|  |
| --- |
| Institute for International Programs – Johns Hopkins University |
| **RADAR Coverage Survey****Sampling guidelines** |
| *Version [DATE]* |

****

Adapted from DHS / MICS manuals

**Table of Contents**

[1. Overview of sample designs for household surveys 2](#_Toc529868543)

[2. Stratification 2](#_Toc529868544)

[3. First-stage sampling: sampling clusters 3](#_Toc529868545)

[3.1. What constitutes a cluster? 3](#_Toc529868546)

[3.2. How to sample clusters 3](#_Toc529868547)

[4. Second stage sampling: sampling households 4](#_Toc529868548)

[4.1. What constitutes a household? 4](#_Toc529868549)

[4.2. Mapping and listing households 5](#_Toc529868550)

[4.3. Alternative approaches 5](#_Toc529868551)

[5. Third stage sampling: sampling individuals 6](#_Toc529868552)

[6. More information 6](#_Toc529868553)

[7. References 6](#_Toc529868554)

This document provides guidance to help you to develop the sampling design for your survey. The document should be used in conjunction with two excel workbooks – “Sampling clusters with PPS” and “Sampling households.” This document does not address sample size; we suggest that you consult the Coverage Survey Sample Size Table to help you estimate the appropriate sample size for your survey.

# Overview of sample designs for household surveys

The most important aspect of the sampling design for a household survey is that it should be a *probability sample*. A probability sample means that every household in your survey area has a known, non-zero probability of being selected (Levy 2008). A probability sample will allow you to estimate the standard errors and confidence intervals around your coverage measures. This will allow the ability to draw a sample that is representative of the entire population with good statistical inference. A well-done probability sample also has less bias than non-probability methods such as convenience, purposive, or quota sampling. **The RADAR coverage survey is intended for use with probability sampling designs**. **Non-probability designs will not allow you to accurately measure coverage or to make inferences about changes in coverage over time.**

To obtain a probability sample, you will need a *sampling frame* – a list of all the households in the area that you are surveying. For surveys conducted in large areas, it is usually impossible to develop a sampling frame for the entire survey area, so we use multi-stage *cluster samples*. For the first stage, we sample small areas (*clusters*), such as villages, and create a sampling frame for each cluster. In the second stage, we can sample households from these cluster sampling frames. The last stage of sampling occurs at the household level. When the survey interviewer visits a sampled household, the first thing he/she will do (after obtaining consent) is to draw up a list of the household members along with other demographic information, such as their sex and age. This list allows the interviewer to identify women, children and men who meet the survey’s eligibility criteria and who should be interviewed.

# Stratification

Cluster samples can be stratified or non-stratified. The goal of stratification is to divide the sample into groups that are relatively different from each other, but relatively homogeneous within each group. From a statistical perspective, stratification can reduce variability (and therefore increase precision). From a practical perspective, in population surveys we often stratify by administrative or residence area (for example, by region or district) with the goal of ensuring that we sample a certain number of households in each administrative area so that we can provide precise estimates for that area. Therefore, if you are planning to report on intervention coverage by administrative area, we recommend that we stratify by administrative area. If you are planning to report separately on rural vs. urban levels of coverage, you should also plan to stratify by urban/rural to ensure that you have a large enough sample in urban areas. As an example, if you were implementing a program in 6 districts and wanted to report coverage levels in each district as well as for rural and urban areas, you would have 12 strata, with one urban and one rural stratum in each district. The cluster sampling methods described in the following sections should be carried out in each stratum.

# First-stage sampling: sampling clusters

## What constitutes a cluster?

Cluster sampling typically makes use of ‘natural’ population groupings such as villages or neighborhoods. Large household survey programs such as DHS and MICS use census enumeration areas (EAs), each of which is generally a village or neighborhood. You may use EAs for your survey or you may define your own clusters. EAs are defined during census mapping, contain an average number of inhabitants or households depending on country context, and are limited by natural or man-made features (streams, roads, etc.). Using EAs can make sampling easier because lists and maps of census EAs are maintained by the National Institute of Statistics in each country. If you plan to use census EAs as your clusters, you should plan to liaise early on with the country’s National Institute of Statistics to determine the process for obtaining EA lists and maps for your survey area.

Whether you use EAs or you define your own clusters, the clusters you use for your survey should have the following characteristics:

* You should be able to establish a list (sampling frame) of all the clusters in your survey area;
* Every household in your survey area should be included in one (and only one) cluster;
* The clusters should have known boundaries; and
* For logistical reasons, clusters should have no more than 150-200 households. Clusters larger than this size will be difficult to map and enumerate.

