



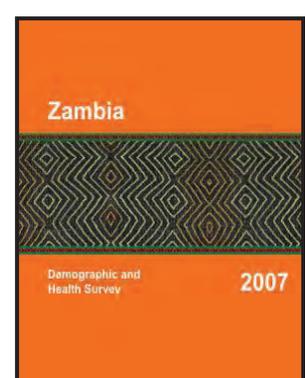
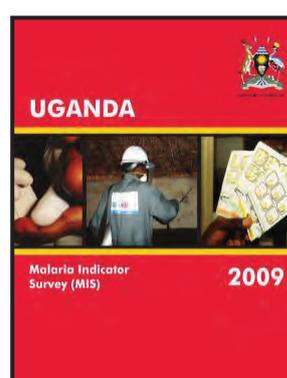
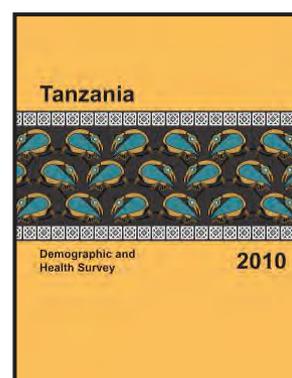
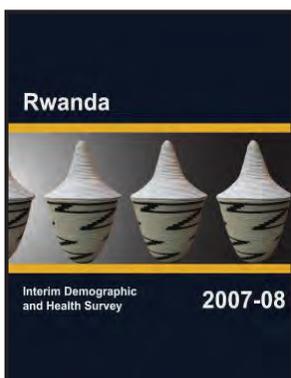
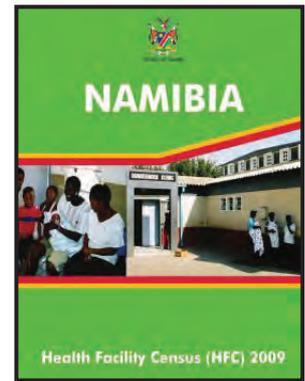
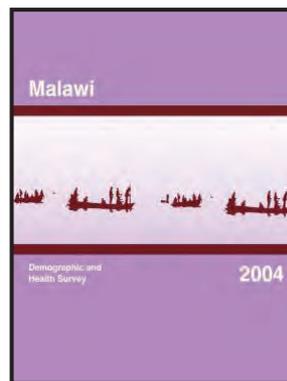
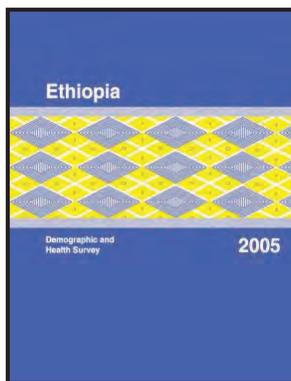
USAID
FROM THE AMERICAN PEOPLE



MODULE 4

Understanding and Using the Demographic and Health Surveys

DHS Curriculum
Facilitator's Guide
July/August 2011



About the DHS Curriculum Facilitator's Guide

The following persons (in alphabetical order) have contributed to developing, reviewing, and editing *Understanding and Using the Demographic and Health Surveys – DHS Curriculum Facilitator's Guide*: Thada Bornstein, Sarah Bradley, Anne Cross, Joy Fishel, Debbie Gachuhi, Hannah Guedenet, Kiersten Johnson, Shane Khan, Laurie Liskin, Erica Nybro, Sarah Schneider

The DHS Curriculum **Facilitator's Guide** is a comprehensive package of ready-made training materials about understanding and using Demographic and Health Survey reports. The curriculum is designed for use in African universities and with public health program staff. Over 25 hours of instruction are divided into seven stand-alone modules designed to be a course on its own or customized and integrated into existing curricula. Each module is complete with instructor guides, Power Point slides, exercises, handouts, pre and post tests and answer keys. The DHS Curriculum **Facilitator's Guide is available in both** print and electronic versions.

Questions and comments regarding the DHS Curriculum can be sent to curriculum@measuredhs.com

About MEASURE DHS

MEASURE DHS assists countries worldwide in the collection and use of data to monitor and evaluate population, health, and nutrition programs. Funded by the U.S. Agency for International Development (USAID) under the terms of Contract No. GPO-C-00-08-00008-00, MEASURE DHS is implemented by ICF Macro in Calverton, Maryland. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Agency for International Development.

The main objectives of the MEASURE DHS project are:

- 1) to provide decision makers in survey countries with information useful for informed policy choices
- 2) to expand the international population and health database
- 3) to advance survey methodology
- 4) to develop in participating countries the skills and resources necessary to conduct quality demographic and health surveys

Information about the MEASURE DHS project or the status of MEASURE DHS surveys is available on the Internet at <http://www.measuredhs.com> or by contacting:

ICF Macro
11785 Beltsville Drive, Suite 300
Calverton, MD 20705 USA
Telephone: 301-572-0200
Fax: 301-572-0999
Email: info@measuredhs.com

Suggested citation

2011. *Understanding and Using the Demographic and Health Surveys – DHS Curriculum Facilitator's Guide: Module 4, July/August 2011*. Calverton, Maryland, USA: ICF Macro.

Module 4: Conducting a DHS

PREPARATION

Review Instructor Guide

Equipment, Materials, Supplies

- LCD projector and screen
- Flipchart or writing board
- Markers
- Slips and container for exercise
- PowerPoint presentation
- Copies of DHS Final Reports

Exercise Preparation

Review *Two-Stage Probability Sampling Exercise* on p. 4-10

Prepare 40 slips of white paper, numbered 1 to 40

Prepare 15 slips of colored paper (all the same color), numbered 1 to 15

Get container (bag or box) to hold the slips of paper

Handouts

Make copies for each participant of:

- Handout 4.1, *Stages in Conducting a DHS*
- *DHS Sampling Manual*

Room Arrangements

Participants should be seated at tables in groups of four to seven, if possible.

PURPOSE

The purpose of this module is to provide an overview of how a DHS survey is conducted; the basic concepts of sampling used in the DHS, including sample design and sampling procedures; and the general principles of sampling weights.

OBJECTIVES

By the end of the module, participants should be able to:

- Describe the steps in conducting a DHS
- Explain the basic concepts of sampling
- Discuss the sampling procedures used in the DHS
- Explain the general principles of sampling weights

TIME

3 hours

MODULE OVERVIEW

Session 1	Steps in Conducting a DHS	1 hour
Session 2	Sampling Procedures in the DHS	1 hour
Session 3	Principles of Sampling Weights	1 hour

Session 1

1 hour

Session Objective**Steps in Conducting a DHS**

Describe the steps in conducting a DHS

STEP 1PRESENT **Slides 1 and 2.**

DESCRIBE the objectives for Module 4 and the focus of the three sessions.

PRESENT **Slide 3.**

TELL participants that the first session will focus on the steps in conducting a DHS.

ASK participants if they have ever participated in a survey. TELL them to think about their experience while we discuss how a DHS survey is conducted.

PRESENT **Slide 4**

EXPLAIN that there are four main stages in conducting a DHS that take about two years in total. TELL participants the main stages, and EXPLAIN that this session will review the steps within each stage.

EXERCISE

The purpose of this exercise is to get participants thinking and talking about what has to be done to conduct a survey and to remind them of their experiences, if any, with surveys.

DIVIDE participants into four groups and assign a stage to each group. TELL each group to write down all the activities they think need to be done during that stage and to estimate how long these activities will take. TELL each group to select a recorder to take notes and a spokesperson to present the group's conclusions to the rest of the participants. Give them about 10 minutes to work on this. Then BRING everyone together, and ASK the spokespersons to briefly share each group's responses with the rest of the participants. (Alternatively, you may allow

around 20 minutes for the small group work, and ask the groups to record the steps and times on a flipchart.)

EXPLAIN that each stage builds on and is entirely dependent on the quality of the work done in the preceding stages, beginning with the sampling procedures. If the sample is not a scientifically selected random sample, then the data collected will not be representative of the population—no matter how well the questionnaire is designed, how well the staff is trained, how well the data are collected, or how expertly the data are analyzed. Similarly, if interviewers are not well trained, the data collected will be of poor quality. No statistical methods of analysis can make up for the lack of good quality data. Dissemination of the results of a poorly conducted survey can result in faulty or even harmful program or policy recommendations.

STEP 2

PRESENT **Slide 5**.

DISTRIBUTE **Handout 4.1**, which describes the stages and related activities in further detail.

EXPLAIN that this timeline provides a general overview of how a survey takes place. This timeline is subject to many changes, and the actual timeline will differ in each country due to challenges or delays that unexpectedly arise along the way.

ASK for examples of what some delays might be.

PRESENT **Slide 6**.

TELL participants that most surveys are funded by several development partners and by the national government. As the DHS has expanded and added more questions and more biomarkers, the time needed for each interview has gotten longer, requiring more interview teams or more time in the field. This increases the overall cost.

ASK participants to define a biomarker (covered in Module 3). If they cannot, REMIND them that biomarkers are biological measures of health, such as height and weight or hemoglobin levels in the blood. Collecting biomarkers takes additional time.

PRESENT **Slide 7.**

EXPLAIN that during the initial visits, representatives from MEASURE DHS and the country assess the needs of the country, devise an overall plan for the survey, and prepare initial budgets and a survey timeline. Often, a government ministry or agency is appointed to be the implementing agency. The implementing agency serves as the direct contact with ICF Macro staff and takes charge of daily activities in the field.

EXPLAIN how questionnaires are developed. Questionnaires are survey instruments that interviewers read questions from and record answers on. As discussed in Module 1, the MEASURE DHS project uses a common set of questionnaires to interview households, women, and men. These are referred to as “core questionnaires.” Because the questions are very similar from survey to survey, results can be compared in one country over time and between countries.

DHS core questionnaires are adapted to the needs of the country or, more specifically, the needs of government ministries, country programs, NGOs, and funding agencies. For example, the Ministry of Gender may want to include the extra questions on domestic violence. Or if a large-scale behavior change campaign has been broadcasting radio programs about HIV, officials may want to include a question to determine how many people have heard the programs. Questionnaire content is usually discussed at a meeting or a series of meetings with survey stakeholders. ASK participants who are the stakeholders? Answer: Major users of the data.

Questionnaires are translated into the primary languages of the country and pre-tested to make sure that respondents understand the questions. After any necessary revisions, the questionnaires are finalized.

PRESENT **Slide 8.**

This slide shows how questions appear on the DHS questionnaire. Questionnaires vary by country, and the specific questionnaires used in each country can be found in the appendix section of the country final report.

