

# CHAPTER 3

## *Sampling Instructions*



The two sampling methods described in this chapter are appropriate for a service organization that wants to collect information from women of reproductive age living in a defined geographic area. Thus, the geographic bounds of the survey area will need to be clearly identified. Normally, the area of interest is the area served by the organization and where programs and services for women and their families will be developed or improved. This is also sometimes called the “catchment” area. Some examples of catchment areas include a camp of refugees, internally displaced persons, or returnees; a group of camps; or villages or towns where displaced populations are living among local populations. In preparation for sampling, obtain or create a map that represents the geographic area to be surveyed. An example of a map of a hypothetical survey area is below (Figure 3.1).

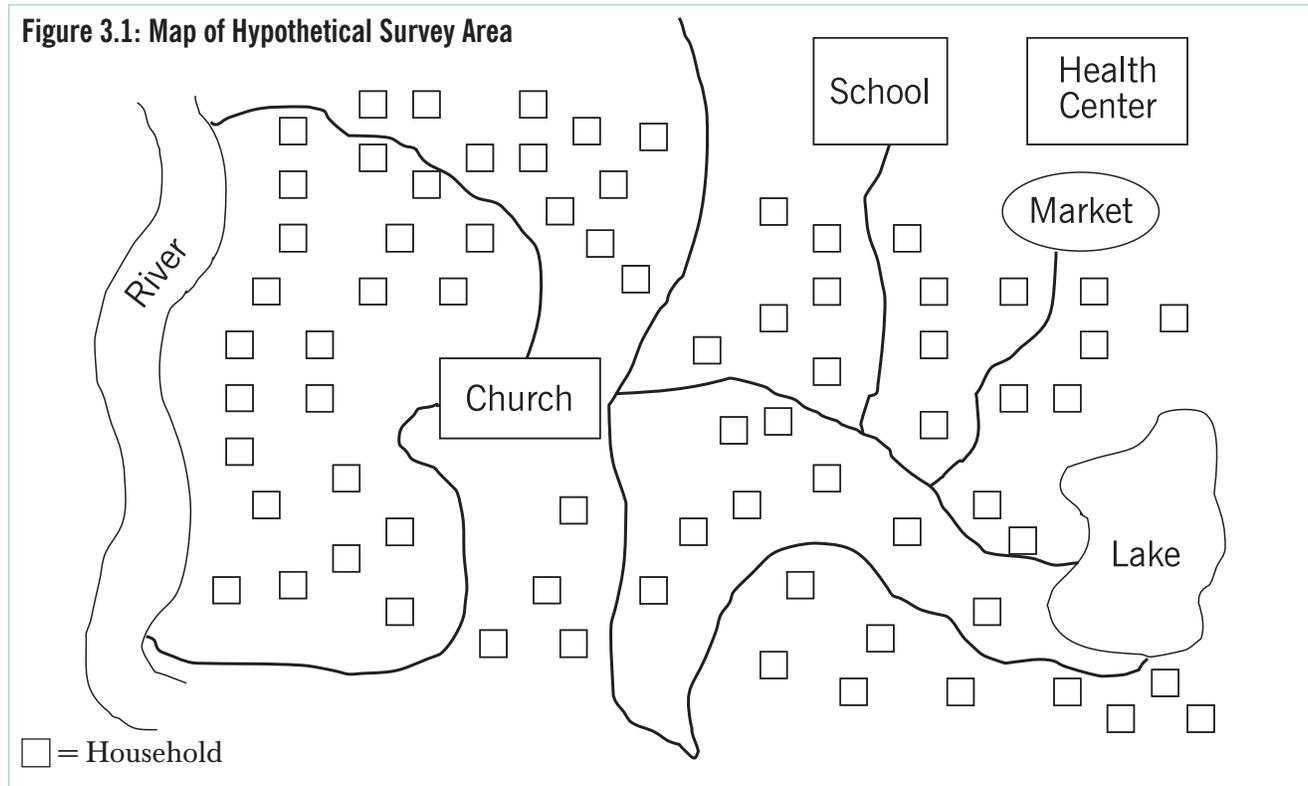
Sampling is needed when it is not feasible to interview every household with a woman of reproductive age in the catchment area in a timely fashion. Sampling means that only some of the households in the catchment area are selected for the survey. Selected households are meant to

be representative of all eligible households in the area. The concept of “representativeness” means that the selected sample of households reasonably represents the entire group. In addition, each household must have an equal chance of being selected to participate in the survey.

To accomplish this, households can be sampled using one of two suggested sampling methods: random sampling or cluster sampling. Both methods have specific requirements and each have their own advantages and disadvantages, which are discussed. The user must weigh these factors when deciding which sampling strategy to use.

