CHAPTER IV

CHOOSING THE SAMPLE

This chapter is written for survey coordinators and technical resource persons. It will enable you to:

- ✔ Understand the basic concepts of sampling.
- ✔ Calculate the required sample size for national and subnational estimates.
- ✔ Determine the number of clusters to be used.
- ✔ Choose a sampling scheme.

UNDERSTANDING THE BASIC CONCEPTS OF SAMPLING

In the context of multiple-indicator surveys, sampling is a process for selecting respondents from a population. In our case, the respondents will usually be the mothers, or caretakers, of children in each household visited,[1] who will answer all of the questions in the Child Health modules. The Water and Sanitation and Salt Iodization modules refer to the whole household and may be answered by any adult. Questions in these modules are asked even where there are no children under the age of 15 years.

In principle, our survey could cover *all* households in the population. If all mothers being interviewed could provide perfect answers, we could measure all indicators with complete accuracy. However, interviewing all mothers would be time-consuming, expensive and wasteful. It is therefore necessary to interview a *sample* of these women to obtain *estimates* of the actual indicators.

The difference between the estimate and the actual indicator is called *sampling error*. Sampling errors are caused by the fact that a sample—and not the entire population—is surveyed.

Sampling error can be minimized by taking certain precautions:

- Choose your sample of respondents in an unbiased way.
- Select a large enough sample for your estimates to be precise.

Choose your sample from all the households. Avoid choosing samples which might result in *biased* estimates.

[1] A *household* is a group of persons who live and eat together.

✎ **EXAMPLE:**

Do not choose samples exclusively from particular groups, such as children coming to clinics.

Do not ask mothers to bring their children to a central point in the community, because some of them will not come; you will not be able to find out how many failed to appear and how different they may be from those who came.

Do not use samples chosen at will by the interviewer, field supervisor or field director.

Do not restrict your sample to families living in easily accessible households, such as those close to a main road or near a village center; families living in less accessible areas may be poorer and less healthy.

Do not omit households where no one is at home the first time you call. Find out if the household is inhabited, and revisit at a later time.

If a sample does not accurately represent the whole population of interest, the estimate will be *biased*. That is, the estimate will be shifted in one direction or another from the true value.

To avoid bias you should use *probability sampling* to select your sample of respondents. Bias depends on the selection procedure, not on sample size. Larger samples tend to be more precise but are not necessarily less biased.

Probability sampling ensures that all individuals in the target population (for example, all mothers of young children) have a known chance of being interviewed.

In order to take a probability sample, you must have a *sampling frame*. This is a list of all the units (states or provinces, districts, communities, households, individuals, etc.) in the population from which you will choose your sample. Virtually no country has updated listings of all individuals and only very few have household listings. Your sampling frame, therefore, will typically include listings of larger population units such as states or provinces, districts, towns, villages or census enumeration areas. You must also have an estimate of the population or number of households for each of these units as well as ensuring that *all* the units are included in the sampling frame.

Since in multiple indicator surveys we are usually interested in obtaining national-level indicators, it is necessary to employ *multi-stage sampling*. For example, you might start by selecting a number of provinces (first stage), then a number of districts in each of the chosen provinces (second stage), a number of communities in each selected district (third stage) and, finally, a number of households in each chosen community (fourth stage). Each of these selections may use one of the different types of sampling listed below. Further on in this chapter, we will provide some actual examples of how different types of sampling were combined in multiple stages.

Simple Random Sampling

Simple random sampling is the simplest form of probability sampling. Random numbers are chosen using a calculator, a computer program or a random number table. Alternatively, the names or identification numbers of all communities, households or individuals could be written on pieces of paper and the desired sample could be selected by picking the required number of papers. Each community, person or household corresponding to the numbers chosen is then included in the sample. In simple random sampling, the selection of one individual is *independent* of the selection of another individual.

Systematic Sampling

Systematic sampling is a modification of simple random sampling. It consists of picking every *n*th (5th, or 10th, or 50th and so on) household from a complete list of households. When employing systematic sampling, you must ensure that the list you use is not ordered in any regular way which would bias your sample.

> ✎ **EXAMPLE:**
> If the list is ordered village head, subhead, other households, then another village head, subhead, other households, and so on, systematic sampling may bias your results. For instance, if you always start at the fifth household in each village you will never include any village heads or subheads.

Cluster Sampling

As already noted, in countries where surveys are necessary to measure goal indicators, lists of households or individuals are not usually available. It is usually too expensive and time-consuming to construct such a list and then to locate the households. The logical choice, then, is to use cluster sampling—that is, to select groups of households which are geographically close to one another.

Cluster sampling usually involves at least two stages of selection.

✎ **EXAMPLE:**

A number of communities[2] may be initially selected from a sampling frame (list of communities), and then a cluster is randomly chosen within each of these communities. All households in those clusters are then interviewed.

Households are the *basic sampling unit* (that is, the smallest unit to be sampled). One might also use children under age five as the basic sampling unit, but households are much easier to identify and to list than children under five years of age. The final choice of households is made in the community. Different ways to choose households within a community are discussed in chapter 6.

Cluster sampling is always used in multiple-indicator surveys.

Cluster sampling has many advantages. It reduces time and travel costs and simplifies field work. It also eases field supervision and survey administration. This is important since better supervision of interviewers will result in better data quality. However, cluster sampling implies that each respondent is not chosen independently of the other respondents—after all, they are neighbors—and this may increase the sampling error.

Sampling with Probability Proportional to Size

Communities may vary considerably in population. If simple random or systematic sampling is used, both large and small communities will have the same probability of being included, which is incorrect. One way of compensating for these differences is to choose clusters from the sampling frame with probability proportional to size (PPS).

Communities with larger populations should have a proportionately greater chance of containing a selected cluster than smaller communities.

One advantage of PPS is that, if properly used, each household in the sample will have an equal chance of being selected. The sample is then said to be *self-weighting*, which will simplify the analysis.

