It should be pointed out that when one of the primary goals of a survey is to measure change in particular indicators over time, there is an advantage of using overlapping samples. In this case the correlation between the two surveys will improve the precision of the estimates of trend. The maximum correlation would be provided by a longitudinal panel of sample households, although this will add a much higher level of complexity to the survey methodology. However, even using the same sample enumeration areas will improve the comparability of the results and the precision of the estimates of change in the indicators.

Regarding subgroup analyses, such as indicators by gender or socio-economic group, the indicator estimates will be less precise than those for the whole sample.

The example below shows how the margins of error increase for ever-smaller subgroups.

EXAMPLE:

Based on the full (national) sample, if the calculated precision is, for example, plus or minus five percentage points for a 50-percent coverage rate at a 95 percent confidence level, the margin of error would be, approximately plus or minus

- 6.3 percentage points for gender-specific indicators, assuming 50 percent boys and 50 percent girls in the sample
- 8.6 percentage points for a subgroup comprising 20 percent of the overall sample.²⁰

Thus, reasonably precise results can be obtained for gender-specific indicators as well as for other subgroups making up one fifth or more of the whole sample.

²⁰ See the unpublished note to UNICEF by Graham Kalton, 'Some Proposed Modifications for the WHO Simplified Cluster Sampling Method for Estimating Immunization Coverage', p. 10, September 1988.

WATER AND SANITATION INDICATORS

Gathering data on water and sanitation is an important component of MICS4. There are statistical issues when using or evaluating the data, however, which ought to be kept in mind.

As we have stated before, the sample design for MICS is based on collecting personal variables as opposed to household variables. Access to water and sanitation are household characteristics, in the sense that the entire household has the same type of access. Moreover, in an area cluster it is often the case that all households in the cluster have the same type of access. For these reasons, the sample design effect, *deff*, which we have previously discussed, is considerably higher for water and sanitation access indicators than the value of 1.5 that we have assumed for the sample size calculations. As a result, the sampling errors on the water and sanitation indicators will be much larger than those based on individuals.

We should point out that if the main purpose of MICS4 were to gather data on types of water and sanitation access, the sample design would be much different. It is likely that the sample plan would entail, more simply, a community survey in which a single informant in the sample communities (area clusters) would be asked about water and sanitation, as opposed to the MICS4 approach where those questions are posed to every household in the sample.

Despite the expected large sampling errors for water and sanitation indicators in MICS4, the results should be useful nevertheless, especially in determining *trends*. This is because data have been collected on these indicators in other household surveys, such as DHS and Living Standards Measurement Surveys or Living Condition Surveys, whose designs are similar to MICS4. Comparison of the MICS4 results with those of other surveys should be highly useful in determining trends because the sample design effects, while high in both surveys, have less impact on the precision when estimating change.

WEIGHTING THE DATA AND PREPARING SAMPLING ERRORS

In this section we discuss the weighting alternatives for preparing the estimates, as well as details on calculating sampling errors.

SAMPLE WEIGHTING

Two types of weighting, if appropriate, may be applied in sequence in producing the estimates of the indicators. Unless the sample households have been selected with uniform overall probabilities (that is, a self-weighting design), all sample data should be weighted using the inverse of the overall selection probabilities – the so-called *design* weights. The design weights should be adjusted to account for non-response, however, even if the sample is self-weighting. This might be done in a variety of ways, including weighting up the respondents in each PSU (or cluster) to represent the non-respondents in that PSU. The main advantage of this approach is

that it does not require external data. These two steps, applying design weights and non-response adjustments, may be all the weighting that is necessary for your survey.

A second type of weighting may be undertaken by adjusting the design weights to make the weighted sample distribution for some key variables, such as urban-rural or region, conform to an external population distribution, such as the most recent population census. This type of post-stratification weighting should be considered when there have been significant departures from the design at the implementation stage, when approximate procedures have been used because of deficiencies in the sampling frame, or when the sample departs from strict probability criteria.

The actual formulas and calculations of the design weights, non-response adjustment factors and, if necessary, post-stratification adjustments are highly dependent on the sample design used – the number of sampling stages, self-weighting versus non-self-weighting plans, whether non-response is uniform or widely variable across population subgroups or areas, availability of external data for post-stratification, etc. Therefore, it is not practical to present detailed steps on how to calculate the weights for your MICS4.

Nevertheless there are two Excel weighting templates in the documentation of MICS on the website that you may be able to use as is or with minor modifications. One is for the case where your sample is self-weighting within domains; the other is for samples that have different weights for each cluster in each domain. (An example of the calculation of weights for a specific MICS, however, can be found in the country example for Lebanon in the last section of this chapter). Using the first template, once you enter your country data in several of the columns, the template automatically calculates design weights, weights adjusted for non-response and so-called normalized weights for 3 tabulation groupings - households, women and children. The template suffices when you have a straightforward sample design that enables you to calculate the sampling fraction (probability of selection) in each of 8 domains. It is easily modified if you have a different number of domains. Your sample plan must be self-weighting at the domain level, however, in order to use the template; in other words the sampling fraction must be the same for the sample households in every cluster in the domain. The most difficult column to fill for the template is column B, the sampling fraction. Typically, however, whenever you have a basic, two-stage design, a given cell in that column is equal to [number of PSUs selected in the domain multiplied by the cluster size and divided by the total number of households in the domain population.]

It should be pointed out that most household surveys such as MICS are generally not completely self-weighting by domain given the variability between the number of households in the frame and the listing. In this case the weights are mostly calculated at the cluster level and the second template must be used. Instead of having 8 or so domain weights, you would then have several hundred cluster weights. The sampling statistician who designs your sample should be fully capable and responsible for determining the appropriate weighting scheme and modifying the template, if necessary, to make it reflect your sample design.

The survey weights in the two templates are normalized, which shows their relative or differential contribution to the sample. Normalized weights are appropriate whenever rates, proportions or percentages are tabulated but not when estimated totals are wanted. Most of the data needs in MICS do not require estimated totals. However, the relative weights may hide the fact the sampling frame is incomplete, or there may be listing problems or other issues with the weights. It is very useful, therefore, to calculate the weighted total number of households and population, using design weights or design weight adjusted for non-response, to compare to census results or projections in order to evaluate the coverage of the frame as well as to validate the calculation of the weights. This should be done before the weights are normalized. It is recommended therefore that both population weighted data and normalized weighted data be produced in the country data sets.