ASK participants to break into pairs, and DISTRIBUTE

copies of DHS final reports. ASK participants to spend 10 minutes looking through the survey questionnaires found in the appendix.

PRESENT **Slide 9**.

EXPLAIN that the training of interviewers is very important to the quality of the data collected. Training can last four to five weeks or even longer, depending on how many modules are added to the questionnaire. Every interviewer goes through the questionnaire very carefully and practices administering it, first with other trainees and later in a formal pretest.

Interviewers are also responsible for taking weight and height measurements and for taking drops of blood for HIV and other tests. This requires about a week of training for interviewers who have no previous medical training.

TELL participants that data collection takes about three to four months, depending on the size of the sample, the number of interview teams, and the size of the country. An interview team includes both male and female interviewers (the DHS requires that men interview men and women interview women), a supervisor, a field editor, and a driver. The supervisor manages the team, assigns interviewers to houses, talks with village or community officials, and responds to problems. The field editor reviews every questionnaire to check that no questions have been missed, that all the information is recorded legibly, and that there are no inconsistencies or errors.

DISCUSS possible inconsistencies with participants. For example, the interviewer may note in one section of the questionnaire that a woman's age at first sex was 19, but another part of the questionnaire may list her age at the birth of her first child as 16. Clearly, one of these responses is incorrect. The field editor is responsible for ensuring internal consistency within each questionnaire and also for packing up the questionnaires from a cluster and sending them to the implementing agency office for data input. Lists of interview teams are always included in the appendix of DHS final reports.

EXPLAIN that when the interviewer reaches a household, s/he asks permission to interview household members. Once permission is granted, the interviewer begins by listing all household members and visitors and identifying

which ones are eligible for the individual interview. Eligible women are defined as those who are in the specified age group (15-49) and who slept in the household the night before the interviewer's visit. Depending on the survey, eligible men will be in the same age range or older. If an interviewer visits a household and does not make contact, she/he is required to make two return visits before the interview is abandoned.

Interviewers begin with the household questionnaire before completing individual men's and women's questionnaires. Any usual member of the household age 15 or older can respond to the household questionnaire.

Most questionnaires are on paper, but the MEASURE DHS project is starting to use PDAs (personal digital assistants), also known as handheld computers. PDAs have been used in several countries. It is likely that in the future most surveys will be carried out with handheld computers.

PRESENT **Slide 10**.

EXPLAIN that after interviews are finished, information from the completed questionnaires is entered into databases and rigorously checked for consistency of answers. This is another method that the DHS uses to ensure data quality. Data entry is done at the same time as data collection. This lets the DHS monitor teams in the field by checking response rates and various measures of data quality. In the event that a team is performing poorly, survey managers can respond appropriately.

After all data are entered, the data are tabulated to produce the tables in the final report. Weights are added to the data. Weighting, or mathematically adjusting the numbers, will be discussed in Session 3.

PRESENT **Slide 11**.

EXPLAIN that disseminating the survey results is a critical, but often neglected part of the survey process. Too often projects spend a lot of time and money conducting a survey, but set aside too little time and money to get the results out to the people who are responsible for improving programs and policies.

After the data are put into tables, the authors of the final report draft the chapters. Reports are usually written by local experts or staff from the implementing agencies, or both. Often the project convenes an in-country report-writing workshop.

The final report is edited, published, and distributed by the implementing agency. Often the MEASURE DHS project develops other print materials such as fact sheets, summary reports, and PowerPoint presentations to complement the distribution of the final report.

In almost every country, a national seminar is held to launch the survey results. Between 100 and 500 people attend these events. Authors of the final report present the findings using PowerPoint presentations, followed by question-and-answer sessions. Many countries also hold dissemination seminars at the state or regional level, if funding and time permit.

The MEASURE DHS project works hard to help various stakeholders make use of the survey findings. How they do that will be discussed in Module 7.

Final note: A major contribution of the MEASURE DHS project is that it makes datasets from each country available at the MEASURE DHS website. This allows researchers to work with the datasets and conduct their own analyses.

STEP 3

End the session by REMINDING participants that a survey is a systematic and complex process that requires time, expertise, political will, and resources. Each survey is different, but the first chapter of every DHS report contains a brief summary of the survey timeline and a discussion of how sampling was done for that survey.

Session 2

1 hour

Session Objectives**Sampling Procedures Used in the DHS**

Explain the basic concepts of sampling.

Discuss the sampling procedures used in the DHS.

STEP 1

If this session immediately follows the previous one, CONDUCT an energizer to get participants up and moving.

PRESENT **Slide 12.**

REVIEW with participants the definitions of the following terms, which were introduced in Module 2:

- Population
- Sampling
- Sample
- Representative

STEP 2

PRESENT **Slide 13.**

TELL participants to recall the earlier discussion of what a sample is. EMPHASIZE that it is:

- A group of people selected for a study, and
- A group meant to represent a larger population.

PRESENT **Slide 14.**

EXPLAIN that sample sizes in the DHS are large, in the thousands, which produces estimates for indicators that we can be confident in. Some indicators, such as mortality and fertility, require larger sample sizes than others in order to obtain a reliable estimate.

The DHS provides estimates at the national level, for urban and rural areas, and usually for about five to ten sub-national administrative areas. Providing representative samples at the sub-national level requires a larger sample and increases the cost of the survey. Many more

households, women, and men need to be included in the survey.

The sample sizes in the DHS depend on funding and the needs of the government. DHS sampling attempts to balance methodological sampling concerns against cost effectiveness; in other words, the DHS strives to get the “best” indicators for the best price.

PRESENT **Slide 15.**

This table shows the minimum sample sizes for some basic indicators. These are the ideal numbers that are required to achieve a reasonable level of accuracy when estimating the indicator. If sample sizes are lower than those indicated in the table, the level of accuracy is reduced and the confidence intervals are larger.

In the DHS, the domain usually refers to the sub-national regions or provinces. If a given country has 8 sub-national regions, an ideal sample size for total fertility rate is approximately 8,000 women (1,000 women for each region).

PRESENT **Slide 16.**

EXPLAIN the example shown on the slide: The 2003 Nigeria DHS was representative at the national level and for six sub-national zones; each zone included about 6 states. (Nigeria has 37 states in total.) The sample size for this survey was over 7,200 households; it included 7,600 women and 2,350 men. During preparations for the 2008 Nigeria DHS, the government asked for separate estimates for each of the country’s 37 states. This meant having a much larger sample, with over 34,000 households; it included 33,385 women and 15,486 men.

PRESENT **Slide 17.**

EXPLAIN each of the following terms in the slide.

National coverage: DHS surveys cover the entire country.

Two-stage probability sampling: In simple terms, this means that the selection of the sample is done in two separate stages or steps.

Pre-existing sampling frame: The sampling frame is a list of the entire population. DHS prefers to use an existing frame—most often a national census.

A **probability sampling** method is any method of sampling that uses some form of **random selection**. It means that the different units in the population have equal probabilities of being chosen; for example, each woman age 15-49 has the same chance of being selected for an interview in a DHS survey. This approach reduces the possibility of bias and increases the representativeness of the survey sample.

Simplicity of design: A simple design is the easiest and least expensive to implement.

Separate household listing: A numbered list of every house in a cluster is created before final sampling.

STEP 3

PRESENT **Slide 18**, and EXPLAIN the following terms:

Sampling frame: Ideally there should be a complete list of the entire population from which to draw a sample. This is why the DHS uses census data as the sampling frame, because theoretically all residents of a country should be accounted for in the national census.

Clusters: During the first stage of sampling, a group of areas (known as clusters) is selected from the entire sampling frame.

Stratification: This process groups members of the population into comparatively homogeneous subgroups or domains (for example, urban and rural, or by geographic region) before sampling. This allows a representative sample to be drawn for each strata, or subgroup.

The word stratum (plural: strata) means a division or a subgroup. In this instance, we are dividing the entire sample into subgroups.

Household listing: After clusters are selected, a survey team visits each one to make a complete, numbered list of

EXERCISE

every household or dwelling in the cluster. Although this is time-consuming, it is important because it makes it more likely that every household in the cluster has an equal probability of being selected to participate in the survey.

Two-stage probability sampling exercise

Materials:

- 40 slips of white paper numbered 1 to 40. (Each slip represents one cluster in the sampling frame.)
- 15 slips of colored paper (all the same color) numbered from 1 to 15. (Each slip represents one household from a cluster.)
- Box or large bag

REMIND participants that the DHS uses a two-stage sampling process, and EXPLAIN that this exercise is an example of how it is done.

To orient participants to the exercise, TELL them that they are going to select a random sample for a survey using the two-stage process. REMIND them that the first stage randomly selects clusters from the census, while the second stage randomly selects households from each cluster. Clusters are the primary sampling units, and the randomly selected households are the secondary sampling units.

In this case, TELL the participants that you began with a list of 40 clusters, or enumeration areas (EAs), taken from the country's population census or some other sampling frame. Each cluster was given a unique number. For our example, we decide that 20 clusters are needed to produce a random sample of the right size. To get the 20 clusters, we must randomly select 20 numbers. SHOW participants a box or bag with the 40 numbered slips of white paper inside. ASK one participant to stir or shake the container to mix up the slips, close his/her eyes, and pull one slip at a time out of the container until he/she has 20 slips. TELL participants that not looking is important to ensure that the slips are chosen randomly.

TELL participants that these are the 20 clusters from which you will select households. TELL participants that you have now completed the first stage of the two-stage sampling process.

EXPLAIN that the next step is to send teams to each of the 20 clusters to make a map of the area with the location of all the dwellings in that cluster (or to update an existing map). Each dwelling on the map is given a number and those numbers will be used to select households in the second stage.

Once the cluster maps have been drawn, we are ready for the second stage of sampling, that is, selecting which households to interview. From each of our 20 randomly selected clusters, we will randomly select 5 households.

SHOW one of the 20 slips of white paper selected earlier, in the first stage of the sampling process. EXPLAIN that this slip of paper represents one cluster in which the households have now been mapped and numbered by field teams. This particular cluster contains 15 households, and the sample requires 5 households. We are now going to select which 5 households to interview.

SHOW participants the set of 15 colored paper slips, and PLACE the slips in the box or bag.

REMIND participants that the sample size needed is 5. ASK them if they know what the next step is.

TELL a different participant to mix up the slips of paper and pull 5 slips out of the container without looking. REMIND participants that not looking is important to ensure that the slips were chosen randomly.