The procedure for PPS sampling is described in the following section. You may use PPS selection at most stages of your sampling scheme. For example, you may use PPS sampling to select states or provinces in a country,

[2] A *community* is defined as a grouping within a population, such as a town, village or census enumeration area. In this manual we use the term *cluster* to mean a randomly selected sample from the community. This is the EPI usage of the term; in statistical literature, the cluster *is* the community, and other publications might refer to it in that way.

municipalities within a state or communities or villages within a district. The example below deals with selecting communities from a district but the procedure is exactly the same whatever the sampling units you choose.

Procedure for Sampling with Probability Proportional to Size

To select the communities to be used in the survey you will need:

- a list of all communities in the district to be surveyed (in this example, every household in the district belongs to one—and only one—community)
- an estimate of population size (or the number of households) in each community[3]
- the number of clusters you want to have in the district (discussed later)

To sample the communities with PPS, complete the following seven steps:

1. **Make a table with three columns, as shown in the illustrative table on the following page.**

Column (1): Assign a number to each community. List the communities.
Column (2): List the population of each community.
Column (3): List the cumulative population of each community—that is, the sum of the population of that community plus the populations of all the communities above it in the table.

☞ When listing the communities, try to use a logical order, such as a geographical pattern. See the discussion about *implicit stratification*, below.

2. **Calculate the *sampling interval*** using the following formula:

Sampling interval = cumulated total population ÷ number of clusters required

Using the example above, and supposing we need to select three clusters, this would give:

Sampling interval = 6,700 ÷ 3 = 2,233

[3] If the average size of households does not vary greatly from one community to another, you may use the following formula to estimate the population in each community: *population size = number of households × average size of a household*.

ILLUSTRATIVE TABLE:
Cumulating Community Populations

Village	Population size	Cumulative
1	1,000	1,000
2	400	1,400
3	200	1,600
4	300	1,900
5	1,200	3,100
6	1,000	4,100
7	1,600	5,700
8	200	5,900
9	350	6,250
10	450	6,700

3. **Select a random number which is equal to or less than the sampling interval**. For the above example, we need to choose a random number between 1 and 2,233. You can select this random number by using a calculator, a computer, a table of random numbers from a statistics textbook or even by asking someone (who is not familiar with the survey) to choose a number within this range. Let us suppose that the chosen number is 1,814.

4. **Look back at the table**. Locate the first cluster by finding the community whose cumulative population exceeds this random number. In the example, the first cluster would be located in village 4, where the cumulative population is 1,900.

5. **Add the sampling interval to the random number.** In the example, 2,233 + 1,814 = 4,047.

6. **Choose the community whose cumulative population just exceeds this number**. The second cluster will be located in this community. In the example, the second cluster will be located in village 6.

7. **Identify the location of each subsequent cluster** by adding the sampling interval to the number which located the previous cluster. Stop when you have located as many clusters as you need.

If a community has a population which is larger than the sampling interval, it may be selected twice. Two independent clusters must then be selected from that same community. This is perfectly valid.

Once you have done your PPS sampling and ensured that there are no errors, resist the temptation to do it again if your first sample does not please you.

Stratified Sampling (Stratification)

You may want to measure indicators for subgroups within your national population. Since some of these subgroups may have fewer individuals than others, you will want to ensure that all of them are adequately represented in your sample. Using PPS may result in small samples for groups representing a small proportion of the overall population. Stratification solves this problem. You will create one stratum for each subgroup and sample each stratum separately.

Stratification is commonly used to produce subnational data—for example, at the state or province level. If you do not stratify, your large states or provinces will have samples many times larger than the small states or provinces. To achieve equal precision for all states (see discussion below) you may select the same number of households in each state, and later weight the results according to the actual proportion of the national population in each state (see chapter 7).

In addition to providing subnational indicators, stratification can be used to generate data for subgroups living under specific conditions.

✎ **EXAMPLE:**

Urban slum dwellers may constitute only a small fraction of the population in some countries but may present special needs. Therefore, at the beginning of the sampling procedure one might list urban slum areas separately from other areas and ensure that a sufficient number of slum dwellings are included in the sample by giving them a higher probability of selection. *This sample would then have two strata: slum dwellers and all others.*

When stratification is used, the sample size in each stratum is not usually proportional to its actual population. This imbalance has to be corrected in the data analysis. For example, if slum dwellers constitute 5 per cent of a country's population but your stratification procedure results in a sample with 10 per cent belonging to that category, the sample data will have to be *weighted* to produce national results.[4]

---

[4]In this example, each slum-dwelling child would be given a weight of 0.5 (since 5 per cent is one half of 10 per cent); also, non-slum dwellers would be given a weight of 0.95/0.90 because they account for 95 per cent of the population but only for 90 per cent of the sample. The general formula for the weight is equal to (proportion in the population)/(proportion in the sample). See chapter 6.

Implicit Stratification

Implicit stratification is a special form of stratification. It consists of listing the units to be sampled (for example, clusters) according to a logical order and then sampling systematically or with PPS from that listing.

> ✎ **EXAMPLE:**
> You could follow a geographical pattern from north to south or west to east, or you could start from the district capital and move away from it.

This ensures that the units will be more or less evenly spread within the listing and avoids the possibility that, due to chance, one type of community ends up being under-represented. This logical order may be geographical or may follow the distribution of a relevant variable, such as proximity to health services or socioeconomic indicators.

In addition to helping produce reliable results at the subnational level, stratification is a useful tool for improving the overall precision of your national estimate.

The above-discussed types of sampling—simple random, systematic, cluster, PPS and stratification—are by no means exclusive. As the examples below illustrate, they may be combined at different stages of sampling to produce national estimates.

HOW TO CALCULATE THE SAMPLE SIZE

Sample size estimation is a simple procedure, yet many people are put off by its apparent complexity. This section will lead you through the various steps required for this calculation. At the end, a spreadsheet program is introduced which carries out these calculations. Understanding the basic principles of sample size estimation, presented below, is essential for the proper use of the spreadsheet.