CALCULATING SAMPLING ERRORS

As has been emphasized throughout this chapter, the sample size in your survey is based on a key indicator for which we have pre-specified the expected precision level. However, the survey will produce hundreds of estimates, each of which will have its own precision (sampling error) and these will vary by whether they are national or subnational, as well as by the size of the p , r and d_{eff} associated with each (see sample size formula). As noted above, for example, the water and sanitation estimates are expected to have much higher sampling errors than other indicators due to the higher design effects. This is why it is important to calculate sampling errors for the most important (not all) of the indicators.

As part of the routine preparation of the survey results, therefore, sampling errors and associated variables, such as d_{eff} s, are to be estimated for the main indicators – perhaps 30 to 50 different estimates. This is essential in order to evaluate the reliability of the indicator estimates. To that end, *the confidence intervals incorporating the margin of error around the survey estimates cannot be constructed unless the sampling errors are estimated. Otherwise, interpretation of the estimates will be severely hampered.*

Calculation of sampling errors, or *standard errors*, can be a fairly complicated part of the survey operation. Standard errors must be calculated in a way that takes account of the complex sample design (clustering, stratification and weighting). The inappropriate application of simple random sampling formulas will, as a rule, seriously underestimate the standard errors.

Information to help you calculate sampling errors can be found on the MICS4 website, www.childinfo.org.

In addition, there are several software packages that have been developed and can be adapted for variance estimation. They include the CLUSTERS computer program, originally developed for the World Fertility Survey and available from the University of Essex; CENVAR, a software product available from the U.S. Census Bureau without charge; and WesVar, a program produced by WESTAT for use with SPSS. Packages such as SAS, SUDAAN and Epi-Info can

also deal with complex designs. The Stata software and the Complex Samples add-on to recent versions of SPSS are also capable of calculating sampling errors for complex designs. Some of the packages are free and can even be downloaded from the Internet, while others are commercially sold.²¹

The particular sample used in your MICS is one of a large number of all possible samples of the same size that could have been selected using the same sample design. The particular value of an estimated indicator, r , (so-called *point* estimate), derived from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is the *sampling* error. While it is not possible to calculate the actual sampling error since we only have data from one of the possible samples, the *standard* error of a given estimate calculated using one of the above-mentioned computer packages is nevertheless an *estimate* of the sampling error. The standard error also partially measures the effect of some nonsampling errors such as that which can be attributed to the *variability* among interviewers and coders but does not measure any systematic *biases* in the data.

The point estimate from the sample for a given indicator and an estimate of its standard error permit us to construct interval estimates with a prescribed level of confidence that the interval includes the average result of all possible samples. To illustrate, if all possible samples were selected, each were surveyed under the same conditions and an indicator and its estimated standard error were calculated from each sample, then approximately 95-percent of the intervals from two standard errors below to two standard errors above the indicator would include the average value of all possible samples – the so-called 95-percent confidence interval.

The standard error tables usually provide the following information:

Variable name: name of the indicator

Value (r): point estimate of the indicator

Standard Error (se): see text above for description

Coefficient of variation (se/r): sometimes abbreviated as *cv*; it is a measure that shows the ratio of the standard error to the value of the indicator, also called the relative standard error

Design Effect (deff): the factor by which the MICS sampling variance is larger than that for a simple random sample of the same sample size

²¹ For a comprehensive review of these programs, see ‘Sampling Error Software for Personal Computers’, by Jim Lepkowski and Judy Bowles of the University of Michigan. The article appears in *The Survey Statistician*, No. 35, December 1996, pp. 10–17 (see website: www.fas.harvard.edu/~stats/survey-soft/iass.html). A more recent review can be found in a UN Statistics Division publication, ‘Household Sample Surveys in Developing and Transition Countries’, March 2005, particularly Chapter 21, on ‘Sampling Error Estimation for Survey Data’ by Donna Brogan (http://unstats.un.org/unsd/HHsurveys/pdf/Household_surveys.pdf).

Square root of Design Effect (deft): the factor by which the MICS standard error is larger than that for a simple random sample of the same sample size

Weighted count: weighted total population estimate, or sum of weights for the sample observations used to estimate the *value* of the point estimate.

Unweighted count: number of sample observations (not weighted), upon which the *value* of the point estimate is based

Confidence Limits: range of the 95-percent confidence interval; the lower end of the range shows the point estimate minus twice the standard error ($r - 2se$) and the upper end shows the point estimate plus twice the standard error ($r + 2se$).

In general the following observations hold concerning the sampling errors for MICS, though there are occasional exceptions:

1. As the number of sample cases increases, the coefficient of variation (relative standard error) decreases. When an indicator is based on a small number of cases it is likely to have a large *cv*. Moreover, when it is very small the *cv* may be very large. In general when the *cv* is greater than 20-percent the indicator is considered too unreliable to be meaningful or informative. However, in the case of proportions such as the rate indicators, the *cv* may be high for smaller proportions, so it is more relevant to examine the standard error or margin of error. For example, if a point estimate for a proportion is 0.04 and the standard error is 0.01, the corresponding *cv* would be 25 percent. However, the 95% confidence interval of 0.02 to 0.06 may still be considered to provide a sufficient level of precision.
2. Similarly, as the number of sample cases increases, the confidence limits around the point estimate become narrower – another way of saying that the estimate is more reliable the larger the number of cases in sample.
3. Related to points (1) and (2) above, the estimates at the national level are more reliable (have smaller standard errors) than those for urban and rural, which, in turn, are more reliable than those for regions, provinces or other sub-national categories. Further, region totals are more reliable than their urban and rural components.

DETAILS OF SAMPLE SELECTION TECHNIQUES

In this section we provide detailed illustrations of how to apply the sample selection techniques that have been discussed in this chapter.

PROCEDURES FOR SAMPLING WITH PPS – OPTION 2

The first technique to be illustrated is how to select the first-stage units using *pps*. The illustration also shows you how to combine systematic *pps* sampling with geographic arrangement of the sampling frame to achieve *implicit* stratification.

For the purposes of the illustration, we will use Option 2 from the chapter, the standard segment design, and select a national sample. Suppose (1) the standard segment size under Option 2 is to be 500 persons, or about 100 households; (2) census enumeration areas (EAs) are to be the sample frame; and (3) the number of PSUs to be selected is 300. The steps of the first-stage selection, which follow, should be done as a computer operation, although it is possible to do them manually.