TELL participants that these are the 5 households that will be interviewed in the cluster and that this represents the second stage of the sampling process.

(You can repeat this exercise to demonstrate the random nature of the ultimate selection of households for interview. If you repeat it, compare the five numbers selected the first time and the second time.)

DEBRIEF the exercise, and ASK why the selection procedure had to be done that way.

Make sure the following points come out:

- If a sample is not selected randomly, applying the rules of probability, then the sample will not be representative of the entire population and you cannot generalize the results of the data collected to

the entire population. DISCUSS with participants why this is true. Answer: Because the sample and the data it produces would reflect only the characteristics of the individuals selected to answer the survey questions—not the entire population.

- The point of drawing a sample for a population-based survey is to ensure that the information you collect from a few people is indeed representative of all of the people. Therefore, it is extremely important to ensure that random sampling procedures are strictly adhered to in order to ensure the representativeness of the data.

PRESENT **Slide 19.**

EXPLAIN that the graphic on this slide gives a basic visual overview of the two-stage sampling process that is used for the majority of DHS surveys. The exercise you just completed was an example of a two-stage probability sample where clusters were randomly selected from the sampling frame. The DHS follows this same process except that the sampling frame is stratified into subgroups such as urban/rural or geographic regions/provinces before the clusters are selected from the sampling frame.

As already explained, the population sampling frame comes from the most recently completed national census. The sampling frame provides a complete list of clusters, or enumeration areas, for the country.

The sampling process starts with stratifying the sampling frame by geographic area to ensure that enough households are chosen in each region or province.

During the first stage of the two-stage sample selection process, clusters are selected for the survey sample from each strata. Most DHS surveys include 300 to 500 clusters. The number of clusters depends on how many sub-national areas (i.e., regions or provinces) need to be included.

After the clusters are selected, survey staff go into the field to map villages and to collect a complete listing of households or dwellings in each cluster.

For the second stage, the household listings are returned to the implementing agency, and 20 to 30 households are randomly selected from the entire list of households in

each cluster.

The selected households in each cluster are listed by number. The survey team responsible for the cluster receives this list and so knows which households to visit.

This may seem complex, but specific country examples will help to clarify the process.

PRESENT **Slide 20.**

TELL participants that this is an example of a household listing and map used for a DHS survey. Before households are chosen, field staff visit each cluster, map it, and make a listing of all the households. From this listing, households are randomly selected.

PRESENT **Slide 21.**

TELL participants that in order to better understand how the two-stage sampling process works, we will see how the sample was selected for the 2007 Zambia Demographic and Health Survey (ZDHS).

The Year 2000 Census of Zambia created 16,757 enumeration areas or clusters. In the first stage of sampling for the ZDHS, 320 clusters were selected from these 16,757 clusters.

In the second stage, a household listing and map was made for each of these 320 clusters. An average of 25 households was selected to participate in the survey from each of these clusters, resulting in a representative sample of 8,000 households.

When interviewing took place, all women age 15–49 who were permanent residents of the household or visitors present in the household the night before the survey were eligible to be interviewed. Additionally, in a subsample of one-third of all households, syphilis testing was performed among eligible women and men. A subsample of one eligible woman per household was randomly selected to be asked extra questions about domestic violence.

PRESENT **Slide 22.**

EXPLAIN that this figure shows the geographical distribution of the 320 clusters in the Zambia DHS.

As you can see from this map, the clusters are spread throughout the country. This map also gives a good picture of the level of effort involved in doing a DHS. Interviewers visited 25 or more households in each cluster. It's a credit to their hard work that interviews were completed in almost all of the households.

PRESENT **Slide 23.**

EXPLAIN that a total of 7,969 households were selected for the ZDHS. Of those, 7,326 (or about 98%) were occupied. Interviews were completed with about 90% of all the households selected. ASK how the 90% was calculated. Answer: 7,164 divided by 7,969.

Note that among the 7,408 eligible women in these households, over 7,146 (or 97%) agreed to be interviewed. ASK participants what this percentage is called. Answer: This is called the **response rate**.

ASK participants what the response rate among men is. Answer: 91% ($6,500/7,146 \times 100$).

ASK why response rates are lower among men in most surveys. Answer: The response rate is lower for men in almost all surveys because men are more likely to be working away from home.

PRESENT **Slide 24**

TELL participants we are going to look at a second example, how the sample was selected for the 2008 Ghana Demographic and Health Survey (GDHS).

The Year 2000 Ghana Population and Housing Census served as the sampling frame for the 2008 GDHS. In the first stage of sampling for the GDHS, 412 clusters were selected.

In the second stage, a household listing and map was made for each of these 412 clusters. An average of 30

households was selected to participate in the survey from each of these clusters, resulting in a representative sample of 12,323 households.

In half of the households selected for the survey, all women age 15–49 and all men age 15–59 who were permanent residents of the household or visitors present in the household the night before the survey were eligible to be interviewed. Height and weight measurements of female respondents and children under the age of five years were done only in the households selected for the

individual interview. Eligible women and children age 6 to 59 months in the households selected for individual interview were also tested for anemia.

PRESENT **Slide 25**

EXPLAIN that this figure shows the geographical distribution of the 412 clusters in the Ghana DHS.

As you can see from this map, the clusters are spread throughout the country. This map also gives a good picture of the level of effort involved in doing a DHS. Interviewers visited 30 or more households in each cluster. It's a credit to their hard work that interviews were completed in almost all of the households.

PRESENT **Slide 26**

EXPLAIN that a total of 12,323 households were selected for the GDHS. Of those, 11,913 (or about 97%) were occupied. Interviews were completed with about 90% of all the households selected. ASK how the 90% was calculated. Answer: 11,778 divided by 12,323.

Note that among the 5,096 eligible women in these households, over 4,916 (or 96%) agreed to be interviewed. ASK participants what this percentage is called. Answer: This is called the response rate.

ASK participants what the response rate among men is. Answer: 96% ($4,568/4,769 \times 100$).

STEP 4

End this session by REFERRING participants to the *DHS Sampling Manual*, which they are welcome to read in their free time. EXPLAIN that the manual provides significantly more detail on DHS sampling than was covered in this session.

Session 3

1 hour

Session Objective**Principles of Sampling Weights**

Explain the general principles of sampling weights

STEP 1PRESENT **Slide 27.**

(NOTE to the instructor: More information on how sampling weights are calculated is available in the Sampling Manual. Please familiarize yourself with this section of the manual before presenting this session)

ASK participants what they think sampling weights are and why they are part of the DHS.

PRESENT **Slide 28.**

(NOTE for the instructor: A technical definition of sampling weights is: Adjustment factors applied to each case in tabulations to adjust for differences in probability of selection and interview between cases in a sample, either due to design or happenstance. Be aware that many students struggle with the terms “weight” and “weighting.” While these are the technically correct terms to use, you may want remind students that they are mathematical adjustment factors.)

EXPLAIN that, as discussed in previous modules, the goal of DHS surveys is to provide representative data at the national and sub-national levels. This can be challenging because some regions may be much larger than others. To solve this problem, the DHS oversamples in regions with small populations; this ensures that each region—even the smallest—has a large enough sample to be representative. And since we cannot always sample as many people as we want (as discussed in Session 2), the DHS undersamples in regions with large populations.

To correct for this deliberate oversampling and undersampling, the DHS applies **sampling weights** to the sample.

PRESENT Slide 29.

TELL participants that the following slides give an example of how sampling weights were applied in the 2005 Ethiopia DHS. Let's assume that we need to interview about 14,500 households in order to get reliable estimates of our indicators at the national and sub-national levels. We select these households from each region EXACTLY in proportion to the distribution of households in the country; in other words, we take a simple random sample of households in Ethiopia.

PRESENT Slide 30.

EXPLAIN that this pie chart shows the distribution of households by region in Ethiopia. It is clear that some regions have much larger populations than others.

Click to cue the animation. ASK What problems might we encounter if we took a simple random sample of 14,500 households based on this distribution? Answer: There will be very few respondents from certain regions, such as Gambela, Harari, and Dire Dawa. Too few respondents in each of the regions will not give precise data.

PRESENT Slide 31.

EXPLAIN that this is how a simple random sample would be divided across Ethiopia's 11 regions. TELL participants to look closely at the numbers for Gambela, one of the smallest regions. Only 0.4% of Ethiopia's population lives in Gambela. Therefore, in a simple random sample only 58 households—or 0.4% of the total sample of 14,500 households—would be selected from Gambela.

ASK participants if we can really get enough information about Gambela by selecting only 58 households. Answer: No. But remember, we only have enough resources to select 14,500 households. ASK what should be done.

PRESENT Slide 32.

EXPLAIN that in order to include enough households to provide representative data at the regional level, we must

oversample in regions with smaller populations and undersample in regions with larger populations. The total number of households in the sample—14,645—remains the same, but the households are redistributed so as to select more households in smaller regions.

This table shows the number of households selected in the actual 2005 EDHS sample. You can see that the survey selected 925 households in Gambela instead of 58. This provides enough information about households in Gambela to describe their health and population status accurately. In contrast, the EDHS undersampled in the largest region, Oromiya. So many households from Oromiya would be included in a simple random sample (over 2,000), that their number could be reduced without affecting the representativeness of the regional sample.

PRESENT **Slide 33.**

EXPLAIN that, after oversampling, the households selected in Gambela make up 6% of the total DHS sample—a much higher percentage than their 0.3% share of Ethiopia's population. In contrast, after undersampling, the households selected in Oromiya make up 15% of the DHS sample, which is much less than their 36% share of Ethiopia's population.

ASK participants why this is a problem. ANSWER: Data from these two regions will not have the correct proportion of influence on the national figures. The next slide will clarify this through an example.

PRESENT **Slide 34.**

TELL participants that we will now examine why oversampling and undersampling can be a problem when calculating national averages for indicators. Consider, for example, the percent of households in Ethiopia who own an ITN.

Let's assume that after a mass ITN distribution campaign, households in Gambela have an unusually high rate of ITN ownership of 70%, compared with just 20% for households in other regions. ASK why this is a problem, and what we can do about it.

Answer: This is a problem because Gambela will have too much influence on the national average. The households in Gambela account for 6% of the EDHS sample when, in reality, they make up less than 1% of the population. Therefore, the high rate of ITN ownership in Gambela would skew the national average.