Surveying a sample of children will give you an *estimate* of the actual goal indicator, such as the proportion of households using iodized salt. Due to sampling error, this estimate is not an exact measurement. The use of probability sampling, however, allows us to calculate a range in which the actual indicator can be reasonably expected to fall. The upper and lower limits of this range are known as the *margin of error*.[5]

> ✎ **EXAMPLE:**
> The classical EPI surveys estimated vaccine coverage with a margin of error of plus or minus 0.10, or 10 percentage points. This means that, if our survey found a BCG

---

[5]The term ‘‘margin of error,” used in this document, may be called ‘‘level of precision” or ‘‘confidence interval’’ in other publications.

coverage of 75 per cent, we could be reasonably confident that the actual population value would fall between 65 and 85 per cent.

By "reasonably confident," we mean that there will be only a 5 per cent chance that the actual coverage in this population is outside the margin of error determined by the survey. In other words, we can be confident that in 95 out of 100 surveys the true rate in the population would lie within this margin.

For sample size estimation, you must obtain or estimate some basic figures. Follow the steps below to calculate the required sample for each goal indicator. We will use BCG vaccine coverage as an example. Remember that *in the calculations you must always enter percentages as decimal numbers* (for example, enter 0.6 for 60 per cent).

1. **Guess/anticipate the *proportion* you are about to measure.** Using any available information, make a rough estimate of the proportion of children who are vaccinated. This is called the *anticipated proportion*.

✎ **EXAMPLE:**
You may guess that 80 out of every 100 children born have been vaccinated. Therefore, your anticipated proportion will be 0.8.

☞ You probably do not have good estimates of current coverages. (That is exactly the reason you are doing the survey!) However, even approximate information is adequate. When in doubt, err toward 0.5 because estimates close to 0.5 require larger samples for a given margin of error than either larger or smaller coverages.

✎ **EXAMPLE:**
If you think coverage is between 60 and 70 per cent, choose 0.6.

2. **Decide on the acceptable *margin of error* for your estimate.** That is, decide how close to the actual rate your estimate should be. For national goal monitoring, a reasonable margin of error is plus or minus 5 percentage points,[6] but for subnational estimates you may be satisfied with 10 percentage points.

---

[6]In this text, margins of error are expressed in terms of an absolute number of percentage points. For example, a coverage estimated at 60 per cent with an error of plus or minus 5 percentage points means that the actual coverage should lie between 55 and 65 per cent.

✎ **EXAMPLE:**

You may want to estimate BCG coverage within plus or minus 5 percentage points. Therefore, if your surveys finds a coverage of 82 per cent, the actual population coverage is likely to be between 77 and 87 per cent.

☞ The smaller the acceptable margin of error, the larger the sample size used in the survey must be.

You must ensure that your sample is large enough to achieve the absolute margin of error that you require. For national goal monitoring, margins of error of, say, 15 percentage points or greater are probably too wide to be of any use. But you must also balance the need for precision against the costs involved in doing a very extensive survey.

✎ **EXAMPLE:**

If you want to estimate BCG coverage with an error of plus or minus 1 percentage point, you would require over 10,000 children aged 12–23 months. This would imply visiting 70,000 households in a typical developing country, a task which most countries would not dare to attempt.

3. **Choose an appropriate *design effect.*** When cluster sampling is used, respondents are not chosen completely independently of the other respondents in the same neighborhood.

✎ **EXAMPLE:**

If children within a cluster are like each other with regard to their vaccine status, we say they are "homogeneous." This means that if one given child has been vaccinated it is likely that his or her neighbors will also have been vaccinated. Homogeneity also applies to unvaccinated children: other children in the same cluster will also be likely to be unvaccinated.

If children are more homogeneous (or alike) within a cluster than in the whole population, it will result in a larger sampling error than if the same children had been obtained through a strict random sample—that is, if they had been selected completely independently of each other.

✎ **EXAMPLE:**

A sample of 1,000 children obtained through cluster sampling would lead to a less precise coverage estimate than another sample of 1,000 children chosen through simple random sampling.

Some indicators will be more affected by clustering than others. Vaccine coverage may be affected because of distance from health centers and by local immunization campaigns.

Homogeneity is particularly severe for indicators of incidence of infectious diseases such as measles (which is spread from child to child) as well as variables related to water and sanitation (since neighbors often share the same sources of water and sanitation facilities).

The mathematical expression of this clustering is called the *design effect*. It depends both on the degree of similarity among children within a cluster and on the size of the clusters. Other things being equal, large clusters will lead to a greater design effect than small ones.

Since sample size calculations are usually based on simple random samples, the size of the sample must be increased to compensate for a large design effect. Using experience from other cluster surveys, we usually allow for a design effect of 2.0 for most variables, which requires us to double the sample size in comparison with a simple random sample. For water and sanitation variables, we have used a design effect of 10.

☞ It is very hard to predict what your design effect will be before carrying out the study. Unless you have good reason to the contrary, use the figures suggested above. After your survey is finished, you can calculate the design effect (see chapter 7) and use it to calculate your actual margin of error.

4. **Calculate the sample size using the formula:**

$$\text{sample size} = \frac{4 \times \text{proportion} \times (1 - \text{proportion}) \times \text{design effect}}{\text{margin of error} \times \text{margin error}}$$

✎ **EXAMPLE:**

Using this formula, and the values given in the examples above for BCG coverage, where:

proportion = 0.8 (that is, 80 per cent)
design effect = 2
margin of error = 0.05 (±5 percentage points)

we have:

$$\text{sample size} = \frac{4 \times 0.8 \times (1 - 0.8) \times 2}{(0.05 \times 0.05)} = 512$$

So the sample size should be 512 children aged 12–23 months. That is, to obtain an estimate with a margin of error of plus or minus five percentage points (a 95 per cent chance that the true rate falls within 75 per cent and 85 per cent), you should interview mothers of 512 one-year-old children.

These calculations must be repeated for each of the indicators being measured in the survey. This can be done with the help of a spreadsheet program (see Box 4.1, beginning on page 4.14).