- Step 1:** Sort the file of enumeration areas by urban and rural.
- Step 2:** Within the urban category, further sort the file in geographic serpentine order according to the administrative subdivisions of your country (for example, province or state, district, commune, etc.).
- Step 3:** Repeat step 2 for the rural category.
- Step 4:** In one column, show the census population count of the EAs.
- Step 5:** In the next column, compute the number of standard segments, which is equal to the population count divided by 500 and rounded to the nearest integer. This is the measure of size for the EA.
- Step 6:** Cumulate the measures of size in the next column.
- Step 7:** Compute the sampling interval, I , by dividing the total cumulant by 300, to one decimal place. In this illustration suppose the total cumulant is 5,281. Then the sampling interval, I , would be equal to $5,281/300$, or 17.6.
- Step 8:** Select a random start between 0 and 17.6. The way to do this, in practice, is to use a table of random numbers and select a three-digit number between 001 and 176 and insert the decimal afterward. Suppose you select 042; then your random start is 4.2. Then your first sample PSU would be the one for which the cumulant measure of size is the smallest value equal to or greater than 4.2.²²
- Step 9:** Add 4.2 to I , or $4.2 + 17.6 = 21.8$; then your next sample PSU would be the one whose cumulant corresponds to the smallest value equal to or greater than 21.8.
- Step 10:** Add 21.8 to I , or $21.8 + 17.6 = 39.4$; the next sample PSU is the one with the cumulant corresponding to the smallest value equal to or greater than 39.4.
- Step 11:** Continue as above, through the urban EAs followed by the rural ones, until all 300 PSUs have been selected.

This procedure is further demonstrated in Table 15.

The two sample PSUs that are depicted in the illustration are those in EA 003 of commune 01 and EA 002 of commune 03, both in district 01 and province 01. In the case of the first EA, its measure of size is 3. This would mean that three segments would have to be created, each of

²² Kish recommends rounding down when the sampling interval is fractional. See Kish, L. 1965. *Survey Sampling*, p. 116. New York: Wiley.

which is roughly 540 persons (1,630 divided by 3), with one of the segments selected at random for listing and sub-sampling of households. In the second sample EA, two segments would be created, each containing about 590 persons, before selecting one of them at random.

Table 15
Illustration of Systematic *pps* Sampling and Implicit Stratification – Sample Option 2

Urban	Population	Measure of size (segments of 500 population)	Cumulative
Province 01			
District 01			
Commune 01			
EA 001	1,470	3	3
EA 002	562	1	4
EA 003	1,630	3	7 selected
EA 004	1,006	2	9
Commune 02			
EA 001	412	1	10
EA 002	1,537	3	13
EA 003	1,312	3	16
EA 004	397	1	17
Commune 03			
EA 001	1,540	3	20
EA 002	1,181	2	22 selected
EA 003	1,025	2	24
District 02			
Commune 01			
EA 001	567	1	25
EA 002	1,111	2	27
EA 003	409	1	28
*			
*			
etc.			
Rural			
Province 12			
District 05			
Commune 05			
EA 001	512	1	5,280
EA 002	493	1	5,281

The illustration demonstrates the many advantages of implicit stratification: First, it is very easy to achieve, merely requiring that the frame of enumeration areas be sorted geographically before selecting the sample systematically with *pps*. Second, it automatically provides a sample of PSUs that is proportionately distributed by urban and rural areas and by province (or other geographic subdivisions). For example, if 10 percent of your population is located in province 12, then 10 percent of your sample will also be selected in that province. Third, it can be easily implemented on the computer.

Once the PSUs have been selected, under Option 2, segmentation will have to be carried out in those PSUs where the measure of size (number of segments) is two or more, followed by one

segment being selected at random in each PSU. Then, a new household listing will have to be made in the selected segments plus the one-segment PSUs. The final step in the selection procedure for Option 2 is to select the sample households within the selected segments. This procedure is described in Table 16 with an illustration.

Table 16
Selecting the Households – Option 2

Suppose your standard segment size is 500 persons. Let your desired cluster size for the survey be designated as \bar{n} households.

1. Calculate the average number of households per segment by dividing 500 by the average household size in your country. Let this be s_h .
2. Divide s_h by \bar{n} . This is your sampling interval, I , for selecting households within each sample segment.

(Note: If your standard segment size is other than 500, that value must be used.)

Illustration:

Suppose your average household size is 5.5. Then s_h is $500/5.5$, or 90.9. Suppose you want your cluster size, \bar{n} , to be 25. Divide 90.9 by 25 (1 decimal place) = $90.9/25$, or 3.6. Then, select households in each segment at the rate of 1 in 3.6, starting with a random number between 01 and 36 (inserting the decimal after selecting the number).

PROCEDURES FOR SAMPLING WITH PPS – OPTION 3

If Option 3, the modified segment design described in this chapter, is used instead of Option 2, implicit stratification is done in the same way, *although the measure of size is different*. Under Option 3, if we suppose, as an example, that our segment size is going to be 20 households (on average), then the measure of size would be calculated by dividing the census count of households by 20, rounded to the nearest whole number. Note that under Option 3 the second column in Table 17 must be the number of households rather than population. You would calculate the sampling interval, I , by dividing the total cumulant – suppose it is 26,425 – by the desired number of PSUs, again, assume it to be 300. So, you would have $26,425/300 = 88.1$. If the random start is chosen to be 19.4, the first two PSUs selected, as illustrated in Table 17, would be those corresponding to the smallest cumulants exceeding or equal to the values, 19.4 and 107.5 ($88.1 + 19.4$), respectively. They are EA 002 in commune 01 and EA 002 in commune 03 of province 01, district 01.

Table 17
Illustration of Systematic *pps* Sampling and Implicit Stratification – Sample Option 3

Urban	Households	Measure of size (segments of 20 hhs)	Cumulative
Province 01			
District 01			
Commune 01			
EA 001	290	14	14
EA 002	120	6	20 selected
EA 003	325	16	36
EA 004	200	10	46
Commune 02			
EA 001	81	4	50
EA 002	307	15	65
EA 003	261	13	78
EA 004	80	4	82
Commune 03			
EA 001	308	15	97
EA 002	236	12	109 selected
EA 003	205	10	119
*			
*			
*			
etc.			
Rural			
Province 12			
District 05			
Commune 05			
EA 001	102	5	26,400
EA 002	99	5	26,405

Recall that, under Option 3, the measure of size is equivalent to the number of segments of pre-designated size that must be created (20 in our illustration). So, for the first two sample PSUs chosen, 6 segments of an approximate size of 20 households each must be formed in the first PSU and 12 in the second. Again, one of the segments would then be selected at random within each sample PSU, *and all of the households within that segment would be interviewed for the survey*, even if the actual number of households in the segment departs significantly from its expected size.