To fix this problem, the DHS applies sampling weights. Sampling weights are designed to correct for the fact that people in some areas or subgroups are more likely than others to be selected for interviews and included in the survey sample. Because of over- and undersampling in the 2005 EDHS, for example, households in Gambela had a greater probability of being selected than households in Oromiya.

PRESENT **Slide 35.**

EXPLAIN that, in essence, weighting takes all of the households in a sample and recalculates or “weights” the number of households from each region. Weighting makes each region’s contribution to the total sample proportional to the actual distribution of households across the country. The DHS does this so that the impact of any one region on the national average will not be out of proportion to its actual share of the population.

EXPLAIN to participants that once the final number of households interviewed is complete and available, weights are applied in order to reduce or increase the size of the region in proportion to the total population.

In areas where we oversampled (like Gambela), one household should not have as much influence on the national average. And areas that were undersampled, households should have more influence. So, the weight in oversampled households would be less than 1 and undersampled households would be more than 1 to account for this.

For example, in Gambela, if a household owns an ITN, they will actually only count for 0.057 of a household because it was oversampled. And in Oromiya (where we undersampled), one household will count for 2.22 households.

PRESENT Slide 36.

TELL participants to look carefully at the numbers for Gambela in this table. After weights are applied, the weighted number of households from Gambela in the 2005 EDHS sample is only 47 out of 13,721—which is close to Gambela’s true percentage share of the population. Weighting is also done for regions with large populations, such as Oromiya, that were undersampled.

Each final country report shows the weighted and unweighted numbers of women and men interviewed for the survey in Table 3.1.

(NOTE: The numbers in this slide are showing the actual and weighted number of households *interviewed*, which is less than the number *selected*, as shown on slide 33. Remember that not all households selected provide an interview.)

PRESENT Slide 37.

EXPLAIN that in the end after applying weights we have a distribution that looks very much like the distribution we began with. But the advantage is that we are confident that every region has a large enough sample for the results to be representative at the sub-national level. This is not true for a simple random sample.

PRESENT Slide 38.

TELL participants that many people have been taught to worry about small sample sizes. However, that refers to the unweighted number of respondents, which is generally not shown in DHS tables. Most DHS tables show the weighted number of households and respondents, and this sometimes may appear to be rather small. But you have to take into account that the unweighted number was much larger.

To help you assess the sample size, the tables in the final report use certain symbols to let you know if the number of unweighted cases is too small to be considered reliable. Parentheses indicate that a number is based on 25-49 unweighted cases. This means the number should be

interpreted with caution. An asterisk indicates that a number is based on less than 25 unweighted cases; this is too small to be reliable, so DHS tables do not even show these numbers.

(NOTE for instructor: Weights have another important function: they can also correct for non-response. For example, a certain percentage of selected respondents probably will not complete the survey, and that percentage may vary between subgroups. Weights can account for this. Weights are calculated after all the data are collected.)

STEP 2

End this session by ASKING participants if they have any questions about applying sampling weights. TELL them that applying weights is a standard statistical process that is done with many different types of data. REMIND participants that the *DHS Sampling Manual* and the *Guide to DHS Statistics* have additional information on the mathematical equations used for sampling weights and that they can read this in their free time.



USAID
FROM THE AMERICAN PEOPLE



Module 4

Conducting a DHS

Objectives for Module 4

By the end of this module, students should be able to:

- Describe the steps in conducting a DHS
- Explain the basic concepts of sampling
- Discuss the sampling procedures used in the DHS
- Explain the general principles of sampling weights

Module 4, Slide 2

DESCRIBE the objectives for Module 4 and the focus of the three sessions.

Module 4

Session 1

Steps in Conducting a DHS

TELL participants that the first session will focus on the steps in conducting a DHS.

ASK participants if they have ever participated in a survey. TELL them to think about their experience while we discuss how a DHS survey is conducted.

Conducting a DHS

DHS surveys are undertaken in four main stages. From initial contact to final procedures, they take about two years.

The four main stages are:

1. Survey preparation & questionnaire design
2. Training and fieldwork (data collection)
3. Data processing and analysis
4. Data dissemination and use of data

Module 4, Slide 4

EXPLAIN that there are four main stages in conducting a DHS that take about two years in total. TELL participants the main stages, and EXPLAIN that this session will review the steps within each stage.

Timeline for Conducting a DHS

Stage	Timeline	Activities
1	Month 1	Survey design visit
	Month 2	Sample design
	Month 3	Questionnaire design
	Month 3 – 4	Household listing
	Month 5	Pretest
	Month 6	Revision of questionnaires and manuals
2	Month 7	Training of field personnel
	Month 8	Data processing set up
	Month 8 – 11	Fieldwork
3	Month 9 – 12	Data entry and editing
	Month 13	Preparation of the preliminary report
	Month 14 – 16	Tabulation, analysis, and preparation of final report
	Month 17	First draft of the report
	Month 18	Review and revision of report
4	Month 19	Printing of the final report
	Month 20	National seminar
	Month 20+	Further analysis and/or data dissemination activities

Module 4, Slide 5

DISTRIBUTE **Handout 4.1**, which describes the stages and related activities in further detail.

EXPLAIN that this timeline provides a general overview of how a survey takes place. This timeline is subject to many changes, and the actual timeline will differ in each country due to challenges or delays that unexpectedly arise along the way.

ASK for examples of what some delays might be.

Preliminary Preparation

- A USAID Mission or the national government usually requests the DHS survey.
- The first visit identifies an implementing agency (typically the National Bureau of Statistics or a similar agency) to be the primary group to carry out the survey.
- The team assesses available funding. Options include:
 - National government
 - USAID or other US government partners
 - Various international organizations (e.g., UNICEF, UNFPA, World Bank, DFID, and WHO)

Module 4, Slide 6

TELL participants that most surveys are funded by several development partners and by the national government. As the DHS has expanded and added more questions and more biomarkers, the time needed for each interview has gotten longer, requiring more interview teams or more time in the field. This increases the overall cost.

ASK participants to define a biomarker (covered in Module 3). If they cannot, REMIND them that biomarkers are biological measures of health, such as height and weight or hemoglobin levels in the blood. Collecting biomarkers takes additional time.

Stage 1: Survey Preparation and Questionnaire Design

- Assess needs and information gaps, budgets, and time frame
- Begin questionnaire development with stakeholders
- Adapt DHS core questionnaires to country, translate, pre-test, and finalize

NOTES:

Questionnaires are survey instruments that interviewers read questions from and record answers on.

DHS uses standard core questionnaires to interview households, women, and men.

Using standard questionnaires permits comparisons across countries and over time.

Module 4, Slide 7

EXPLAIN that during the initial visits, representatives from MEASURE DHS, USAID, and the country assess the needs of the country, devise an overall plan for the survey, and prepare initial budgets and a survey timeline. Often, a government ministry or agency is appointed to be the implementing agency. The implementing agency serves as the direct contact with ICF Macro staff and takes charge of daily activities in the field.

EXPLAIN how questionnaires are developed. Questionnaires are survey instruments that interviewers read questions from and record answers on. As discussed in Module 1, the MEASURE DHS project uses a common set of questionnaires to interview households, women, and men. These are referred to as “core questionnaires.” Because the questions are very similar from survey to survey, results can be compared in one country over time and between countries.

DHS core questionnaires are adapted to the needs of the country or, more specifically, the needs of government ministries, country programs, NGOs, and funding agencies. For example, the Ministry of Gender may want to include the extra questions on domestic violence. Or if a large-scale behavior change campaign has been broadcasting radio programs about HIV, officials may want to include a question to determine how many people have heard the programs. Questionnaire content is usually discussed at a meeting or a series of meetings with survey stakeholders. ASK participants who are the stakeholders? Answer: Major users of the data.

Questionnaires are translated into the primary languages of the country and pre-tested to make sure that respondents understand the questions. After any necessary

Sample DHS Questions

310	Are you currently doing something or using any method to delay or avoid getting pregnant?	YES 1 NO 2	→ 322
311	Which method are you using? CIRCLE ALL MENTIONED. IF MORE THAN ONE METHOD MENTIONED, FOLLOW SKIP INSTRUCTION FOR HIGHEST METHOD ON LIST.	FEMALE STERILIZATION A MALE STERILIZATION B PILL C IUD D INJECTABLES E IMPLANTS F MALE CONDOM G FEMALE CONDOM H DIAPHRAGM I FOAM/JELLY J LACTATIONAL AMEN. METHOD K RHYTHM METHOD L WITHDRAWAL M OTHER X (SPECIFY)	→ 316 → 312 → 315 → 311E → 315 → 313 → 313 → 315 → 315 → 319A
311A	CIRCLE 'A' FOR FEMALE STERILIZATION.		
311E	What are the reasons of injectables use (continued)?	NO/DON'T KNOW / (S) MONTHS	

This slide shows how questions appear on the DHS questionnaire. Questionnaires vary by country, and the specific questionnaires used in each country can be found in the appendix section of the country final report.

ASK participants to break into pairs, and DISTRIBUTE copies of DHS final reports. ASK participants to spend 10 minutes looking through the survey questionnaires found in the appendix.

Stage 2: Training and Fieldwork (Data Collection)

- Data are collected primarily by interviewing respondents and members of households.
- For biomarkers, data collection may also involve taking height, weight, and blood samples.
- Interviewers are trained on the questionnaires and then sent to the field to conduct interviews and collect data.
- Women always interview women, and men always interview men.
- Media outreach informs the public about the survey and encourages them to participate.

Module 4, Slide 9

EXPLAIN that the training of interviewers is very important to the quality of the data collected. Training can last four to five weeks or even longer, depending on how many modules are added to the questionnaire. Every interviewer goes through the questionnaire very carefully and practices administering it, first with other trainees and later in a formal pretest.

Interviewers are also responsible for taking weight and height measurements and for taking drops of blood for HIV and other tests. This requires about a week of training for interviewers who have no previous medical training.

TELL participants that data collection takes about three to four months, depending on the size of the sample, the number of interview teams, and the size of the country. An interview team includes both male and female interviewers (the DHS requires that men interview men and women interview women), a supervisor, a field editor, and a driver. The supervisor manages the team, assigns interviewers to houses, talks with village or community officials, and responds to problems. The field editor reviews every questionnaire to check that no questions have been missed, that all the information is recorded legibly, and that there are no inconsistencies or errors.

