Data: What is the DHS?
Topics: Purpose, history, and design of the Demographic and Health Surveys

What is the DHS?
- Nationally & subnationally representative
- Cross-sectional, ~ every 5 years
- Complex sampling
- 6,000-30,000 households
  - Primary respondents: women 15-49
  - Secondary respondents: men 15-59
- Implementers: National Government with support by USAID’s MeasureDHS

1. What is a DHS?

The Demographic and Health Surveys (DHSs) are nationally and sub-nationally representative household surveys. I start this lecture by summarizing the study design of typical DHS, then back up and explain each study design concept in more detail by looking at an example DHS.

Usually DHSs are representative at the provincial-level, but increasingly surveys are representative at a district-level in countries where health system decision are made at the district-level, for example, in Rwanda and Malawi.

DHSs are cross-sectional, and repeated approximately every five years in low- and middle-income countries. This means that entirely unique sets of villages and neighborhoods are sampled each time a DHS is conducted.

Complex sampling methods are used. Urban areas, for example, often comprise a small portion of the overall population in low-income countries, but are oversampled in a DHS to ensure that there is sufficient sample size to make urban population estimates.

In a typical DHS, 6,000 to 30,000 households will be sampled across a country. In the selected households, all women age 15-49 are invited to complete an interview. In many countries, a subsample of households also include male respondents age 15-59.

DHSs are implemented jointly by the country’s National Institute of Statistics, or similar government agency, with technical and financial support from ICF International under the US Agency for International Development (USAID) project, MeasureDHS. See www.measuredhs.com for more information.
2. DHS Sample Selection Process

I just gave an overview of the DHS study design, now let me review an example: the 2010 Rwanda DHS with more explanation about each aspect of the study design.

2.a Stratification

The first step is stratification. The Rwandan government decided to make estimates of demographic and health outcomes at the district-level since their health system administers care and implements policy by district. This means that the sample was stratified by district, or in other words, an independent sample of households was drawn in each of the 30 districts.

2.b First stage sampling

Within each district, all primary sampling units, or PSUs, are listed with their population size. Typically PSUs are the enumeration areas of the last census. In the 2010 Rwanda DHS, villages were used as the PSUs since the previous census was out of date at that time. Then a sample of PSUs are selected with probability proportionate to population size. Here I show a map of Kirehe District – these are not actually village boundaries, they are sector boundaries, but I use this map to show how all PSUs are listed, then sampled in the first stage of sampling. The number of PSUs – or the size of the first stage sample is often more important to the statistical power of the final sample, than the total number of households or individuals sampled. I talk more about this in the lectures on “Survey Analysis in Stata” and “Summary Statistics”.

2.c Oversampling

The next step it to check how many PSUs were selected in urban versus rural areas across the national sample. Urban and rural populations tend to be different in terms of their demographic and health characteristics, so it is important to be able to summarize the national statistics in these two subpopulations. In low-income countries like Rwanda, the proportion of the national population living in urban areas is small, so the total urban sample might not provide a sufficient sample size to
make urban estimates. Therefore, the survey implementers oversample in urban areas (urban areas are shown on this map in orange). How exactly was oversampling performed in the Rwanda 2010 DHS? The implementers increased the number of sampled PSUs from 16 to 20 in each of the three districts that comprise the capital city, Kigali. Oversampling a subpopulation leads to unequal probability of selection which is why it is important to use sampling weights when you analyze DHS data. People from Kigali are overrepresented in this sample. Sampling weight will be covered in greater detail later. Here is a map of all 492 PSUs that were included in the final 2010 Rwanda DHS, with urban neighborhoods displayed in orange, and rural villages displayed in blue.

2.d Household listing

After the final sample of PSUs were selected, the implementers send listing (mapping) teams to each of the 492 PSUs. These teams hand-sketched maps of each PSU; each map showed all households, main roads, and key landmarks in the village/neighborhood.

2.e Second Stage Sampling

In the second stage of sampling, these maps were brought back to the central office where the survey planners selected which households would be included in the final sample. It is important that this step is done outside of the field context by someone not directly involved in the interviewing process to prevent any conscious or subconscious bias, for example, selecting households that are easier to access or households that look more inviting to visit. Sampling households from the maps is usually done systematically, for example: selecting a household in a corner of the map, counting 10 households, selecting another household, counting 10 households, selecting another household, and so on. After this point the sampling is done and the interview team enters to field to implement the survey.

3. DHS Questionnaires

DHS surveys are comprised of multiple questionnaires: household questionnaire, women’s questionnaire, and men’s questionnaire. These
questionnaires are administered by an interviewer in the respondent’s home, and the respondent’s answers are recorded on paper by the interviewer.

How many questions do you think are in a typical DHS? Depending on how you count, during a typical DHS interview, women respondents answer about 115 questions about their households and household members, and 352 questions about themselves and their children. A single interview can easily take two hours to complete! Men respondents are asked roughly 165 questions. Additional time is needed in many surveys to collect blood samples, and measure the height and weight of respondents and their children.

4. Quantitative vs Qualitative Information

Is the DHS a qualitative or quantitative survey? The answer is, DHSs only collect quantitative data.

Quantitative data can be used to answer questions that start with the words When...?, How much...?, or How often...?, and the results can be generalized to the whole population.

Qualitative data, on the other hand, can answer in-depth questions about How...? and Why...?, however the results from qualitative data collection are usually specific to the respondents interviewed and may be difficult to generalize to the population.

DHS interviewers follow a specific protocol that guides them through hundreds of questions in a specific order and skip pattern. Although many questions are about sensitive topics and respondents may share deeply personal details from their lives, all data are coded according to predetermined response categories.