COUNTRY EXAMPLES

In the final section of this chapter we describe the sample designs that were used in two countries – Lebanon and Mozambique. The Lebanon case comes from their participation in the year 2000 round of MICS, while Mozambique is from MICS3. Each one illustrates a different aspect of the sampling points that have been discussed in this chapter.

LEBANON

We begin with Lebanon and provide a fairly detailed description in order to compare the sample design with the year 2000 MICS manual. In addition, we show how the weights were calculated to illustrate how this is done in a specific survey.

Sample Plan

MICS was conducted by the Central Administration of Statistics. The principal measurement objectives of MICS were (1) to obtain the relevant, end-of-decade World Summit for Children indicators to profile the situation of women and children in Lebanon and (2) to provide these estimates at the national level and for five subnational areas (domains), which include the four largest *muhafazas* (governorates) plus the combination of Sud and Nabatieh. The sampling methodology for MICS in Lebanon was to select a probability sample of 1,625 households in each of the five domains using a stratified, clustered design. It was a two-stage selection process using the 1995-1996 Census of Buildings as the sampling frame. Altogether, 8,125 households were to be selected nationwide.

The first-stage clusters (that is, the primary sampling units – PSUs) were defined as *ilots*, and these were identical in concept and construction as census enumeration areas. They were selected with probability proportionate to their measures of size (or *pps*), the latter of which are the 1995-1996 counts of occupied dwellings from the Census of Buildings. A total of 65 *ilots* were selected in each domain, or 325 altogether. *Ilots* selected at the first stage were to be listed anew, that is, a fresh listing of households prepared in a field operation. *Ilots* that are very large were divided into geographical segments with one segment selected at random for listing. The new listing constituted the sampling frame for the second stage of selection.

A fixed sample size of 25 households from the new listings was selected systematically in each sample *ilot* for the MICS interview. In a few of the sample *ilots*, that is, those that contained fewer than 25 households, all the households were surveyed for MICS.

It was thought that non-response in the survey could be as high as 10 or 12 percent, in which case the number of interviewed households would range from about 1,430 to 1,465 in each domain, or 7,150-7,325 nationwide.

Discussion of Sample Plan – MICS Philosophy

It is significant to note that the sample plan described above is consistent with the overall MICS philosophy for sample design and implementation, according to the previous edition of the MICS manual. That edition, like the current one, promotes several features for sample design, all of which are faithfully followed in the plan undertaken by Lebanon's Central Administration of Statistics. They include the following:

- Use of simple sampling methodology
- Use of the latest census as the sample frame
- Sampling in two stages
- Use of probability techniques at every stage of selection
- Use of clusters of moderate size
- Use of an adequate sample size.

The sampling methodology is simple and straightforward, as opposed to complex; for this reason, it was expected that non-sampling error would be kept to a minimum. The Census of Buildings in 1995-1996 was the appropriate frame to use because (1) it was the latest available and (2) maps were available, so that those *ilots* selected for MICS could be easily located and their internal boundaries identified. This cartographic aspect is also extremely convenient for the segmentation operation that was required for large *ilots*. The age of the frame meant that natural population movement and change could decrease the accuracy of the estimates, and for this reason, a new listing of households at the second stage was necessary to bring the frame up to date in the selected areas.

Strict probability selection methods were used at both stages. This helped ensure that the results would be representative of the various target populations, including households, women of childbearing age and children in Lebanon. The cluster size of 25 households is moderate, which ensured that the sample design effect was not so large as to make the results unreliable for the important indicators.

The national sample size of 8,125 households was consistent with the recommendations in the 2000 edition of the MICS manual regarding this highly important survey parameter. As that manual pointed out, the sample size must be large enough to accomplish various measurement objectives. These include being able to measure key target indicators with a margin of error of 3-5 percentage points, depending upon the indicator. A second important objective was to be able to obtain reasonably reliable estimates of the indicators for the domains of interest (as mentioned in the summary, these included the four large governorates of Beyrouth, Mont Liban, Beqaa and Nord plus the combination, Sud/Nabatieh, as well as the national level).

Like this edition of the MICS manual, the 2000 manual suggested that the calculated sample size may need to be adjusted upward to account for loss of cases due to non-response. This feature was taken into account.

Sample Size

The indicator chosen for establishing the sample size in Lebanon was the proportion of children under age five with diarrhoea in the past 2 weeks. This was expected to be about 15 percent, for which the margin of error, or precision, was set at plus or minus 3 percent. Thus the confidence interval, at the 95 percent level of confidence, would yield a survey value in the range of 12 to 18 percent if the true value in the population is 15 percent.

The target population for the diarrhoea indicator is children under five, and they constituted just under 10 percent of the population. Sample design effect was presumed to be about 1.5. This was the factor by which the sample size was increased in order to counterbalance the fact that a cluster design was used instead of a simple random sample with regard to its effect on sampling reliability.

The number of domains was five, as discussed previously. The expected rate of non-response was taken to be 12.5 percent. That is a conservative number, intended to yield a larger sample than necessary. It was hoped that actual survey response would be much higher than 87.5 percent. The average household size in Lebanon at that time was 4.8. The factor necessary to achieve the 95 percent confidence level is 4.

