**Information gathered: 12th – 17th October**

**(Understanding DHS Files & Documentation)**

**DATA COLLECTION**

1. **Survey Questionnaires (DHS, AIDS indicator surveys, Malaria indicator surveys)**

**DHS Surveys**

Demographic and Health Surveys (DHS) are nationally representative population-based surveys with large sample sizes (usually between 5,000 and 30,000 households). In all households, women age 15-49 are eligible to participate; in many surveys men age 15-54 from a sub-sample are also eligible to participate. Individual questionnaires include information on ***fertility, mortality, family planning, marriage, reproductive health, child health, nutrition, and HIV/AIDS.*** There are three core questionnaires in DHS surveys:

* **Household** **Questionnaire**: Information on characteristics of the household's dwelling unit and characteristics of usual residents and visitors. Contains information about **household schedule** (For usual members of the household and visitors, information is collected about age, sex, relationship to the head of the household, education, parental survivorship and residence, and birth registration) and **household characteristics** (the source of drinking water, toilet facilities, cooking fuel, assets of the household, and exposure to second-hand smoke. Cooking salt is tested for iodine content. In areas with a high prevalence of malaria, questions are asked about the use of mosquito nets). **The main purpose of the Household Questionnaire is to provide the mechanism for identifying women eligible for individual interview and children under five who are to be weighed, measured, and tested for anaemia.**

**Summary**: HR has household basic data (**HV0**), household schedule (**HV1:** for each household member), household characteristics (**HV2**), Malaria Info (**HVML**), HIV & Anaemia testing of eligible women/men (**HA/HB**), Anaemia testing children (**HC**), Basic Malaria Info (**HVML**), Malaria for household members (**HML**), etc.

* **Women’s Questionnaire:** For eligible respondents of household questionnaire. Contains information about **contraception, children’s health, HIV & STDs, reproductive behaviour, background characteristics, etc.**
* **Men's questionnaire:** For eligible respondents of household questionnaire. Shorter than women’s questionnaire, contains **background characteristics, reproduction, contraception (knowledge and use), HIV & STDs, health issues, etc.**

**AIDS Indicator Surveys**

Standardized tool to obtain indicators for effective monitoring of HIV/AIDS programmes. Information on AIS indicators are available for **HIV, HIV prevalence, etc.**

**Malaria Indicator Surveys**

A stand-alone household survey that collects data on recognized malaria indicators such as ownership of mosquito nets, diagnostic blood testing of children under 5, use of insecticides, etc. Also includes measurement of malaria parasites and anaemia (common result of malaria). MIS carried out to time with **high malaria transmission season**. Specially trained interviewers take drops of blood from eligible respondents who consent to test. Two tests carried out as follows:

* Immediate blood test for **anaemia** taken in the field with results provided in few minutes (rapid diagnostic test)
* Blood tested for **anaemia** using microscope in laboratory

It is recommended that both laboratory and field tests are taken on blood.

1. **BIOMARKER COLLECTION**

The DHS program is able to collect biomarker data relating to a wide range of health conditions including infectious and sexually transmitted diseases, chronic illnesses, etc. Data in DHS surveys is self-reported; however, biomarkers provide an objective profile of a specific diseases or health condition in a population. Test for HIV, Malaria, Syphilis, Chlamydia, Gonorrhoea, etc. The data can provide information to track trends in prevalence over time, improve training for health care providers, identify risk factors for certain conditions, etc. Many tests are carried out using **only a few drops of blood from a finger or heel. The results of some tests such as anaemia can be provided to the respondent within minutes or will be transported to lab for analysis. Dried blood spots on filter paper are used for HIV tests.** The biomarker questionnaire contains information about the height and weight of men, women and children to determine nutritional status, anaemia and HIV.

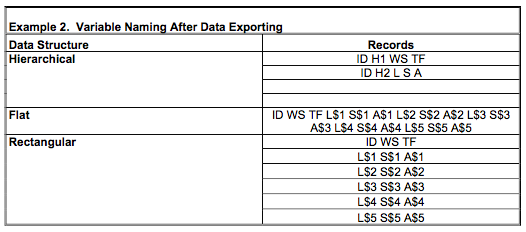
**Structure of DHS Data Files**

DHS uses ISSA (special software package) to process its surveys. It provides a mechanism to export the data to statistical packages SPSS, SAS and STATA. Data files exported using ISSA are stored using ASCII representation and they are provided in **flat or rectangular format.** Along with the data file, there is a **syntax file that described the data file.** These files have extensions – **SAS, DCT, DO, or SPS depending on the software the data file was exported to.** DHS produces **flat and rectangular files** for households, women, men and couples when the men questionnaire was included in the survey. **These files are mainly in recode format (standardized format allowing easy comparison among countries or different DHS phases in a country).**

**Hierarchical Files**

All data is stored in one ASCII file. The hierarchical file produced by ISSA has a two-level structure reflecting the relationship between the Household (Level 1) and Woman questionnaire (Level 2). Each household questionnaire has 0 or several women questionnaires depending on the number of eligible women listed in the household questionnaire. In each level, there can be one or more different records. For example, records in the household level in a typical DHS file can either be single (household characteristics) or multiple (household member listings).

