

# **Sub-Saharan Africa**

## **Reference Manual for**

### **Harmonizing Household Surveys**

The World Bank  
1818 H Street, N.W.  
Washington, D.C.  
20433, USA.

April 2019

## Table of Contents

<b>Chapter 1: Overall harmonization guidelines .....</b>	<b>4</b>
1.1 Datalibweb .....	5
1.1 Folder and file structure .....	6
1.4 Guidelines across modules .....	6
1.4 Qcheck.....	7
<b>Chapter 2: Module P – Poverty-related Variables.....</b>	<b>8</b>
Table 2.1 Sample, Geography and Basic Household Identifier.....	9
Table 2.2 Consumption Expenditure Variables .....	12
<b>Chapter 3: Module H – Household-Level Variables .....</b>	<b>18</b>
Table 3.1 Sample, Geography and Basic Household Identifier.....	18
Table 3.2 Housing and Utilities .....	19
Table 3.3 Access to Social Amenities.....	28
Table 3.4 Ownership of Durable Assets.....	29
Table 3.5 Household Remittances .....	32
<b>Chapter 4: Module I - Individual-Level Variables .....</b>	<b>35</b>
Table 3.0 Sample, Geography and Basic Household Identifier.....	35
Table 3.1 Basic Demographic Characteristics.....	35
Table 3.2 Literacy and Education .....	40
Table 3.5 Migration.....	46
<b>Chapter 5: Module L - Labor Force Variables.....</b>	<b>49</b>
Table 4.0 Sample, Geography and Basic Household Identifier.....	49
Table 4.1 Household Chores .....	50
Table 4.2 Labor Screening Questions last 7-days .....	52
Table 4.3 Primary Employment last 7-days .....	54
Table 4.4 Secondary Employment last 7-days .....	60
Table 4.5 Employment last 12-months.....	65
<b>Annex I: The Methodology of constructing employment variables .....</b>	<b>68</b>
<b>Annex II: International Standard Industrial Classification of All Economic Activities (ISIC).....</b>	<b>74</b>
ISIC Rev. 4.0 categories .....	74
ISIC Rev. 3.1 categories .....	77
<b>Annex II: International Standard Classification of Occupations (ISCO) .....</b>	<b>79</b>
<b>Annex III: Data checks and evaluation .....</b>	<b>80</b>
(a) Validation and certification.....	80

<b>(b) Sources of error reviews</b> .....	<b>82</b>
(i) Sampling errors .....	83
(ii) Non-sampling errors.....	83
<b>Annex IV: ISO 3166-1 Alpha-3 Country Codes (Sub-Saharan Africa) .....</b>	<b>84</b>

## CHAPTER 1: OVERALL HARMONIZATION GUIDELINES

As Sub-Saharan African economies become more open and globalized, huge opportunities are created for individuals and families. Yet, a large fraction of households has not benefited sufficiently, and economic and social inequality is real and, in some cases, growing. Household surveys are a data source providing rich information on living standards and the impact of economic changes on individuals and households. Unfortunately, this source of information is largely underutilized due to the complexity of household surveys and the significant time required to prepare the survey data for analytical work.

The Sub-Saharan Team for Statistical Development (SSATSD) seeks to eliminate the bottleneck of analyzing household survey data by extracting about 200 variables from existing household surveys and ensuring that have the same definition and variable names. These variables include household consumption, access to infrastructure (water, electricity, etc.), employment status, education, and health. Invariably, in each survey, questions will be asked in a different manner, which poses challenges to consistently define harmonized variables. The harmonized household survey data presents the best available variables with harmonized definitions.

This manual presents detailed guidelines for harmonizing household survey data into a set of commonly defined variables that are available in most types of household surveys. To ensure the quality and transparency of the final harmonized data, it is critical to document the harmonization process and check the final data for quality concerns. This assures that the results can be replicated from the original household survey data with ease and that the final data provides reliable temporal and cross-country comparisons.

Four harmonized modules are prepared for each survey. Each of these modules contain a theme of harmonized variables that have the same variable names and definitions. The four harmonized modules are:

1. **Module P: Poverty-related variables:**

This module contains consumption variables, regional identifiers, spatial/temporal prices indices, variables indicating national poverty lines, and variables indicating whether households are classified as poor.

2. **Module H: Household-level variables (except for poverty-related variables):**

This module contains information on housing amenities, ownership of assets, access to infrastructure and services, and household remittances.

3. **Module I: Individual-level variables (except labor force variables):**

This module contains basic characteristics of individuals such as age, sex, literacy, education, and migration status.

4. **Module L: Labor force variables:**

This module contains information on labor force variables, such as labor force status, industry, sector of employment, wages, etc.

## 1.1 DATALIBWEB

In order to ensure the transparency and replicability of the harmonized data, a strict method of organizing folders and files is used. This method ensures that different versions of harmonizations are kept track of, and that users and future members of the harmonization team can run the harmonization .do-files without changing file paths.

The method applied for directory organization and file name conventions follows a practice adopted across regions and implemented through datalibweb. Datalibweb is a data system specifically designed to enable users to access the most up to date versions of non-harmonized (original/raw) and harmonized datasets of different collections across Global Practices. It can easily perform computations relevant for poverty and shared prosperity analysis based on the micro data from different harmonized collections: EAPPOV, ECAPOV, MNAPOV, SARMD, SEDLAC, SSAPOV, and the global collection GPWG.

Datalibweb can be installed in two ways

1. Directly from Stata: In order to get install to Datalibweb command in Stata, type the following code, and click on the datalibweb (hyperlink) to install in your computer.
  - Close all Stata sessions
  - Enter this line in Stata “[net from http://eca/povdata/datalibweb/\\_ado](http://eca/povdata/datalibweb/_ado)”
2. Manual installation: In addition, users can install the package the manual way.
  - Get the file from this link: [http://eca/povdata/datalibweb/\\_ado/datalibweb.zip](http://eca/povdata/datalibweb/_ado/datalibweb.zip)
  - Copy with replacement all the files into c:/ado, without changing the folder structure.

Once datalibweb is installed, and access to data has been granted, all raw data for a survey can be access with the following command:

```
datalibweb, country(CCC) year(YYY) type(SSARAW) surveyid(SURVEYNAME) clear,
```

where CCC stands for ISO 3 letter country code (see Annex III), YYY is the survey year according to IHSN standards, which is when the fieldwork started, and SURVEYNAME is the survey acronym.

When harmonizing surveys, harmonizers should always load data this way through datalibweb. This assures that no local file paths are used to load the data, and thus that other individuals who have access to the raw data can run the .do-files.

All documents related to a survey, such as questionnaires and technical reports, can be accessed through the following command:

```
datalibweb, country(CCC) year(YYY) type(SSARAW) surveyid(SURVEYNAME) request(doc)
```

Once a harmonization is done, the final harmonized files will be stored in datalibweb and can be accessed through the following command:

```
datalibweb, country(CCC) year(YYY) type(SSARAW) surveyid(SURVEYNAME) mod(MODULENAME), where MODULENAME takes the value, P, H, I or L.
```

## 1.1 FOLDER AND FILE STRUCTURE

The back-end of datalibweb contains a very specific folder structure and file naming convention. Although we do not work in these folders directly when working with data, it is useful for each harmonizer to copy the folder structure locally. As such, before harmonizing a survey using this manual, the harmonizer should first create sub-directories as instructed below. Additionally, all harmonization files must be named per this manual. This rigorous procedure is to ensure a seamless integration with datalibweb and that different versions of the harmonizations are kept track of.

All harmonizers will get assigned a folder on a server, \\WBGMSAFR1001\AFR\_Database\SSAPOV-Harmonization, with his/her name. This should be the parent directory from which all harmonizations are saved and from which all work is conducted. This folder should contain subfolders with the ISO3 country codes of the countries with which the harmonizer is working. Within each country-folder, there should be a folder with the name CCC\_YYYY\_SURVEYNAME for each of the surveys the harmonizer has been working on. For example, if a person is working on harmonizing the 2015 HICES survey of Ethiopia, then all material related to this should be saved in this path: \\WBGMSAFR1001\AFR\_Database\SSAPOV-Harmonization\[Name of harmonizer]\ETH\ETH\_2015\_HICES\. This folder should also be saved as a global in the beginning of each .do-file.

Each survey-specific folder should have two subfolders with the following content:

- “01.Programs”: This folder should contain the 4 .do-files used to construct each module, respectively. Each .do-file should match to a module, and there should be no .do-file except for the four used to generate the modules. If some preliminary data cleaning is needed, this should be included in the other .do-files. The .do-files should not call each other or any other .do-files.
- “02.Output”: This folder should contain the 4 .dta-files with the harmonized modules

All .do-files that do the harmonization, and each .dta that contains a harmonized module should be named according to the following convention.

CCC\_YYY\_SURVEYNAME\_v0x\_M\_v0y\_A\_SSAPOV\_MODULENAME.do

CCC\_YYY\_SURVEYNAME\_v0x\_M\_v0y\_A\_SSAPOV\_MODULENAME.dta

Here “v0x” is the version of the raw data. This will almost always be v01, but if errors were found in the original data and a new version of data is received from the National Statistical Office, then it will be called v02, etc. “v0y” is the version of the harmonized data. This will often be v01, but if an error is found in harmonized data, and the .do-file needs to be updated then the new .do-file and .dta-file will carry v02, etc.

All of this assures that any team member can run the .do-file without any changes and code and that, if the path becomes outdated, only one line of code needs to be changed.

## 1.4 GUIDELINES ACROSS MODULES

A number of harmonization guidelines are applicable across the four modules:

- In all .do-files, to the extent possible, the variables should be created and coded in the order that they appear in this manual.

- Frequently, surveys do not have information on all variables that we seek to harmonize. In this case, the variables should still be created as missing such that all variables appear in all modules.
- In the P and H-module it is important that the household identifier *hid* uniquely identifies observations. That is *isid hid* should not return an error. Likewise, it is important in the I and L-modules that *hid* and *pid* uniquely identifies observations. That is *isid hid pid* should not return an error. An implication of this is that *hid* (and *pid* in the I and L-modules) should have no missing values.
- The exact same households should appear in all four modules. As a user, it is confusing and can be frustrating when all households do not merge between different modules. If some households do not exist in the P module, but they do exist in the H, I or L module, then they should be removed from the H, I and L modules. If some households exist in the P module but not in the H, I, or L module, then they should be created (with missing values) in the H, I and L module. In general, the households that should appear should be the ones that are used for national poverty estimation. Although a few country specific cases may not be able to follow this rule, it should apply in general.
- Any critical assumption that is made in the course of the harmonization should be stated clearly in the .do-file
- For each module a labelling .do-file exists. At the end of each .do-file this labelling .do-file should be inserted. This ensures that all variables have the exact same labels and formats across surveys. The labelling .do-file also create a few variables that are a function of some of the other harmonized variables.

#### 1.4 QCHECK

Once a module is harmonized, a quality check will be performed on the harmonized data. A quality check program called *qcheck*, has been created to this end. *Qcheck* tests if all variables are in the dataset, if all variables have the correct format, if the variables take plausible values, and if some of the variables are mutually inconsistent. For example, *qcheck* will test if the age variable in the I-module takes any negative values or values above 120, both of which indicate an error. It will also flag if someone is coded to have no education in one education variable, but have completed secondary education in another variable.

## CHAPTER 2: MODULE P – POVERTY-RELATED VARIABLES

The most common measures used for living standards are consumption and income. Income refers to actual earnings from productive activities and transfers while consumption refers to resources consumed. While income may be used as an indicator to measure welfare, it is not ideal in countries where the majority of the population works in informal sectors, such as small business, work on land, etc., as net income becomes very difficult to measure in these cases. In addition, for self-employed incomes may be zero or negative for a given period, even though these individuals could have wealth to draw upon. In these cases, income is a poor proxy for welfare. Consumption is therefore thought to provide a better picture of a household's standard of living than a measure of current income.

For these reasons, the vast majority of countries in Sub-Saharan Africa use consumption to measure poverty. The p-module contains a list of variable related to consumption, such as it's breakdown by food and non-food consumption, consumption per capita and per adult equivalent, as well as indicators for whether a household's consumption falls short of the poverty line.

There are limitations of household surveys in measuring household consumption: -

- A household survey is an instrument relying mostly on self-reported data and on household members' memory. This latter makes the estimates heavily dependent on the length of the recall period.
- Although consumption is the best household welfare indicator, it is impossible to distinguish between consumption and expenditure (a bulk purchase could cause overestimation of household welfare). What was bought may also not necessarily be consumed by households in its entirety and thus it becomes difficult to separate consumption and expenditure.
- The duration of the recall period may lead to under- or over-estimation of the reported data and expenditure consumption surveys should be designed to envisage such a problem.
- A perennial issue relating to national income in any country has been the difference between the System of National Accounts (SNA) Statistics and National Sample Survey estimates on consumption expenditure. The SNA private household consumption expenditure is available as a macro estimate and a scalar for the nation as a whole while the National Sample Survey consumption estimates are available separately by different sub-groups such as provinces, rural and urban areas among others, which can be aggregated to derive a national estimate. The estimates of private consumption from these two sources are different, primarily as these are derived from different concepts and estimation approaches.

Consumption aggregates are not comparable across households if prices differ across time and space. For this reason, a lot of effort goes into adjusting the consumption aggregates temporally and spatially. The P-module contains several variables trying to document whether spatial and or temporal deflation was used for a particular survey, both for purposes of national poverty estimation and for purposes of international poverty estimation.

**TABLE 2.1 SAMPLE, GEOGRAPHY AND BASIC HOUSEHOLD IDENTIFIER**

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	harmonization	Type of harmonization	<p><b>String variable</b> Should equal SSAPOV. This variable is automatically generated in the labeling file: gen harmonization = "SSAPOV"</p>
2	country	Country code	<p><b>String variable</b> 3-character length (Annex IV)</p>
3	survey	Type of survey	<p><b>String variable</b> Specifies the type of survey. Possible names are: HBS, LSMS, IS, CWIQ, etc. Upper-case letters should be used.</p>
4	survey_coverage	<p><b>Survey coverage</b> 1 = National 2 = Urban 3 = Rural 4 = Other</p>	<p><b>Numeric variable</b></p>
5	usemicrodata	<p><b>Use of microdata</b> 0 = Grouped 1 = Micro</p>	<p><b>Numeric variable</b></p>
6	year_IHSN	4-digit year of survey based on IHSN standards	<p><b>Numeric variable</b> This is the start year of survey based on the IHSN standards. It should be identical to the year used for file-naming purposes.</p>
7	region1	Subnational ID – highest level	<p><b>String variable</b> This variable should contain the first-level administrative divisions of a country. It should contain numeric entries in string format using the following naming convention: "1 – Hatay" (as string). The code below shows how to turn a numeric variable with labels into the format required: gen region1="" qui levelsof inputvar, local(lev) foreach cc of local lev { cap loc la_`cc': label(inputvar) `cc' if !_rc { qui replace region1=""`cc' - `la_`cc'"" if inputvar == `cc' } } <a href="#">This link</a> may be helpful in terms of identifying the right variables.</p>
8	region2	Subnational ID – second highest level	<p><b>String variable</b> This variable should contain the second-level</p>

			<p>administrative divisions of a country. It should contain numeric entries in string format using the following naming convention: "1 – Hatay" (as string). The code below shows how to turn a numeric variable with labels into the format required:</p> <pre>gen region2="" qui levelsof inputvar, local(lev) foreach cc of local lev { cap loc la_`cc': label(inputvar) `cc' if !_rc { qui replace region2="`cc' - `la_`cc'" if inputvar == `cc' } } }</pre> <p><a href="#">This link</a> may be helpful in terms of identifying the right variables</p>
9	<b>region3</b>	<b>Subnational ID – third highest level</b>	<p><b>String variable</b></p> <p>This variable should contain the third-level administrative divisions of a country. It should contain numeric entries in string format using the following naming convention: "1 – Hatay" (as string). The code below shows how to turn a numeric variable with labels into the format required:</p> <pre>gen region3="" qui levelsof inputvar, local(lev) foreach cc of local lev { cap loc la_`cc': label(inputvar) `cc' if !_rc { qui replace region3="`cc' - `la_`cc'" if inputvar == `cc' } } }</pre> <p><a href="#">This link</a> may be helpful in terms of identifying the right variables</p>
10	<b>lev_agg</b>	<b>Level at which data are disaggregated for analysis in country</b>	<p><b>Numeric variable</b></p> <p>Lowest level at which data is considered to be representative. It usually corresponds to the lowest level of poverty reporting areas. It will often (but not always) be identical to one of the region variables but in numeric format with labels rather than in string format.</p>
11	<b>strata</b>	<b>Strata</b>	<p><b>Numeric variable</b></p>
12	<b>rururb</b>	<p><b>Area of residence</b>  <i>0 = Rural</i>  <i>1 = Urban</i></p>	<p><b>Numeric variable</b></p> <p>Each country defines this jurisdiction according to a certain criterion. In transition economies where 'semi-urban' is a recognized category which includes 'villages of the town type' this will be collapsed into the 'urban' category unless if the country defines</p>

			these as rural towns.
13	<b>capital</b>	<b>Capital/city, other urban and rural classification</b>	<b>Numeric variable</b> This is a country-specific variable which may indicate capital city or a different urban/rural classification than the one in <i>rururb</i> . Each numeric code should have a label.
14	<b>cluster</b>	<b>Primary sampling unit (enumeration area)</b>	<b>Numeric variable</b> Primary sampling unit based on country requirements.
15	<b>hhno</b>	<b>Household number</b>	<b>Numeric variable</b> Household number
16	<b>hid</b>	<b>Household unique identification</b>	<b>String variable</b> This variable should uniquely identify observations and cannot be missing, i.e. <i>isid hid</i> should return no error.
17	<b>int_month</b>	<b>Month of interview visit</b>	<b>Numeric variable</b> The month when the survey questionnaire was administered to the household.
18	<b>int_year</b>	<b>Year of interview visit</b>	<b>Numeric variable</b> The year when the survey questionnaire was administered to the household.
19	<b>hhsiz</b>	<b>Household size</b>	<b>Numeric variable</b> Total number of residents (regular members). The definition of regular member is country-specific.
20	<b>ctry_adq</b>	<b>Adult equivalent scale</b>	<b>Numeric variable</b> Definition varies from country to country, as different adult scales exist worldwide. Total number of adult equivalent people in household: <ul style="list-style-type: none"> <li>• Must be greater 0.</li> <li>• Must be less than or equal to <i>hhsiz</i> (household size).</li> </ul> Can be provided by the NSO.
21	<b>wta_hh</b>	<b>Household weights</b>	<b>Numeric variable</b> To obtain household estimates, this is the weight to be used in all computations referring to household-level estimates. This variable cannot be used for poverty estimation. The interpretation is the proportion of households with a certain characteristic is XX%.
22	<b>wta_pop</b>	<b>Population weights</b>	<b>Numeric variable</b> This variable should be used for poverty estimation. The interpretation is the proportion of individuals with

			a certain characteristic is XX%. This variable is automatically generated in the labelling file: <code>gen wta_pop = wta_hh*hhsz</code>
23	<b>wta_cadq</b>	<b>Adult equivalent weights</b>	<b>Numeric variable</b> In a number of countries, this weight is used to derive the proportion of poor population. The interpretation is the proportion of adult equivalent population with a certain characteristic is XX%. This variable is automatically generated in the labelling file: <code>gen wta_cadq = wta_hh*c_adq</code>