5. **Estimate the number of households you must visit.** The above formula and the spreadsheet provide the number of members of the target population required for measuring the indicator. Note that the target populations vary according to each indicator—for example, children aged 12–23 months for vaccine coverage, children with recent diarrhoea for oral rehydration therapy (ORT) use rate, households for water and sanitation, and so on—so that a common denominator is necessary.

☞ By calculating the number of households that you must visit to find the required number in the target population, the sample size requirements for all indicators may be compared. The largest required number of households should then be chosen. This guarantees that all other indicators will also be measured with equal or smaller margins of error.

To estimate the number of households required for finding a given number of children, you must know the average household size and the proportion of children in the population. Note that the larger the households and the greater the birth rate, the fewer households you will have to visit to find a sufficient number of children in any age range.

✎ **EXAMPLE:**

If the average household size is six persons, and 3 per cent of the population is made up of children aged 12–23 months, in order to find 512 children in this age range you must visit 512 ÷ (6 × 0.03) = 2,844 households.

☞ Check the last available census to find out the average household size and the proportion of children under 5 years of age in the population. If you do not have such data, typical figures are six persons per household and 3 per cent of the population being children in each one-year age group (e.g., aged 0–11 months, 12–23 months, etc.), so that 15 per cent of the population will consist of under-fives. You may also use 25 per cent as the frequency of diarrhoea in the preceding two weeks. Table 4.1 provides guidance on estimating the number of households for measuring the mid-decade goal indicators.

Add at least 10 per cent to your estimated sample size to allow for losses.

6. **Allow for possible losses.** In all surveys, some of the planned interviews fail to take place because people are absent from their homes and because of refusals. Strategies for minimizing these losses are discussed in chapter 6. When calculating the sample size, however, it is wise to add to the sample the expected proportion of losses. These may vary widely from country to country, and from urban to rural areas, particularly where maternal work outside the home is common. You must estimate what this proportion is from previous surveys in the country.

✎ **EXAMPLE:**

If your estimated sample is 2,844 households, add about 10 per cent and try to reach 3,100 households.

Table 4.1. Typical numbers of households to be visited for finding, on average, one individual belonging to the target population for each mid-decade goal indicator

Indicator	Target population	Per cent of whole population[a]	Number of households to be visited to find one individual[b]
TT2 coverage	Mothers of children 0–11 months	3%	5.56
Other vaccine coverages	Children 12–23 months	3%	5.56
Vitamin A coverage	Children 0–23 months	6%	2.78
Iodized salt consumption	Households	—	1.00
Per cent with low weight or height for age	Children 0–59 months	15%	1.11
Use of ORT in diarrhoea	Children 0–59 months with diarrhoea in the previous two weeks	4% (25% of 15%)	4.17
School enrollment	Children 5–9 years	15%	1.11
Water and sanitation	Whole population	100%	0.17

[a]Typical estimates for less developed countries.
[b]Assuming six individuals per household.

The discussion on sample-size estimation is followed below by three specific instances of sample sizes: for estimating trends (Box 4.2), for subnational analyses (Box 4.3) and for subgroup—for example, gender-specific—indicators (Box 4.4).

SUMMARY OF SAMPLE-SIZE CALCULATIONS

Based on a number of simulations using data from different countries, measuring Mid-Decade Goals (MDGs) with a precision of 5 percentage points will require a sample of about 4,000–5,000 households.

The total population of your country is not important for defining the sample size.

Box 4.1
SAMPLE SIZE ESTIMATES FOR MEASURING MID-DECADE GOALS

Instructions for using the Spreadsheet SAMPGOAL.WK1

Data Entry

Please follow these steps to enter your data:

1. Run Lotus or Quattro-Pro (or any common spreadsheet program) and open the SAMPGOAL.WK1 file.
2. Enter data in the highlighted cells only. Note that all percentages or proportions should be entered as decimals (e.g., 15% should be entered as 0.15).
3. Provide country-specific data on the following basic assumptions:

   Cells C4 and D4. These are the estimated design effects due to the use of cluster sampling. The program defaults are 2 for most variables and 10 for the water/sanitation indicators (since the latter are likely to be more affected by clustering). Unless you have local information on these, please leave them as they are.
   Cell C5. Average number of persons per household, obtained from local census or survey.
   Cell C6. Per cent of the population under 5 years of age, also from local census or survey.
   Cell C7. Prevalence of diarrhoea in the preceding 15 days among children under 5 years of age. Obtain this number from previous studies in the country. If you do not have any data, use the default value of 0.25.
4. Provide the data on the estimated coverage level for each goal indicator in cells D12 through D24. You may change the indicators (columns A and B), but please make sure that they refer to the same target population (column C) or you are likely to get wrong results.
5. Enter the margin of error you are prepared to accept in cells E12 through E24. This should be expressed as percentage points.

Results

The spreadsheet calculates the target sample (in column F)—for example, the number of children aged 12–23 months required to measure each indicator. It also uses the information provided in the top section of the spreadsheet to calculate how many households you will have to visit (column G) to find the required number of persons in the target sample. Finally, it picks the largest of these numbers of households to be the required sample size (cell G26).

Perhaps the most useful application of this spreadsheet is to perform repeated simulations, changing the parameters in order to observe the effect on sample size. In particular, small changes in the margin of error may have important effects on the required sample.

The following table shows an example of the spreadsheet.