Taking all the above parameters into account, the formula for sample size calculation below (see Appendix 7 of the 2000 MICS manual) gives the number of households needed in the sample for one domain. It had to be modified for five domains, which is discussed below.

$$n = \frac{4(1-r)(1.125)f}{e^2(p)(n_h)} = \frac{4(.15)(.85)(1.125)(1.5)}{(.03)(.03)(.1)(4.8)} = 1992$$

To adjust the sample size for domains, two alternatives are suggested in the 2000 MICS manual. One is to multiply n that is needed for one domain by the number of domains. This gives $5 \times 1,992$, or 9,961. Because the survey budget could not accommodate that large a sample size, another approach that would give somewhat higher margins of error is to multiply n by the factor D raised to the power of 0.65, where D is the number of domains. This is approximately equivalent to taking the cube root of D squared, which, in our case, is the cube root of 25, or about 2.92. That number multiplied by 1,992 would give a sample size of 5,817 for five domains. It was decided that the average of the two alternatives, 7,889, would be used. It was further adjusted upward in order to accommodate 65 clusters per domain (7,889 would require about 63 clusters, that is $7,889/25$, or about 315 or 316 total and thus $315/5$, or 63 in each domain). It is useful to see the effect on the margin of error with this sample size. The calculated value of n for diarrhoea is 1,992. Recall above that this would provide a margin of error of plus or minus 3 percentage points. The final sample size, however, is $8,125/5$ or 1,625 per domain. Thus the margin of error is somewhat higher. The increase is calculated as the square root of the ratio of sample sizes, or the square root of $1,992/1,625$, which is about 1.033. Thus the expected margin

of error for this particular indicator at the governorate level would be about 3.3 percentage points, if the prevalence rate for diarrhoea was approximately 15 percent.

Sample Allocation

The measurement objectives must be taken into account when deciding how to properly allocate the sample. When national estimates take priority, the sample should be allocated proportionately among the various governorates. For example, if a governorate contains 35 percent of the nation's population, then 35 percent of the sample would be selected in that governorate under the proportional allocation scheme. The problem with this approach, however, is that small governorates (say, one containing only 10 percent of the population) would have too small a sample size to provide reliable estimates. The latter becomes a serious problem if the governorate level, as opposed to the national level, is the main priority.

The governorate level was in fact the main priority for the government and for UNICEF/Lebanon. This is because programme planning and policy implementation projects to improve the situation of women and children must proceed at the local level. For that reason, estimates that were as reliable as possible were wanted for each of the governorates. To accomplish this, the sample had to be allocated equally among the governorates (the four large ones plus the southern combination).

Table 18 displays the allocation of the sample and the rest of its parameters by governorate.

Table 18
Sample Parameters – MICS 2000 in Lebanon

Muhafaza (governorate)	Sample size – number of households (hhs)	Sample size – number of clusters (<i>ilots</i>)	Cluster size – no. of sample hhs per cluster	Total number of frame units (occupied dwellings)	Sample interval to select <i>ilots pps</i>
Beyrouth	1,625	65	25	101,707	1 in 1,564.7
Mont Liban	1,625	65	25	356,517	1 in 5,484.9
Nord	1,625	65	25	150,018	1 in 2,308.0
Beqaa	1,625	65	25	97,638	1 in 1,502.1
Sud/Nabatieh	1,625	65	25	149,093	1 in 2,293.7
Lebanon total	8,125	325	25	854,973	-

Implementation – Sample Selection Procedures

The first stage of selection was a systematic *pps* sample of census *ilots*, where the measure of size was the count of occupied dwellings. Prior to selection, the *ilots* were sorted by governorate and within governorate by *caza* (district), grand zone and *ilot*, the latter of which were geographically sequenced. This was all done using the seven-digit geographic identification code of the geographic database. In this way, the systematic selection that was carried out ensured implicit stratification of the sample for each of the domains, or governorates. Sixty-five *ilots* were thus selected in a computer operation for each domain.

To implement the second stage of selection, a current field listing of households was undertaken. Many of the sample *ilots* were very large and had to be segmented before conducting the listing. About 180 of the 325 *ilots* were segmented because they contain more than 150 households each, according to the 1995-1996 sampling frame.

Segmentation was carried out in accordance with Table 19.

Table 19
Number of Segments to Create

Number of households (according to frame)	Number of segments
Under 150	No segmentation
150-249	2
250-349	3
350-449	4
450-549	5
550-649	6
650-749	7
Etc.	Etc.

Some of the sample *ilots*, about 12 of them, contained fewer than 25 households, according to the frame. For these *ilots*, no field listing was necessary. Instead, all households contained within the boundary of the *ilot* were interviewed for MICS, even if the number of such households at the time the MICS was conducted in 2000 was greater than 25.

For segmentation, Table 19 was used to figure how many segments to create. For example, if the frame showed that the sample *ilot* contained 580 households, six segments were created. Procedures about segment construction outlined in Chapter 6 of the 2000 edition of the MICS manual were followed. After segmentation, one segment was selected at random from each *ilot* that had been segmented.

For all *ilots* (under 150 households plus the selected segments for those greater than 150 households), another visit to the field was made in order to make a complete listing of the then-

current households. Next, sampling of households was systematic within each *ilot*, or segment, at the rate of 1 in I, where I was the sample interval, equal to $N_i \div \eta$; N_i was the actual number of households in the listing for the i-th sample cluster, and η was 25, the desired cluster size. The value of the sampling interval, I, varied for each sample *ilot* or segment and it was calculated to one decimal place.

Weighting and Estimation

To prepare the survey estimates, weighting had to be used because the sample was not self-weighting; in other words, the sample cases did not all have the same probability of selection. Weighting necessitated multiplying the raw sample data by the sample weight. Weights were variable by cluster since the weight for households in each cluster depends on the measure of size and the number of households obtained in the listing operation for that cluster.

The weights are equal to the *inverses* of the probabilities of selection. The probability for the households, persons, children, women or any other target population group in a particular cluster is given by the following:

$$P_{ih} = \frac{65m_{ih}}{M_h} \times \frac{1}{s_{ih}} \times \frac{25}{N_{ih}} = \frac{(65)(25)m_{ih}}{(M_h)(N_{ih})(s_{ih})}, \text{ where}$$

- P_{ih} is the probability of selecting a household (or person) from the i-th cluster of the h-th domain
- **65** is the number of clusters in each domain
- **25** is the cluster size
- m_{ih} is the measure of size for the i-th cluster of the h-th domain, that is, the number of occupied dwellings according to the sample frame
- N_{ih} is the number of households listed in the i-th cluster of the h-th domain
- s_{ih} is the number of segments created in an *ilot* (the value of s_{ih} is 1 for *ilots* that are not segmented)
- M_h is the cumulative total measure of size in the h-th domain, that is, the sum of the m_{ih} values, or $\sum m_{ih}$; this is equal to the total number of occupied dwellings in the sampling frame for the h-th domain

Note that the value of N_{ih} is the number of households listed in the cluster, which in non-segmented *ilots* is the entire *ilot*, while in segmented *ilots* it is the segmented part (see examples below).