DISCUSS possible inconsistencies with participants. For example, the interviewer may note in one section of the questionnaire that a woman's age at first sex was 19, but another part of the questionnaire may list her age at the birth of her first child as 16. Clearly, one of these responses is incorrect. The field editor is responsible for ensuring internal consistency within each questionnaire and also for packing up the questionnaires from a cluster and sending them to the implementing agency office for data input. Lists of interview teams are always included in the appendix of DHS

Stage 3: Data Processing

- Information from completed questionnaires is entered into databases and rigorously checked as soon as it is entered.
- Weights (mathematical adjustments) are added to the data file.
- Data tabulation produces the tables for the preliminary and final reports.

Module 4, Slide 10

EXPLAIN that after interviews are finished, information from the completed questionnaires is entered into databases and rigorously checked for consistency of answers. This is another method that the DHS uses to ensure data quality. Data entry is done at the same time as data collection. This lets the DHS monitor teams in the field by checking response rates and various measures of data quality. In the event that a team is performing poorly, survey managers can respond appropriately.

After all data are entered, the data are tabulated to produce the tables in the final report. Weights are added to the data. Weighting, or mathematically adjusting the numbers, will be discussed in Session 3.

Stage 4: Data Dissemination and Use of Data

- The final country report is published.
- National and regional seminars disseminate results.
- Data are used to:
 - Guide policy development and recommendations
 - Design interventions
 - Monitor project activities
 - Conduct advocacy
 - Further analyze trends or make comparisons with other countries

Module 4, Slide 11

EXPLAIN that disseminating the survey results is a critical, but often neglected part of the survey process. Too often projects spend a lot of time and money conducting a survey, but set aside too little time and money to get the results out to the people who are responsible for improving programs and policies.

After the data are put into tables, the authors of the final report draft the chapters. Reports are usually written by local experts or staff from the implementing agencies, or both. Often the project convenes an in-country report-writing workshop.

The final report is edited, published, and distributed by the implementing agency. Often the MEASURE DHS project develops other print materials such as fact sheets, summary reports, and PowerPoint presentations to complement the distribution of the final report.

In almost every country, a national seminar is held to launch the survey results. Between 100 and 500 people attend these events. Authors of the final report present the findings using PowerPoint presentations, followed by question-and-answer sessions. Many countries also hold dissemination seminars at the state or regional level, if funding and time permit.

The MEASURE DHS project works hard to help various stakeholders make use of the survey findings. How they do that will be discussed in Module 7.

Final note: A major contribution of the MEASURE DHS project is that it makes

Module 4

Session 2

Sampling Procedures Used in the DHS

REVIEW with participants the definitions of the following terms, which were introduced in Module 2:

Population

Sampling

Sample

Representative

What is a Sample?

- Group of people selected for a study
- Group meant to represent a larger population



Module 4, Slide 13

TELL participants to recall the earlier discussion of what a sample is. EMPHASIZE that it is:

- A group of people selected for a study, and
- A group meant to represent a larger population.

Sample Sizes in the DHS

- Often large — in the thousands
- When the sample size is small, survey findings may not be reliable
- Need to balance sample size with budgetary constraints
- The more sub-national areas (e.g., provinces or districts) for which data are desired, the larger the sample must be

Module 4, Slide 14

EXPLAIN that sample sizes in the DHS are large, in the thousands, which produces estimates for indicators that we can be confident in. Some indicators, such as mortality and fertility, require larger sample sizes than others in order to obtain a reliable estimate.

The DHS provides estimates at the national level, for urban and rural areas, and usually for about five to ten sub-national administrative areas. Providing representative samples at the sub-national level requires a larger sample and increases the cost of the survey. Many more households, women, and men need to be included in the survey.

The sample sizes in the DHS depend on funding and the needs of the government. DHS sampling attempts to balance methodological sampling concerns against cost effectiveness; in other words, the DHS strives to get the “best” indicators for the best price.

Ideal Sample Sizes for Indicators

Indicator	Ideal sample size per domain
Infant mortality rate (IMR)	800 to 1,000 women
Total fertility rate (TRF)	800 to 1,000 women
Contraceptive prevalence rate (CPR)	400 to 500 women

Module 4, Slide 15

This table shows the minimum sample sizes for some basic indicators. These are the ideal numbers that are required to achieve a reasonable level of accuracy when estimating the indicator. If sample sizes are lower than those indicated in the table, the level of accuracy is reduced and the confidence intervals are larger.

In the DHS, the domain usually refers to the sub-national regions or provinces. If a given country has 8 sub-national regions, an ideal sample size for total fertility rate is approximately 8,000 women (1,000 women for each region).

Sample Sizes at the Sub-national Level

- The 2003 Nigeria DHS was representative at the national level and for six zones.
- Each zone included about six states.
- The sample included 7,225 households, with 7,600 women and 2,350 men.
- The 2008 Nigeria DHS wanted to report results **for each of the nation's 37 states.**
- Therefore, the sample had to be larger: it include more than 34,000 households, with 33,385 women and 15,486 men.

Module 4, Slide 16

EXPLAIN the example shown on the slide: The 2003 Nigeria DHS was representative at the national level and for six sub-national zones; each zone included about 6 states. (Nigeria has 37 states in total.) The sample size for this survey was over 7,200 households; it included 7,600 women and 2,350 men. During preparations for the 2008 Nigeria DHS, the government asked for separate estimates for each of the country's 37 states. This meant having a much larger sample, with over 34,000 households; it included 33,385 women and 15,486 men.

DHS Sampling Principles

General principles of DHS sample design:

- National coverage
- Two-stage probability sampling
- Pre-existing sampling frames
- Simplicity of design
- Separate household listing

Module 4, Slide 17

EXPLAIN each of the following terms in the slide.

National coverage: DHS surveys cover the entire country.

Two-stage probability sampling: In simple terms, this means that the selection of the sample is done in two separate stages or steps.

A **probability sampling** method is any method of sampling that uses some form of **random selection**. It means that the different units in the population have equal probabilities of being chosen; for example, each woman age 15-49 has the same chance of being selected for an interview in a DHS survey. This approach reduces the possibility of bias and increases the representativeness of the survey sample.

Simplicity of design: A simple design is the easiest and least expensive to implement.

Separate household listing: A numbered list of every house in a cluster is created before final sampling.

Key Sampling Terms

- *Sampling frame*: The list of the entire population from which the study sample is drawn, often a census.
- *Clusters*: Groups into which the sampling frame is divided; they are sometimes called *primary sampling units* (PSUs) or *enumeration areas* (EAs).
- *Stratification*: The process of grouping members of the population into homogeneous subgroups or domains (for example, urban/rural or by geographic region) before sampling; this allows a representative sample to be drawn for each stratum, or subgroup.
- *Household listing*: A numbered list of every house in a cluster; it is created before final sampling.

Module 4, Slide 18

EXPLAIN the following terms:

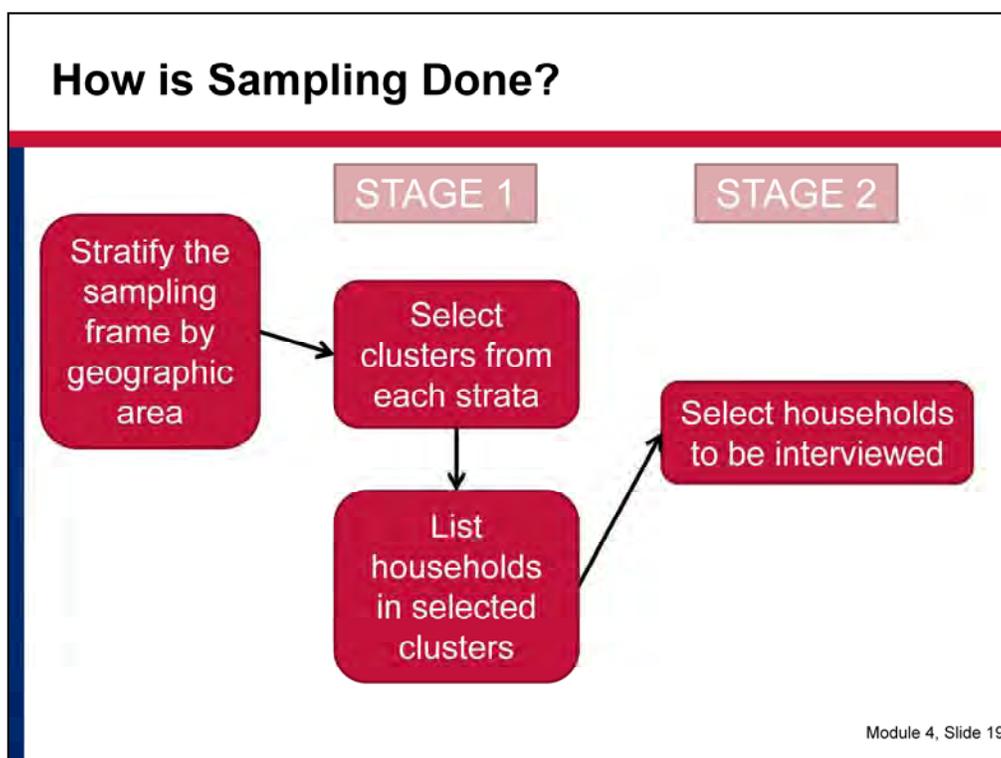
Sampling frame: Ideally there should be a complete list of the entire population from which to draw a sample. This is why the DHS uses census data as the sampling frame, because theoretically all residents of a country should be accounted for in the national census.

Clusters: During the first stage of sampling, a group of areas (known as clusters) is selected from the entire sampling frame.

Stratification: This process groups members of the population into comparatively homogeneous subgroups or domains (for example, urban and rural, or by geographic region) before sampling. This allows a representative sample to be drawn for each stratum, or subgroup.

The word stratum (plural: strata) means a division or a subgroup. In this instance, we are dividing the entire sample into subgroups.

Household listing: After clusters are selected, a survey team visits each one to make a complete, numbered list of every household or dwelling in the cluster. Although this is time-consuming, it is important because it makes it more likely that every household in the cluster has an equal probability of being selected to participate in the survey.



EXPLAIN that the graphic on this slide gives a basic visual overview of the two-stage sampling process that is used for the majority of DHS surveys. The exercise you just completed was an example of a two-stage probability sample where clusters were randomly selected from the sampling frame. The DHS follows this same process except that the sampling frame is stratified into subgroups such as urban/rural or geographic regions/provinces before the clusters are selected from the sampling frame.

As already explained, the population sampling frame comes from the most recently completed national census. The sampling frame provides a complete list of clusters, or enumeration areas, for the country.

The sampling process starts with stratifying the sampling frame by geographic area to ensure that enough households are chosen in each region or province.