Box 4.1 (continued)

Sample Size Calculations for Measuring Mid-Decade Goals

Basic Assumptions	**Low**	**High**
Design effect	2	10
Persons per household	6	
Pct of population <5 years	0.15	
Prevalence of diarrhoea 15 days	0.25	

Goal	**Indicator**	**Target population**	**Estimated prevalence**	**Margin of error**	**Required target sample**	**Required number of households**
1.1	DPT3 coverage	12–23 mos	0.7	0.05	672	3,733
1.2	Measles coverage	12–23 mos	0.4	0.05	768	4,267
1.3	OPV3 coverage	12–23 mos	0.5	0.05	800	4,444
1.4	BCG coverage	12–23 mos	0.8	0.05	512	2,844
1.6	TT2 coverage (pregnancy)	0–11 mos	0.3	0.05	672	3,733
5.1	Vitamin A coverage	0–23 mos	0.3	0.05	672	1,867
6.1	Iodized salt consumption	Households	0.2	0.05	512	512
7.1	Use of ORT(1) in diarrhoea	Diarrhoea <5 yrs	0.4	0.05	768	3,413
7.2	Use of ORT(2) in diarrhoea	Diarrhoea <5 yrs	0.5	0.05	800	3,556
11.1	Percent low weight/age	All <5 yrs	0.4	0.05	768	853
12.4	School enrolment	5–9 yrs	0.7	0.05	672	747
13.1	Safe water	Population	0.6	0.05	3840	640
13.2	Sanitation	Population	0.2	0.05	2560	427
Required number of households						**4,444**

Box 4.2
SAMPLE SIZE CALCULATIONS FOR MEASURING TRENDS

The foregoing discussion refers to sample size calculations necessary for estimating a proportion within a given precision. Such calculations are adequate for many Mid-Decade and Year 2000 goals, such as reaching 80 per cent immunization and ORT use rates. Some of the goals, however, are expressed as expected reductions, such as decreasing the prevalence of malnutrition by 20 per cent between 1990 and 1995, or closing the education gap by one-third in the same five-year period. This assessment requires at least two surveys at different points in time. Since many countries do not have reliable data for 1990, they will not be able to measure these goals.

The sample size calculations for measuring such *trends* are somewhat more complex and are not be discussed in detail. There are two useful rules of thumb, however:

- If your two samples are completely unrelated (e.g., different clusters are used in the two surveys) and if your estimate has an error of plus or minus five percentage points, then the smallest trend difference you should be able to detect would be seven percentage points, or 1.4 times greater than your original margin of error.
- If you use the same clusters in both surveys, your precision improves and you should be able to measure a change of the same size (for example, five percentage points) as your margin of error.*

Using the same clusters in repeated surveys has advantages and disadvantages. In addition to reducing sample size, it simplifies field work because the areas will already have been mapped in the first survey. On the negative side, surveys may include educational messages or raise the community's awareness of health problems and therefore might lead to changes in health behavior (known as the "Hawthorne effect"). Subsequent surveys in the same areas may, therefore, be misleading because these communities may no longer be representative of the country. When considering repeated surveys, each country should weigh these advantages and disadvantages.

***Technical note**: This assumes a correlation of 0.5 between the cluster-specific estimates in the first and second surveys.

Box 4.3
SAMPLE SIZES FOR SUBNATIONAL ESTIMATES

Thus far, we have been concerned with sample sizes necessary for generating national estimates of indicators for the Mid-Decade Goals. Several of these indicators are based on specific (and often small) population subgroups such as children aged 12–23 months, or children who recently had diarrhoea. Most countries, however, will also want to use the surveys to provide subnational estimates—for example, at the level of regions, states, provinces or districts. These data may be used for detecting areas where greater efforts are required, as well as for programming and evaluation purposes.

The most limiting factors affecting national sample size are that immunization is based solely on children aged 12–23 months and that ORT use refers to children with diarrhoea in the preceding two weeks.

A possible solution is to report on indicators at subnational levels for *all under-fives* (or children aged 12–59 months for immunization)* as well as to accept a margin of error of plus or minus 10 percentage points. This would reduce the required sample sizes at subnational levels to about 300 households per district/region. A separate version of the spreadsheet, named DISTRICT.WK1, is provided for these district-level calculations.

With 300 households, ORT use would still not be measured properly. Two options are:

1. Extend the 15-day recall period to one month, thus raising the percentage of under-fives with diarrhoea from about 25 per cent to about 40 per cent in a “typical” district. This might lead to some bias due to selective recall of more severe episodes, which tend to be more intensively treated, but it should not affect the ranking of the districts. Even with one-month recall, you would still need 500–600 households per district for a margin of error of 10 percentage points, which will require changing the first question in the diarrhoea module to refer to the last month. An extra question will be required for separating episodes starting up to 15 days before the interview (for reporting on national Mid-Decade Goals) from the remainder.
2. A second option is to ask local health planners to group districts which they believe have similar ORT use rates and to analyse the data at this multi-district level.

By expanding the target populations for mid-decade goal indicators, you will be able to produce locally useful information for programming at low cost, and yet be able to report nationally on the narrow subgroups required.

*Widening the age groups may lead to some loss of validity in selected indicators due to poor recall, but this possible bias is likely to be small from the health planner's point of view. Also, this bias might be expected to operate uniformly in all districts so that ranking of areas for prioritizing interventions would not be affected.

Box 4.4
SAMPLE SIZES FOR SUBGROUP ANALYSES

Subgroup analyses may include breakdowns of the indicators by gender, socioeconomic group and so forth. Indicators based on subgroups will be less precise than those calculated for the whole sample.

The smaller the subgroup, the less precise the estimate.

The examples below show how the margins of error increase for smaller subgroups. Based on an overall sample with a margin of error of plus or minus five percentage point for a given indicator,* one would have:

- A margin of error of approximately ±6.3 percentage points for gender-specific indicators (given 50 per cent boys and 50 per cent girls in the sample).
- A margin of error of approximately ±8.6 percentage points for a subgroup making up 20 per cent of the sample (for example, a given socioeconomic category).

These results show that if the overall margin of error is about five percentage points, reasonably precise results will also be obtained for gender-specific indicators as well as for other subgroups making up 20 per cent or more of the whole sample.

*__Technical note:__ Based on estimating a coverage of 50 per cent and a design effect of 2.0. Also assumes that the coverage in each subgroup will be close to 50 per cent.

Decide on the Total Number of Clusters

A basic consideration in deciding the size of your clusters is to estimate the number of households which can be covered in one day by your team of interviewers. You do not want your field workers to spend more than a day in a given cluster, nor do you want them to finish their interviews at, say, 3 P.M. and have no further work to do for the rest of the day.