**TABLE 2.2 CONSUMPTION EXPENDITURE VARIABLES**

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	<b>welfaretype</b>	<b>Type of welfare measure (income, consumption, expenditure)</b> "CONS" "INC" "EXP"	<b>String variable</b> Specifies the type of welfare aggregate used for poverty estimation in a country. CONS=consumption INC=income EXP=expenditure
2	<b>fdtexp</b>	<b>Purchased and auto-consumption food expenditure, nominal (annual)</b>	<b>Numeric variable</b> Country-derived by the NSO.
3	<b>nfdtexp</b>	<b>Purchased &amp; auto-consumption non-food expenditure, nominal (annual)</b>	<b>Numeric variable</b> Country-derived by the NSO.
4	<b>hhtexp</b>	<b>Household food and non-food consumption expenditure, nominal (annual)</b>	<b>Numeric variable</b> Country-derived by the NSO. This variable is automatically generated in the labelling file: <code>gen hhtexp = fdtexp+nfdtexp</code> If the raw data does not separate between food and non-food consumption, create this file instead of letting it be created in the labelling file.
5	<b>pc_fd</b>	<b>Per capita food consumption expenditure, nominal (annual)</b>	<b>Numeric variable</b> Country-derived by the NSO. This variable is automatically generated in the labelling file: <code>gen pc_fd=fdtexp/hhsz</code>
6	<b>pc_hh</b>	<b>Per capita food and non-food consumption, nominal (annual)</b>	<b>Numeric variable</b> Country-derived by the NSO. This variable is automatically generated in the labelling file: <code>gen pc_hh=hhtexp/hhsz</code>

7	padq_fd	Per adult equivalent food consumption expenditure, nominal (annual)	<b>Numeric variable</b> Country-derived by the NSO. This variable is automatically generated in the labelling file: gen padq_fd = fdtxp/ctry_adq
8	padq_hh	Per adult equivalent food and non-food consumption, nominal (annual)	<b>Numeric variable</b> Country-derived by the NSO. This variable is automatically generated in the labelling file: gen padq_hh=fdtxp/ctry_adp
9	fdspindex	Food spatial price index	<b>Numeric variable</b> Country-derived by the NSO.
10	nfdspindex	Non-food spatial price index	<b>Numeric variable</b> Country-derived by the NSO.
11	spindex	Spatial price index	<b>Numeric variable</b> Country-derived by the NSO.
12	fdtpindex	Food temporal price index	<b>Numeric variable</b> Country-derived by the NSO.
13	nfdtpindex	Non-food temporal price index	<b>Numeric variable</b> Country-derived by the NSO.
14	tpindex	Temporal price index	<b>Numeric variable</b> Country-derived by the NSO.
15	fdpindex	Spatial/temporal food index	<b>Numeric variable</b> Country-derived by the NSO. This variable should never be missing. If no separate food spatial/temporal price index is used, set this equal to sptpindex.
16	nfdpindex	Spatial/temporal non-food index	<b>Numeric variable</b> Country-derived by the NSO. This variable should never be missing. If no separate non-food spatial/temporal price index is used, set this equal to sptpindex.
17	pindex	Final spatial/temporal price index	<b>Numeric variable</b> Country-derived by the NSO. This variable should be the one used to derive <i>wel_ppp</i> and <i>wel_abs</i> . Should never be missing. If no temporal/spatial deflation is used, generate a column of 1's.
18	fdtxpdr	Purchased and auto-consumption food expenditure, deflated (annual)	<b>Numeric variable</b> This variable is automatically generated in the labelling file: gen fdtxpdr = fdtxp/fdpindex
19	nfdtxpdr	Purchased & auto-	<b>Numeric variable</b>

		<b>consumption non-food expenditure, deflated (annual)</b>	This variable is automatically generated in the labelling file: <code>gen nfdtexpdr = nfdtexp/nfdpindex</code>
20	<b>hhtexpdr</b>	<b>Household food and non-food consumption expenditure, deflated (annual)</b>	<b>Numeric variable</b> This variable is automatically generated in the labelling file: <code>gen hhtexpdr = hhtexp/pindex</code>
21	<b>pc_fddr</b>	<b>Per capita food consumption expenditure, deflated (annual)</b>	<b>Numeric variable</b> This variable is automatically generated in the labelling file: <code>gen pc_fddr = fdtexpdr/hhsize</code>
22	<b>pc_hhdr</b>	<b>Per capita food and non-food consumption expenditure, deflated (annual)</b>	<b>Numeric variable</b> This variable is automatically generated in the labelling file: <code>gen pc_hhdr = hhtexpdr/hhsize</code>
23	<b>padq_fddr</b>	<b>Per adult equivalent food consumption expenditure, deflated (annual)</b>	<b>Numeric variable</b> This variable is automatically generated in the labelling file: <code>gen padq_fddr = fdtexpdr/ctry_adq</code>
24	<b>padq_hhdr</b>	<b>Per adult equivalent food &amp; non-food consumption expenditure, deflated (annual)</b>	<b>Numeric variable</b> This variable is automatically generated in the labelling file: <code>gen padq_hhdr = hhtexpdr/ctry_adq</code>
25	<b>wel_abs_deflation</b>	<b>Spatial/temporal deflation used for national poverty estimation</b> 0 = Neither spatially nor temporally deflated 1 = Spatially deflated 2 = Temporally deflated 3 = Both spatially and temporally deflated	<b>Numeric variable</b>
26	<b>wel_abs_pcpadq</b>	<b>Per adult equivalent or per capita adjustment used for national poverty estimation</b> 0 = Per capita 1 = Per adult equivalent	<b>Numeric variable</b>
27	<b>wel_abs</b>	<b>Welfare aggregate used for national poverty estimation (annual)</b>	<b>Numeric variable</b> This is the welfare aggregate used by the country to estimate its national poverty. This aggregate can be nominal or spatially/temporally deflated. It should equal one of

			<p>these four variables: pc_hh, padq_hh, pc_hhdr, padq_hhdr.</p> <p>This variable is automatically generated in the labelling file:</p> <pre> gen wel_abs = . if wel_abs_deflation==0 &amp; wel_abs_pcpadq==0 { replace wel_abs = pc_hh } if wel_abs_deflation==0 &amp; wel_abs_pcpadq==1 { replace wel_abs = padq_hh } if inlist(wel_abs_deflation,1,2,3) &amp; wel_abs_pcpadq==0 { replace wel_abs = pc_hhdr } if inlist(wel_abs_deflation,1,2,3) &amp; wel_abs_pcpadq==1 { replace wel_abs = padq_hhdr } </pre>
28	wel_fd	<b>Food part of welfare aggregate used for national poverty estimation (annual)</b>	<p><b>Numeric variable</b></p> <p>This is the food part of the welfare aggregate used by the country to estimate its national poverty. This aggregate can be nominal or spatially/temporally deflated. It should equal one of these four variables: pc_fd, padq_fd, pc_fddr, padq_fddr.</p> <p>This variable is automatically generated in the labelling file:</p> <pre> gen wel_fd = . if wel_abs_deflation==0 &amp; wel_abs_pcpadq==0 { replace wel_fd = pc_fd } if wel_abs_deflation==0 &amp; wel_abs_pcpadq==1 { replace wel_fd = padq_fd } if inlist(wel_abs_deflation,1,2,3) &amp; wel_abs_pcpadq==0 { replace wel_fd = pc_fddr } if inlist(wel_abs_deflation,1,2,3) &amp; wel_abs_pcpadq==1 { replace wel_fd = padq_fddr } </pre>
28	pl_abs	<b>National Absolute Poverty line (annual)</b>	<p><b>Numeric variable</b></p> <p>Country-derived by the NSO.</p>
27	pl_fd	<b>National Food Poverty line (annual)</b>	<p><b>Numeric variable</b></p> <p>Country-derived by the NSO.</p>

29	<b>pl_ext</b>	<b>National Hardcore poverty line (annual)</b>	<b>Numeric variable</b> Country derived by the NSO. This line may be identical to the food poverty line or may be different.
30	<b>poor_abs</b>	<b>Absolute poor based on pl_abs</b> 1 = Poor 0 = Non-poor	<b>Numeric variable</b> This variable is automatically generated in the labelling file: <code>gen poor_abs = wel_abs&lt;pl_abs</code>
31	<b>poor_fd</b>	<b>Food poor based on pl_fd</b> 1 = Poor 0 = Non-poor	<b>Numeric variable</b> This variable is automatically generated in the labelling file: <code>gen poor_fd = wel_fd&lt;pl_fd</code>
32	<b>poor_ext</b>	<b>Hard core (extreme) poor based on pl_ext</b> 1 = Poor 0 = Non-poor	<b>Numeric variable</b> This variable is automatically generated in the labelling file: <code>gen poor_ext = wel_abs&lt;pl_ext</code>
33	<b>converfactor</b>	<b>Conversion factor</b>	<b>Numeric variable</b> Specifies value for additional conversion factors if needed (e.g. from US\$ to LCUs; currency change).
34	<b>wel_PPPnom</b>	<b>Welfare aggregate used for international poverty estimation (nominal, annual)</b>	<b>Numeric variable</b> This is the nominal expenditure welfare aggregate. This should equal <code>pc_hh</code> . This variable is automatically generated in the labelling file: <code>gen wel_PPPnom = pc_hh</code>
35	<b>wel_PPPdr</b>	<b>Welfare aggregate used for international poverty estimation (deflated, annual)</b>	<b>Numeric variable</b> This is the spatial and/or temporal deflated expenditure welfare aggregate. This should equal <code>pc_hhdr</code> . This variable is automatically generated in the labelling file: <code>gen wel_PPPdr = pc_hhdr</code>
36	<b>wel_PPP_deflation</b>	<b>Spatial/temporal deflation used for international poverty estimation</b> 0 = Neither spatially nor temporally deflated 1 = Spatially deflated 2 = Temporally deflated 3 = Both spatially and temporally deflated	<b>Numeric variable</b>
37	<b>wel_PPP</b>	<b>Welfare aggregate used for international poverty estimation (annual)</b>	<b>Numeric variable</b> This is the final welfare variable used for international poverty monitoring purposes, that feeds into the GMD. It should equal either <code>wel_PPPnom</code> or <code>wel_PPPdr</code> .

			<p>This variable is automatically generated in the labelling file:</p> <pre>gen wel_PPP = . if wel_PPP_deflation==0 { replace wel_PPP = wel_PPPnom } if inlist(wel_PPP_deflation,1,2,3) { replace wel_PPP = wel_PPPdr }</pre>
--	--	--	---

### CHAPTER 3: MODULE H – HOUSEHOLD-LEVEL VARIABLES

The H-module records household-level information and includes information on housing characteristics and utilities, access to various amenities measured in terms of distances/time, and ownership of durable goods among others.

**TABLE 3.1 SAMPLE AND BASIC HOUSEHOLD IDENTIFIER**

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	country	Country code	<b>String variable</b> 3-character length (Annex IV)
2	year_IHSN	4-digit year of survey based on IHSN standards	<b>Numeric variable</b> This is the start year of survey based on the IHSN standards. It should be identical to the year used for file-naming purposes.
3	hhno	Household number	<b>Numeric variable</b> Household number
4	hid	Household unique identification	<b>String variable</b> This variable should uniquely identify observations and cannot be missing, i.e. <code>isid hid</code> should return no error.
5	wta_hh	Household weights	<b>Numeric variable</b> To obtain household estimates, this is the weight to be used in all computations referring to household-level estimates. The interpretation is the proportion of households with a certain characteristic is XX%.

**TABLE 3.2 HOUSING AND UTILITIES**

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	<b>ownhouse</b>	<b>Ownership of dwelling unit</b> <i>1 = Yes</i> <i>0 = No</i>	<b>Numeric variable</b> Refers to occupancy status of the dwelling unit by the residing household. Yes refers to actual ownership or if household is in the process of purchasing the dwelling unit living in. No refers to renters, squatters, free housing among others.
2	<b>rooms</b>	<b>Number of habitable rooms</b>	<b>Numeric variable</b> This refers to number of rooms in the dwelling unit and may consist of one or more structures. Must be $\geq 1$ . Includes all rooms used for living, sleeping and eating. Excludes store, bathrooms and kitchens.
3	<b>roofcs</b>	<b>Main material used for roof (country specific)</b>	<b>String variable</b> This refers to the variable on roof material (if any), as it comes in the survey. If more than one material is used for structure, the dominant material is the information required. The format should be code and value label. For example, "1 - Stone"; "2 - Mud"; etc.
4	<b>roof</b>	<b>Main material used for roof</b> <i>1 = Thatch (bamboo/grass)</i> <i>2 = Earth (adobe, mud, clay)</i> <i>3 = Wood</i> <i>4 = Iron/Metal sheets</i> <i>5 = Concrete/cement/stone</i> <i>6 = Tiles/bricks</i> <i>9 = Other</i>	<b>Numeric variable</b> This variable must be coded from <i>roofcs</i> . - <i>Earth</i> includes adobe, mud. Includes all building technique that relies on earth or mud put over a frame or mixed with other materials for strength. - <i>Thatch</i> includes grass or any form of natural vegetation for roofing. - <i>Iron sheets</i> are processed or galvanized iron or steel sheets. Does not include tins. - <i>Cement</i> includes concrete and stone blocks. - <i>Tiles/bricks</i> are a thin, flat or slab of hard material and include baked/unbaked bricks made of clay or other human-made building blocks. - <i>Other</i> includes tin from cans, cardboard among others.
5	<b>wallcs</b>	<b>Main material used for external walls (country specific)</b>	<b>String variable</b> This refers to the variable on external wall material (if any), as it comes in the survey. If more than one material is used for structure, the dominant material is the information required. The format should be code and value label. For example, "1 - Stone"; "2 - Mud"; etc
6	<b>wall</b>	<b>Main material used for external walls</b> <i>1 = Earth (adobe, mud, clay)</i>	<b>Numeric variable</b> This variable must be coded from <i>wallcs</i> .

		<p>2 = <i>Thatch (bamboo/grass)</i></p> <p>3 = <i>Bricks</i></p> <p>4 = <i>Wood panels</i></p> <p>5 = <i>Iron/metal sheets</i></p> <p>6 =</p> <p><i>Concrete/cement/stone</i></p> <p>9 = <i>Other</i></p>	
7	<b>floorcs</b>	<b>Main material used for floor (country specific)</b>	<b>String variable</b> This refers to the variable on floor material (if any), as it comes in the survey. If more than one material is used for structure, the dominant material is the information required. Format should be code and value label. For example, "1 - Stone"; "2 - Mud"; etc
8	<b>floor</b>	<b>Main material used for floor</b> 1 = <i>Earth (adobe, mud, clay)</i> 2 = <i>Bricks</i> 3 = <i>Wood planks</i> 4 = <i>Polished wood/tiles</i> 5 = <i>Cement</i> 9 = <i>Other</i>	<b>Numeric variable</b> This variable must be coded from <i>floor</i> . - <i>Earth</i> includes adobe, mud. - <i>Bricks</i> slab of hard material and include baked/unbaked bricks made of clay or other human-made building blocks - <i>Cement</i> includes concrete and stone.
9	<b>watercs_type</b>	<b>Type of water questions used in the survey</b> 1 = <i>Drinking water</i> 2 = <i>General water</i> 3 = <i>Both</i> 4 = <i>Other</i>	<b>Numeric variable</b> This variable records the type of question(s) asked about access to water in the survey. For example, if the survey had a specific question on the water source on drinking water, or on water source on general water, or both. Subsequent question on water will depend on this response.
10	<b>watercs</b>	<b>Main source of water (country specific)</b>	<b>String variable</b> This refers to the variable on the main water source (if any), as it comes in the survey. If more than one water source, only main source required. In some surveys, drinking water is asked and is differentiated from other water uses. In these cases, use the drinking water source to code this variable. If two sources of water are available (water source during the wet and dry season), use water source during dry season. The reason for using water during the dry season is that the world is experiencing global warming and the climate is changing rapidly. The format should be code and value label. For example, "1 - Pipe"; "2 - Spring"; etc.
11	<b>watercs_d</b>	<b>Main source of water during the dry season (country specific)</b>	<b>String variable</b> Question must be explicitly asked in survey on water source during the dry season.