As mentioned, the weight, W_{ih} , is the inverse of the probability of selection, or $1/P_{ih}$.

The M_h values are those in the fourth column of Table 18 for each of the domains. The product of 65 and 25 is 1,625. Weights, therefore, for households or persons for the five domains are as

shown in Table 20.

Table 20
Survey Weights by Domain (for *ilots* with 25 or more households)

Governorate	Weight
Beyrouth	$(62.6N_i s_i)/m_i$
Mont Liban	$(219.4N_i s_i)/m_i$
Nord	$(92.3N_i s_i)/m_i$
Beqaa	$(60.1N_i s_i)/m_i$
Sud/Nabatieh	$(91.7N_i s_i)/m_i$

EXAMPLE:

Suppose in Beyrouth the measure of the size, m_i , in the sampling frame for the first sample cluster selected is 612. Thus it is segmented into six segments, which is the value of s_i . Suppose further that the number of households listed in the cluster is 110. Then the weight for all households and persons in sample for that cluster would be equal to $(62.6)(110)(6)/(612)$, or 67.5.

EXAMPLE:

Suppose in Mont Liban the measure of the size, m_i , in the sampling frame for the first sample cluster selected is 106. Hence it is not segmented and the value of s_i is 1. Suppose further that the number of households listed in the cluster is 98. Then the weight for all households and persons in sample for that cluster would be equal to $(219.4)(98)(1)/(106)$, or 202.8.

It was expected that it would be necessary to adjust the *design* weights discussed above by an additional factor to account for non-response. This was simply a matter of multiplying the weight, separately for each domain, by the ratio of sampled households to interviewed households. The product of the design weight and non-interview factor then becomes the final weight used to multiply the raw sample counts to produce the estimates.

MOZAMBIQUE

The 2008 Mozambique MICS is a very good example of a cost-effective MICS3 sample design based on using a current sample of EAs and listing of households for a larger national household survey that was being conducted during the same period. This MICS also had the benefit of a recent census frame (the 2007 Mozambique Census of Population and Housing, conducted in August 2007). At the same time, the sample design for the 2008 MICS has some features in common with the Lebanon MICS example. The Mozambique MICS has the same six features of an effective MICS sample design listed for the Lebanon example under the *Discussion of the Sample Plan*.

Use of Sample Enumeration Areas for a Larger National Household Survey

The Mozambique *Instituto Nacional de Estatística* (INE) conducted the 2008 Mozambique MICS in October and November 2008. Around the same time INE started conducting the *Inquérito sobre o Orçamento Familiar* (IOF) 2008/09, a national household budget survey, with data collection spread over a 12-month period (October 2008 to September 2009) to represent seasonality in the income and expenditure data as well as other characteristics. It was possible to coordinate both the first and second stage sampling between these two surveys in order to reduce the cost of implementing both surveys.

First it was necessary to examine the IOF 2008/09 sampling frame, sample size and distribution in order to determine whether it was possible to use the IOF sample EAs or a subsample for the 2008 MICS. The geographic domains of estimation for IOF were the 11 provinces of Mozambique, including the City of Maputo, as well as the urban and rural domains at the national level. Fortunately these were the same geographic domains defined for the MICS. The total sample size for IOF was 1,060 EAs and 10,858 households. Table 21 presents the allocation of the sample EAs and households for IOF 2008/09 by province, urban and rural strata.

Table 21
Allocation of Sample EAs and Households for IOF 2008/09 by Province, Urban and Rural Strata

Province	Total		Urban		Rural	
	No. Sample EAs	No. Sample Households	No. Sample EAs	No. Sample Households	No. Sample EAs	No. Sample Households
Niassa	80	816	32	384	48	432
Cabo Delgado	80	780	20	240	60	540
Nampula	160	1,584	48	576	112	1,008
Zambézia	160	1,524	28	336	132	1,188
Tete	80	768	16	192	64	576
Manica	80	804	28	336	52	468
Sofala	80	852	44	528	36	324
Inhambane	80	804	28	336	52	468
Gaza	80	816	32	384	48	432
Maputo Província	80	900	60	720	20	180
Maputo Cidade	100	1,200	100	1,200	0	0
Mozambique	1,060	10,848	436	5,232	624	5,616

Given the nature of the data collection for IOF, with multiple visits to each sample household over a two-week period, 12 households were selected in each sample urban EA and 9 households

in each rural EA. However, the most important consideration for the MICS was to ensure that the number of sample EAs selected for IOF was enough for using the same sample PSUs. In this case it was determined that the number of IOF sample EAs was more than sufficient, so the MICS could actually use a subsample of these EAs. The IOF sample EAs were selected in such a way that there were four nationally-representative subsamples of EAs, one assigned to each quarter of the survey data collection, as explained later. This made it convenient to select a subset of the IOF quarterly subsamples for MICS.

Sampling Frame for IOF and MICS

Fortunately the 2007 Mozambique Census had been conducted recently (August 2007), so it provided an updated sampling frame for the two national surveys. However, the census data entry had not been completed before the sampling frame was needed, so it was necessary to use the preliminary census summary information on the number of households and population in each census EA, with the corresponding geographic codes, found in the EA summary sheets from the census enumeration. This information was made available in a database that could also be used for the first stage selection of EAs.

For the second stage of selection, a new listing of households was conducted in each sample EA. This listing was used for selecting the sample households for both the IOF and MICS. In order to avoid increased response burden and potential conditioning bias from having some households included in the two surveys, INE decided to select a separate sample of households for each survey from the listing in the subsample of EAs selected for MICS. This was done by selecting the two samples of households at the same time, as explained later.

Stratification

Given that the provinces were geographic domains for the 2008 MICS tables, the first level of stratification of the sampling frame was by the 11 provinces, including Maputo City. Within each province the sampling frame was further stratified by urban and rural domains. Since Maputo City only has urban EAs, there were a total of 21 strata, as shown in Table 21. Within each province, urban and rural stratum, the EAs were sorted by geographic code in order to provide implicit stratification from the first stage systematic selection of EAs with PPS within each stratum, as described previously in this chapter.