✎ **EXAMPLE:**
Let us assume that your interview takes 10 minutes for each child under 5 years and 3 minutes for children aged 5–9 years (education module only), plus 5 minutes for the household questions. A typical household will include one child in each of these groups (see Table 4.1) and therefore last 18 minutes. Allow 5 more minutes to move from house to house. Considering a 6-hour working day (to allow for transportation to and from the selected area and for a midday break), one interviewer would cover about 15 households a day [(6 hours × 60 minutes) ÷ 23 minutes]. A two-interviewer team will cover 30 households a day, which then becomes the cluster size.

☞ You should do a pre-test to estimate how long each interview takes. Field workers should be well familiar with the questionnaire before the pre-test since training greatly reduces the duration of the interview. Due to distance, field work in rural areas is always much slower than in urban areas.

Typical cluster sizes range from 20 to 40 households. Do not use a cluster greater than 100 households. It is always better to use *more small clusters* than fewer, larger clusters.

The number of clusters is given by the formula:

$$\textit{Number of clusters} = \frac{\textit{Total number of households in sample}}{\textit{cluster size}}$$

✎ **EXAMPLE:**

Assuming that the interviewing team can cope with 30 households in a day, and aiming at a national sample of 4,500 households, you would need 4,500/30 = 150 clusters.

Most national samples will include 100 or more clusters. For subnational estimates, try to have 15 to 20 clusters in each province, state or district. If in doubt, increase the number of clusters to be sampled.

Use a Flow Chart to Choose Your Sampling Scheme

When choosing a sampling scheme, it is desirable to aim at the most accurate and least biased sample. However, this choice is also affected by the availability of population data for building a sampling frame. Finally, time, feasibility and cost considerations also play a major role.

Unless the survey can be finished within a few months at reasonable cost, it cannot be used for goal monitoring and subnational planning.

A large number of combinations of different types of sampling may be used to produce a multi-stage sample for a multiple-indicator survey. The flow chart in Figure 4.1 (see page 4.21) shows the decisions that have to be made. This example refers to a three-stage sample in which large areas are selected first; the second stage includes selection of small areas within the large areas; and the third stage consists of the selection of households within each of the small areas. The final alternatives, labeled A to D in the bottom right corner of the flow chart, are ranked in order of preference.

☞ In order to design the sampling scheme, you will need the help of professionals who are familiar with the types of population data available in the country (such as census bureau officers or demographers), as well as sampling experts. If the latter are not available locally, international help will have to be arranged.

The ample experience with EPI surveys in the 1980s showed that a survey does not have to be perfect to produce usable data. Survey methods have to be sufficiently good for the purposes for which the results will be used.

How to Select the Large Areas

The first item on the flow chart is whether recent data are available on the population of large areas. These may be administrative divisions such as states or provinces, districts, municipalities and so on. Large areas will typically have populations ranging from several thousands to millions. To use probability sampling, it is necessary to have appropriate data on the population of these large areas—that is, a sampling frame. You have several options (in order of accuracy):

1. You can go to a recent census (say, within the last 5–10 years).

2. A recent Demographic and Health Survey (DHS) or other national household survey is another option. Although these do not usually provide population data per se, the national sampling frames have to be updated before they are carried out.

3. If no recent census or sampling frame is available but population growth since the last census has been fairly uniform throughout the country, population data from the existing census may be used.

4. If there is no recent census or other sampling frame and population growth has been variable (due to migration or disasters, for example), you should try to estimate regional population changes by consulting with local experts, including demographers, planners or workers at statistical bureaus.

Although you should attempt to have the best possible sampling frame, it does not have to be completely accurate. Later stages of the sampling scheme may help correct some of the inaccuracies in your frame.

Options 1 through 3 above provide data for using a large-area population to carry out the first stage of sampling procedure. Option 4, if completed successfully, will also allow it.

Figure 4.1. Flowchart for selecting the sampling scheme.

FIGURE MISSING

Once the sampling frame is available, you can achieve sampling at the first stage through stratification, through probability proportional to size (PPS) or with a combination of the two. This stage may also be subdivided into two or more substages.

✎ **EXAMPLE (1):**

Suppose you want separate estimates for each province in the country. You would stratify the country into its provinces and take equal-sized samples in each to obtain estimates at the provincial level. For estimating the indicators at the national level, it would be necessary to weight the provincial results according to the population in each province.

✎ **EXAMPLE (2):**

Suppose that you cannot afford to obtain separate estimates for each province, but that your country is divided into three broadly similar regions, the capital (with 30 per cent of the population), the coastal provinces (with 50 per cent) and the interior provinces (with 20 per cent). You might then select equal-sized samples in each region in order to obtain regional estimates. By later weighting the results according to the population, you would produce the national estimates.

In a few countries, particularly those under emergency situations, none of the options presented above for obtaining a sampling frame may be available (see flow chart). One should then work with local informants to stratify the country into broad categories of large areas. This approach does not provide a strict probabilistic sample due to the lack of population data. However, if no proper sampling frame is available, this option may at least provide approximate information on the levels of indicators.

✎ **EXAMPLE:**

Following the example above, the country may be divided into the capital, the coastal region and the interior region, and equal-sized samples taken in each region. Unlike the last example, it would not be possible to weight the results since you would not know the population in each stratum.

Having chosen the large areas, the next step is to select small areas.

How to Select the Small Areas

The second question in the flow chart refers to small areas. These areas include census enumeration tracts or small administrative units such as communities, villages or submunicipal units. Small areas will typically have less than 1,000 households. Data on population and boundary maps are required if one is to use probability sampling at this stage. The chart assumes that if population data are available, boundary maps will also be present.

If a recent census was carried out, data on enumeration tracts should be available. However, these may not be readily accessible to the researchers. If available, enumeration tracts have the advantages of (1) dividing the population into mutually exclusive tracts so that each household is assigned to one and only one tract, and (2) being usually of uniform size since they are often defined as the area that one enumerator can cover.