			<p>Labels must be translated to English.</p> <p>If more than one water source, only main source required.</p> <p>In some surveys, drinking water is asked and is differentiated from other water uses. Use the drinking water source to code this variable.</p> <p>For each value label, there should be a space between the hyphen.</p> <p>Format should be code and value label. For example, “1 – Pipe”; “2 – Spring”; etc.</p>
12	<b>water14</b>	<b>Main source of drinking water (14 categories)</b> <i>1 = Piped water into dwelling</i> <i>2 = Piped water to yard/plot</i> <i>3 = Public tap or standpipe</i> <i>4 = Tube well or borehole</i> <i>5 = Protected dug well</i> <i>6 = Protected spring</i> <i>7 = Bottled water</i> <i>8 = Rainwater</i> <i>9 = Unprotected spring</i> <i>10 = Unprotected dug well</i> <i>11 = Cart with small tank/drum</i> <i>12 = Tanker-truck</i> <i>13 = Surface water</i> <i>14 = Other</i>	<p>Must be coded from WATERCS.</p> <p><i>Piped into dwelling</i>, also called a household connection is defined as water service pipe connected with in-house plumbing to one of more taps (e.g. in kitchen, bathroom, etc.). Privacy is the criterion here.</p> <p><i>Piped water to yard/plot</i>, also referred as a yard connection. This is defined as a piped water connection to a tap placed in the yard or plot but outside the house.</p> <p><i>Public standpipe</i> refers to water delivered via pipe but may or may not be within compound (water point shared among households). This refers to public stand-taps or community water points.</p> <p><i>Tubewell or borehole</i> is a deep hole that has been drilled with the purpose of reaching groundwater supplies. Boreholes/tubewells are constructed with casing or pipes, which prevent the small diameter hole from caving in and protects the water source from infiltration by run-off water. Water is delivered from a tubewell or borehole through a pump, which may be powered by human, animal, wind or electric, diesel or solar means. Boreholes/tubewells are usually protected by a platform around the well, which leads spilled water away from the borehole and prevents infiltration of run-off water at the well head.</p> <p><i>Protected dug well</i> is a dug well that is protected from run-off water by a well lining or casing that is raised above the ground level and a platform that diverts spilled water away from the well. A protected dug well</p>

			<p>is also covered to prevent any infiltration.</p> <p><i>Protected spring</i> is typically protected from any run-off infiltration by a “spring box”, which is constructed of brick or concrete and is built around the spring so that water flows directly out of the box into a pipe without being exposed to outside pollution.</p> <p><i>Surface water</i> is water located above the ground and includes lakes, rivers, ponds, streams, canals and irrigation canals.</p> <p>Cart with a small tank/drum refers to water sold by a provider into a community. The types of transportation used include donkey carts, motorized vehicles and other means.</p> <p><i>Tanker-truck</i> is water trucked into a community and sold from a water truck. The water source unknown.</p> <p><i>Other</i> includes other water sources not mentioned above.</p>
13	<b>water8</b>	<p><b>Main source of drinking water (8 categories)</b>  1 = Piped water (own tap)  2 = Public tap or standpipe  3 = Protected well  4 = Unprotected well  5 = Surface water  6 = Rainwater  7 = Tanker-truck, vendor  8 = Other</p>	<p><i>Wells</i> include springs, boreholes but <b>must</b> be protected from any possible sources of contamination such as surface water or seepage.</p> <p>recode water14 (1=1) (2 3=2) (4 5 6=3) (9 10=4) (13=5) (8=6) (11 12=7) 14=8),gen(water8)  ta water14 water8</p>
14	<b>waterpipe</b>	<p><b>Household has piped water</b>  0 = No  1 = Yes, in premise  2 = Yes, but not in premise  3 = Yes, unstated whether in or outside premise</p>	<p>Main water source is piped water which can be within household, plot or public standpipe.</p> <p>“Piped” is the condition.</p> <p>recode water14 (1=1) (2=2) (3=3) (else=0), gen(waterpipe)  If water14 is missing but you have the information to code waterpipe, do not use the red code above.</p>
15	<b>waterimp</b>	<p><b>Household has improved water sources</b>  1 = Yes  0 = No</p>	<p>An improved drinking water source, by nature of its construction and design, is likely to protect the source from outside contamination, from fecal matter.</p>

			<p>Improved drinking water sources include:</p> <ul style="list-style-type: none"> <li>• Piped water into dwelling, plot or yard</li> <li>• Public tap/stand pipe</li> <li>• Tube well/borehole</li> <li>• Protected dug well</li> <li>• Protected spring and</li> <li>• Rainwater collection</li> </ul> <p>On the other hand, unimproved drinking water sources are:</p> <ul style="list-style-type: none"> <li>• Unprotected dug well,</li> <li>• Unprotected spring,</li> <li>• Cart with small tank/drum,</li> <li>• Tanker truck,</li> <li>• Surface water (river, dam, lake, pond, stream, canal, irrigation channel and any other surface water), and</li> <li>• Bottled water (if it is not accompanied by another improved source)</li> </ul> <p>Source: (WHO &amp; UNICEF, 2010)  <a href="http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=8">http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=8</a></p>
			<p>recode water14 (1/6 8=1) (else=0),gen(waterimp)</p>
16	adiswat_d	<b>Actual distance to main water point (kms) during the dry season</b>	<p>This refers to actual distance to water point (<b>one way</b>) used by household in kms during the dry season.</p> <p>If no season is specified, use this variable.</p> <p>By convention: 1 km = 1000 m  1 km = 5/8 mile.</p> <p>If within dwelling, code zero.</p>
17	adiswat_w	<b>Actual distance to main water point (kms) during the wet season</b>	<p>This refers to actual distance to water point (<b>one way</b>) used by household in kms.</p> <p>By convention: 1 km = 1000 m  1 km = 5/8 mile.</p> <p>If within dwelling, code zero.</p> <p>If no season is specified, code this as missing.</p>
18	atimwat_d	<b>Actual time taken to</b>	<p>This refers to actual time taken to water point used by</p>

		<b>main water point (mins) during the dry season</b>	household. If roundtrip provided, divide by 2.
19	<b>atimwat_w</b>	<b>Actual time taken to main water point (mins) during the wet season</b>	This refers to actual time taken to water point used by household. If roundtrip provided, divide by 2.
20	<b>toiletcs</b>	<b>Main toilet facility (country specific)</b>	<b>String variable</b>  Labels must be translated to English. Make sure translation is correct from a language expert.  For each value label, there should be a space between the hyphen.  Format should be code and value label. For example, “1 – Flush”; “2 – VIP”; etc.
21	<b>toilet14</b>	<b>Main toilet facility (14 categories)</b> <i>1 = A flush toilet</i> <i>2 = A piped sewer system</i> <i>3 = A septic tank</i> <i>4 = Pit latrine</i> <i>5 = Ventilated improved pit latrine (VIP)</i> <i>6 = Pit latrine with slab</i> <i>7 = Composting toilet</i> <i>8 = Special case</i> <i>9 = A flush/pour flush to elsewhere</i> <i>10 = A pit latrine without slab</i> <i>11 = Bucket</i> <i>12 = Hanging toilet or hanging latrine</i> <i>13 = No facilities or bush or field</i> <i>14 = Other</i>	Must be coded from TOILETCS.  If several types of toilet are used, only main source required.  Helps to identify the efforts that are needed to prevent common and basic diseases, in particular water-borne diseases.  <i>Flush toilet</i> also referred as a Water Closet (WC) is a toilet that disposes waste matter by using water to flush it through a drainpipe to a main sewer or septic tank or pit latrine. This excludes a <ul style="list-style-type: none"> <li>• pour flush uses a water seal, but unlike a flush toilet, it uses water poured by hand for flushing (no cistern is used)</li> <li>• flush toilet to “somewhere else” such a flushed to a river, hanging toilet or some place</li> </ul> <i>Ventilated Improved Pit latrine (VIP):</i> The primary features of VIP latrines consist of an enclosed structure (roof and walls) with a large diameter (110mm), PVC vertical ventilation pipe running outside the structure from the pit of the latrine to vent above the roof. They often will have concrete slabs containing the latrine hole.  <i>A composting toilet</i> is a type of dry toilet that uses a predominantly aerobic processing system to treat human excreta, by composting or managed aerobic

			<p>decomposition. These toilets generally use little to no water and may be used as an alternative to flush toilets.</p> <p><i>Pit latrine</i> is a simple pit latrine but covered or with a slab.</p> <p><i>No facility</i> includes, open fields, bush.</p> <p><i>Other</i> includes bucket, pan, and open/uncovered pit latrines among others.</p>
22	<b>toilet6</b>	<p><b>Main toilet facility (6 categories)</b>  <i>1 = Flush toilet</i>  <i>2 = Ventilated Improved Pit (VIP) latrine</i>  <i>3 = Composting toilet</i>  <i>4 = Pit latrine with slab</i>  <i>5 = No facility</i>  <i>9 = Other</i></p>	<p>Must be coded from TOILETCS.</p> <p>recode toilet14 (1/4=1) (5=2) (7=3) (6=4) (13=5) (else=9),gen(toilet6)</p>
23	<b>toiletflush</b>	<p><b>Access to flushed toilet</b>  <i>0 = No</i>  <i>1 = Yes, in premise</i>  <i>2 = Yes, but not in premise including public toilet</i>  <i>3 = Yes, unstated whether in or outside premise</i></p>	<p>Must be asked in survey explicitly.</p> <p>Do not guestimate.</p>
24	<b>toiletshared</b>	<p><b>Is toilet facility shared with other households?</b>  <i>1 = Yes</i>  <i>0 = No</i></p>	<p>This question must have been asked in the survey.</p> <p>If question not asked leave as missing.</p>
25	<b>toiletimp</b>	<p><b>Does household have access to improved sanitation?</b>  <i>1 = Yes</i>  <i>0 = No</i></p>	<p>This includes TOILET6&lt;=4 and not shared.</p> <p>An improved sanitation facility is one that likely hygienically separates human excreta from human contact. Improved sanitation facilities include:</p> <ul style="list-style-type: none"> <li>• Flush or pour-flush to piped sewer system, septic tank or pit latrine,</li> <li>• Ventilated improved pit latrine,</li> <li>• Pit latrine with slab and</li> <li>• Composting toilet</li> </ul> <p>Sanitation facilities are not considered improved when shared with other households, or open to public use.</p>

			<p>While, unimproved sanitation include:</p> <ul style="list-style-type: none"> <li>• Flush or pour-flush to elsewhere,</li> <li>• Pit latrine without slab or open pit,</li> <li>• Bucket, hanging toilet or hanging latrine and</li> <li>• No facilities or bush or field (open defecation)</li> </ul> <p>If question of shared toilet facility is asked, use the variable to recode appropriately.</p> <p>Source: (WHO &amp; UNICEF, 2010)  <a href="http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=9">http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=9</a></p>
26	<b>fuelcookcs</b>	<b>Main cooking fuel (country specific)</b>	<p><b>String variable</b></p> <p>If several fuels asked in survey, only main source required.</p> <p>Labels must be translated to English. Make sure translation is correct from a language expert.</p> <p>For each value label, there should be a space between the hyphen.</p> <p>Format should be code and value label. For example, “1 – Electricity”; “2 – Firewood”; etc.</p>
27	<b>fuelcook</b>	<p><b>Main cooking fuel</b>  <i>1 = Electricity</i>  <i>2 = Gas</i>  <i>3 = Kerosene</i>  <i>4 = Charcoal</i>  <i>5 = Firewood</i>  <i>9 = Other</i></p>	<p>If several fuels asked in survey, only main source required.</p> <p><i>Firewood</i> includes both purchased and collected.</p> <p><i>Electricity</i> refers to mains, generator and solar energy provided by the government or private entity.</p> <p><i>Other</i> includes fuel derived from coffee waste, saw dust, crop residue, cow dung among others.</p>
28	<b>fuellighcs</b>	<b>Main lighting fuel (country specific)</b>	<p><b>String variable.</b></p> <p>If several fuels asked in survey, only main source required.</p> <p>Labels must be translated to English. Make sure translation is correct from a language expert.</p> <p>For each value label, there should be a space between the hyphen.</p>

			Format should be code and value label. For example, “1 – Electricity”; “2 – Firewood”; etc.
29	<b>fuelligh</b>	<b>Main lighting fuel</b> 1 = <i>Electricity</i> 2 = <i>Gas</i> 3 = <i>Kerosene</i> 4 = <i>Candles</i> 9 = <i>Other</i>	If several fuels asked in survey, only main source required.  <i>Electricity</i> refers to mains, generator and solar energy provided by the government or private entity.  <i>Other</i> includes fuel derived from coffee waste, saw dust, crop residue, cow dung among others.
30	<b>elecsorce</b>	<b>Main source of electricity</b> 1 = <i>Mains</i> 2 = <i>Solar</i> 3 = <i>Generator</i> 4 = <i>Other</i> 5 = <i>No electricity</i>	Use both FUELCOOK and FUELLIGH.  FUELLIGH should be the main one to use.  If electricity source not specified, code “other” but this should be on a country-to-country situation.
31	<b>electricity</b>	<b>Connection of electricity in dwelling from mains only</b> 1 = <i>Yes</i> 0 = <i>No</i>	This specifies access to electricity connection to a main grid only (ELECSOURCE=1).  Note: having an electrical connection says nothing about the actual electrical service received by the household in a given country or area.
			<a href="#">Check consistency by type ta electricity elecsorce</a>
32	<b>kitchen</b>	<b>Separate kitchen in dwelling</b> 1 = <i>Yes</i> 0 = <i>No</i>	
33	<b>bath</b>	<b>Bathing facility such as shower or bathtub in the dwelling</b> 1 = <i>Yes</i> 0 = <i>No</i>	
34	<b>garbdispcs</b>	<b>Garbage and trash disposal (country specific)</b>	<b>String variable</b>  Labels must be translated to English. Make sure translation is correct from a language expert.  For each value label, there should be a space between the hyphen.  Format should be code and value label. For example, “1

			– Collected”; “2 – Buried”; “3 - Street”; etc.
35	<b>garbdisp</b>	<b>Garbage and trash disposal</b> <i>1 = Collected</i> <i>2 = Buried/burned</i> <i>3 = Discarded in empty lots, street, rivers</i> <i>9 = Other</i>	Refers to only garbage or trash generated by household.

**TABLE 3.3 ACCESS TO SOCIAL AMENITIES**

In some surveys this may not be available for each household but will be present in the community survey. The distances and time are to the nearest services from the household irrespective of whether the household uses these services.

All distances and times refer to *single/one way journeys*. Please note that all data for distances and time that are not categorized (continuous) are to the nearest 2 decimal places.

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	<b>dispsch</b>	<b>Distance to nearest elementary/primary school (kms)</b>	One way.  This refers to distance to nearest primary school in kms.  By convention 1 km = 1000 meters 1 km = 5/8 mile  If roundtrip provided, divide by 2.  This is a continuous variable.  If survey question is pre-coded, do not guestimate this into a continuous variable. Leave as missing.
2	<b>timpsch</b>	<b>Time taken to nearest elementary/primary school (minutes)</b>	One way.  This refers to time taken to reach nearest primary school in mins.  By convention 1 hr = 60 min.  If roundtrip provided, divide by 2.  This is a continuous variable.

			If survey question is pre-coded, do not guestimate this into a continuous variable. Leave as missing.
3	<b>disheal</b>	<b>Distance to nearest health facility (kms)</b>	<p>One way.</p> <p>This refers to distance to nearest health facility in kms.</p> <p>By convention 1km = 1000 meters 1 km = 5/8 mile</p> <p>If roundtrip provided, divide by 2.</p> <p>This is a continuous variable.</p> <p>If survey question is pre-coded, do not guestimate this into a continuous variable. Leave as missing.</p>
4	<b>timheal</b>	<b>Time taken to nearest health facility (minutes)</b>	<p>One way.</p> <p>This refers to time taken to reach nearest primary school in mins.</p> <p>By convention 1hr = 60 min.</p> <p>If roundtrip provided, divide by 2.</p> <p>This is a continuous variable.</p> <p>If survey question is pre-coded, do not guestimate this into a continuous variable. Leave as missing.</p>

**TABLE 3.4 OWNERSHIP OF DURABLE ASSETS**

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	<b>radio</b>	<b>Ownership of radio</b> <i>1 = Yes</i> <i>0 = No</i>	Functioning radio includes a radio, radio cassette, and 3-in-1-radio cassette and regardless of what condition the asset is in.
2	<b>television</b>	<b>Ownership of television</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning television in house and regardless of what condition the asset is in.
3	<b>television_cable</b>	<b>Ownership of television cable</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning television cable in house and regardless of what condition the asset is in.
4	<b>video</b>	<b>Ownership of video</b> <i>1 = Yes</i>	Presence of a functioning video in house and regardless of what condition the asset is in.