These sampling instructions may be adapted to your setting and needs. Sampling plans must be developed by the study coordinator prior to carrying out the field work. Because sound sampling methods are essential to obtaining representative data from your population, DRH will review your sampling plan, if requested, before you begin data collection. DRH can also assist you in adapting the instructions, if necessary. Refer to Chapter 2, Planning Checklist, for the logistical steps of the sampling process.

**Figure 3.1: Map of Hypothetical Survey Area**



## Random Sampling

Random sampling requires household lists that contain the minimum elements described in Step 1. Though random sampling may provide more precise estimates than other sampling methodologies, it may be more difficult to implement in some situations. For example, if the area of interest covers a large geographic area, then survey teams will be required to travel greater distances to reach the selected households. This would require additional resources for staffing or fuel and may prolong the data collection period.

For random sampling, households are randomly selected, and then one woman of reproductive age is randomly selected from each household. This is called “two-stage random sampling.” In general for random sampling, you will need to have completed interviews for at least 400 women of reproductive age. By following these sampling instructions, you will meet DRH’s standards of point estimates within +/- 5% of the true population prevalence, with 95% confidence. We made the following assumptions when estimating needed sample size:

- We estimate a prevalence rate of 50%, the most conservative estimate, for all reproductive health outcomes.
- We want a 95% confidence interval  $\pm 5.0$  percentage points (based on exact binomial confidence intervals).
- We estimate a response rate of 80%, in which case 500 households must be contacted to obtain 400 completed interviews. Please note that this estimate is based on household lists where only households with women of reproductive age are identified. If women of reproductive age can not be identified before the sampling, the response rate would need to be adjusted to account for sampled households without women of reproductive age.

To conduct random sampling, you must work through the steps in Box 3.1.

### Box 3.1: Random sampling steps

- Step 1: Obtain household lists.
- Step 2: Select the households to be sampled.
- Step 3: Select one woman of reproductive age within each selected household to be surveyed.

#### *Step 1: Obtain household lists.*

Household lists may be available from registration files, census lists, ration card lists, community leaders, and other sources. The term “household” should be clearly defined and understood by all members of the survey team and should match how household is defined in the lists used for the survey. Household lists should have the following minimum elements:

- Total population
- Total number of households
- Breakdown of households by categories (camp, neighborhood blocks, etc.), if needed for stratification

Household lists may need to be updated (e.g., recording who has moved in and out of the community since the lists were created) to ensure reasonable accuracy and to eliminate ineligible households. Community leaders, community health workers, and traditional birth attendants are all examples of people in the community who may be able to help you update household lists. In some instances, lists will have detailed information that will allow you to eliminate some households before selection of your sample. For example, if the age and sex of all residents of the household is provided, you should eliminate all households with no women of reproductive age or male-only households.

#### *Step 2: Select the households to be sampled.*

Once the household lists are obtained and updated (if necessary), you are ready to randomly select the households to be surveyed. You will need to assign a number in consecutive order to each household on the lists for the selection process.

### Step 2a: Stratifying the sample

To ensure the sample of households is more representative, you can stratify your sample by dividing the sample into groups, such as ethnic groups or zones (camps or villages). Stratification reduces the risk of drawing an extreme sample that is unrepresentative of the population. For example, if there are two major ethnic populations in your camp, you may want to stratify by ethnicity, provided that your household lists include information on ethnicity. This will ensure that you have enough participants from each ethnic group in your sample to make valid comparisons. If you do not plan to stratify your sample, you can skip the rest of Step 2a and go to Step 2b.

After stratifying your sample, you can either continue to number households consecutively or restart the numbering for each group. Next, you will need to determine how many households to sample from each group. To do so, you will need to create a table similar to Table 3.1. The first column of the sample size table is filled in according to your stratification groups. For example, if you are surveying a series of camps, column 1 will list each camp. If you are sampling from one large community, column 1 will list the different neighborhoods or zones in the community. If you are stratifying by ethnic groups, column 1 would list those groups. For the following example, we will be looking at a geographic area consisting of refugees in different zones (Zone 1, Zone 2, etc.).

The second column lists the number of households within each stratification group (in this example, the number of households within each zone are listed). This number is obtained from the household list. The third column, “household proportion,” is calculated by dividing the number of households in each zone by the total sample size. The fourth column, “sample size,” is calculated by multiplying the third column by the total survey sample size to get the sample size for each zone. Box 3.2 provides an example of how to calculate the numbers in each column.

**Table 3.1: Sample size selection of eligible households by zone for random sampling.**

Zone (Camp, neighborhood, village, etc. . .)	Households	Household proportion	Sample size
Zone 1	228	15.06%	75
Zone 2	344	22.72%	114
Zone 3	223	14.73%	74
Zone 4	314	20.74%	104
Zone 5	405	26.75%	134
Total	1514	100.00%	501*

\* Total may be higher due to rounding

### Box 3.2: Example for calculating sample size for Zone 1.

Refer to Table 3.1 to work through this exercise. In this example, there are 228 eligible households in Zone 1.