If you plan to sample clusters proportional to population size (PPS; see next section), you should also have an estimate of the total number of households in each cluster.

## How to sample clusters

You have several options for sampling clusters. All of them require you to have a list of all the clusters in your survey area (divided by strata, if you are stratifying). As probability sampling methods, you may use simple random sampling, systematic random sampling, stratified sampling or a sampling approach combining two or more designs. In this regard, the preferred sampling approach is **systematic sampling proportional to population size (PPS)**.

PPS sampling has the advantage of giving every household in your survey area (or in a stratum, if you are stratifying) the same probability of selection. This also simplifies the calculation of sampling weights at the analysis phase. The drawback of PPS sampling is that it requires you to have an estimation of the size (number of households) of each cluster, and there may be situations in which this is not feasible. In these cases, you can use simple random sampling or systematic sampling (not proportional to population size) to sample clusters.

To conduct PPS sampling, you must have a list of clusters with the estimated number of households in each cluster and the total number of households in the survey area (or stratum, if you are stratifying). This information may be provided from general population census or data from specific demographic surveys or health system routine data collection. You then calculate the sampling interval by dividing the total number of households by the number of clusters, *n*, that you want to sample. This is operationalized in the “Sampling clusters with PPS” Excel workbook, and an example is provided in Table 1 below.

**Table 1. Example of systematic random sampling of clusters with PPS**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Admin level 1** | **Admin level 2** | **Admin level 3** | **EA number** | **Village** | **Number of households** | **Cumulative number of households** | **Survey cluster number** | **Sampled HH number** | **Number of clusters to sample** | 4 |
| Region 1 | District 1 | Commune 1 | Ex1 | Village 1 | 101 | 101 | 1 | 74 | **Sampling interval (total pop/number clusters)** | 345.25 |
| Region 1 | District 1 | Commune 1 | Ex2 | Village 2 | 129 | 230 |  | 419 | **Random number** | 0.21448877 |
| Region 1 | District 1 | Commune 1 | Ex3 | Village 3 | 195 | 425 | 2 | 765 | **Starting point** | 74.0522486 |
| Region 1 | District 1 | Commune 1 | Ex4 | Village 4 | 172 | 597 |  | 1110 |  |  |
| Region 1 | District 1 | Commune 1 | Ex6 | Village 5 | 86 | 683 |  |  |  |  |
| Region 1 | District 1 | Commune 1 | Ex7 | Village 5 | 143 | 826 | 3 |  |  |  |
| Region 1 | District 1 | Commune 1 | Ex8 | Village 5 | 160 | 986 |  |  |  |  |
| Region 1 | District 1 | Commune 1 | Ex9 | Village 5 | 187 | 1173 | 4 |  |  |  |
| Region 1 | District 1 | Commune 1 | Ex10 | Village 5 | 208 | 1381 |  |  |  |  |
|  |  |  |  |  | **Total Households** | **1381** |  |  |  |  |

In our example in Table 1, there are 10 clusters with a total of approximately 1381 households, and we are looking to sample four clusters. The sampling interval is thus approximately 345 households. We generate a random number (0.21448877) that we then multiply by the sampling interval to get our starting point of household number 74. In our sampling frame, household 74 is in EA Ex1. We then add 345 to 74 to get household 419, which falls in EA Ex3, and so forth until we have sampled 4 clusters.

# Second stage sampling: sampling households

## What constitutes a household?

The definition of a household is tricky and can vary somewhat by setting. In general, we define a household as a group of persons who live in the same housing dwelling (which could be a single unit, a compound, or part of a compound), who have common arrangements for cooking and

eating, and who recognize one member as the head of household (distinct from a family patriarch). A household could consist of a single person, but usually it consists of a husband, his wife or wives, their children, and other relatives, servants, etc. A dwelling may contain one or several households.

It is essential that survey interviewers and teams mapping and listing households be well-trained on the definition of a household in that setting. This is particularly critical in settings where extended families made up of multiple households live in large compounds.

## Mapping and listing households

Best practice second-stage sampling (sampling households within clusters) involves creating or updating a map of the sampled cluster that shows all of the dwellings, and listing (enumerating) all of the households in the cluster along with their dwelling number. The households can then be sampled from this list. If you are using census EAs as your cluster, you can start with the EA base maps provided by the National Institute of Statistics and simply update the maps.