		<i>0 = No</i>	
5	<b>landphone</b>	<b>Ownership of landline (fixed) phone</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning fixed land line telephone in house and regardless of what condition the asset is in.
6	<b>cellphone</b>	<b>Ownership of at least one cellular phone</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning cellular in house and regardless of what condition the asset is in.
7	<b>phone</b>	<b>Ownership of at least phone</b> <i>1 = Yes</i> <i>0 = No</i>	Where not specified as landline or cellphone, presence of a functioning land/cellular in house and regardless of what condition the asset is in.
8	<b>fridge</b>	<b>Ownership of refrigerator</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning refrigerator in house and regardless of what condition the asset is in.  Does not include a food freezer
9	<b>sewmach</b>	<b>Ownership of sewing machine</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning sewing machine in house and regardless of what condition the asset is in.
10	<b>washmach</b>	<b>Ownership of washing machine</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning washing machine in house and regardless of what condition the asset is in.
11	<b>fan</b>	<b>Ownership of fan</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning fan in house and regardless of what condition the asset is in.
12	<b>airconditioner</b>	<b>Ownership of air conditioner</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning air conditioner in house and regardless of what condition the asset is in.
13	<b>computer</b>	<b>Ownership of computer</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning computer in house.  Can be desktop or laptop.
14	<b>etablet</b>	<b>Ownership of an electronic tablet</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning tablet in house and regardless of what condition the asset is in.
15	<b>stove</b>	<b>Ownership of stove</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning stove or cooker in house and regardless of what condition the asset is in.

16	<b>oxcart</b>	<b>Ownership of animal cart</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning animal cart, which is used as a means of transport or a farm tool and regardless of what condition the asset is in.
17	<b>bicycle</b>	<b>Ownership of bicycle</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning bicycle and regardless of what condition the asset is in.
18	<b>boat</b>	<b>Ownership of boat</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning boat and regardless of what condition the asset is in.
19	<b>canoe</b>	<b>Ownership of canoe</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning canoe and regardless of what condition the asset is in.
20	<b>motorcycle</b>	<b>Ownership of motorcycle</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning motorcycle and regardless of what condition the asset is in.
21	<b>car</b>	<b>Ownership of private car</b> <i>1 = Yes</i> <i>0 = No</i>	Presence of a functioning car is important and regardless of what condition the asset is in.  This refers to car for household use and NOT a commercial vehicle.
22	<b>internet</b>	<b>Access to internet inside the house</b> <i>1 = Yes</i> <i>0 = No</i>	This variable should indicate whether the household can access the internet within their home. This could be from a computer, a phone, a tablet etc.  If the survey asks each individual in the household <i>separately</i> whether they have access to the internet, then the household is considered to have access if at least one individual has access.  If the survey asks if the internet connection is <i>working</i> , only consider the household to have access if the connection is working.  If the survey asks if individuals are <i>using</i> the internet, then this cannot be included as an indicator of internet access unless the survey also asks <i>where</i> internet is used, and one of the options is in the home.  If the survey asks where internet is ' <i>most often</i> ' used, then this is not good enough to identify internet access, as individuals may use the internet at home but use it more frequently at work.

**TABLE 3.5 HOUSEHOLD REMITTANCES**

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	hh_remit	<b>Did household receive any remittances?</b> <i>1 = Yes</i> <i>0 = No</i>	Did the household receive any remittances?  Source of remittances not important here.  If HH_REMIT=0 then subsequent questions are null and void
2	sex_rmt_1	<b>Sex of the 1<sup>st</sup> remittance sender</b> <i>1 = Male</i> <i>0 = Female</i>	The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).  In some countries, the remittances is by number of transactions, enter each transaction as a unique identifier. The reason being one cannot tell if this is the same sender or not. This applies to all questions in this section.
3	sex_rmt_2	<b>Sex of the 2<sup>nd</sup> remittance sender</b> <i>1 = Male</i> <i>0 = Female</i>	
4	sex_rmt_3	<b>Sex of the 3<sup>rd</sup> remittance sender</b> <i>1 = Male</i> <i>0 = Female</i>	
5	relat_rmt_1	<b>Relationship to the household head of the 1<sup>st</sup> remittance sender</b> <i>2 = Spouse</i> <i>3 = Son/daughter</i> <i>4 = Parents/parents-in-law</i> <i>5 = Grandchild</i> <i>6 = Son-in-law/daughter-in-law</i> <i>7 = Other relative</i> <i>9 = Non-relative</i>	
6	relat_rmt_2	<b>Relationship to the household head of the 2<sup>nd</sup> remittance sender</b> <i>2 = Spouse</i> <i>3 = Son/daughter</i> <i>4 = Parents/parents-in-law</i> <i>5 = Grandchild</i> <i>6 = Son-in-law/daughter-in-law</i> <i>7 = Other relative</i> <i>9 = Non-relative</i>	
7	relat_rmt_3	<b>Relationship to the</b>	

		<b>household head of 3<sup>rd</sup> remittance sender</b> <i>2 = Spouse</i> <i>3 = Son/daughter</i> <i>4 = Parents/parents-in-law</i> <i>5 = Grandchild</i> <i>6 = Son-in-law/daughter-in-law</i> <i>7 = Other relative</i> <i>9 = Non-relative</i>	
8	<b>des_mig_1</b>	<b>Destination of migration of the 1<sup>st</sup> remittance sending member</b> <i>1 = Capital</i> <i>2 = Within the country (but not capital)</i> <i>3 = Abroad</i>	The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).
9	<b>des_mig_2</b>	<b>Destination of migration of the 2<sup>nd</sup> remittance sending member</b> <i>1 = Capital</i> <i>2 = Within the country (but not capital)</i> <i>3 = Abroad</i>	
10	<b>des_mig_3</b>	<b>Destination of migration of the 3<sup>rd</sup> remittance sending member</b> <i>1 = Capital</i> <i>2 = Within the country (but not capital)</i> <i>3 = Abroad</i>	
11	<b>origin_rmt</b>	<b>Origin of the remittance senders</b> <i>1 = Domestic</i> <i>2 = Abroad</i> <i>3 = Both</i>	<b>Numeric variable</b> This variable is automatically generated in the labelling file: <pre>gen origin_rmt = 1 if inlist(des_mig_1,1,2) &amp; inlist(des_mig_2,1,2) &amp; inlist(des_mig_3,1,2) replace origin_rmt = 2 if des_mig_1==3 &amp; des_mig_2==3 &amp; des_mig_3==3 replace origin_rmt = 3 if origin_rmt==. replace origin_rmt = . if des_mig_1==. &amp; des_mig_2==. &amp; des_mig_3==.</pre>
12	<b>amt_rmt_1</b>	<b>Amount of annual remittance by the 1<sup>st</sup> remittance sender</b>	<b>Numeric variables</b> The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).
13	<b>amt_rmt_2</b>	<b>Amount of annual remittance by the 2<sup>nd</sup></b>	

		<b>remittance sender</b>	
14	<b>amt_rmt_3</b>	<b>Amount of annual remittance by the 3<sup>rd</sup> remittance sender</b>	
15	<b>amt_rmt_fd</b>	<b>Total amount of annual remittances received in food (annual)</b>	The total includes the remittances received in the form of food from all remittance senders.
16	<b>amt_rmt_oth</b>	<b>Total amount of annual remittances received in other forms (annual)</b>	The total includes the remittances received in other forms (cash, etc.) from all remittance senders.

## CHAPTER 4: MODULE I - INDIVIDUAL-LEVEL VARIABLES

This module extracts variables of individuals in the household and covers approximately forty quantitative variables. The information is organized in 5 tables that provide variables on basic household identification, demographic characteristics, education, health, and child’s vaccination and anthropometry.

**TABLE 3.0 SAMPLE AND BASIC HOUSEHOLD IDENTIFIER**

All Table 3.0 variables will be extracted from the Poverty file Table 2.0.

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	country	Country code	To be merged from the p-file. If you don’t have the p-file, create the variable.
2	year_IHSN	4-digit year of survey based on IHSN standards	To be merged from the p-file. If you don’t have the p-file, create the variable.
3	hhno	Household number	To be merged from the p-file. If you don’t have the p-file, create the variable.
4	hid	Household unique identification	To be merged from the p-file. If you don’t have the p-file, create the variable.
5	wta_hh	Household weights	<b>Numeric variable</b> To obtain household estimates, this is the weight to be used in all computations referring to household-level estimates. This variable cannot be used for poverty estimation. The interpretation is the proportion of households with a certain characteristic is XX%.

**TABLE 3.1 BASIC DEMOGRAPHIC CHARACTERISTICS**

The file may have different household size when compared to the poverty-level file. Make sure that the regular household members are selected in the same criterion as the Poverty-level file. Secondly, households that do not match the Poverty-level file must be dropped as they do not have the consumption component.

All variables are numeric unless specified.

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
----	---------------	-----------------	----------------------------------

1	pid	<b>Individual identification</b>	<p>Uniquely identifies the regular household members in each household.</p> <p>Sequentially numbered from 1 to <math>N</math> (household size).</p> <p>If the PID is a concatenation of HID and person ID, concatenate HID and leave PID only.</p> <p>Check that each household member ID is unique.  duplicates tag (hid pid),gen(dup)  tab dup</p>
2	ageyrs	<b>Age in completed years (continuous)</b>	<p>Age is an important variable for most socio-economic analyzes and must be established as accurately as possible.</p> <p>Missing ages must be left as missing.</p> <p>If 99=missing, recode to missing.</p> <p>If date of birth is provided, derive age and compare with the given recorded age.</p> <p>If age of Household head is missing, use the var=hhagey in the poverty file to replace the missing age of household head only.</p> <p>For children aged less than 5 years, this is used to interpret child malnutrition and survival data. Check consistency with age in months (AGEM) to get correct age in completed years.</p> <p>For older surveys, check consistency and maintain AGEYRS.</p>

			<p>This can only be done if date of birth and date of interview are provided.</p> <pre> gen bday=mdy(month,day,year) gen iday=mdy(imonth,iday,iyear) format bday iday %d gen age = (iday - bday)/365.25 gen ages=trunc(Age) gen diff=ages-recorded_age tab diff </pre>
3	agecat	Age intervals (string)	<p><b>String variable</b> Country specific categorical variable. It will only be created only when the country does not report the age of the interviewed people but intervals years of their age.</p> <p>Otherwise leave as missing.</p> <pre> gen outputvar="" qui levelsof inputvar, local(lev) foreach cc of local lev { cap loc la_`cc': label(inputvar) `cc' if !_rc { replace outputvar = "`cc' - `la_`cc'" if inputvar == `cc' } } </pre>
4	sex	<p><b>Sex</b> <i>1 = Male</i> <i>0 = Female</i></p>	Sex of the individual
5	relathcs	Relationship to household head (country-specific)	<p><b>String variable</b> Country-specific.</p> <p>For each value label, there should be a space between the hyphen (before and after).</p> <p>Code and name: Example: "1 - Head"; "2 - Spouse";</p>

			<p>“3 – Child”; etc.</p> <pre> gen relathhcs="" qui levelsof inputvar, local(lev) foreach cc of local lev { cap loc la_`cc': label(inputvar) `cc' if !_rc { qui replace relathhcs = "`cc' - `la_`cc'" if inputvar == `cc' } } </pre>
6	<b>relathh9</b>	<p><b>Relationship to household head (9 categories)</b></p> <p><i>1 = Head</i>  <i>2 = Spouse</i>  <i>3 = Child</i>  <i>4 = Parents/parents-in-law</i>  <i>5 = Grandchild</i>  <i>6 = Son-in-law/daughter-in-law</i>  <i>7 = Other relative</i>  <i>8 = Domestic help/paying boarder</i>  <i>9 = None relative</i></p>	<p>This refers to the relationship of each household member to the household HEAD.</p> <p>Must have one and only one head in each household.</p> <p>Child refers to biological child or adoptive children by either marriage or other reason.</p> <p>Domestic help (servant, guard, cook, baby-sitter among others) refers to a person who is paid for services rendered (cash or in-kind e.g. training skills, board and lodging) even if they are related to the head of household.</p> <p>Paying boarder is someone who pays the household for room and/or board.  None relative include friends living in household regularly.</p> <p>Use relathhcs to derive this variable after the edits.</p> <p><b>If all categories are not present in the questionnaire, leave this variable as missing</b></p>
7	<b>relathh6</b>	<p><b>Relationship to household head (6 categories)</b></p> <p><i>1 = Head</i>  <i>2 = Spouse</i>  <i>3 = Child</i>  <i>4 = Parents</i>  <i>5 = Other relative</i>  <i>6 = Non-relative</i></p>	<p>This refers to the relationship of each household member to the household HEAD.</p> <p>Must have one and only one head in each household.</p> <p>Other includes grandchild, in-laws, etc.</p> <p>Non-relative includes domestic help, paying</p>

			boarder, etc.
			recode relathh9 (1=1) (2=2) (3=3) (4=4) (5/7=5) (8/9=6), gen(relathh6)
8	marital6	<b>Marital status (6 categories)</b> <i>1 = Married monogamous</i> <i>2 = Married polygamous</i> <i>3 = Never married</i> <i>4 = Living together</i> <i>5 = Divorced/separated</i> <i>6 = Widowed</i>	<p>Polygamous unions exclude relationships that are not officially recognized such as mistresses, concubines.</p> <p>Check for consistency in married unions. Marital status for couples must be identical.</p> <p>Do not derive polygamous unions if survey does not ask. Leave variable as missing.</p> <p>If marital asked for persons only above 12 years, one can confidently guestimate that the children are "Never married".</p> <p><b>If all categories are not present in the questionnaire, leave this variable as missing</b></p>
9	marital5	<b>Marital status (5 categories)</b> <i>1 = Married</i> <i>2 = Never Married</i> <i>3 = Living together</i> <i>4 = Divorced/Separated</i> <i>5 = Widowed</i>	<p>The term 'married' may have different meanings in different countries. Married refers to both formal and informal unions such as common-law marriages, union coutumiere, free unions, living together.</p> <p>Check for consistency in married unions. Marital status for couples must be identical.</p> <p>Not all can be imputed but for children less than say 10 years, one can assume with some level of accuracy and certainty that they are never married.</p>
			recode marital6 (1 2=1) (3=2) (4=4) (5=5),gen(marital5) tab marital6 marital5

10	sp_pres	<b>Spouse of household head living in household</b> <i>1 = Yes</i> <i>0 = No</i>	<p>Code based on a question that asks whether the household head spouse lives in the household. Otherwise leave as missing.</p> <p>Only if MARITAL6&lt;=4 or MARITAL5&lt;=3</p> <p>DO NOT TRY TO DEDUCE FROM HOUSEHOLD MEMBERSHIP.</p> <p>Whether or not the member of household has a spouse (formal marriage or union/common law spouse) who lives in the household.</p> <p>However, under some special circumstances, a couple may be divorced/separated but living in the same household (dwelling unit) but in separate rooms. In this instance, sp_pres=1.</p> <hr/> <p>Check the ages and see if consistent. If a child is a spouse, go back to varname=relathhcs, relathh9, relathh6 and edit accordingly.</p> <p>tab ageyrs if sp_pres==1</p>
----	---------	--	--

**Note:**

- ✓ For any variable not collected in a country, variable should be created and left as missing (.) in the final harmonized file.
- ✓ Variables in the data files must follow the sequence in which they appear in the manual.

**TABLE 3.2 LITERACY AND EDUCATION**

All variables are numeric unless specified.

