IPUMS Data Training Exercise: 
An introduction to IPUMS DHS 
(Exercise on child health and demographics)

Learning goals

- Understand how IPUMS DHS dataset is structured
- Understand how to choose between related variables based on your research question
- Understand how IPUMS DHS employs composite coding to handle differences in response codes.
- Develop a research question and hypothesis using IPUMS DHS dataset

Summary

In this exercise, you will gain basic familiarity with IPUMS DHS data exploration to answer research questions about health and children. After completing this exercise, you will have experience navigating the IPUMS DHS website and should be able to leverage these data to explore your own research interests.
Exploring the IPUMS DHS Website

On the IPUMS DHS website, https://www.idhsdata.org/idhs/, go to the "Select data" page.

Select Unit of Analysis

- Chose "Children" as the unit of analysis.
- Note that by selecting children as the unit of analysis, you will now only see variables available for children.

Select Variables

On the drop-down menu under "Topics," the first groups of variables relate to the characteristics of the child's mother. Variables about the child him/herself begin with the "Child demography" group.
1. Women of childbearing age are asked some basic questions about all of their children born in the 3-5 years (reference period varies across samples) before the survey, but many questions focus on a subset of their children.

   a. Which children are excluded from the variable KIDLIVESWITH (Child lives with female respondent or with others)? Use the Search tool to locate the variable and check the Universe tab in its variable description.

   b. Which children are excluded from the variable ANDRGPARA (Drugs for intestinal parasites taken during pregnancy)? Check the universe tab for the variable.

2. Go to the variable description for VACBCG.

   a. According to the Description tab, what illness does the BCG vaccine protect the child from?

   b. According to the frequencies on the Codes page, what was the primary source of information about whether a child had received the BCG vaccination, health cards recording the dates of vaccinations or mothers' verbal reports?

   c. According to the Comparability tab, what feature makes some samples from the 1980s different for VACBCG, and how should researchers deal with the problem if comparing data from these 1980 samples to later samples? Which countries' samples are affected by this issue?
3. Based on the variables displayed under "Misc. childhood illnesses," which sample collected the most information about parasitic diseases and their treatment in young children? What variable included in multiple samples collects information about drug treatment of intestinal parasites in young children?

4. Use the Select Samples tab to select only the most recent sample for each country. Then look at the frequencies in the variable DIAFLUIDUPDN. In general, is it more common for mothers to increase or decrease fluid intake when young children have diarrhea? Then identify a survey in which the opposite pattern holds true.

5. Open-ended question: Focusing again on the most recent samples only, what dimensions of the mother's socioeconomic status could you use to identify the most and least privileged children? What SES variable would you choose to use to stand in for stratification, and why?

6. From the DHS homepage, click on "Sample Descriptions." Focusing on the most recent samples for each country, which countries have the largest and second largest sample sizes of children?
Answers

1. Which children are excluded from the variable KIDLIVESWITH (Child lives with female respondent or with others)?
   a. Children who had died before the time of the survey are excluded from KIDLIVESWITH.
   b. Children who were not the mother’s last-born child but who were born in the five years before the survey are excluded from ANDRGPARA.

2. VACBCG
   a. Tuberculosis
   b. Dated entries on vaccination cards
   c. Some 1980s samples (for Ghana, Kenya, Mali, Uganda, and Zimbabwe) collected information about BCG vaccinations only from vaccination cards, not from verbal reports of mothers. To maximize comparability, studies of change in BCG vaccination rates that include these samples should exclude affirmative responses based on the mother’s report (code 24) in later samples for those countries.

3. Which sample collected the most information about parasitic diseases and their treatment in young children? What variable included in multiple samples collects information about drug treatment of intestinal parasites in young children? Ghana 1988 includes questions about both bilharzia (also called schistosomiasis, caused by parasitic worms) and Guinea worms and the treatment of these conditions. WORMTREKID reports, in multiple samples, whether the child was given drugs for intestinal parasites in the last 6 months.

4. Use the Select Samples tab to select only the most recent sample for each country. Then look at the frequencies in the variable DIAFLUIDUPDN. In general, is it more common for mothers to increase or decrease fluid intake when young children have diarrhea? What about the opposite? Decreasing fluids (giving the child somewhat
less, much less, or nothing to drink) seems to be a more common response to diarrheal disease than increasing fluids, according to mothers' reports. However, increasing fluids was reported more often than decreasing fluids for Cote d'Ivoire 2011, Zambia 2007, and Zimbabwe 2010.

5. What dimensions of the mother’s socioeconomic status could you use to identify the most and least privileged children? What SES variable would you choose to use to stand in for stratification, and why? **Answers may vary, but possible dimensions** include household wealth (WEALTHQ), mother's level of education or literacy or years of schooling, her partner's level of education or years of schooling, her husband's occupation, or the mother’s own employment status.

6. From the DHS homepage, click on "Sample Descriptions." Focusing on the most recent samples for each country, which countries have the largest and second largest sample sizes of children? **India 2005 (nearly 49,000 children age 0-4) and Nigeria 2013 (nearly 32,000 children age 0-4).**