During the first stage of the two-stage sample selection process, clusters are selected for the survey sample from each strata. Most DHS surveys include 300 to 500 clusters. The number of clusters depends on how many sub-national areas (i.e., regions or provinces) need to be included.

After the clusters are selected, survey staff go into the field to map villages and to collect a complete listing of households or dwellings in each cluster.

For the second stage, the household listings are returned to the implementing agency, and 20 to 30 households are randomly selected from the entire list of households in each cluster.

The selected households in each cluster are listed by number. The survey team responsible for the cluster receives this list and so knows which households to visit.

This may seem complex, but specific country examples will help to clarify the process.

Sample Household Listing and Map

Form DHS/1 Demographic and Health Survey Map Information Page 1 of 2

IDENTIFICATION		Observations
Province: <u>K. Ayas</u>	Province Code: <u>1</u>	<u>Segment 1 is not accessible from Kibondo road. It is a long distance from the market. It is a residential area. It is a school on Mwanza road. It is a school on Mwanza road. It is a school on Mwanza road.</u>
District: <u>Diana</u>	District Code: <u>04</u>	
Town/Village: <u>Diana</u>	Town/Village Code: <u>02</u>	
Name of Mapper: <u>Hanslow Sibile</u>	Cluster Code: <u>011</u>	
Name of place: <u>Jake Maliki</u>	DHS Cluster Location: <u>011</u>	
	Segment Number: <u>1</u>	
GPS-Unit Tracking Number: <u>13</u>		
Waypoint Address (entered in GPS unit): <u>0001-2</u>		
Latitude (North/South): <u>14 57 52 S</u>		
Longitude (East/West): <u>008 50 42 E</u>		
Altitude / Elevation (Meters): <u>0325</u>		

Here is an example of the household listing and map made for each cluster selected for the survey.

Module 4, Slide 20

TELL participants that this is an example of a household listing and map used for a DHS survey. Before households are chosen, field staff visit each cluster, map it, and make a listing of all the households. From this listing, households are randomly selected.

Sample Selection in the 2007 ZDHS

First stage

- 320 clusters were selected from 16,757 clusters created for the 2000 Census of Zambia.
- For the household listing, a team:
 - Visited each of the 320 clusters,
 - Listed and numbered all households in the cluster, and
 - Made a sketch map of the cluster.

Second stage

- 25 households were randomly selected in each cluster.
- The representative sample totaled 8,000 households.

Module 4, Slide 21

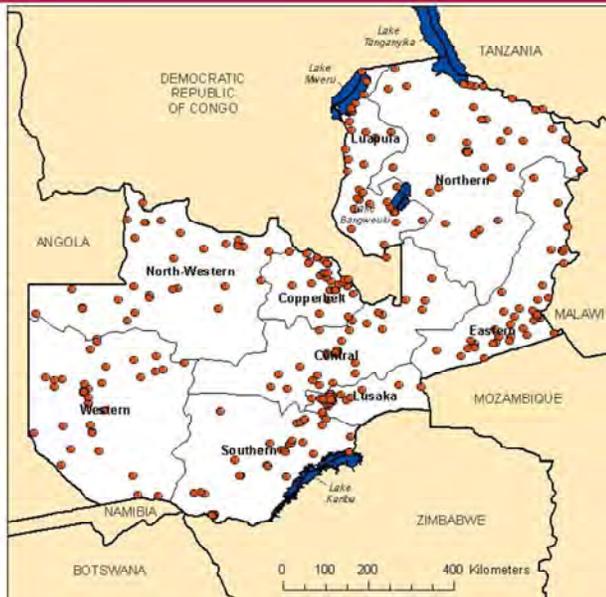
TELL participants that in order to better understand how the two-stage sampling process works, we will see how the sample was selected for the 2007 Zambia Demographic and Health Survey (ZDHS).

The Year 2000 Census of Zambia created 16,757 enumeration areas or clusters. In the first stage of sampling for the ZDHS, 320 clusters were selected from these 16,757 clusters.

In the second stage, a household listing and map was made for each of these 320 clusters. An average of 25 households was selected to participate in the survey from each of these clusters, resulting in a representative sample of 8,000 households.

When interviewing took place, all women age 15–49 who were permanent residents of the household or visitors present in the household the night before the survey were eligible to be interviewed. Additionally, in a subsample of one-third of all households, syphilis testing was performed among eligible women and men. A subsample of one eligible woman per household was randomly selected to be asked extra questions about domestic violence.

Cluster Points Selected in the 2007 ZDHS



In each cluster 25 households were selected for interviews, resulting in a total sample of 8,000 households.

Module 4, Slide 22

EXPLAIN that this figure shows the geographical distribution of the 320 clusters in the Zambia DHS.

As you can see from this map, the clusters are spread throughout the country. This map also gives a good picture of the level of effort involved in doing a DHS. Interviewers visited 25 or more households in each cluster. It's a credit to their hard work that interviews were completed in almost all of the households.

Response Rates in the 2007 ZDHS

Result	Urban	Rural	Total
Household interviews			
Households selected	2,899	5,070	7,969
Households occupied	2,748	4,578	7,326
Households interviewed	2,694	4,470	7,164
Interviews with women age 15-49			
No. of eligible women	3,320	4,088	7,408
No. of eligible women interviewed	3,178	3,968	7,146
Interviews with men age 15-54			
No. of eligible men	3,225	3,921	7,146
No. of eligible men interviewed	2,831	3,669	6,500

EXPLAIN that a total of 7,969 households were selected for the ZDHS. Of those, 7,326 (or about 98%) were occupied. Interviews were completed with about 98% of all the households selected. ASK how the 98% was calculated. Answer: 7,164 divided by 7,326.

Note that among the 7,408 eligible women in these households, over 7,146 (or 96%) agreed to be interviewed. ASK participants what this percentage is called. Answer: This is called the **response rate**.

ASK participants what the response rate among men is. Answer: 91% ($6,500/7,146 \times 100$).

ASK why response rates are lower among men in most surveys. Answer: The response rate is lower for men in almost all surveys because men are more likely to be working away from home.

Sample Selection in the 2008 Ghana DHS

First stage

- 412 clusters were selected from the 2000 Ghana Population and Housing Census.
- For the household listing, a team:
 - Visited each of the 412 clusters,
 - Listed and numbered all households in the cluster, and
 - Made a sketch map of the cluster.

Second stage

- 30 households were randomly selected in each cluster.
- The representative sample totaled 12,323 households.

Module 4, Slide 24

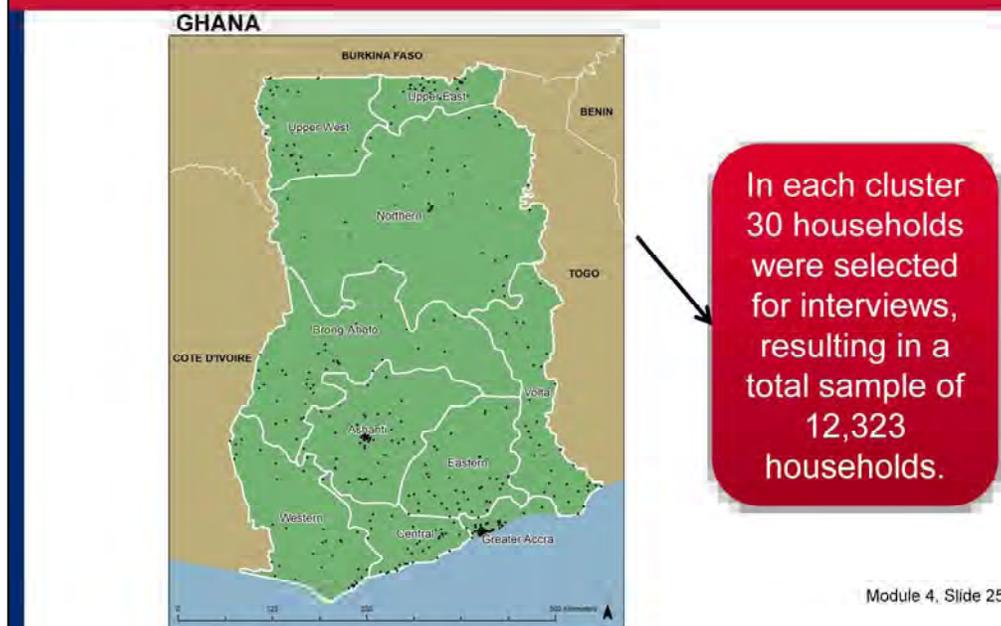
TELL participants we are going to look at a second example, how the sample was selected for the 2008 Ghana Demographic and Health Survey (GDHS).

The Year 2000 Ghana Population and Housing Census served as the sampling frame for the 2008 GDHS. In the first stage of sampling for the GDHS, 412 clusters were selected.

In the second stage, a household listing and map was made for each of these 412 clusters. An average of 30 households was selected to participate in the survey from each of these clusters, resulting in a representative sample of 12,323 households.

In half of the households selected for the survey, all women age 15–49 and all men age 15–59 who were permanent residents of the household or visitors present in the household the night before the survey were eligible to be interviewed. Height and weight measurements of female respondents and children under the age of five years were done only in the households selected for the individual interview. Eligible women and children age 6 to 59 months in the households selected for individual interview were also tested for anemia.

Cluster Points Selected in the 2008 GDHS



EXPLAIN that this figure shows the geographical distribution of the 412 clusters in the Ghana DHS.

As you can see from this map, the clusters are spread throughout the country. This map also gives a good picture of the level of effort involved in doing a DHS. Interviewers visited 30 or more households in each cluster. It's a credit to their hard work that interviews were completed in almost all of the households.

Response Rates in the 2008 GDHS

Result	Urban	Rural	Total
Household interviews			
Households selected	5,458	6,865	12,323
Households occupied	5,252	6,661	11,913
Households interviewed	5,175	6,603	11,778
Interviews with women age 15-49			
No. of eligible women	2,239	2,857	5,096
No. of eligible women interviewed	2,162	2,754	4,916
Interviews with men age 15-59			
No. of eligible men	2,014	2,755	4,769
No. of eligible men interviewed	1,914	2,654	4,568

EXPLAIN that a total of 12,323 households were selected for the GDHS. Of those, 11,913 (or about 97%) were occupied. Interviews were completed with about 90% of all the households selected. ASK how the 90% was calculated. Answer: 11,778 divided by 12,323.