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	literacy	<b>Literacy status</b> <i>1 = Yes, can read and write</i> <i>0 = No, cannot read or write</i>	<p>For individuals aged 5 and above only.</p> <p>Value must be missing for all others.</p> <p>Literacy: Is the ability to both read and write with understanding, a short simple statement on his/her everyday life in any language. It will be useful to align measurements of literacy with this given standard international</p>

			<p>definition.</p> <p>Be careful while coding 1; one must be able to both read and write.</p> <p>If a person can either read or write, he/she will be considered illiterate (LITERACY=0).</p> <p>It can be assumed with some degree of accuracy that if respondent has secondary level and above of education, then must be literate.</p> <p>Also, persons with over 5 years of primary can be assumed literate. Can be programmed with EDUCYRS if literacy is missing for some members.</p>
2	<b>ed_mod_age</b>	<b>Education module application age (country-specific)</b>	<p>Minimum age for which education section is applied in country. The questionnaire and/or manual specifies this.</p> <p>For this reason, the lower age cutoff at which information is collected will vary from country to country.</p>
3	<b>everattd</b>	<p><b>Ever attended school</b>  <i>1 = Yes</i>  <i>0 = No</i></p>	<p>Value must be missing for individuals less than the required age (ed_mod_age).</p> <p>Depends on how school attendance is defined in a country. Example, in some countries, a criterion is placed to decide if ever attended school is valid or not and is determined by number of weeks or months or school term in attendance.</p> <p>Does not require to have completed any level of education.</p> <p>Indirect derivation if not collected by survey would be to program EDUCAT10 and ATSCHOOL. If ATSCHOOL=1 then ever attended=1. If EDUCAT10&gt;=3 and EDUCAT10&lt;=9, ever attended = 1.</p>
4	<b>educat10</b>	<p><b>Highest level of education completed (10 categories)</b>  <i>1 = No education</i>  <i>2 = Preschool</i>  <i>3 = Primary incomplete</i>  <i>4 = Primary complete but less</i></p>	<p>Value must be missing for individuals less than the required age (ed_mod_age).</p> <p>If a person is currently enrolled in the highest year of education, then his/her level of education completed should be determined by</p>

		<p><i>than completed lower secondary</i></p> <p><i>5 = Completed lower secondary (or post-primary vocational education) but less than completed upper secondary</i></p> <p><i>6 = Completed upper secondary (or extended vocational/technical education)</i></p> <p><i>7 = Post-secondary but not university</i></p> <p><i>8 = University and higher</i></p> <p><i>9 = Formal adult education or literacy program</i></p> <p><i>10 = Other</i></p>	<p>minus one year. For example, if a person is currently enrolled in P6, then his/her highest level completed should be coded as 1 (Pre-school/ Primary, not completed).</p> <p>Individuals enrolled in University level are coded as 8 (University and higher) regardless of whether completed or not.</p> <p>Other refers to level of education not defined by the above codes. This may refer to level of education not explicitly defined e.g. person attending a village polytechnic yet level reached not stated. This classification should be documented whenever possible.</p> <p>If Koranic school teaches formal curricula then it will be classified under formal education, then code appropriately.</p> <p>Koranic schools that teach Islamic knowledge with only (a) basic recitation or (b) recitation and Arabic writing or hafeez (memorization and Arabic fluency) are not mainstream formal schools. Code as "Other"</p> <p><i>If education level is missing for any member, do not try to impute but leave it as MISSING.</i></p> <p><b>If all categories are not present in the questionnaire, leave this variable as missing.</b></p>
5	<b>educat7</b>	<p><b>Highest level of education completed (7 categories)</b></p> <p><i>1 = No education</i></p> <p><i>2 = Primary incomplete</i></p> <p><i>3 = Primary complete</i></p> <p><i>4 = Secondary incomplete</i></p> <p><i>5 = Secondary complete</i></p> <p><i>6 = Post-secondary but not university</i></p> <p><i>7 = University (complete or incomplete)</i></p>	<p>Value must be missing for individuals less than the required age (ed_mod_age).</p> <p>Primary complete implies that one completed the stipulated primary education by undertaking an exam or test.</p> <p>Secondary complete implies that one completed the stipulated secondary education by undertaking an exam or test.</p> <p>Post-secondary technical education level refers to any higher education after successfully completing secondary level of education such as higher professional schooling, college, etc.</p> <p>University and higher education level refers undergraduate and higher.</p>

			<p><i>If education level is missing, do not try to impute but leave it as MISSING.</i></p> <p><b>If all categories are not present in the questionnaire, leave this variable as missing.</b></p>
6	<b>educat5</b>	<p><b>Highest level of education completed (5 categories)</b></p> <p>1 = No education  2 = Primary incomplete  3 = Primary complete but Secondary incomplete  4 = Secondary complete  5 = Tertiary/post-secondary (complete or incomplete)</p>	<p>Value must be missing for individuals less than the required age (ed_mod_age).</p> <p><i>If education level is missing, do not try to impute but leave it as MISSING.</i></p> <p><b>If all categories are not present in the questionnaire, leave this variable as missing.</b></p> <hr/> <p>Can be programmed from educat7.  recode educat7 (3 4=3) (5=4) (6 7=5),gen(educat5)  tab ageyrs educat5</p>
7	<b>educat4</b>	<p><b>Highest level of education completed (4 categories)</b></p> <p>1 = No education  2 = Primary (complete or incomplete)  3 = Secondary (complete or incomplete)  4 = Tertiary (complete or incomplete)</p>	<p>Value must be missing for individuals less than the required age (ed_mod_age).</p> <p><i>No education</i> includes people in pre-school and never attended. Pre-school definition is country-specific. This may include baby class, kindergarten and nursery school among others. This is the level before joining the regular stipulated primary level education cycle.</p> <p>At the minimum, educat4 must be available for all countries.</p> <p><i>If education level is missing, do not try to impute but leave it as MISSING.</i></p> <hr/> <p>Can be programmed from educat7.  recode educat7 (2 3=2) (4 5=3) (6 7=4),gen(educat4)  tab ageyrs educat4</p>
8	<b>educat_ISCED</b>	<p><b>ISCED education categories (highest level enrolled in or completed)</b></p> <p>1 = Early childhood education  2 = Primary education  3 = Lower secondary education  4 = Upper secondary education  5 = Post-secondary non-</p>	<p>These are the UNESCO ISCED 2011 education categories. Please note that we use the highest level enrolled in or completed. For example, if you are enrolled in primary education, you should get category 2 even if you have not completed primary yet or never will.</p> <p>Check this link for country ISCED Mappings <a href="#">9</a></p> <p><i>Post-secondary non-tertiary</i> education may be</p>

		<i>tertiary education</i> <i>6 = Short-cycle tertiary education</i> <i>7 = Bachelor's or equivalent level</i> <i>8 = Master's or equivalent level</i> <i>9 = Doctoral or equivalent level</i>	referred in many ways depending on country. However, these are typically vocational programmes that prepare one for the labor market such as technician diploma, electrician diploma.
9	<b>primarycomp</b>	<b>Primary school completion</b> <i>1 = Yes</i> <i>0 = No</i>	Value must be missing for other individual less than the required age (ed_mod_age).  One can assume with a degree of certainty these conditions qualify primary-school completion: <ul style="list-style-type: none"> <li>• EDUCAT10&gt;=4 &amp; EDUCAT10&lt;=8</li> <li>• EDUCAT7&gt;=3 &amp; EDUCAT7&lt;=7</li> <li>• EDUCAT5&gt;=3 &amp; EDUCAT5&lt;=5</li> </ul>
9	<b>educyrs</b>	<b>Years of completed education</b> <i>0 = Pre-school</i> <i>1 = Grade 1</i> <i>2 = Grade 2</i> . . .	Value must be missing for other individual less than the required age (ed_mod_age).  If grade level not listed, leave EDUCYRS=.  For individuals who are currently enrolled in school, their years of education completed correspond to the class <i>currently attending minus one</i> .  For individuals who are not currently enrolled in school, the years of completed education corresponds to the highest level of education completed.  This is a continuous variable of the number of years of formal schooling completed. It is constructed only if the survey asked for the number of year of education or highest grade level completed; otherwise, the values are constructed as missing.  The years of education that each grade corresponds to, varies by country, for example - some countries may have 5 or 6 years of primary school, 3 years of lower-secondary school, while other countries may have 4 years of primary school and 4 years of lower-secondary school. Refer to the UNESCO ISCED

			<p>mappings.<sup>1</sup></p> <p>For higher education, the grades/years may not have been asked explicitly. In such cases, the variable should be constructed based on the following assumptions: -</p> <ul style="list-style-type: none"> <li>• If the individual has completed the tertiary education specified, add to years of completed education - 4 years for BA/BSc, 6 years for MA/MSc, and 8 Years for PhD after the completion of secondary education.</li> <li>• If the individual has not completed tertiary education or completion cannot be ascertained, add to years of completed education – 2 years for BA/BSc, 5 years for MA/MSc, and 7 years for PhD.</li> </ul> <p>The variable does not take into account the actual number of years required to reach this grade level. In other words, first grade repeated three times only counts as 1 year of completed education.</p>
10	<b>atschool</b>	<p><b>Currently enrolled in or attending school</b>  <i>1 = Yes</i>  <i>0 = No</i></p>	<p>Value must be missing for individuals less than the required age (ed_mod_age).</p> <p>Use the question that asks for current attendance.</p> <p>If such a question is missing, use the question that explicitly asks for enrollment over the past 12 months. In such surveys, record this in the comments.</p> <p>Code as 0 if EVERATTD=0.</p>
11	<b>atschtyp</b>	<p><b>Type of school currently enrolled/attending</b>  <i>1 = Public</i>  <i>2 = Private</i>  <i>9 = Other</i></p>	<p>Value must be missing for individuals less than the required age (ed_mod_age).</p> <p>Code only for individuals currently attending school (ATSCHOOL=1).</p> <p><i>Public</i> includes fully government owned as well as semi-public owned.  <i>Private</i> are facilities run by non-governmental organizations (e.g. NGOs, religious institutions) or by private entities.</p>

<sup>1</sup> <http://www.uis.unesco.org/Education/ISCEDMappings/Pages/default.aspx>

			<i>Other</i> refers to schools that cannot be categorized by the above such as community schools which cannot be easily classified if run by either government or private.
12	<b>atslevattd</b>	<b>Level of schooling currently enrolled/attending</b> <i>1 = Preschool</i> <i>2 = Primary</i> <i>3 = Secondary</i> <i>4 = Post-secondary but not university</i> <i>5 = University and higher</i> <i>6 = Formal adult education or literacy program</i> <i>9 = Other</i>	Value must be missing for individuals less than the required age (ed_mod_age).  See EDUCAT10 for definition.  Check for consistency between EDUCAT10. That is EDUCAT10 cannot be university yet current level primary.

**Note:**

- ✓ For any variable not collected in a country, variable should be created and left as missing (.) in the final harmonized file.
- ✓ Variables in the data files must follow the sequence in which they appear in the manual.

**TABLE 3.5 MIGRATION**

Even if the survey does not ask any information on migration create variables as missing.

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	<b>rb_mod_age</b>	<b>Migration module application age (country-specific)</b>	Minimum age for which migration is applied.  For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.
2	<b>rbirth</b>	<b>Was member born in this country?</b> <i>1 = Yes</i> <i>0 = No</i>	Value must be missing for individuals less than the required age (rb_mod_age).
3	<b>rbirth_ctry</b>	<b>In what country was member born?</b>	<b>String variable</b>  Value must be missing for individuals less than the required age (rb_mod_age).  Only if RBIRTH=0.

			<p>If born outside country, enter 3-digit ISO country code (see Annex X).</p> <p>Several codes added for use if country no specified.</p> <p>“Other Africa”  “Other Europe”  “Other America”  “Other (unspecified)”</p>
4	<b>rbirthreg</b>	<p><b>Was person born in this region?</b>  1 = <i>Yes</i>  0 = <i>No</i></p>	<p>Value must be missing for individuals less than the required age (rb_mod_age).</p>
5	<b>rbirth_reg</b>	<b>Region of birth</b>	<p><b>String variable</b></p> <p>Value must be missing for individuals less than the required age (rb_mod_age).</p> <p>Only if RBIRTH_REG==0</p> <p>Use survey region codes. Must entered as “1 – region 1 name”, “2 – region 2 name”, etc.</p>
6	<b>rbirth_prevref</b>	<b>Reference time for previous residence</b>	<p><b>String variable</b></p> <p>Indicates the time reference of the question about migration (or place of residence).</p> <p>For example, RBIRTH_PREV_REF=5, means that the question asks about place of residence 5 years ago.</p>
7	<b>rbirthprev</b>	<p><b>Ever lived in a previous residence than the current one?</b>  1 = <i>Yes, within county</i>  2 = <i>Yes, outside country</i>  3 = <i>No</i></p>	<p>Value must be missing for individuals less than the required age (rb_mod_age).</p> <p>If person lived in several places, only the most recent should be recorded here.</p>
8	<b>rbirth_prev</b>	<b>Region of previous residence</b>	<p><b>String variable</b></p> <p>Value must be missing for individuals less than the required age (rb_mod_age).</p> <p>Only if RBIRTHPREV==1.</p> <p>Code using region codes of survey, must entered as “1 - region name”, etc.</p>

			If survey asks by area of residence, leave this variable as missing.
9	<b>ymove</b>	<b>Year individual moved to current location</b>	Value must be missing for individuals less than the required age (rb_mod_age). Indicates year of most recent move to RBIRTH_PREV.

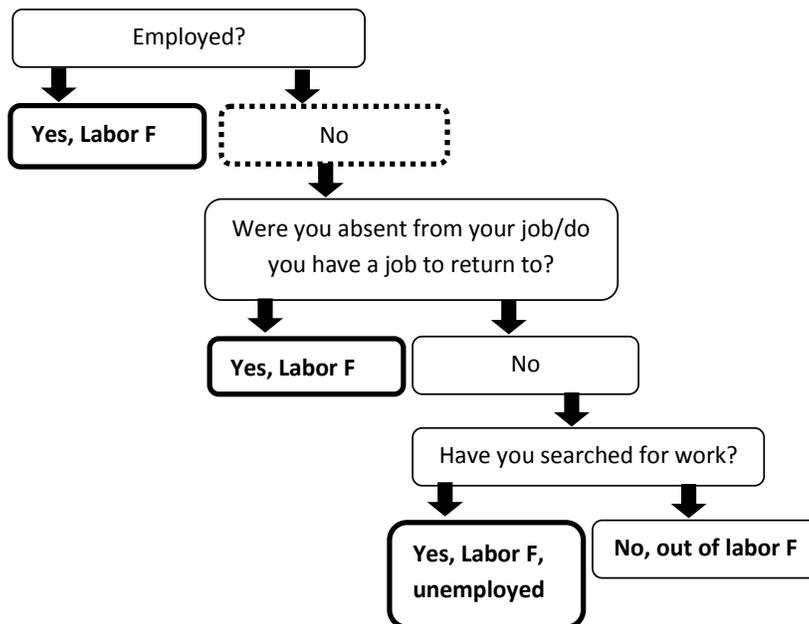
## CHAPTER 5: MODULE L - LABOR FORCE VARIABLES

The construction of employment and labor participation variables is specific to Sub-Saharan African context since over 80 percent of employment activities are in the informal sector. Studying labor participation in Tanzania, it was found that due to poor questionnaire design many unpaid family workers under reported their economic activities, especially women who reported domestic duties as the main activity. These individuals inevitably undertake some unpaid economic activities such as cultivating, and raising livestock, (preparing meals for the family and caring for own children are not classified as economic activities by ILO definition). The SSA harmonization developed complementary steps to capture these under-reported economic activities.

This manual reclassifies the employment status of these individuals, who claim household duties as their main activity, as employed rather than inactive. Because labor force questionnaires are significantly different from one another, it is not possible to provide a set of very specific steps that one can follow to classify employment status in Africa. The diagram below illustrates the logic used to classify unpaid economic activities. For details refer to Bardasi, Beegle, Dillon and Serneels "Do Labor Statistics Depend on How and to Whom the Questions Are Asked", World Bank Policy Research Working Paper 5192.

It is also strongly recommended that the user studies Appendix I carefully before starting to construct labor force variables, for detailed and in-depth explanations of the logic and how to construct various key SHIP labor force variables.

### Definition of Unemployment and Labor Force in Africa Region



**TABLE 4.0 SAMPLE AND BASIC HOUSEHOLD IDENTIFIER**

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
----	---------------	-----------------	----------------------------------

1	<b>country</b>	<b>Country code</b>	To be merged from the p-file. If you don't have the p-file, create the variable.
2	<b>year_IHSN</b>	<b>4-digit year of survey based on IHSN standards</b>	To be merged from the p-file. If you don't have the p-file, create the variable.
3	<b>hhno</b>	<b>Household number</b>	To be merged from the p-file. If you don't have the p-file, create the variable.
4	<b>hid</b>	<b>Household unique identification</b>	To be merged from the p-file. If you don't have the p-file, create the variable.
5	<b>wta_hh</b>	<b>Household weights</b>	<b>Numeric variable</b> To obtain household estimates, this is the weight to be used in all computations referring to household-level estimates. This variable cannot be used for poverty estimation. The interpretation is the proportion of households with a certain characteristic is XX%.

**TABLE 4.1 HOUSEHOLD CHORES**

Create tempfile with HID, PID, AGEYRS from the individual-level file that will be used to create the labor file variables. The individual-level file provides the correct household composition and size.

All variables are numeric unless specified.

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	<b>pid</b>	<b>Individual identification</b>	To be merged from individual-level file
2	<b>hh_mod_age</b>	<b>Household chores module application age</b>	Minimum age for which household chores section is applied.  For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.
3	<b>fetchwood</b>	<b>Fetches wood for the household</b> <i>0= No</i> <i>1= Yes</i>	Value must be missing for individuals less than the required age (hh_mod_age).  Code as 1 (YES) if the individual fetched wood for his/her own household <i>or for others</i> , otherwise code 0 (NO).  Based on UN definition of SSN. Fetching wood ( <b><i>for pay or in-kind</i></b> ) is an economic activity.

4	<b>fetchwater</b>	<b>Fetches water for the household</b> <i>0= No</i> <i>1= Yes</i>	Value must be missing for individuals less than the required age (hh_mod_age).  Code as 1 (YES) if the individual fetched water for his/her own household <i>or for others</i> , otherwise code 0 (NO).  Based on UN definition of SSN. Fetching water ( <b><i>for pay or in-kind</i></b> ) is an economic activity.
5	<b>cooking</b>	<b>Helped cook or prepare meals/drinks for the household</b> <i>0= No</i> <i>1= Yes</i>	Value must be missing for individuals less than the required age (hh_mod_age).  Code as 1 (YES) if the individual helped cook or prepare meals/drinks for his/her own household <i>or for others</i> , otherwise code 0 (NO).
6	<b>cleaning</b>	<b>Helped clean household or wash/iron clothes for the household</b> <i>0= No</i> <i>1= Yes</i>	Value must be missing for individuals less than the required age (hh_mod_age).  Code as 1 (YES) if the individual helped clean household or wash/iron clothes for his/her own household <i>or for others</i> , otherwise code 0 (NO).
7	<b>childcare</b>	<b>Helped take care of children for the household</b> <i>0= No</i> <i>1= Yes</i>	Value must be missing for individuals less than the required age (hh_mod_age).  Code as 1 (YES) if the individual helped take care of children for his/her own household <i>or for others</i> , otherwise code 0 (NO).
8	<b>oldcare</b>	<b>Helped take care of the elderly for the household</b> <i>0= No</i> <i>1= Yes</i>	Value must be missing for individuals less than the required age (hh_mod_age).  Code as 1 (YES) if the individual helped take care of the elderly for his/her own household <i>or for others</i> , otherwise code 0 (NO).
9	<b>hour_necon</b>	<b>Hours spent per week on non-economic activities.</b>	Value must be missing for individuals less than the required age (lb_mod_age).  These include activities such as preparing food, and care for children.  Fetching wood and water are considered economic activities and should not be included here.