For example, a question in the domestic violence module on the women’s questionnaire asks “Does your husband ever do any of the following things to you: push you, shake you, or throw something at you? slap you? twist your arm or pull your hair?”, and so on. Although a respondent may describe an incidence of domestic violence in her personal life, her response is recorded in a quantitative format; with the value of 1=Yes and 2=No to each question.
5. Types of questions asked in the DHS

Here is the list of topics that are included in DHS surveys, and the number of surveys ever conducted on each topic. The main focuses of the DHS are fertility and reproductive health, and nutrition among women and children. Most surveys also collect data about HIV knowledge, behavior, and outcomes; Tuberculosis; malaria; tobacco usage; and health insurance coverage. A small sample of men are included in most surveys now. And a subsample of women are often asked about domestic violence. Most surveys now also collect GPS coordinates from PSUs.

6. Timeline of DHS Questionnaire Development

Over 300 surveys in 90 countries have been conducted since 1984 when USAID launched the MEASURE DHS Project. Initial surveys only included women respondents. From these interviews, four data files were created for each survey: women, kids, households, and household members. You will notice that data are duplicated across these four files; for example, the kids recode file includes mother and household data. Due to cultural taboos against asking unmarried women about sex, many early surveys only included married women. Soon after, though, MEASURE DHS began interviewing all women age 15-49, and included male respondents age 15-59 in a subset of households, which resulted in a fifth men’s data file.

In 1996, DHS field staff began collecting latitude/longitude coordinates in PSUs. Coordinates were created for previous surveys going back to 1987 using a method called gazettering. This resulted in a sixth dataset of PSU latitude and longitude coordinates.

In 2001, Mali was the first country to implement HIV testing on a subsample of DHS respondents, resulting in a seventh dataset with HIV test results. In 2006, Malaria and other blood tests were added to the DHS; but rather than choosing a subsample of respondents, the tests were administered on all respondents and therefore the results are included in existing data files.

Finally, in 2006 the World Health Organization provided new guidelines for estimating child
malnutrition based on a new dataset of “average” height and weight measurements from a global sample of children. Starting in 2006, all new DHS surveys reported stunting, wasting, and undernutrition results according to the old and the new child growth standards. For surveys before 2006, a special data file was added with malnutrition classifications according to the new growth standards. All DHS data files contain common identification variables which allow the files to be merged for analysis.

See the article “A Systematic review of Demographic and Health Surveys: data availability and utilization for research” by Madeleine Fabic and colleagues (published in 2012) for a detailed timeline of when biomarkers and key questions were added to the questionnaires.

7. What happens to DHS data after it is collected?

Data from the questionnaires are double entered into CSPro, a free survey data software by the US Census, and discrepancies are investigated and resolved in-country. The final raw dataset is sent to ICF International where it is “recoded” with standardized variable IDs and definitions. The variable v106, for example, is “Highest Level of Education” among women respondents in all DHS datasets, and the response categories are always “no education, primary, secondary, and higher”. As a result, DHS dataset can be easily combined across time and countries. And the MEASURE DHS Project creates a final report of summary statistics for each DHS.

8. Related Surveys

The MEASURE DHS Project is responsible for other surveys including the Service Provision Assessment (SPA) which provides a comprehensive overview of the country’s health service delivery system by interviewing health care providers, administrators, and patients across a sample of health facilities.

They also perform Malaria Indicator Survey (MIS) which record data about bednet ownership and use, and treatment of fever in young children. This questionnaire is often incorporated in a typical DHS, but may be administered separately.
And thirdly, the AIDS Indicator Survey (AIS) which provides data to monitor national HIV/AIDS programs and outcomes. This questionnaire is usually added to a typical DHS as well. Like Demographic and Health Surveys, these surveys are funded by USAID and implemented jointly by ICF International and the national government.

9. Uses of DHS Data

DHS data are among the most widely used sources of data in low- and middle-income countries. At the MeasureDHS website under the “publications” tab, you will find dozens of reports, working papers, and factsheets for each survey. DHS data are also usually incorporated into a country’s national statistical warehouse. In this way, DHS data are used directly by national governments to monitor health outcomes, and to set national public health agendas. International organizations such as the World Health Organization and United Nations also use DHS data to generate global health statistics. And finally, individual researchers and donor agencies use this robust, extensive dataset in scientific research.

10. Ways to Access DHS Data

There are three ways to access DHS data, and which approach you use will depend on your data needs and skills. The most basic DHS user cites summary statistics in the final DHS report. Users who want to tabulate statistics differently than what is available in the final report, but have limited time or skills, may use DHS’s online StatCompiler tool. StatCompiler is especially helpful when creating summary statistics across multiple surveys, though only a few dozen of the most common variables are available. You might check out MeasureDHS’s five minute video with step-by-step instructions about how to use StatCompiler.

Advanced user: “Recode” files

www.measuredhs.com/data/Using-DataSets-for-Analysis.cfm

The resources here at populationsurveyanalysis.com will help you to become an advanced DHS data user. Learning how to use the DHS recode files will allow you to perform summary statistics and advanced analyses with the 1000s of variables that are available in each DHS.

Basic user: DHS Reports

Publications tab > Search > Country


Intermediate/advanced user: STATcompiler

5 min video and step-by-step instructions:

http://preview.measuredhs.com/data/STATcompiler.cfm

DHS: www.measuredhs.com

WHO: www.who.int/gho/countries/en/

Google Scholar: http://scholar.google.com/