Sample Size and Allocation

One source of information that was very useful for estimating the sample size for MICS and to estimate the expected level of precision was the final report for the 2003 Mozambique Demographic and Health Survey (DHS), which includes an appendix with tables of estimates, standard errors and design effects for key indicators that are the same or similar to the MICS indicators. The geographic domains for the 2003 DHS were also similar to those for the 2008 MICS (that is, 11 provinces, and urban/rural at the national level). First we should examine the

allocation of the sample for the 2003 DHS by province, urban and rural strata, which is presented in Table 22. The total sample size for the DHS was 604 sample EAs and 14,496 households, with 24 households selected per sample EA. Given the objectives of obtaining a similar level of precision for each province, the sample size for the DHS was the same for all provinces except for Nampula and Zambézia, which each have a population more than double that of the other provinces.

Table 22
Distribution of Sample EAs and Households for 2003 Mozambique DHS by Province, Urban and Rural Strata

Province	Total		Urban		Rural	
	No. Sample Segments	No. Sample Households	No. Sample Segments	No. Sample Households	No. Sample Segments	No. Sample Households
Niassa	52	1,248	25	600	27	648
Cabo Delgado	52	1,248	8	192	44	1,056
Nampula	68	1,632	16	384	52	1,248
Zambézia	68	1,632	10	240	58	1,392
Tete	52	1,248	12	288	40	960
Manica	52	1,248	25	600	27	648
Sofala	52	1,248	20	480	32	768
Inhambane	52	1,248	12	288	40	960
Gaza	52	1,248	19	456	33	792
Maputo Província	52	1,248	30	720	22	528
Maputo Cidade	52	1,248	52	1,248	0	0
Mozambique	604	14,496	229	5,496	375	9,000

In order to determine the sample size requirements for the 2008 MICS, the INE statisticians first examined the estimates of sampling errors and design effects for key indicators included in the 2003 DHS report that were also covered by the MICS. The following indicators were selected from Appendix B of the 2003 DHS report for this purpose:

1. Literacy rate for women 15-49 years
2. Proportion of women 15-49 years who are currently using any type of contraceptive
3. Proportion of women 15-49 years who are currently using any modern type of contraceptive
4. Proportion of children under 5 whose birth was attended by medical staff
5. Proportion of births during the past 5 years protected by anti-tetanus vaccinations
6. Proportion of children 12-23 months who received a measles vaccination
7. Proportion of children 12-23 months who were fully immunized
8. Infant mortality (less than 1 year)
9. Child mortality rate (less than 5 years)

It was decided that the level of precision for the 2003 DHS results for these indicators by province was generally satisfactory, so there was no need to increase the overall sample size for the MICS. This sample size was also about the maximum that could be supported with the available resources. However, it was possible to improve the level of precision slightly with a more efficient sample design and allocation, which resulted in a more dispersed sample. The number of sample EAs in most provinces was increased from 52 to 60 compared to the 2003 DHS. As in the case of the 2003 DHS, a larger sample was selected for the provinces of Nampula and Zambézia, which have more than twice the population of the other provinces. For these two large provinces, the number of sample EAs was increased to 80 each. The number of sample EAs for the City of Maputo was increased to 75, given the larger variability of socioeconomic characteristics in this city. The availability of a new listing of households for a larger representative sample of EAs within each stratum for the IOF 2008/09 reduced the cost of increasing the dispersion of the 2008 MICS sample.

While the total number of sample EAs for the 2008 MICS was increased to 715 compared to 604 for the 2003 DHS sample, the number of households selected per sample EA was decreased to 20, compared to 24 for DHS. This should result in a slight reduction in the design effects (due mostly to clustering) for the indicators presented in the DHS report, thus improving the efficiency of the sample design and the level of precision for the 2008 MICS results. The overall number of sample households for the 2008 MICS actually decreased slightly to 14,300 compared to 14,496 for the 2003 DHS. The final allocation of the sample EAs and households for the 2008 MICS by province, urban and rural strata is presented in Table 23.

Table 23
Distribution of Sample EAs and Households for 2008 Mozambique MICS by Province, Urban and Rural Strata

Province	Total		Urban		Rural	
	No. Sample Segments	No. Sample Households	No. Sample Segments	No. Sample Households	No. Sample Segments	No. Sample Households
Niassa	60	1,200	24	480	36	720
Cabo Delgado	60	1,200	15	300	45	900
Nampula	80	1,600	24	480	56	1,120
Zambézia	80	1,600	14	280	66	1,320
Tete	60	1,200	12	240	48	960
Manica	60	1,200	21	420	39	780
Sofala	60	1,200	33	660	27	540
Inhambane	60	1,200	21	420	39	780
Gaza	60	1,200	24	480	36	720
Maputo Província	60	1,200	45	900	15	300
Maputo Cidade	75	1,500	75	1,500	0	0
Mozambique	715	14,300	308	6,160	407	8,140

The allocation of sample EAs to the urban and rural strata within each province for IOF 2008/09 presented in Table 21 was proportional to the number of households in each stratum, but giving a higher weight to the urban stratum. Given the higher transportation costs in rural areas and the higher variability of socioeconomic characteristics in urban areas, it was reasonable to use a higher sampling rate for the urban strata. The number of sample EAs to be allocated to the urban stratum of each province was determined using proportional allocation after the number of households in the urban stratum was given a weight of 1.5, and the result was rounded up to the next multiple of 4. The remaining sample EAs in the province were allocated to the rural stratum. Given the nature of the subsample of EAs for the MICS, it has the same relative distribution of the EAs by urban and rural stratum within province as the IOF.

Given the greater dispersion of the sample EAs and the resulting slightly lower design effects for the 2008 MICS compared to the 2003 DHS, it is expected that the standard errors and corresponding margins of error will be slightly lower for the MICS results. In order to determine the level of precision for indicators other than the list above which had measures of precision in the DHS report, we can use the MICS4 spreadsheet for calculating the sample size, only we would change the formula to input the sample size and determine the resulting margin of error (ME) or relative margin of error (RME). For example, in the case of a rate indicator with an estimate of 0.2 for a subgroup that represents 15 percent of the population, with a design effect of 1.5 and a nonresponse rate of 10 percent, as shown in the example for the sample size template, the resulting RME based on a sample of 14,300 households at the national level would be 0.05, less than half the target of 0.12. With a sample of 1,200 households for a province, the resulting RME would be 0.17, which is reasonable according to the guidelines for subnational estimates presented in this sampling chapter.