Note that among the 5,096 eligible women in these households, over 4,916 (or 96%) agreed to be interviewed. ASK participants what this percentage is called. Answer: This is called the **response rate**.

ASK participants what the response rate among men is. Answer: 96% ($4,568/4,769 \times 100$).

Module 4

Session 3

Principles of Sampling Weights

(NOTE to the instructor: More information on how sampling weights are calculated is available in the Sampling Manual. Please familiarize yourself with this section of the manual before presenting this session)

ASK participants what they think sampling weights are and why they are part of the DHS.

Using Sampling Weights in the DHS

- The goal of DHS surveys is to provide representative data at the national and sub-national levels.
- In many countries, the population is not evenly distributed among different regions. Oversampling in regions with small populations ensures that they have a large enough sample to be representative. Undersampling is done in regions with large populations.
- To correct for oversampling and undersampling, **sampling weights** are applied.

Module 4, Slide 28

(NOTE for the instructor: A technical definition of sampling weights is: Adjustment factors applied to each case in tabulations to adjust for differences in probability of selection and interview between cases in a sample, either due to design or happenstance. Be aware that many students struggle with the terms “weight” and “weighting.” While these are the technically correct terms to use, you may want remind students that they are mathematical adjustment factors.)

EXPLAIN that, as discussed in previous modules, the goal of DHS surveys is to provide representative data at the national and sub-national levels. This can be challenging because some regions may be much larger than others. To solve this problem, the DHS oversamples in regions with small populations; this ensures that each region—even the smallest—has a large enough sample to be representative. And since we cannot always sample as many people as we want (as discussed in Session 2), the DHS undersamples in regions with large populations.

To correct for this deliberate oversampling and undersampling, the DHS applies **sampling weights** to the sample.

Let's Imagine...

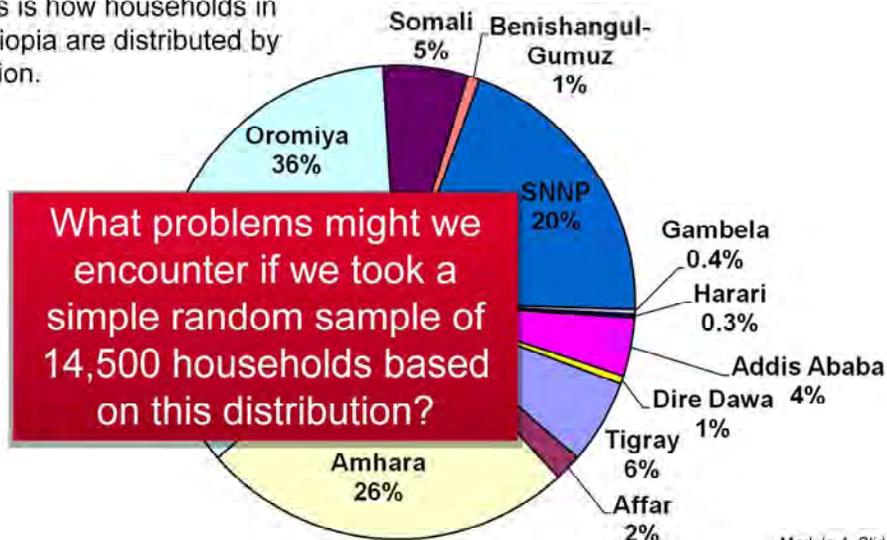
- We need to interview 14,500 households in Ethiopia to have reliable estimates at the national and sub-national levels.
- We select a simple random sample of households based on the regional population distribution.

Module 4, Slide 29

TELL participants that the following slides give an example of how sampling weights were applied in the 2005 Ethiopia DHS. Let's assume that we need to interview about 14,500 households in order to get reliable estimates of our indicators at the national and sub-national levels. We select these households from each region EXACTLY in proportion to the distribution of households in the country; in other words, we take a simple random sample of households in Ethiopia.

Ethiopia 2005: Distribution of Households by Region

This is how households in Ethiopia are distributed by region.



Module 4, Slide 30

EXPLAIN that this pie chart shows the distribution of households by region in Ethiopia. It is clear that some regions have much larger populations than others.

Click to cue the animation. ASK What problems might we encounter if we took a simple random sample of 14,500 households based on this distribution? Answer: There will be very few respondents from certain regions, such as Gambela, Harari, and Dire Dawa. Too few respondents in each of the regions will not give precise data.

Regional Distribution in a Simple Random Sample

Region	% of sample	No. of households
Tigray	6.3	914
Affar	1.6	232
Amhara	25.4	3,683
Oromiya	36.0	5,220
Somali	4.3	624
Benishangul	1.0	145
SNNP	19.9	2,886
Gambela	0.4	58
Harari	0.3	44
Addis Ababa	4.2	609
Dire Dawa	0.5	73
Total	100%	14,500

Notice the sample only includes 58 households in Gambela. Is that enough households to give reliable estimates for Gambela?

Module 4, Slide 31

EXPLAIN that this is how a simple random sample would be divided across Ethiopia’s 11 regions. TELL participants to look closely at the numbers for Gambela, one of the smallest regions. Only 0.4% of Ethiopia’s population lives in Gambela. Therefore, in a simple random sample only 58 households—or 0.4% of the total sample of 14,500 households—would be selected from Gambela.

ASK participants if we can really get enough information about Gambela by selecting only 58 households. Answer: No. But remember, we only have enough resources to select 14,500 households. ASK what should be done.

Actual EDHS Sample		
Region	No. of households	% of sample
Tigray	1,349	9
Affar	935	6
Amhara	2,158	15
Oromiya	2,241	15
Somali	901	6
Benishangul-Gumuz	954	7
SNNP	2,012	14
Gambela	925	6
Harari	960	7
Addis Ababa	1,400	10
Dire Dawa	810	6
Total	14,645	100%

Module 4, Slide 32

If we oversample in Gambela, we have enough households to be confident that estimates will be reliable for Gambela. Conversely, we can undersample in Oromiya and be confident that the sample is still large enough to produce reliable estimates.

EXPLAIN that in order to include enough households to provide representative data at the regional level, we must oversample in regions with smaller populations and undersample in regions with larger populations. The total number of households in the sample—14,645—remains the same, but the households are redistributed so as to select more households in smaller regions.

This table shows the number of households selected in the actual 2005 EDHS sample. You can see that the survey selected 925 households in Gambela instead of 58. This provides enough information about households in Gambela to describe their health and population status accurately. In contrast, the EDHS undersampled in the largest region, Oromiya. So many households from Oromiya would be included in a simple random sample (over 2,000), that their number could be reduced without affecting the representativeness of the regional sample.

But now Gambela's households represent 6% of the DHS sample when they only make up 0.4% of the population. And Oromiya is only 15% of the population. Why is this a problem?

	No. of households selected	% of sample
	1,349	9
	935	6
	2,158	15
Oromiya	2,241	15
Somali	901	6
Benishangul-Gumuz	954	7
SNNP	2,012	14
Gambela	925	6
Harari	960	7
Addis Ababa	1,400	10
Dire Dawa	810	6
Total	14,645	100%

Module 4, Slide 33

EXPLAIN that, after oversampling, the households selected in Gambela make up 6% of the total DHS sample—a much higher percentage than their 0.3% share of Ethiopia's population. In contrast, after undersampling, the households selected in Oromiya make up 15% of the DHS sample, which is much less than their 36% share of Ethiopia's population.

ASK participants why this is a problem. ANSWER: Data from these two regions will not have the correct proportion of influence on the national figures. The next slide will clarify this through an example.

The Problem of Over- and Undersampling

- Suppose we want to know what percent of households in Ethiopia own an insecticide-treated net (ITN).
- Assume, hypothetically, that an NGO did a huge ITN distribution campaign in Gambela, so 70% of households in Gambela own an ITN. But the average in all the other regions is 20%.
- But because of oversampling, households in Gambela make up 6% of the sample instead of their 0.4% share of the population. Why is this a problem? And what can we do about it?

Module 4, Slide 34

TELL participants that we will now examine why oversampling and undersampling can be a problem when calculating national averages for indicators. Consider, for example, the percent of households in Ethiopia who own an ITN.

Let's assume that after a mass ITN distribution campaign, households in Gambela have an unusually high rate of ITN ownership of 70%, compared with just 20% for households in other regions. ASK why this is a problem, and what we can do about it.

Answer: This is a problem because Gambela will have too much influence on the national average. The households in Gambela account for 6% of the EDHS sample when, in reality, they make up less than 1% of the population. Therefore, the high rate of ITN ownership in Gambela would skew the national average.

To fix this problem, the DHS applies sampling weights. Sampling weights are designed to correct for the fact that people in some areas or subgroups are more likely than others to be selected for interviews and included in the survey sample. Because of over- and undersampling in the 2005 EDHS, for example, households in Gambela had a greater probability of being selected than households in Oromiya.

Applying Sampling Weights

- Once we have the final number of households interviewed in a region, we take that number and multiply it by a **weight** that will reduce or increase it to the size it should in proportion to the total population.
- In areas where we oversampled, the weight will be less than 1, and where we undersampled, the weight will be more than 1.
- For example, in Gambela, one household that owns an ITN will count for 0.057 of a household when calculating the national percent of households that own an ITN.
- In Oromiya (where we undersampled), one household that owns an ITN will count for 2.22 households.

Module 4, Slide 35

EXPLAIN that, in essence, weighting takes all of the households in a sample and recalculates or “weights” the number of households from each region. Weighting makes each region’s contribution to the total sample proportional to the actual distribution of households across the country. The DHS does this so that the impact of any one region on the national average will not be out of proportion to its actual share of the population.

EXPLAIN to participants that once the final number of households interviewed is complete and available, weights are applied in order to reduce or increase the size of the region in proportion to the total population.

In areas where we oversampled (like Gambela), one household should not have as much influence on the national average. And areas that were undersampled, households should have more influence. So, the weight in oversampled households would be less than 1 and undersampled households would be more than 1 to account for this.

For example, in Gambela, if a household owns an ITN, they will actually only count for 0.057 of a household because it was oversampled. And in Oromiya (where we undersampled), one household will count for 2.22 households.

To make sure that households in Gambela are not overrepresented when calculating the national average, **sampling weights** are applied to reflect their true percentage of the population.