**Note:**

- ✓ For any variable not collected in a country, variable should be created and left as missing (.) in the final harmonized file.
- ✓ Variables in the data files must follow the sequence in which they appear in the manual.

**TABLE 4.2 LABOR SCREENING QUESTIONS LAST 7-DAYS**

Create tempfile with HID, PID, AGEYRS from the individual-level file that will be used to create the labor file variables. The individual-level file provides the correct household composition and size.

All variables are numeric unless specified.

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	<b>lb_mod_age</b>	<b>Labor force module application age (country-specific)</b>	Age at which the labor module starts being applied (working age: people at which can start legally working) For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.
2	<b>lstatus</b>	<b>Labor status last 7 days</b> <i>1 = Employed</i> <i>2 = Unemployed</i> <i>3 = Not-in-labor force</i>	Value must be missing for individuals less than the required age (lb_mod_age).  All persons are considered active in the labor force if they presently have a job (formal or informal, i.e. are employed) or do not have a job but are actively seeking work (i.e. unemployed). Employed is defined as anyone who worked during the last 7 days or reference week, regardless of whether the employment was formal or informal, paid or unpaid, for a minimum of 1 hour. Individuals who had a job, but for any reason did not work in the last 7 days are considered employed. A person is defined as unemployed if he or she is, presently not working but is actively seeking a job. The formal definition of unemployed usually includes being 'able to accept a job'. This last question was asked in a minority of surveys and is, thus, not incorporated in the present definition. A person presently not working but waiting the start of a new job is considered to be unemployed.

3	<b>nlfreason</b>	<b>Reason not in labor force last 7 days</b> <i>1 = Student</i> <i>2 = Housewife</i> <i>3 = Retired</i> <i>4 = Disabled</i> <i>5 = Other</i>	<p>This variable is constructed for all those who are not presently employed and are not looking for work (Istatus=3) and missing otherwise.</p> <p>Student, the person is studying.</p> <p>Housekeeping is the person takes care of the house, older people, or children. Disabled is the person who cannot work due to physical conditions.</p> <p>Other the person does not work for any other reason</p> <p>Missing value for people below working age, employed, and unemployed. Other missing values allowed.</p>
4	<b>unempldur_l</b>	Unemployment duration (months) lower bracket	<p>Continuous variable</p> <p>Variable is constructed for all persons who are unemployed (Istatus=2, otherwise missing). If continuous records the numbers of months in unemployment. If the variable is categorical it records the lower boundary of the bracket.</p> <p>Missing values are allowed for everyone who is not unemployed. Other missing are also allowed.</p>
5	<b>unempldur_u</b>	Unemployment duration (months) upper bracket	<p>Continuous variable</p> <p>Variable is constructed for all persons who are unemployed (Istatus=2, otherwise missing). If continuous records the numbers of months in unemployment. If the variable is categorical it records the upper boundary of the bracket. If the right bracket is open a missing value should be inputted.</p> <p>Missing values are allowed for everyone who is not unemployed. Other missing are also allowed.</p>

**Note:**

- ✓ For any variable not collected in a country, variable should be created and left as missing (.) in the final harmonized file.
- ✓ Variables in the data files must follow the sequence in which they appear in the manual.

**TABLE 4.3 PRIMARY EMPLOYMENT LAST 7-DAYS**

Create tempfile with HID, PID, AGEYRS from the individual-level file that will be used to create the labor file variables. The individual-level file provides the correct household composition and size.

All variables are numeric unless specified.

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	<b>njobs</b>	<b>Number of total jobs</b>	Missing value for people below working age, unemployed and for people out of the labor force. Other missing values allowed.
2	<b>empstat</b>	<b>Employment status</b> 1 = Paid Employee 2 = Non-Paid Employee 3 = Employer 4 = Self-employed 5 = Other, workers not classifiable by status	Variable is constructed for all persons administered this module in each questionnaire. For this reason the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. Definitions taken from the International Labor Organization’s Classification of Status in Employment with some revisions to take into account the data available. Classifies the main job employment status of any individual with a job ( <i>lstatus=1</i> ) and is missing otherwise. <b>Paid employee</b> includes anyone whose basic remuneration is not directly dependent on the revenue of the unit they work for, typically remunerated by wages and salaries but may be paid for piece work or in-kind. The ‘continuous’ criteria used in the ILO definition is not used here as data are often absent and due to country specificity. <b>Non paid employee</b> includes contributing family workers are those workers who hold a self-employment job in a market-oriented establishment operated by a related person living in the same households who cannot be regarded as a partner because of their degree of commitment to the operation of the establishment, in terms of working time or other factors, is not at a level comparable to that of the head of the establishment. <b>Employer</b> is a business owner (whether alone or in partnership) with employees. If the only people working in the business are the owner and ‘contributing family workers, the person is

			<p>not considered an employer (as has no employees) and is, instead classified as own account.</p> <p><b>Own account or self-employment</b> includes jobs are those where remuneration is directly dependent from the goods and service produced (where home consumption is considered to be part of the profits) and have not engaged any permanent employees to work for them on a continuous basis during the reference period.</p> <p>Members of producers' cooperatives are workers who hold a self-employment job in a cooperative producing goods and services in which each member takes part on an equal footing with other members in determining the organization of production, sales and/or other work of the establishment, the investments and the distribution of the proceeds of the establishment amongst the members.</p> <p><b>Other, workers not classifiable by status</b> include those for whom insufficient relevant information is available and/or who cannot be included in any of the preceding categories.</p> <p>All apprentices should be mapped as unpaid workers</p>
3	ocusec	<p><b>Sector of activity</b></p> <p><i>1 = Public sector, Central Government, Army</i></p> <p><i>2 = Private, NGO</i></p> <p><i>3 = State owned</i></p> <p><i>4 = Public or State-owned, but cannot distinguish</i></p>	<p>Variable is constructed for all persons administered this module in each questionnaire.</p> <p>Classifies the main job's sector of activity of any individual with a job (<i>Istatus=1</i>) and is missing otherwise.</p> <p><b>Public sector</b> includes armed forces.</p> <p><b>Private sector</b> is that part of the economy which is both run for private profit and is not controlled by the state, it also includes non-governmental organizations</p> <p><b>State owned</b> includes para-state firms and all others in which the government has control (participation over 50%).</p> <p><i>Note:</i>  <i>If no such question, leave as missing.</i>  <i>Do not code based on occupation (ISCO) or industry (ISIC) codes.</i></p>
4	industry_orig	Original industry codes	This variable correspond to whatever is in the

			original file with no recoding. Missing value for people below the working age. Other missing values allowed. .a indicates non-response
5	<b>industry</b>	<b>1 digit industry classification</b> <i>1 = Agriculture, Hunting, Fishing, etc.</i> <i>2 = Mining</i> <i>3 = Manufacturing</i> <i>4 = Public Utility Services</i> <i>5 = Construction</i> <i>6 = Commerce</i> <i>7 = Transport and Communications</i> <i>8 = Financial and Business Services</i> <i>9 = Public Administration</i> <i>10 = Others Services, Unspecified</i>	Variable is constructed for all persons administered this module in each questionnaire. For this reason the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. Classifies the main job of any individual with a job (lstatus=1) and is missing otherwise. The codes for the main job are given here based on the UN International Standard Industrial Classification (revision 3.1). The main categories subsume the following codes: 1 Agriculture, Hunting, Fishing (01-05) 2 Mining (10-14) 3 Manufacturing (15-37) 4 Electricity and Utilities (40-41) 5 Construction (45) 6 Commerce (50-55) 7 Transportation, Storage and Communication (60-64) 8 Financial, Insurance and Real Estate (65-74) 9 Services: Public Administration (75) 10 Other Services ( 80 -99) In the case of different classifications (former Soviet Union republics, for example), recoding has been done to best match the ISIC-31 codes. 10 is also assigned for unspecified categories or items.
6	<b>industry1</b>	<b>1 digit industry classification (Broad Economic Activities)</b> <i>1 = Agriculture</i> <i>2= Industry</i> <i>3 = Services</i> <i>4 = Other</i>	This variable is either created directly from the data (if industry classification does not exist for 10 categories) or created from industrycat10. The following convention will be used to get from 10 to 4 categories (based on ISIC): gen industry1= replace industry1=1 if inlist(industry, 1) replace industry1=2 if inlist(industry, 2, 3, 4, 5) replace industry1=3 if inlist(industry, 6, 7, 8, 9) replace industry1=4 if inlist(industry, 10) assert industry1!=. if industry!=.
7	<b>occup_orig</b>	Original occupation code	This variable correspond to whatever is in the original file with no recoding

8	<b>occup</b>	<b>1 digit occupation classification</b> <i>1 = Managers</i> <i>2 = Professionals</i> <i>3 = Technicians and associate professionals</i> <i>4 = Clerical support workers</i> <i>5 = Service and sales workers</i> <i>6 = Skilled agricultural, forestry and fishery workers</i> <i>7 = Craft and related trades workers</i> <i>8 = Plant and machine operators, and assemblers</i> <i>9 = Elementary occupations</i> <i>10 = Armed forces occupations</i> <i>99 = Other/unspecified</i>	<p>Variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.</p> <p>Classifies the main job of any individual with a job (Istatus=1) and is missing otherwise. As most surveys collected detailed information and then coded it, and the original data is not in the data bases, no attempt has been made to correct or check the original coding. The classification is based on the International Standard Classification of Occupations (ISCO) 88.</p> <p>In the case of different classifications re-coding has been done to best match the ISCO-88.</p>
9	<b>wage_no_compen</b>	<b>Last wage payment</b>	<p>Continuous variable: wage in local currency. Variable is constructed for all persons administered this module in each questionnaire. For this reason the lower age cutoff (and perhaps upper age cutoff) will vary from country to country.</p> <p>States the main job's wage earner of any individual (Istatus=1 &amp; empstat=1) and is missing otherwise.</p> <p>Wage from main job (job to which the person dedicated most time in the week preceding the survey). This excludes tips, bonuses, and other payments.</p> <p>For all those with self-employment or owners of own businesses, this should be net revenues (net of all costs EXCEPT for tax payments) or the amount of salary taken from the business. Due to the almost complete lack of information on taxes, the wage from main job is NOT net of taxes.</p> <p>By definition non-paid employees (empstat=2) should have wage=0.</p> <p>Excludes tips, compensations such bonuses, dwellings or clothes, and other payments.</p>

10	<b>bonuses</b>	Tips, compensations such bonuses, dwellings or clothes, and other payments.	Includes tips, compensations such bonuses, dwellings or clothes, and other payments. Please annualize this value considering the number of months working in the firm and the periodicity of the bonuses
11	<b>unitwage</b>	<b>Last wages time unit</b> 1 = Daily 2 = Weekly 3 = Every two weeks 4 = Every two months 5 = Monthly 6 = Quarterly 7 = Every six months 8 = Annually 9 = Hourly 10 = Other	Type of reference for the wage_no_compen variable. States the main job's wage earner time unit measurement of any individual (lstatus=1 & empstat=1) and is missing otherwise.
12	<b>whours</b>	<b>Hours of work in last week</b>	Variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. Classifies the main job of any individual with a job (lstatus=1) and is missing otherwise. This is the number of hours worked in the last 7 days or the reference week in the person's main job. Main job defined as that occupation to which the person dedicated more time. For persons absent from their job in the week preceding the survey due to holidays, vacation or sick leave, the time worked in the last week the person worked is recorded. (Note sometimes the questions are phrased as, on average how many hrs a week do you work). For individuals who only give information on how many hours they work per day and no information on number of days worked a week, multiply the hours by 5 days. In the case of a question that has hours worked per month, divide by 4.2 to get weekly hours.
13	<b>contract</b>	<b>Employment contract</b> 0 = No 1 = Yes	Variable is constructed for all persons administered this module in each questionnaire. Indicates if a person has a signed (formal) contract, regardless of duration. For this reason the lower age cutoff (and perhaps upper age cutoff) at which

			<p>information is collected will vary from country to country.</p> <p>Classifies the contract status of any individual with a job (<i>lstatus</i>=1) and is missing otherwise. This variable is only constructed if there is an explicit question about contracts.</p>
14	<b>healthins</b>	<p><b>Health insurance</b></p> <p><i>0 = No</i> <i>1 = Yes</i></p>	<p>Variable is constructed for all persons administered this module in each questionnaire. For this reason the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.</p> <p>Classifies the health insurance status of any individual with a job (<i>lstatus</i>=1) and is missing otherwise.</p> <p>This variable is only constructed if there is an explicit question about health security.</p>
15	<b>socialsec</b>	<p><b>Social security</b></p> <p><i>0 = No</i> <i>1 = Yes</i></p>	<p>Variable is constructed for all persons administered this module in each questionnaire. For this reason the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.</p> <p>Classifies the social security status of any individual with a job (<i>lstatus</i>=1) and is missing otherwise.</p> <p>This variable is only constructed if there is an explicit question about pension plans or social security.</p>
16	<b>union</b>	<p><b>Union membership</b></p> <p><i>0 = No</i> <i>1 = Yes</i></p>	<p>Variable is constructed for all persons administered this module in each questionnaire. For this reason the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.</p> <p>Classifies the union membership status of any individual with a job (<i>lstatus</i>=1) and is missing otherwise.</p> <p>This variable is only constructed if there is an explicit question about trade unions.</p>
17	<b>firmsize_l</b>	<p><b>Firm size (lower bracket)</b></p>	<p>Variable is constructed for all persons who are employed. If continuous records the number of people working for the same employer. If the variable is categorical it records the lower boundary of the bracket.</p>

18	<b>firmsize_u</b>	<b>Firm size (upper bracket)</b>	Variable is constructed for all persons who are employed. If continuous records the number of people working for the same employer. If the variable is categorical it records the upper boundary of the bracket. If the right bracket is open, a missing value should be inputted.
----	-------------------	----------------------------------	--

**Note:**

- ✓ For any variable not collected in a country, variable should be created and left as missing (.) in the final harmonized file.
- ✓ Variables in the data files must follow the sequence in which they appear in the manual.

**TABLE 4.4 SECONDARY EMPLOYMENT LAST 7-DAYS**

Create tempfile with HID, PID, AGEYRS from the individual-level file that will be used to create the labor file variables. The individual-level file provides the correct household composition and size.

All variables are numeric unless specified.

Check consistency with number of jobs (njobs) from previous Section 4.3. If njobs=1 than this section must be missing.

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	<b>empstat_2</b>	<b>Employment status - second job</b> <i>1 = Paid Employee</i> <i>2 = Non-Paid Employee</i> <i>3 = Employer</i> <i>4 = Self-employed</i> <i>5 = Other, workers not classifiable by status</i>	Variable is constructed for all persons administered this module in each questionnaire. For this reason the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. Definitions taken from the International Labor Organization's Classification of Status in Employment with some revisions to take into account the data available. Classifies the second job employment status of any individual with a job (lstatus=1) and is missing otherwise. Paid employee includes anyone whose basic remuneration is not directly dependent on the revenue of the unit they work for, typically remunerated by wages and salaries but may be paid for piece work or in-kind. The 'continuous' criteria used in the ILO definition is not used here as data are often absent and

			<p>due to country specificity.</p> <p>Non paid employee includes contributing family workers are those workers who hold a self-employment job in a market-oriented establishment operated by a related person living in the same households who cannot be regarded as a partner because of their degree of commitment to the operation of the establishment, in terms of working time or other factors, is not at a level comparable to that of the head of the establishment.</p> <p>Employer is a business owner (whether alone or in partnership) with employees. If the only people working in the business are the owner and 'contributing family workers, the person is not considered an employer (as has no employees) and is, instead classified as own account.</p> <p>Own account or self-employment includes jobs are those where remuneration is directly dependent from the goods and service produced (where home consumption is considered to be part of the profits) and have not engaged any permanent employees to work for them on a continuous basis during the reference period.</p> <p>Members of producers' cooperatives are workers who hold a self-employment job in a cooperative producing goods and services in which each member takes part on an equal footing with other members in determining the organization of production, sales and/or other work of the establishment, the investments and the distribution of the proceeds of the establishment amongst the members.</p> <p>Other, workers not classifiable by status include those for whom insufficient relevant information is available and/or who cannot be included in any of the preceding categories.</p>
2	<b>industry_orig_2</b>	<b>Original industry codes second job</b>	This variable correspond to whatever is in the original file with no recoding.
3	<b>industry_2</b>	<b>1 digit industry classification - second job</b> <i>1 = Agriculture, Hunting, Fishing, etc.</i> <i>2 = Mining</i>	Variable is constructed for all persons administered this module in each questionnaire. For this reason the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