**Calculation 1:**  $\frac{228 \text{ eligible households in Zone 1}}{1514 \text{ total households}} = 15.06\%$  of all households.

**Calculation 2:**  $15.06\% \text{ household proportion} \times 500 \text{ total sample} = \text{sample size of } 75.3 \text{ rounded to } 75 \text{ women to be interviewed for Zone 1.}$

- Therefore, we will randomly select 75 of the possible 228 households in Zone 1.

### Step 2b: Randomly select households

Once you have determined how many households you need from each area, use a computer program or a random numbers table to select which households will be sampled for the survey. Several computer applications, such as Excel, Access, SAS, and CPro, can generate random numbers. If you are familiar with these programs, you can generate numbers electronically rather than using the random numbers table. The random numbers table and instructions on how to use it can be found in Appendix B.

***Step 3: Select one woman of reproductive age within each selected household to be surveyed.***

Once the households have been randomly selected, locators will be responsible for going to each selected household and randomly selecting one respondent from all women of reproductive age that live in that household. The selection of women at the household level is described in detail in the training manual in Module 8: Locator Training–Use of Forms.

## Cluster Sampling

The cluster sampling method has been used widely in developing countries to assess health measures. Cluster sampling is usually selected over (the more statistically precise) random sampling when the geographic area is large, and it will be too difficult, costly, and/or lengthy to cover the entire area with random sampling. This method is also appropriate in cases where household lists are not available or do not meet the criteria needed for random sampling. Cluster sampling can help save time and resources as you need only to create a list of households in the selected clusters rather than for all households in the entire population. However, one disadvantage of cluster sampling is that households in clusters (e.g., neighborhoods, blocks) may share similar characteristics (e.g., income, education, ethnicity). For this reason, the information collected from clusters can be more homogenous than information collected from a random sampling throughout the catchment area, and the sample may not be as representative of the entire population as a sample selected using random sampling.

Because cluster sampling is less precise than random sampling, we must obtain a larger sample size. For cluster sampling, we estimate you will need to have completed interviews from at least 500 women of reproductive age. We anticipate a response rate of 80%, in which case 625 households must be contacted to obtain 500 completed interviews. For cluster sampling, we can not guarantee any precision of the data collected. The sample size is only an estimate, and the true sample size depends upon how different the clusters are from each other, which usually cannot be determined until after data collection is completed.

The cluster survey method in this Toolkit has been designed based on a scientific paper by Steve Bennett and colleagues<sup>6</sup>.

Cluster sampling involves a multi-step process. First, you must select which clusters to survey. Second, within each of the selected clusters, you must select which households to survey. Then one woman of reproductive age is randomly selected from each household. To conduct cluster sampling, you must work through the steps listed in Box 3.3.

### Box 3.3: Cluster sampling steps

- Step 1: Define clusters within the geographic boundaries.
- Step 2: Determine the number of households within each cluster.
- Step 3: Select the clusters to be sampled.
- Step 4: Select the households within each cluster to be sampled.
- Step 5: Select one woman of reproductive age within each selected household to be surveyed.

### ***Step 1: Define clusters within the geographic boundaries.***

This step involves identifying natural groupings of populations, such as camps, villages, neighborhoods, districts, city blocks, or other communities, within the geographic area selected for the survey. When defining groupings, ensure that they do not overlap. These natural groupings will serve as the basis to form clusters. Clusters may vary in size, but each must contain enough households so that the number of eligible households within each cluster can be interviewed. If clusters are too small (i.e., fewer than 25 households), then you must combine them with other neighboring clusters. There is no upper limit on how many households each cluster can contain. We have determined that at least 25 clusters with at least 25 households each are needed for our survey. If you have fewer than 25 clusters, seek technical assistance from DRH to determine the best way to gather information for your population.

Using the map that you created for defining the geographic bounds of your survey area, identify each grouping of households that will serve as a cluster. For example, if refugee tents are organized by blocks, then these blocks may be used to define the clusters.

**Step 2: Determine the number of households within each cluster.**

For each of the clusters, you must know the total number of households. If you have a listing of households, total up the number of households per cluster. If you do not have a listing of all the households, then the total number can be estimated from the total population and the average size of each household. Divide the total population by the average size of households to estimate the number of households, as follows:

$$\text{Total population} \div \text{average size of households} = \text{total number of households}$$

For example, if a camp has an approximate population of 5000 people, and the average size of each household is about 5 people, then there are about 1000 households in this camp.