One alternative to using enumeration tracts is to use small local administrative units, so long as these include all of the population and there is no overlap between different units.

If using villages as sampling units, you should ensure that every scattered rural household belongs to a given village.

Once information about the population of small areas has been obtained, these may be sampled with PPS. The table exemplifying PPS selection in the section above shows how villages are drawn from a district.

If census enumeration areas are used and they are roughly of the same size, you may use simple random sampling instead of PPS, since every household would have approximately the same probability of being selected.

If, following the flow chart, you find that data on small areas are not available, you should work with local informants to divide the large area into broad groupings of small areas, and then sample randomly from within each grouping.

✎ **EXAMPLE:**

Let us suppose that the large area (district) has one city as well as towns, villages and scattered rural households. Let us also suppose that there are highland and lowland populations. Strata would be created according to these characteristics, if possible, with rough population estimates. The local informants would then attempt to make a complete listing of the names of all areas (e.g., towns, villages or rural neighborhoods) in each stratum. A number of small areas—based on the crude population estimates—would then be randomly selected in every stratum.

As noted above, the use of key informants does not provide a strict probabilistic sample but may be the only alternative under some circumstances.

Having selected the small areas, it is then necessary to choose the clusters containing the households to be visited.

How to Select the Clusters

Most small areas will include several times more households than our chosen cluster size. It is then necessary to sample households within these areas. The flow chart shows that there are four basic ways of doing this: enumeration (option A in the flow chart), segmentation (option B), or the EPI random walk method (further subdivided into options C and D). The main advantages and disadvantages of these methods are summarised in Table 4.2.

When choosing a method, you should consider its advantages and disadvantages, as well as local limitations, including availability of maps and population information, time and cost considerations, geographical characteristics and type of field personnel available.

Table 4.2. Some advantages and disadvantages of the main methods of selecting households within a small area

Method of selection	Advantages	Disadvantages
Enumeration	• Strict probability sample. • Adjusts for changes in population. • Does not use contiguous households and thus is closer to simple random sample.	• Needs listing and mapping every household. • Households are not contiguous so that field work takes longer. • Houses must be identified precisely.
Segmentation	• Strict probability sample. • Adjusts for changes in population. • Saves time by using contiguous households.	• Needs extra time for sketch mapping. • Use of contiguous households may affect sample homogeneity. • Supervisors need thorough training for sketch mapping and segmenting the area.
Random walk	• Saves time by using contiguous households. • Does not require mapping. • Widely known and used in many countries by non-experts. • Particularly adequate for dispersed populations and large tracts. • May produce unbiased sample.	• Not a strict probability sample. • Does not account for changes in population. • Use of contiguous households may affect sample homogeneity. • Requires thorough training of interviewers for locating houses and close supervision.

Enumeration. Enumeration consists of mapping and listing all households in the small area and then taking a sample, usually a systematic one. This is the method of choice for countries where listings and maps are available (for example, at health facilities). However, in the context of stand-alone multiple-indicator surveys, we do not recommend this method unless current household listings are available, because of the time required for the initial step of listing and mapping households (see Table 4.2). A "current" household listing should not be more than two years old.

Segmentation. The flow chart shows that the next option is segmentation. Note that, even if population data are not available but boundary maps are, segmentation may still be used.

Segmentation requires driving or walking through the whole extent of the small area and making an approximate map with the location of the households. This map does not have to be fully accurate nor is there a need to have a complete listing of all households. The map, however, should cover the whole of the small area and the segments should be mutually exclusive.

After sketch mapping, the area is divided into approximately equal-sized segments. The size of each segment is given by the formula below:

$$\textit{segment size} = \frac{\textit{actual number of households in area} \times \textit{desired cluster size}}{\textit{estimated number of households in area}}$$

✎ **EXAMPLE:**

Your estimate from the past census tells you that there should be 510 households in the area. Since you want clusters of, on average, 40 households, the area should be divided into 12.75 segments (or 13 in round numbers). Your sketch map shows that the number of households since the census has increased to 600. Divide the area into 13 equal-sized segments with 46 households (rounded from 46.15) each. Select one of these at random and include all households in your sample.

A major advantage of the enumeration and segmentation methods is that they correct for inaccurate sampling frames for the small areas.

Since it provides automatic adjustment for inaccuracies in the sampling frame and also because it is a strict probability sample, segmentation should be used whenever it is feasible to visit the whole selected area, to draw the sketch maps and to segment the area prior to data collection. Feasibility considerations should include time, costs, geographical size and availability of experienced supervisory personnel—for example, from a department of statistics or census bureau. Chapter 6 describes in detail the steps required for segmentation.

Random Walk Methods. The random walk methods are based on the original EPI cluster survey method. These methods do not produce strict probability samples, but in many circumstances may provide virtually unbiased samples of households. They are also very feasible and rapid, and in some countries may constitute the only viable alternative.

Even when maps of the small areas are available, it may not be feasible to sketch-map the households. This may occur when the sampling units at this stage (for example, census tracts or villages) cover large geographical areas and/or when households are dispersed throughout the area. Under such circumstances, the mapping exercise could be very time-consuming. You would then select option C, by randomly choosing one of several starting points on the boundary map and using that point to start the random walk. This method is further described in chapter 6.

If there are no boundary maps, a number of adaptations of the EPI random walk method may be used (option D). The original method and its limitations are discussed in Box 4.5, and chapter 6 presents a modified version of the method which overcomes some of these problems. The advantage of option C, compared with option D, is that you can use an available boundary map to allocate starting points within the whole small area. In option D, no maps are available, so you must resort to alternative ways of finding a starting point, which are described in Box 4.5 as well as in chapter 6.

COUNTRY EXAMPLES

Experience conducting multiple-indicator surveys in three countries—Bangladesh, Brazil and Jordan—is described in the following sections.