This work is typically done by teams of two (one cartographer to map and one enumerator to list the households) and requires approximately one day per cluster per team (somewhat less if clusters are small or are located close together or dwellings are not very dispersed). Mapping and enumeration can be done several weeks before data collection begins, or in parallel with data collection. This second option requires extremely strong coordination, but can save time if there is a need to get the survey into the field quickly. Once the households in the cluster have been listed, you can draw a systematic random sample or simple random sample from the list.

## Alternative approaches

We strongly recommend mapping and listing households for second stage sampling. However, because mapping and enumerating households requires substantial effort, we note that alternative strategies have been used. These are described below, along with their advantages and limitations. RADAR plans to identify and test the feasibility of possible alternative second stage sampling approaches, but this work will require some time to complete.

* **Existing Lists.** One approach is to use an existing list of households (which you may then update). This approach is feasible in settings where reliable, complete lists of households exist at village level. Note that there are many settings in which lists of households are not comprehensive and should not be used. The decision of whether to use an existing household listing should be made in collaboration with partners who have extensive experience in this setting. If you are considering this approach, you should plan to ‘audit’ the household listing in a few clusters by conducting your own independent enumeration of the households. If your results are substantially different than the existing lists, and particularly if the exiting lists appear to systematically exclude certain types of households (for example, households of a certain ethnicity, religion, or neighborhood), you may decide to conduct your own listing. In any case, you should retain the results of your audit to document your decision about second stage sampling.
* **EPI sampling or spin-the-pen sampling**. This method consists of randomly selecting a household in the cluster as your starting point, and then sampling its nearest neighbors. Although this approach has the advantage of being relatively simple to operationalize, it produces a non-probability sample, meaning that standard errors, confidence intervals, and sampling weights cannot be accurately calculated. Another concern is that this sampling approach is difficult to operationalize in settings where dwellings are highly dispersed or are located along roads or streams. In addition, this sampling strategy can lead to sampling bias (Lemeshow 1985, Kok 1986). **For all of these reasons, we do not recommend EPI sampling for the RADAR coverage survey**.

# Third stage sampling: sampling individuals

Once a household has been sampled, an interviewer can visit the household. Upon arriving at the household, after introducing him/herself and obtaining consent, the interviewer must complete a household listing to determine who in the household is eligible for the woman’s, child’s and man’s interview. The household listing collects basic information about each member of the household, including their name, age, relationship to household head, and, if they are aged less than 5 years, the identity of their mother or caregiver. For the RADAR survey, women and men 15-49 years are eligible for the woman’s questionnaire, and mothers or caregivers of children aged less than 5 years are eligible for the child questionnaire. The child’s caregiver must also be aged 15 years or more. **We do not recommend using screening questions (e.g.: Please list all children under 5 in this household) in lieu of a household listing, as respondents may omit eligible household members, particularly more vulnerable members (e.g., foster/orphan children, servants, second wives, etc.) when responding to screening questions**. The process of completing the household listing allows the interviewer to understand the household structure and probe to ensure that all household members are listed.

For household surveys, like RADAR’s, aiming to measure indicators of reproductive, maternal, newborn and child health and nutrition, **we strongly recommend interviewing all eligible women (generally women aged 15-49 years), all eligible children (generally children aged less than 5 years) in the household, and all eligible men (generally men aged 15-49 years),**. This approach reduces the number of households that interviewers need to visit to obtain the necessary sample size, making surveys shorter and cheaper. Including all eligible women, children and men does introduce some complexity at the data management and analysis stage, but the RADAR CAPI and analysis tools are designed to deal with this complexity.

# More information

If you have questions about this material, please contact mmunos1@jhu.edu or amaiga1@jhu.edu.

If you would like a complete text on survey sampling, we recommend Sampling of Populations: Methods and Applications (Levy and Lemeshow, 2008) or Survey Sampling (Kish, 1965)

# References

Bennet S, Radalowicz A, Vella V, Tomkins A. A computer simulation of household sampling schemes for health surveys in developing countries. *International Journal of Epidemiology*, 1994; **23**: 1282-1291.

Kish L (1965). Survey Sampling. New York: John Wiley & Sons, Inc.

Kok PW. Cluster sampling for immunization coverage. *Soc Sci Med,* 1986; **2**: 781-783.

Lemeshow S, Tserkovnyi A, Tulloch J, Dowd E, Lwanga S, Kejas J. A computer simulation of the EPI survey strategy. *International Journal of Epidemiology 1985;* ***14****:473-481.*

Levy PS and Lemeshow S (2008). Sampling of Populations: Methods and Applications. New York: John Wiley & Sons, Inc.