**Step 3: Select the clusters to be sampled.**

Before you select the clusters, you must determine what your sample size will be, since you will only be surveying a randomly selected sample of the clusters. It is important to determine the right number of clusters and households within those clusters so that the information you collect will accurately represent the entire population and provide the level of accuracy you wish to obtain. As previously mentioned, we have concluded that in most cases, 25 clusters of 25 households each is an appropriate sample size for this survey.

To select the clusters to be surveyed, you will be using a method for sampling called probability proportional to size. This means that a cluster with more households will have a greater chance of being picked for the sample than a cluster with fewer households. This will help assure that the sample is representative. To choose the 25 clusters, refer to the map of your geographic area and clusters within that area. You will be randomly selecting 25 clusters out of all possible clusters within the survey area.

First, create a table of possible clusters in your area. List the clusters in the first column and their estimated number of households in the second column. The list should include communities that are not on official lists (new settlements, refugee camps, etc.). In the third column, which lists the cumulative number of households, add each number of households per community as you go down the list.

Table 3.2 is an example of a cluster list created to facilitate the sampling. This list represents a geographic area with 30 clusters from which you would pick 25. The clusters vary in number of households:

**Table 3.2: Number of households per cluster.**

Cluster #	Number of households in cluster	Cumulative number of households
1	28	28
2	32	60 (28 + 32)
3	65	125 (60 + 65)
4	48	173 (and so forth...)
5	25	198
6	29	227
7	37	264
8	34	298
9	29	327
10	27	354
11	25	379
12	30	409
13	29	438
14	32	470
15	40	510
16	33	543
17	28	571
18	31	602
19	25	627
20	27	654
21	25	679
22	31	710
23	27	737
24	31	768
25	44	812
26	28	840
27	39	879
28	26	905
29	34	939
30	31	970

To select 25 clusters, divide the total number of households in all clusters by 25. This will be the sampling interval:

$$\text{Cumulative number of households} \div 25 \text{ (the number of clusters to be selected)} = \text{sampling interval}$$

From the random number table provided in Appendix B, choose a random number between 1 and your sampling interval, following the directions in Appendix B.

To select the first cluster, refer to the table you created of the number of households per cluster and look to see where this number falls within your cumulative number column. The cluster selected is where the number falls.

To select the second cluster, add the sampling interval to your original random number, and determine where this number falls within your cumulative number column. The second cluster selected is where the number falls.

Continue adding the sampling interval to the number that identified the previous cluster, and choose clusters until you have identified all 25 needed. Because we are using probability proportional to size, it is possible that the same cluster will be selected more than once. Each time a cluster is chosen, it counts as 1 cluster. If a cluster is chosen twice, then 50 (2 x 25) households will be selected. Box 3.4 provides an example of how to select clusters.

### Box 3.4: Example of calculating sampling interval and selecting clusters.

Use Table 3.2 to work through this exercise.

**Calculate sampling interval:** 970 cumulative number of households  $\div$  25 clusters = 38.8, rounded to 39. Thus, 39 is our sampling interval.

**Select first cluster:** Using the random numbers table, let's say you randomly selected 29. Look to see where this number falls within your cumulative number column. 29 falls between 28 and 60, so you would choose cluster #2 as your first sampled cluster.

**Select second cluster:** To choose the second cluster, add the sampling interval (39) to your original random number (29), and the sum of 39 + 29 equals 68. Again, find where this number lies in the cumulative number column (in this case, 68 falls between 61 and 125, so you would choose cluster #3 as your second sampled cluster.

**Select third cluster:** To choose the third cluster, add the sampling interval (39) to the number that identified your previous cluster (68), and the sum of 39 + 68 equals 107. Again, find where this number lies in the cumulative number column (in this case, 107 falls between 61 and 125, so you would choose cluster #3 as your third sampled cluster. Because cluster #3 has now been selected twice, you will select a total of 50 households (2 x 25 households) out of cluster #3's 65 households.

You would then continue this process until you have selected all 25 clusters.

***Step 4: Select the households within each cluster to be sampled.***

Once you have selected the clusters, randomly select households within the selected clusters. You will need some way of identifying the individual households in the selected clusters. If household lists are not available, you can easily and quickly develop a basic list of all the households in the cluster. Often, people familiar with the area, such as community leaders or health workers, can help you develop a list of households.

Once this is done, assign each household a number and then choose your 25 households per cluster using a computer program or the random numbers table (Appendix B).

***Step 5: Select one woman of reproductive age within each selected household to be surveyed.***

Once the clusters and their households have been randomly selected, the locators will be responsible for going to every selected household in that cluster and randomly selecting one respondent from all women of reproductive age who live in that household. The selection of women at the household level is described in detail in the training manual in Module 8: Locator Training—Use of Forms.