Bangladesh

Nationwide multiple-indicator surveys were carried out in Bangladesh in 1992, 1993 and 1994. These surveys were aimed at providing national data on mid-decade goal indicators as well as district-level data for programmatic purposes. Based on the sampling frame from the 1991 Census, the country was divided into 75 strata (or large areas): the rural areas of 64 districts and 11 urban domains including the four large metropolitan areas—two of which were further divided into slum and non-slum strata—and samples of other urban areas.

In each large area, all unions (administrative units of 3,000–4,000 households) were listed following a geographical order, and 13 were selected with PPS. For each selected union, all *mouzas* (in rural areas) or *mahallahs* (in urban areas) were listed. *Mouzas* and *mahallahs* are administrative subdivisions, each consisting of about 250–300 households; they are referred to as "small areas" throughout this Bangladesh survey description.

After the listing, one small area was selected in each union, again using PPS. Therefore, 13 small areas were chosen in each district or urban domain. In each of these 13 areas, the segmentation method was used. Most areas already had updated maps showing all households which had been drawn by health workers, so that sketch mapping was necessary in only a few instances. Each segment contained, on average, 40 households. One segment was randomly selected in each small area and all its households were included in the sample. Each district, therefore, included, on average, 520 households (13 × 40). This was sufficient for estimating all mid-decade goal indicators with margins of error of 10 percentage points or less.

Data from the different rural and urban samples were weighted to provide national estimates of the indicators. Since the national sample comprised 39,000 households (520 × 75), its margin of error was much smaller than that of five percentage points used in the examples above. This large sample resulted from the great number of districts in the country and from the need for district-specific data. Most countries will require much smaller samples, as noted above in the section on sample size.

Another exception in the Bangladesh surveys is the use of 13 clusters per district, while above we recommended at least 20. This choice was based on practical considerations resulting from the large number of districts and urban strata, as well as on the fact that these areas were geographically small and that the characteristics of the population within each area appeared to be reasonably uniform. The national sample included 975 clusters (13 × 75).

The Bangladesh example shows how it is possible to combine the process of obtaining national estimates for monitoring goals with the generation of district-level data for programming purposes.

Brazil

Multiple-indicator surveys were carried out in several of states in Northeast Brazil in the early 1990s. They were aimed at measuring child health and nutrition indicators for monitoring and programmatic purposes. The example that follows is based on the 1991 survey in the State of Pernambuco, with a population of approximately seven million.

A sample of 1,000 children under five years of age was required for measuring indicators with a margin of error of plus or minus three percentage points, using a design effect of 2.0. Based on information on household size and the proportion of the population under the age of five years, 1,920 households had to be visited. Since the questionnaire was extensive—covering children under the age of five as well as women aged 15–49 years—and the households were often dispersed, the cluster size was set at 12 households.

In the first stage of sampling, all municipalities ("large areas") were listed by starting with the capital city and following a geographical order. Using PPS, 20 numbers were selected. Since the capital city included 15 per cent of the population, it was selected three times (.15 × 20).

In the second sampling stage, eight census enumeration areas ("small areas") were chosen in each selected municipality through simple random sampling, totalling 160 clusters (20 × 8) for the state. In the capital, which had been selected three times, 24 enumeration areas (3 × 8) were taken. Each enumeration area includes 200–300 households, and since their size is uniform there was no need for PPS at this stage. These areas cover the whole municipality, including both urban and rural households. Four to six arbitrary starting points were plotted on the maps of each enumeration area and one of these was selected at random. Twelve households closest to this point and belonging to the enumeration area were included in the sample.

Note that this sample was self-weighting, so that there was no need to weight the estimates.

Box 4.5
THE STANDARD EPI CLUSTER SURVEY METHOD

In the standard EPI method, 30 villages were selected with PPS within a district. A starting point was selected in a central location such as a health facility. A bottle was spun at this central point and the interviewer then counted the number of households until the edge of the village and selected one at random. This is the first household in the sample. Subsequent households are selected by going to the house whose front door is closest to that house which has just been visited, until seven children aged 12–23 months (the age range of interest for vaccine coverage) were reached. This method is known at the “random walk.”

A number of concerns were raised relative to this method, including:

1. In most countries, rural populations are not neatly organised into villages with well-defined edges or boundaries, but live in more or less scattered households.
2. The scheme may oversample households close to village centers which might not be representative of the whole village. By moving from one household to the next closest one, scattered rural households are likely to be underrepresented.
3. By stopping when reaching a fixed number of children (not households), the method oversamples areas with few children (where many households had to be visited for finding seven children) and underrepresents areas with many children (where seven children were found within a few households).
4. It is very difficult to ensure that field workers have followed the instructions regarding the “random walk.”
5. The method of selection does not guarantee that every household has a known probability of selection. Therefore, it is not a strict probability sample. This is a somewhat technical point but it has caused concern among many statisticians because, strictly speaking, you cannot calculate the margins of error of the indicators.

These issues do not undermine the usefulness of the EPI-type survey for estimating vaccine coverage for programmatic purposes, where a wider range of errors can be tolerated. They show, however, that the method can be improved.

Jordan

A national survey for assessing immunization coverage, diarrhoea management and infant mortality was carried out in Jordan in 1991 following the EPI cluster survey methodology. A sample of 12,600 households was required for estimating cause-specific infant mortality. Note that this sample is considerably larger than that usually required for estimating vaccine or ORT coverage, but mortality studies often require visiting many households for identifying a sufficient number of deaths.

In the first stage of sampling, the Kingdom was divided into its seven governorates. Since separate estimates were desired for each of these, the standard EPI sample of 30 clusters per governorate was adopted. Each cluster included 60 households, which ensured that the national sample would be of the required size ($30 \times 60 \times 7 = 12{,}600$). The use of fixed samples in all governorates, independently of their population, required later weighting in the analysis.

Within each governorate, multi-stage sampling was used. First, localities were stratified into nine groups according to population size, and a number of clusters was allocated to each stratum based on its total population. The required number of localities was then selected within each stratum with PPS. Next, blocks were chosen within each selected locality and, finally, households were selected within each of these blocks.