**Flat Files**

This file has one record for each case. All variables in each case are placed one after the other on the same record. The multiple/repeating records of the file are placed one after the other in the record, with the maximum number of occurrences of each section being represented in the data file. Each variable in a repeating section is placed immediately after the preceding variable of the same occurrence, such that all variables for occurrence 1 precede all variables for occurrence 2 of a section. The length of the record in the flat data file is fixed.

Record type H1 is a single record with general household information, (e.g., on the source of water (SW) and toilet facilities (TF)). Record type H2 is a multiple record in which the same information (e.g., line number (L), sex (S) and age (A)) is entered for each members of the household. When exporting data, ISSA adds the dollar ($) sign followed by the occurrence number to variables belonging to repeating records. If the maximum number of a repeating record is less than 10, ISSA will add just one digit. If it is 10 or more but less than 100, it will add two digits and so on. **Before using an exported DHS data file, it is important to check for duplicate variable names. Software like SPSS allows only a maximum of eight characters for variable names. If variable Q1005A is defined in a record with a maximum number of occurrences of 20, it will be exported as Q1005A$01, Q1005A$02, Q1005A$20. When this syntax is read by SPSS it will chop the last character with the result of nine ($01 thru $09) variables with the same Q1005A$0 name.**

**Recode Files and Recode Variable Naming Conventions**

The following are units of analysis –

* HR – Household
* IR – Women
* MR – Men
* KR – Children
* CR – Couples

**File types**: Flat (FL) or Rectangular (RT)

**Variables in recode files**:

* HVxxx – Household standard variables
  + HAxx - Anthropometry & Anaemia for women
  + HCxx - Anthropometry & Anaemia for children
  + SHxx – Household, country-specific
* Vxxx – Women standard variables
  + Bxx – Birth history
  + Mxx – Pregnancy, post-natal care & breastfeeding
  + Hxx – Immunization and health
  + HWxx - Anthropometry & Anaemia of eligible women
  + MMxx – Maternal mortality (optional)
  + DVxx – Domestic Violence (optional)
  + Sxx – Women, country-specific
* MVxxx – Men standard variables
  + SMxx – Men, country-specific

**Men, women & household country specific variables, the xx = actual question number in the questionnaire.**

**Missing Value Codes:** 9, 99, 999, etc.

**Inconsistent Codes:** 7, 97, 997, etc.

**‘Don’t Know’ response:** 8, 98, 998, etc.

**Negative Response:** 0 i.e. “No education”, “No”, “No problem”, etc.

**Blank:** Not applicable to this respondent

**Missing, inconsistent, and “don’t know” codes must be excluded when calculating statistics such as means or medians**; otherwise they will be treated as real values. For example, if they are not excluded to calculate the mean age at first sex, eventually the mean will be inflated by ages 97, 98, and 99.

**Sampling Weights**

Sampling weights are 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. In the DHS surveys, many times the sample is selected with unequal probability to expand the number of cases available (and hence reduce sample variability) for certain areas or subgroups for which statistics are needed. In this case, weights need to be applied when tabulations are made of statistics to produce the proper representation. **Sample weights are calculated to six decimals but are presented in the standard recode files without the decimal point. They need to be divided by 1,000,000 before use to approximate the number of cases.**

**Variable Descriptions**

**MAP file:** Contains all variables in recode, variable codes, lengths of values corresponding to the variables, etc. **Very clear to understand in text file.**

**SAS file:** Contains same as MAP file, harder to understand as it is meant to be opened using a software.

**Cote D’Ivoire Data**

**Available Files:**

* **Household Recode**

One record for each household. Includes member’s rooster but no information from individual men/women questionnaire. **Unit of analysis:** Household.

* **Person Recode**

One record for each household member. **Unit of analysis:** Household member.

* **Individual Recode**

One record for every eligible woman defined in the household schedule. Has variables from women’s questionnaire and also household questionnaire.

* **Child Recode**

One record for each child of interviewed women, born 5 years before the survey. Data for the mother of the child is included.

* **Men Recode**

One record for every eligible man as defined in the household schedule.

* **Couple Recode**

One record for each couple.

* **Birth Recode**

One record for every child born to an interviewed woman. It is the full birth history of all interviewed women. **Can be used to calculate mortality rate.**

* **AIDS Recode**

One record for every individual whose blood was drawn for HIV testing. Can be linked to household, men and individual recode files.

* **Geographic Data**

One record for every cluster the survey was conducted. Each cluster latitude and longitude of the center of the cluster.

* **Geospatial Covariates**

One record per cluster the survey was conducted containing most common geospatial covariates. Contains population count, population density and malaria (number of people showing clinical symptoms of malaria in cluster) for years 2005, 2010 and 2015, etc.

**Files Needed for DHS Metrics:**

1. **HIV Rate:** AIDS Recode
2. **Child Mortality Rate:** Birth Recode, Child Recode (if necessary)
3. **Malaria Rate:** Household Recode
4. **Women’s Access to Health:** Individual Recode

**Observations:**

* Surveys done over 352 clusters

Week 1:

Background readings

Getting access to data

Downloading data

Week 2:

Understanding data

How data is gathered

Data available

Variables and their meanings

Which recodes are useful

Which DHS metrics to chose

Week 3:

qGIS, shape file, administrative areas, clusters

which variables are required for each metric

sample weight

cutting down data