		<p>3 = Manufacturing  4 = Public Utility Services  5 = Construction  6 = Commerce  7 = Transport and Communications  8 = Financial and Business Services  9 = Public Administration  10 = Other Services, Unspecified</p>	<p>Classifies the second job of any individual with a job (Istatus=1) and is missing otherwise. The codes for the second job are given here based on the UN International Standard Industrial Classification (revision 3.1). The main categories subsume the following codes:  1 Agriculture, Hunting, Fishing (01-05)  2 Mining (10-14)  3 Manufacturing (15-37)  4 Electricity and Utilities (40-41)  5 Construction (45)  6 Commerce (50-55)  7 Transportation, Storage and Communication (60-64)  8 Financial, Insurance and Real Estate (65-74)  9 Services: Public Administration (75)  10 Other Services ( 80 -99)</p> <p>In the case of different classifications (former Soviet Union republics, for example), recoding has been done to best match the ISIC-31 codes.  10 is also assigned for unspecified categories or items.</p>
4	<b>industry1_2</b>	<p><b>1 digit industry classification (Broad Economic Activities) - second job</b>  1 = Agriculture  2 = Industry  3 = Services  4 = Other</p>	<p>This variable is either created directly from the data (if industry classification does not exist for 10 categories) or created from industrycat10. The following convention will be used to get from 10 to 4 categories (based on ISIC):  gen industry1_2=  replace industry1_2=1 if inlist(industry_2, 1)  replace industry1_2=2 if inlist(industry_2, 2, 3, 4, 5)  replace industry1_2=3 if inlist(industry_2, 6, 7, 8, 9)  replace industry1_2=4 if inlist(industry_2, 10)  assert industry1_2!=. if industry_2!=.</p>
5	<b>occup_2</b>	<p><b>1 digit occupational classification - second job</b>  1 = Managers  2 = Professionals  3 = Technicians and associate professionals  4 = Clerical support workers  5 = Service and sales workers</p>	<p>Variable is constructed for all persons administered this module in each questionnaire. For this reason the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.  Classifies the main job of any individual with a job (Istatus=1) and is missing otherwise. As most surveys collected detailed information and then coded it, and the original data is not in the data bases, no attempt has been made</p>

		<p>6 = Skilled agricultural, forestry and fishery workers</p> <p>7 = Craft and related trades workers</p> <p>8 = Plant and machine operators, and assemblers</p> <p>9 = Elementary occupations</p> <p>10 = Armed forces occupations</p> <p>99 = Other/unspecified</p>	<p>to correct or check the original coding.</p> <p>The classification is based on the International Standard Classification of Occupations (ISCO) 88.</p> <p>In the case of different classifications re-coding has been done to best match the ISCO-88.</p>
6	wage_no_compen_2	<b>Last wage payment</b>	<p>Continuous variable: wage in local currency. Variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country.</p> <p>States the main job's wage earner of any individual (lstatus=1 &amp; empstat=1) and is missing otherwise.</p> <p>Wage from main job (job to which the person dedicated most time in the week preceding the survey). This excludes tips, bonuses, and other payments.</p> <p>For all those with self-employment or owners of own businesses, this should be net revenues (net of all costs EXCEPT for tax payments) or the amount of salary taken from the business. Due to the almost complete lack of information on taxes, the wage from main job is NOT net of taxes.</p> <p>By definition non-paid employees (empstat=2) should have wage=0.</p> <p>Excludes tips, compensations such bonuses, dwellings or clothes, and other payments.</p>
7	bonuses_2		<p>Includes tips, compensations such bonuses, dwellings or clothes, and other payments.</p> <p>Please annualize this value considering the number of months working in the firm and the periodicity of the bonuses</p>
8	unitwage_2	<p><b>Last wages time unit second job</b></p> <p>1 = Daily</p> <p>2 = Weekly</p>	<p>Type of reference for the wage variable.</p> <p>States the second job's wage earner time unit measurement of any individual (lstatus=1 &amp; empstat=1) and is missing otherwise.</p>

		<p>3 = <i>Every two weeks</i>  4 = <i>Every two months</i>  5 = <i>Monthly</i>  6 = <i>Quarterly</i>  7 = <i>Every six months</i>  8 = <i>Annually</i>  9 = <i>Hourly</i>  10 = <i>Other</i></p>	
9	<b>whours_2</b>	<b>Hours worked last week in secondary job</b>	<p>Continuous variable: hours worked in last week</p> <p>Variable is constructed for all persons administered this module in each questionnaire. For this reason the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.</p> <p>Classifies the main job of any individual with a job (Istatus=1) and is missing otherwise.</p> <p>This is the number of hours worked in the last 7 days or the reference week in the person's main job. Main job defined as that occupation to which the person dedicated more time.</p> <p>For persons absent from their job in the week preceding the survey due to holidays, vacation or sick leave, the time worked in the last week the person worked is recorded. (Note sometimes the questions are phrased as, on average how many hrs a week do you work).</p> <p>For individuals who only give information on how many hours they work per day and no information on number of days worked a week, multiply the hours by 5 days.</p> <p>In the case of a question that has hours worked per month, divide by 4.2 to get weekly hours.</p>
10	<b>firmsize_l_2</b>	<b>Firm size (lower bracket)</b>	<p>Variable is constructed for all persons who are employed. If continuous records the number of people working for the same employer. If the variable is categorical it records the lower boundary of the bracket.</p>
11	<b>firmsize_u_2</b>	<b>Firm size (upper bracket)</b>	<p>Variable is constructed for all persons who are employed. If continuous records the number of people working for the same employer. If the variable is categorical it records the upper boundary of the bracket. If the right bracket is open, a missing value should be inputted</p>

--	--	--	--

**Note:**

- ✓ For any variable not collected in a country, variable should be created and left as missing (.) in the final harmonized file.
- ✓ Variables in the data files must follow the sequence in which they appear in the manual.

**TABLE 4.5 EMPLOYMENT LAST 12-MONTHS**

Create tempfile with HID, PID, AGEYRS from the individual-level file that will be used to create the labor file variables. The individual-level file provides the correct household composition and size.

All variables are numeric unless specified.

No	VARIABLE NAME	LABEL AND CODES	FORMAT, DESCRIPTION AND COMMENTS
1	<b>lstatus_year</b>	<b>Labor force status last 12-months</b> <i>1 = Employed</i> <i>0 = Not Employed</i>	Variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. All persons are considered active in the labor force if they presently have a job (formal or informal, i.e. are employed) or do not have a job but are actively seeking work (i.e. unemployed). Variable is constructed for all persons administered this module in each questionnaire. For this reason the age cutoffs at which information is collected will vary from country to country.
2	<b>empstat_year</b>	<b>Employment status</b> <i>1 = Paid Employee</i> <i>2 = Non-Paid Employee</i> <i>3 = Employer</i> <i>4 = Self-employed</i> <i>5 = Other, workers not classifiable by status</i>	Variable is constructed for all persons administered this module in each questionnaire. For this reason the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. Definitions taken from the International Labor Organization's Classification of Status in Employment with some revisions to take into account the data available. Classifies the main job employment status of any individual with a job ( <i>lstatus=1</i> ) and is

			<p>missing otherwise.</p> <p><b>Paid employee</b> includes anyone whose basic remuneration is not directly dependent on the revenue of the unit they work for, typically remunerated by wages and salaries but may be paid for piece work or in-kind. The 'continuous' criteria used in the ILO definition is not used here as data are often absent and due to country specificity.</p> <p><b>Non paid employee</b> includes contributing family workers are those workers who hold a self-employment job in a market-oriented establishment operated by a related person living in the same households who cannot be regarded as a partner because of their degree of commitment to the operation of the establishment, in terms of working time or other factors, is not at a level comparable to that of the head of the establishment.</p> <p><b>Employer</b> is a business owner (whether alone or in partnership) with employees. If the only people working in the business are the owner and 'contributing family workers, the person is not considered an employer (as has no employees) and is, instead classified as own account.</p> <p><b>Own account or self-employment</b> includes jobs are those where remuneration is directly dependent from the goods and service produced (where home consumption is considered to be part of the profits) and have not engaged any permanent employees to work for them on a continuous basis during the reference period.</p> <p>Members of producers' cooperatives are workers who hold a self-employment job in a cooperative producing goods and services in which each member takes part on an equal footing with other members in determining the organization of production, sales and/or other work of the establishment, the investments and the distribution of the proceeds of the establishment amongst the members.</p> <p><b>Other, workers not classifiable by status</b> include those for whom insufficient relevant information is available and/or who cannot be included in any of the preceding categories.</p>
--	--	--	---

3	<b>njobs_year</b>	Number of total jobs in last year	Continuous variable
4	<b>firmsize_l_year</b>	Firm size (lower bracket)	Variable is constructed for all persons who are employed. If continuous records the number of people working for the same employer. If the variable is categorical it records the lower boundary of the bracket.
5	<b>firmsize_u_year</b>	Firm size (upper bracket)	Variable is constructed for all persons who are employed. If continuous records the number of people working for the same employer. If the variable is categorical it records the upper boundary of the bracket. If the right bracket is open, a missing value should be inputted
6	<b>whours_year</b>	Hours of work in typical week in the 12-month job	

## ANNEX I: THE METHODOLOGY OF CONSTRUCTING EMPLOYMENT VARIABLES

The construction of employment and labor participation variables is specific to the Sub-Saharan African context since over 80 percent of employment activities are in the informal sector. Studying labor participation in Tanzania, Bardasi, Beegle, Dillon and Serneel found that due to poor questionnaire design, many unpaid family workers under reported their economic activities, especially women who reported domestic duties as the main activity. These individuals inevitably undertake some unpaid economic activities such as cultivating, raising livestock, fetching water, and collecting wood (preparing meals for the family and caring for own children are not classified as economic activities by ILO definition.) The Sub-Saharan Team for Statistical Development developed complementary steps to capture these under-reported economic activities.

Additionally, due to informal economic activities and under-reported employment, there are often many missing values for the industry of employment based on ISIC code. To remedy this situation, we create a variable to classify industry into farm and non-farm sectors that can be gleaned from other modules of the survey, such as farm, household enterprises and time use modules.

The steps outlined below are designed to capture, to the greatest extent possible, the actual employment status, including women who work from home and take care of household responsibilities at the same time, and students who help with fetching wood and water.

### Construction of labor variables

In the SHIP manual, we construct labor force participation first (Table 2). Please note that we code SHIP employment variables based on 7-day information and by 12-month information. We follow ILO definition broadly with supplementary steps to capture under-reported employment. It is important to keep in mind that supplementary steps only replace missing values generated from previous steps.

The final employment variable of interest is EMP\_CAT\_1 (Table 5), which provides statistics to monitor structural changes in employment, classified into five categories as follows:

<b>EMP_CAT_1 (Table 5)</b>	}	11 Wage public
		12 Wage private non-agriculture
		13 Wage private agriculture
		21 Self-employed non-agriculture (household enterprises)
		22 Self-employed agriculture (farmers).

It should be noted that there is a small overlap between employment by SHIP definition and unemployment by ILO definition. By ILO definition, anyone who is without a job and looking for a job in a reference period (normally from 7 to 2 weeks) should be classified as unemployed. Because of the large proportion of informal employment and the almost absence of unemployment benefits, the status between employed and unemployed is often blurred. However, this small overlap has little significance in analyzing labor force participation.

Below are summary tables on construction of most important SHIP employment variables. The sequence corresponds to the precedence of information, i.e. the later steps only replace missing values generated by previous steps.

Table 1: Construction of **LABFORCE\_WB**

Variable Names		Description	
<b>LABFORCE_WB</b>	<b>YES</b> , in the labor force, if <b>ANY</b> of the conditions is YES	WORKED_7=YES	Did the Individual have any kind of employment for any duration of time within the last 7 day recall period?
		WORKED_12=YES	Did the individual have any kind of employment for any duration of time within the last 12 month recall period?
		ABSENT=YES	If the individual did not work in the past 7 days did he/she have a job to return to?
		LOOKJOB=YES	Did the individual look for a job in the reference period (7days, 2 weeks or 4 weeks)?
		EMPTYTYPE_WB_1≠99	Type of employment based on information in the employment section and other sections of the survey. (Table 3 and Flowchart 1).
	<b>NO</b>	WORKED_7=NO WORKED_12=NO ABSENT=NO LOOKJOB=NO EMPTYTYPE_WB_1=99	If <b>ALL</b> the conditions are satisfied.
		<b>MISSING</b>	WORKED_7=. WORKED_12=. ABSENT=. LOOKJOB=. EMPTYTYPE_WB_1=99

Table 2: Construction of **EMPLOYED\_WB**

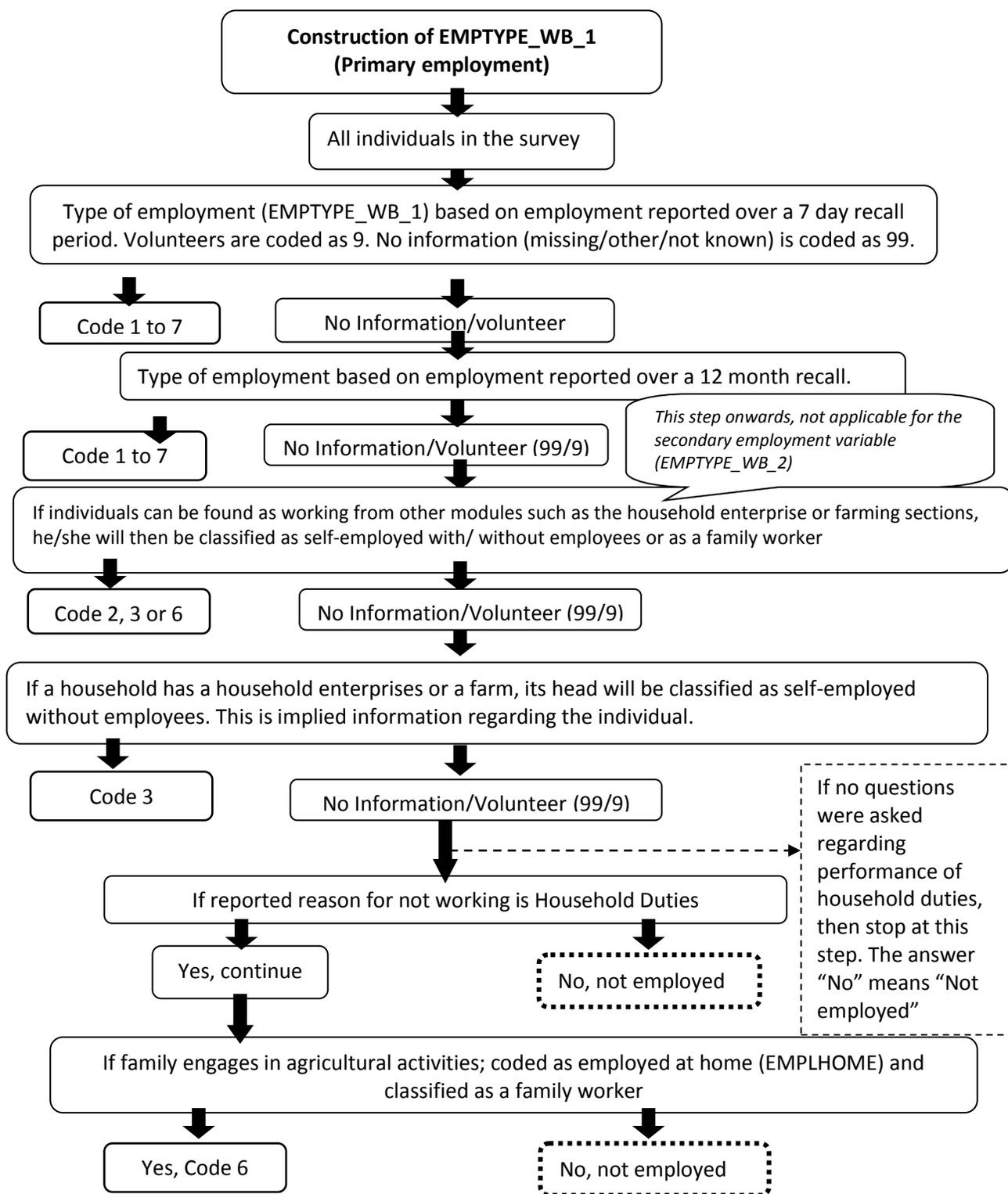
Variable Names		Description	
<b>EMPLOYED_WB</b>	<b>YES, if ANY of the conditions are satisfied</b>	WORKED_7=YES	Did the Individual have any kind of employment for any duration of time within the last 7 day recall period? It is based on an employment screening question, otherwise missing.
		WORKED_12=YES	Did the individual have any kind of employment for any duration of time within the last 12 month recall period? It is based on an employment screening question, otherwise missing.
		ABSENT=YES	If the individual did not work in the past 7 days did he/she have a job to return to?
		EMPTYTYPE_WB_1≠99	Type of employment based on information in the employment section and other sections of the survey. (Table 3 and Flowchart 1).
	<b>NO</b>	LABFORCE_WB=YES EMPTYTYPE_WB_1=99	If the individual is part of the labor-force but does not have a job as identified by the above four variables.
<b>MISSING</b>	LABFORCE_WB=.	If <b>no</b> information on labor force participation.	

#### Description of **EMPTYTYPE\_WB\_1**

EMPTYTYPE\_WB\_1 is the most important variable in the SHIP Labor-force module. Information is sought from all sections of the survey including the household enterprise and the farming sections. By design it has no missing values. Individuals for whom no information can be found (include “not known,” “other”, and no response/not applicable) are coded as 99. EMPTYTYPE\_WB\_1 classifies the type of employment into 8 different types.