	Actual no. of households	Weighted no. of households
	1,282	940
	806	138
	2,066	3,709
	2,155	4,790
	796	540
Amhara-Gumuz	869	128
SNNP	1,933	2,802
Gambela	820	47
Harari	904	39
Addis Ababa	1,333	525
Dire Dawa	757	64
Total	13,721	13,721

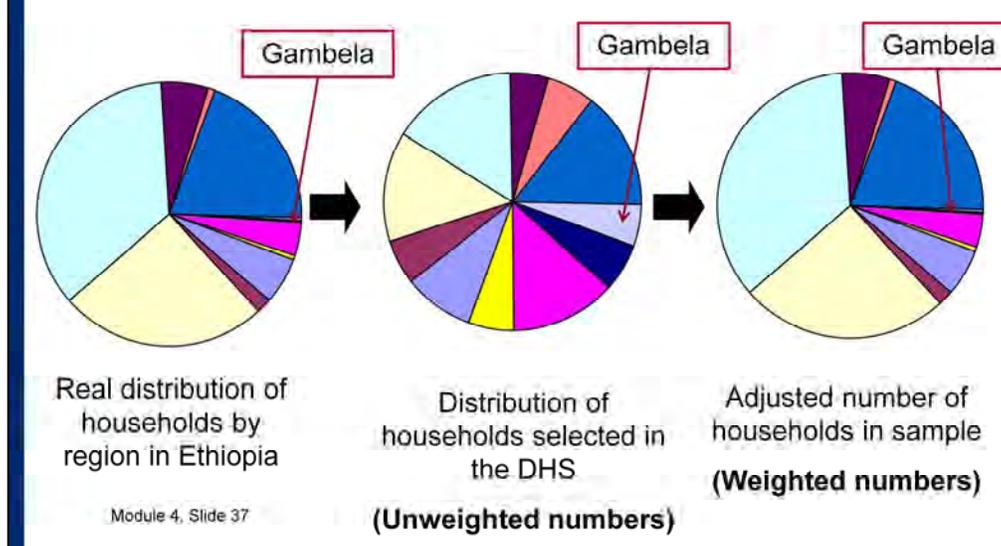
Module 4, Slide 36

TELL participants to look carefully at the numbers for Gambela in this table. After weights are applied, the weighted number of households from Gambela in the 2005 EDHS sample is only 47 out of 13,721—which is close to Gambela’s true percentage share of the population. Weighting is also done for regions with large populations, such as Oromiya, that were undersampled.

Each final country report shows the weighted and unweighted numbers of women and men interviewed for the survey in Table 3.1.

(NOTE: The numbers in this slide are showing the actual and weighted number of households *interviewed*, which is less than the number *selected*, as shown on slide 33. Remember that not all households selected provide an interview.)

Applying Sampling Weights



EXPLAIN that in the end after applying weights we have a distribution that looks very much like the distribution we began with. But the advantage is that we are confident that every region has a large enough sample for the results to be representative at the sub-national level. This is not true for a simple random sample.

Weighted and Unweighted Numbers

- As a rule, DHS tables show the weighted number of respondents for each background characteristic.
- Even if you see a small number in the table, don't worry because this number has likely been adjusted to match the population distribution.
- DHS tables let you know if the number of unweighted cases (the number actually interviewed) is too small to be reliable.
- Figures in parentheses () are based on 25 to 49 unweighted cases. This means "interpret with caution."
- An asterisk * means that the number of unweighted cases is less than 25 and too small to be reliable, so the DHS tables do not even show the number.

Module 4, Slide 38

TELL participants that many people have been taught to worry about small sample sizes. However, that refers to the unweighted number of respondents, which is generally not shown in DHS tables. Most DHS tables show the weighted number of households and respondents, and this sometimes may appear to be rather small. But you have to take into account that the unweighted number was much larger.

To help you assess the sample size, the tables in the final report use certain symbols to let you know if the number of unweighted cases is too small to be considered reliable. Parentheses indicate that a number is based on 25-49 unweighted cases. This means the number should be interpreted with caution. An asterisk indicates that a number is based on less than 25 unweighted cases; this is too small to be reliable, so DHS tables do not even show these numbers.

(NOTE for instructor: Weights have another important function: they can also correct for non-response. For example, a certain percentage of selected respondents probably will not complete the survey, and that percentage may vary between subgroups. Weights can account for this. Weights are calculated after all the data are collected.)

Handout 4.1

Stages in Conducting a DHS

Stage	Timeline	Activities
1	Month 1	Survey design visit
	Month 2	Sample design
	Month 3	Questionnaire design
	Month 3 - 4	Household listing
	Month 5	Pretest
	Month 6	Revision of questionnaires and manuals
2	Month 7	Training of field personnel
	Month 8	Data processing set up
	Month 8 - 11	Fieldwork
3	Month 9 - 12	Data entry and editing
	Month 13	Preparation of the preliminary report
	Month 14 - 16	Tabulation, analysis, and preparation of final report
	Month 17	First draft of the report
	Month 18	Review and revision of report
4	Month 19	Printing of the final report
	Month 20	National seminar
	Month 20	Further analysis and/or data dissemination activities

Preliminary Preparation

During initial visits to a country, the MEASURE DHS representative and the national implementing agency (often a ministry or the central bureau of statistics within the country) assess the needs of the country and devise an overall plan for the DHS. They discuss funding and specific country needs and prepare initial budgets and a schedule. The implementing agency is in charge of daily activities in the field and is the direct contact with MEASURE DHS.

Stage 1: Survey Preparation and Questionnaire Design

The DHS uses a common set of questionnaires to interview households, women, and men. This allows comparison across countries and over time. The team adapts the MEASURE DHS core questionnaires to ensure that survey results will be most relevant and useful to the country. They hold discussions with government ministries, directors of major programs, and funding agencies. For example, if the country has recently rolled out a new mosquito net campaign, they may add additional questions on mosquito net ownership and use. If a national HIV/AIDS television soap opera has been produced recently, stakeholders may want to track how

Handout 4.1

many people have viewed it. The team may add such questions if space and time allow.

The questionnaires are translated into the major languages of the country and pre-tested to ensure that respondents understand the questions. After the pre-test and any necessary revisions, the questionnaires are finalized. The questionnaires used in a country can be found in the appendices of the final report. The core questionnaires are available at www.measuredhs.com.

Stage 2: Training and Fieldwork/Data Collection

Once the questionnaires are finalized, interviewers are trained to use the questionnaires and are then sent to collect information from respondents. The interviewers read the questions from the questionnaire to the respondents and record their answers. The data collection process may also involve collecting biological data, such as height, weight, and blood samples. MEASURE DHS has used paper questionnaires for years. Now the project is beginning to use handheld computers in some countries.

The DHS has many strategies to ensure that the data collected are of the best possible quality. One such strategy is the careful selection and training of interviewers. The DHS also matches interviewers with respondents based on gender: many questions asked in the DHS are of a sensitive or personal nature, and respondents are likely to feel more comfortable sharing this kind of information with someone of the same sex. Therefore, men interview men, and women interview women.

Before and during the fieldwork, the implementing agency often uses various strategies to inform the public about the upcoming survey. Radio spots and television and newspaper advertisements, as well as in-person teams, may be used to sensitize the population about the survey and encourage their participation.

Stage 3: Data Processing

As soon as possible after data collection has been completed in a cluster of about 30 households, the field editor sends the completed questionnaires back to the main survey headquarters so that data entry can begin while the teams are still in the field. Data entry personnel enter information from the completed questionnaires into computers and the data are checked rigorously for consistency and quality. Field check tables are also produced during data entry; these show whether the teams are collecting the data correctly by checking response rates and various measures of data quality. If it is discovered that a team is performing poorly, the survey managers can work with the team to improve the way they are collecting the data. If problems with the data are too bad or too

Handout 4.1

widespread, data collection teams will have to revisit the clusters involved and redo the fieldwork. When all data are entered, analysis begins. This involves the tabulation of data to produce the tables for the final report.

Stage 4: Data Dissemination and Use of Data

DHS survey results are published in a country final report which includes text and tables. In addition, the MEASURE DHS project works with the country implementing agency to develop a summary Key Findings report, PowerPoint presentations, and other print materials. A national seminar is held to launch the survey results and disseminate the key findings. Countries may also opt to hold regional seminars. The survey dataset is made available to researchers worldwide for further analysis at no charge. The datasets can be requested at the DHS website: www.measuredhs.com.

Module 4 Pre-Test

1. **DHS survey questionnaires are pre-tested before the survey is implemented.**
 - a. True
 - b. False

2. **Write the number next to the stage of the DHS so they are in the correct order.**

_____ Training and fieldwork (data collection)

_____ Data dissemination and use of data

_____ Survey preparation & questionnaire design

_____ Data processing and analysis

3. **If an interviewer visits a household and does not make contact, how many return visits is she/he required to make?**
 - a. One
 - b. Two
 - c. Three
 - d. None

4. **Some indicators, such as mortality and fertility, require larger sample sizes than others in order to obtain a reliable estimate.**
 - a. True
 - b. False

5. **The percentage of eligible individuals who agree to participate in the survey is called:**
 - a. Error rate
 - b. Sampling frame
 - c. Response rate
 - d. Probability sampling

Module 4 Post-Test

1. **DHS survey questionnaires are pre-tested before the survey is implemented.**
 - a. True
 - b. False

2. **Write the number next to the stage of the DHS so they are in the correct order.**

_____ Training and fieldwork (data collection)

_____ Data dissemination and use of data

_____ Survey preparation & questionnaire design

_____ Data processing and analysis

3. **If an interviewer visits a household and does not make contact, how many return visits is she/he required to make?**
 - a. One
 - b. Two
 - c. Three
 - d. None

4. **Some indicators, such as mortality and fertility, require larger sample sizes than others in order to obtain a reliable estimate.**
 - a. True
 - b. False

5. **The percentage of eligible individuals who agree to participate in the survey is called:**
 - a. Error rate
 - b. Sampling frame
 - c. Response rate
 - d. Probability sampling

ANSWER KEY

Module 4

1. Survey questionnaires are pre-tested before the survey is implemented.

a. True

b. False

2. Write the number next to the stage of the DHS so they are in the correct order.

2 Training and fieldwork (data collection)

4 Data dissemination and use of data

1 Survey preparation & questionnaire design

3 Data processing and analysis

3. If an interviewer visits a household and does not make contact, how many return visits is she/he required to make?

a. One

b. Two

c. Three

d. None

4. Some indicators, such as mortality and fertility, require larger sample sizes than others in order to obtain a reliable estimate.

a. True

b. False

5. The percentage of eligible individuals who agree to participate in the survey is called:

a. Error rate

b. Sampling frame

c. Response rate

d. Probability sampling