In comparing the final allocation of sample EAs by stratum for the 2008 MICS in Table 23 to the corresponding sample distribution for the IOF 2008/09 in Table 21, we see that the number of MICS sample EAs in most strata is 75 percent of the corresponding number of IOF sample EAs. The only exception is in the provinces of Nampula and Zambézia, where the number of MICS sample EAs is half that for IOF. Given that the IOF had four nationally-representative subsamples of EAs assigned to the different quarters, it was practical to use the first 3 quarterly subsamples of IOF for the MICS in all provinces except for Nampula and Zambézia, where the first two quarterly subsamples of IOF were used. In this way it was also easier to coordinate the timing of the listing of households in the overlapping sample EAs that would be used for selecting the households for both surveys.

Sample Selection Procedures

For the first stage of selection the EAs within each stratum (province, urban/rural) were selected systematically with PPS for the IOF 2008/09, where the measure of size for each EA was based on the number of sample households in the preliminary 2007 Mozambique Census frame. The EAs within each stratum were first ordered geographically to provide implicit stratification. The

first stage selection of EAs in each stratum was implemented using an Excel spreadsheet, based on the systematic PPS selection procedures presented earlier in this chapter. This methodology was very similar to that described for the Lebanon MICS in the previous example. The IOF sample EAs within each stratum were then systematically subdivided into four representative subsamples, so that the implicit stratification was maintained for each subsample. Then one subsample was randomly assigned to each of the four quarters for the IOF data collection. In the case of all provinces except for Nampula and Zambézia, the subsamples of EAs for the first three quarters were chosen for the MICS, while for Nampula and Zambézia the subsamples of EAs for the first two quarters were chosen for the MICS. Based on these sampling procedures, the resulting sample for MICS is similar to having selected the EAs systematically with PPS within each stratum from the original frame.

At the second sampling stage a systematic sample of 20 households was selected for MICS in each EA. This was coordinated with the selection of households for the IOF so that a different set of households was selected for each survey. Given that the MICS fieldwork was generally conducted in a sample EA prior to the IOF data collection, it was necessary to select the sample households in the field soon after the listing operation. For this reason a practical approach was developed for pre-selecting the systematic sample of households for both surveys, using a household selection table. This table specified the household serial numbers to be selected based on the total number of households listed in the EA. The first column of this table identified the total number of households listed. Following the listing operation in the sample EA, the MICS field supervisor determined the total number of households listed and identified the corresponding row in the household selection table, which would identify the serial numbers of the 20 selected households. This avoided the complication of having the field staff use a table of random numbers in the field and possibly introduce bias in the sample selection. The household selection table was generated using an Excel spreadsheet, which calculated the sampling interval, and generated the random start and selection numbers.

In order to coordinate the selection of a different sample of households for the MICS and IOF in the overlapping sample EAs, the systematic sample of households in each EA was actually pre-selected for both surveys at the same time. This was accomplished by selecting a systematic sample of households for both surveys, and then a systematic subsample of households for the individual surveys. Given that the number of households selected per EA for IOF was 12 for the urban EAs and 9 for the rural EAs, the total number of households to be selected in each EA varied by stratum. For IOF it was also necessary to select a systematic sample of 4 reserve households for each EA to replace any non-interview households. In the case of IOF a different household selection table was developed for the urban and rural EAs, while one household selection table was produced for selecting the 20 households in a sample EA for MICS.

The procedures for coordinating the selection of households for the two surveys are described here for the urban EAs. A total of 36 households had to be selected from the listing in each urban EA that is included in the sample for both IOF and MICS: 20 sample households for MICS, 12 for IOF, and 4 reserve households for IOF. First a systematic sample of 36 households

was selected from the listing. Then a systematic subsample of 20 of these households was selected for the MICS, and the remaining 16 sample households were designated for the IOF sample. Finally, a systematic subsample of 4 of the 16 IOF households was selected for the IOF reserve sample, so that the remaining subsample of 12 households was designated as the “original” IOF sample. Excel spreadsheets were used to pre-select the systematic sample of households for both surveys based on the total number of households listed, followed by the systematic selection of subsamples for each survey. This spreadsheet with the combined sample was then used to produce the separate household selection table for each survey.

Weighting Procedures

The weighting procedures used for the MICS were similar to those used for the Lebanon MICS, given the similarity of the two sample designs. However, the MICS was strictly a two-stage sample, since it was not necessary to subdivide any large EAs into smaller segments. Using the terminology used in the survey documentation, the probability of selection can be expressed as follows:

$$p_{hi} = \frac{n_h \times M_{hi}}{M_h} \times \frac{m_{hi}}{M'_{hi}},$$

where:

p_{hi} = probability of selection for the sample households in the i-th sample EA in stratum (province, urban/rural) h

n_h = number of sample EAs selected in stratum h for MICS, shown in Table 23

M_h = total number of households in the frame for stratum h (based on the preliminary 2007 Mozambique Census frame)

M_{hi} = number of households in the frame for the i-th sample EA in stratum h (based on the preliminary 2007 Mozambique Census frame)

m_{hi} = 20 = number of sample households selected in the i-th sample EA in stratum h for MICS 2008

M'_{hi} = number of households listed in the i-th sample EA in stratum h

The two components of this probability of selection correspond to the individual sampling stages. The basic design weight is calculated as the inverse of this probability of selection. Based on the above expression for the probability, the weight can be simplified as follows:

$$W_{hi} = \frac{M_h \times M'_{hi}}{n_h \times M_{hi} \times m_{hi}},$$

where:

W_{hi} = design weight for the sample households in the i-th sample EA in stratum h

A non-interview adjustment factor was applied to the design weight for the sample households at the stratum level, using an Excel spreadsheet similar to the MICS3 weighting template. An additional nonresponse adjustment factor was applied at the stratum level to the weights for the data from the women and child sections of the questionnaire based on the inverse of the corresponding response rates, also consistent with the weighting template.

The adjusted household design weights were used to estimate the total number of households and population in each stratum, and this weighted total was compared to the distribution of the population from the preliminary 2007 Mozambique Census results. These weighted MICS estimates were consistent with the census results (taking into account sampling errors and minor listing problems). This was a way of validating the calculation of the weights and the coverage of the sampling frame.

Finally, the household, women and children weights were normalized so that the sum of the relative weights is equal to the corresponding total number of sample records, according to the standard weighting template for MICS.