Table 3: Description of codes in **EMPTYTYPE\_WB\_1**

Code	Label	Definition
1	Wage and salaried worker	An individual employed by non-household members who are paid in cash or in kind on a regular basis or based on a task. Agricultural and non-agricultural laborers are included in this category.
2	Self-employed with employees	An individual who runs a farm or a non-agricultural enterprise and employs at least one non-household member. Some surveys have information on employment of non-household members and is used to define this category.
3	Self-employed without employees	An individual who runs a farm or a non-agricultural enterprise and DOES NOT employ any non-household member. If the survey does not have information on employment of non-household members then individuals are by default classified in this category. Some surveys have a category called 'own account worker'; such individuals are categorized in this section.
4	Employer	Refers to the owner of a business with employees, irrespective of agricultural or non-agricultural sector. Individuals are classified only if the survey explicitly has 'employer' as a category.
5	Domestic employees	An individual who works for a domestic household. Some surveys that have a question on job description may have this information.
6	Family worker	An individual who is a paid or unpaid worker who assists in the work on a farm or a non-farm enterprise.
7	Apprentice	Individuals who are apprentices, irrespective of whether paid or unpaid
9	Volunteer	An individual who is a volunteer, stated as an explicit category in the questionnaire.
99	Other/Not known/missing	All individuals who respond as other or not known or have a non-response value. Additional information from other sections of the survey is gleaned to reclassify these individuals into one of the above mentioned categories where-ever possible.



**Table 4: Description of EMPFRM and EMPSEC**

<b>EMPFRM</b>	Coded as 'yes' if the individual is employed in agricultural activities. Activities in the agriculture sector include regular farming, sharecropping, raising livestock, bee-keeping, fishing, logging and hunting. In addition, if the household engages in agricultural activities,
---------------	---

	and the head is reported as not working, he/she then is classified to be in the agriculture sector.
EMPSEC	Captures the type of the employment establishment – 1) public, 2) private and 3) State-owned enterprise. It is based only on an explicit question in the survey to this effect.

**EMP\_CAT\_1** is derived from the following underlining variables:

1. EMPTYTYPE\_WB\_1 (Table 3 and Flowchart 1)
2. EMPFARM\_1 (Table 4)
3. EMPSEC\_1 (Table 4)

Table 5: Construction of **EMP\_CAT\_1**

Variable Names		Description	
EMP_CAT_1	EMP_CAT_1=11 Wage public	EMPTYTYPE_WB_1 = 1, 4, 5, 7, 9	Type of employment is wage or employer or domestic employee or apprentice or volunteer.
		EMPSEC_1= 1, 2	Individual worked in either the public sector or in a state owned company.
	EMP_CAT_1=12 Wage private non-agriculture	EMPTYTYPE_WB_1 = 1, 4, 5, 7, 9	Type of employment is wage or employer or domestic employee or apprentice or volunteer.
		EMPSEC_1= 3	Individual worked in the private sector.
		EMPFARM_1≠YES	Individual was <b>NOT</b> employed in agricultural activities.
	EMP_CAT_1=13 Wage private agriculture	EMPTYTYPE_WB_1 = 1, 4, 5, 7, 9	Type of employment is wage or employer or domestic employee or apprentice or volunteer.
		EMPSEC_1= 3	Individual worked in the private sector.
		EMPFARM_1=YES	Individual was employed in agricultural activities.
	EMP_CAT_1=13 Family enterprise non -agriculture	EMPTYTYPE_WB_1 = 2, 3, 6	Type of employment is self-employed with or without employees or family worker.
		EMPFARM_1≠YES	Individual was <b>NOT</b> employed in agricultural activities.
	EMP_CAT_1=13 Family farmer	EMPTYTYPE_WB_1 = 2, 3, 6	Type of employment is self-employed with or without employees or family worker.
		EMPFARM_1 =YES	Individual was employed in agricultural activities.

## ANNEX II: INTERNATIONAL STANDARD INDUSTRIAL CLASSIFICATION OF ALL ECONOMIC ACTIVITIES (ISIC)

### ISIC REV. 4.0 CATEGORIES

#### **A - Agriculture, hunting and forestry**

- 01 - Agriculture, hunting and related service activities
- 02 - Forestry, logging and related service activities
- 03 - Fishing and aquaculture

#### **B - Mining and quarrying**

- 05 - Mining of coal and lignite
- 06 - Extraction of crude petroleum and natural gas
- 07 - Mining of metal ores
- 08 - Other mining and quarrying
- 09 - Mining support service activities

#### **C - Manufacturing**

- 10 - Manufacture of food products
- 11 - Manufacture of beverages
- 12 - Manufacture of tobacco products
- 13 - Manufacture of textiles
- 14 - Manufacture of wearing apparel
- 15 - Manufacture of leather and related products
- 16 - Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
- 17 - Manufacture of paper and paper products
- 18 - Printing and reproduction of recorded media
- 19 - Manufacture of coke and refined petroleum products
- 20 - Manufacture of chemicals and chemical products
- 21 - Manufacture of basic pharmaceutical products and pharmaceutical preparations
- 22 - Manufacture of rubber and plastics products
- 23 - Manufacture of other non-metallic mineral products
- 24 - Manufacture of basic metals
- 25 - Manufacture of fabricated metal products, except machinery and equipment
- 26 - Manufacture of computer, electronic and optical products
- 27 - Manufacture of electrical equipment
- 28 - Manufacture of machinery and equipment n.e.c.
- 29 - Manufacture of motor vehicles, trailers and semi-trailers
- 30 - Manufacture of other transport equipment
- 31 - Manufacture of furniture
- 32 - Other manufacturing
- 33 - Repair and installation of machinery and equipment

#### **D - Electricity, gas, steam and air conditioning supply**

- 35 - Electricity, gas, steam and air conditioning supply

#### **E - Water supply; sewerage, waste management and remediation activities**

- 36 - Water collection, treatment and supply
- 37 - Sewerage
- 38 - Waste collection, treatment and disposal activities; materials recovery
- 39 - Remediation activities and other waste management services

**F - Construction**

- 41 - Construction of buildings
- 42 - Civil engineering
- 43 - Specialized construction activities

**G - Wholesale and retail trade; repair of motor vehicles, motorcycles**

- 45 - Wholesale and retail trade and repair of motor vehicles and motorcycles
- 46 - Wholesale trade, except of motor vehicles and motorcycles
- 47 - Retail trade, except of motor vehicles and motorcycles

**H - Transportation and storage**

- 49 - Land transport and transport via pipelines
- 50 - Water transport
- 51 - Air transport
- 52 - Warehousing and support activities for transportation
- 53 - Postal and courier activities

**I - Accommodation and food service activities**

- 55 - Accommodation
- 56 - Food and beverage service activities

**J - Information and communication**

- 58 - Publishing activities
- 59 - Motion picture, video and television programme production, sound recording and music publishing activities
- 60 - Programming and broadcasting activities
- 61 - Telecommunications
- 62 - Computer programming, consultancy and related activities
- 63 - Information service activities

**K - Financial and insurance activities**

- 64 - Financial service activities, except insurance and pension funding
- 65 - Insurance, reinsurance and pension funding, except compulsory social security
- 66 - Activities auxiliary to financial service and insurance activities

**L - Real estate activities**

- 68 - Real estate activities

**M - Professional, scientific and technical activities**

- 69 - Legal and accounting activities
- 70 - Activities of head offices; management consultancy activities
- 71 - Architectural and engineering activities; technical testing and analysis
- 72 - Scientific research and development
- 73 - Advertising and market research
- 74 - Other professional, scientific and technical activities
- 75 - Veterinary activities

**N - Administrative and support service activities**

- 77 - Rental and leasing activities
- 78 - Employment activities
- 79 - Travel agency, tour operator, reservation service and related activities
- 80 - Security and investigation activities
- 81 - Services to buildings and landscape activities
- 82 - Office administrative, office support and other business support activities

**O - Public administration and defence; compulsory social security**

- 84 - Public administration and defence; compulsory social security

**P - Education**

- 85 - Education

**Q - Human health and social work activities**

- 86 - Human health activities
- 87 - Residential care activities
- 88 - Social work activities without accommodation

**R - Arts, entertainment and recreation**

- 90 - Creative, arts and entertainment activities
- 91 - Libraries, archives, museums and other cultural activities
- 92 - Gambling and betting activities
- 93 - Sports activities and amusement and recreation activities

**S - Other service activities**

- 94 - Activities of membership organizations
- 95 - Repair of computers and personal and household goods
- 96 - Other personal service activities

**T - Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use**

- 97 - Activities of households as employers of domestic personnel
- 98 - Undifferentiated goods- and services-producing activities of private households for own use

**U - Activities of extraterritorial organizations and bodies**

- 99 - Activities of extraterritorial organizations and bodies

When the ISIC 4.0 categories are used in a survey, we use the following mapping:

A	=	Agriculture and fishing
B	=	Mining and quarrying
C	=	Manufacturing
D + E	=	Electricity, gas and water supply
F	=	Construction
G + I	=	Commerce
H + J N (code=79)	=	Transport, storage and communication
K + L	=	Financial, insurance and real estate
O	=	Public administration
M + N (excl. 79) P + Q + R + S + T + U	=	Other services

## **ISIC REV. 3.1 CATEGORIES**

### **A - Agriculture, hunting and forestry**

- 01 - Agriculture, hunting and related service activities
- 02 - Forestry, logging and related service activities

### **B - Fishing**

- 05 - Fishing, operation of fish hatcheries and fish farms; service activities incidental to fishing

### **C - Mining and quarrying**

- 10 - Mining of coal and lignite; extraction of peat
- 11 - Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction, excluding surveying
- 12 - Mining of uranium and thorium ores
- 13 - Mining of metal ores
- 14 - Other mining and quarrying

### **D - Manufacturing**

- 15 - Manufacture of food products and beverages
- 16 - Manufacture of tobacco products
- 17 - Manufacture of textiles
- 18 - Manufacture of wearing apparel; dressing and dyeing of fur
- 19 - Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
- 20 - Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
- 21 - Manufacture of paper and paper products
- 22 - Publishing, printing and reproduction of recorded media
- 23 - Manufacture of coke, refined petroleum products and nuclear fuel
- 24 - Manufacture of chemicals and chemical products
- 25 - Manufacture of rubber and plastics products
- 26 - Manufacture of other non-metallic mineral products
- 27 - Manufacture of basic metals
- 28 - Manufacture of fabricated metal products, except machinery and equipment
- 29 - Manufacture of machinery and equipment n.e.c.
- 30 - Manufacture of office, accounting and computing machinery
- 31 - Manufacture of electrical machinery and apparatus n.e.c.
- 32 - Manufacture of radio, television and communication equipment and apparatus
- 33 - Manufacture of medical, precision and optical instruments, watches and clocks
- 34 - Manufacture of motor vehicles, trailers and semi-trailers
- 35 - Manufacture of other transport equipment
- 36 - Manufacture of furniture; manufacturing n.e.c.
- 37 - Recycling

### **E - Electricity, gas and water supply**

- 40 - Electricity, gas, steam and hot water supply
- 41 - Collection, purification and distribution of water

### **F - Construction**

- 45 - Construction

### **G - Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods**

- 50 - Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel
- 51 - Wholesale trade and commission trade, except of motor vehicles and motorcycles
- 52 - Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods

### **H - Hotels and restaurants**

55 - Hotels and restaurants

**I - Transport, storage and communications**

60 - Land transport; transport via pipelines

61 - Water transport

62 - Air transport

63 - Supporting and auxiliary transport activities; activities of travel agencies

64 - Post and telecommunications

**J - Financial intermediation**

65 - Financial intermediation, except insurance and pension funding

66 - Insurance and pension funding, except compulsory social security

67 - Activities auxiliary to financial intermediation

**K - Real estate, renting and business activities**

70 - Real estate activities

71 - Renting of machinery and equipment without operator and of personal and household goods

72 - Computer and related activities

73 - Research and development

74 - Other business activities

**L - Public administration and defence; compulsory social security**

75 - Public administration and defence; compulsory social security

**M - Education**

80 - Education

**N - Health and social work**

85 - Health and social work

**O - Other community, social and personal service activities**

90 - Sewage and refuse disposal, sanitation and similar activities

91 - Activities of membership organizations n.e.c.

92 - Recreational, cultural and sporting activities

93 - Other service activities

**P - Activities of private households as employers and undifferentiated production activities of private households**

95 - Activities of private households as employers of domestic staff

96 - Undifferentiated goods-producing activities of private households for own use

97 - Undifferentiated service-producing activities of private households for own use

**Q - Extra-territorial organizations and bodies**

99 - Extra-territorial organizations and bodies

When the ISIC 3.1 categories are used in a survey, we use the following mapping:

A + B	=	Agriculture and fishing
C	=	Mining and quarrying
D	=	Manufacturing
E	=	Public utility services
F	=	Construction
G + H	=	Commerce
I	=	Transport, storage and communication
J + K	=	Financial and business services
L + M + N	=	Public administration
O + P + Q	=	Other services

## **ANNEX II: INTERNATIONAL STANDARD CLASSIFICATION OF OCCUPATIONS (ISCO)**

The International Standard Classification of Occupations (ISCO) is one of the main international classifications the ILO. It belongs to the international family of economic and social classifications. ISCO is a tool for organizing jobs into a clearly defined set of groups per the tasks and duties undertaken in the job. Its main objectives is to provide:

- a basis for the international reporting, comparison and exchange of statistical and administrative data about occupations;
- a model for the development of national and regional classifications of occupations; and
- a system that can be used directly in countries that have not developed their own national classifications.

It is intended for use in statistical applications and in a variety of client oriented applications. Client oriented applications include the matching of job seekers with job vacancies, the management of short or long term migration of workers between countries and the development of vocational training programmes and guidance.

The first version of ISCO was adopted in 1957 by the Ninth International Conference of Labour Statisticians (ICLS). It is known as ISCO-58. This version was superseded by ISCO-68, which was adopted by the Eleventh ICLS in 1966. The third version, ISCO-88, was adopted by the Fourteenth ICLS in 1987. Many current national occupational classifications are based on one of these three ISCO versions.

ISCO has recently been updated to take into account developments in the world of work since 1988 and to make improvements in light of experience gained in using ISCO-88. The updating did not change the basic principles and top structure of ISCO-88 but significant structural changes were made in some areas. The updated classification was adopted in December 2007 and is known as ISCO-08. Many countries are now updating their national classification either based on ISCO-08 or to improve alignment with the new international statistical standard.

The resolution adopting ISCO-08, the classification structure and correspondence tables with ISCO-88 are available on the ISCO Website in English, French and Spanish. Final definitions of the ISCO-08 groups are currently available on this Website in English only.

The structure, definitions, correspondence tables, and an Introduction summarising the updating process, outlining the methodology and conceptual model used and describing the main differences between ISCO-88 and ISCO-08 will be released in book form as ISCO-08 Volume 1. An index of occupation titles in both alphabetical and numerical order will also be made available on the Website and subsequently in book form as ISCO-08 Volume 2.

### ANNEX III: DATA CHECKS AND EVALUATION

Statistics have been recognized as playing a key role in the work of many organizations. The production of good statistics implies that data evaluation should be one of the key phases in any statistical operation. This refers to the process of assessing the statistical final product in terms of accuracy and reliability. This should be done before the creation of the standardized files, as data should be edited accordingly to a certain standard of accuracy. Two common types of evaluation involve: -

(a) *Validation and certification*

This ensures that erroneous data are not released. This is conducted in conjunction with an interpretative analysis of the data. Due to time constraints, basic methods include: -

- Checks of consistency with external sources of data or from other surveys;
- Internal data consistency checks;
- Unit-by-unit reviews for aggregate estimates especially consumption, expenditure and income data;
- “Reasonableness” or “rational” checks by subject matter specialists;
- Calculation of data quality checks such as non-response rates.

**Eliminating duplicates:** Before the aggregation, one should make sure that there is no duplicated household or individual id (This problem should have been eliminated at the data entering stage, this is just a quick check). There are two types of duplicate household id. One is that one household was entered twice, in which case all other variables should be also the same, such as physical features of the housing, rent paid, etc. In such case, one of the duplicated observations should be eliminated (STATA) can perform this task easily). Second type of duplicate is two or more households have identical households ID. This type of duplicate is hard to discover because one often takes two households as one household. However, unusual large household size, such as over 30, should warrant a check of the household to make sure that they are not two households. When there are two or more same ids signed to household members, if the two or more members have exact same information, such as age, sex, status enrollment, etc., a double entry of the same member is usually the case, and one should eliminate one of the observations. However, if they are two different individuals, one should reassign a new id to one of the members.

For any data set the following minimum standards should be reviewed. This list is not exhaustive but provides a starting point for data editing and analyzes.

- i) All **coded variables** should have their labels within range, valid answers.  
For example, if sex has two value labels - male = 1 and female = 2 or vice versa, no other values are permitted nor missing values allowed.
- ii) The **serial number** or identification number for each member should be unique.

Any two members cannot have the same unique identifier. The range is 1-n, where n is the number of household members. Member numbers must be consecutive starting with 1.

A quick check can be to count the number of household members' vis-à-vis the n, which is the nth person in the household. If these do not correspond to actual household size, then an error may have occurred during data entry or there could be an omission in the number of household members or inclusion of persons belonging to other households.

iii) There must be **one and only one head of household**.

iv) The spouse of the head of household should be of the opposite sex of the head.

Sex of Head  $\neq$  Sex of Spouse

(v) Check **age of household head**. Persons less than 15 years should not be heads. Age and head should be crosschecked.

Head                      age > minimum (say 15 years)

Child of Head        age < age of head - minimum

Parent of Head age > age of head + minimum

However, there may be situations of heads that are less than 15 years. This must be documented.

v) **Marital status** for head and spouse should be identical. If number of spouse > 1, implies that head of household and spouses are polygamous.

Alternatively, the presence of a spouse implies that the head of household is married.

vi) For persons, greater than 99, this should be recoded appropriately as missing values not permitted.

Only the Household heads will have imputed age if missing.

vii) Check age versus marital status of household members. Firstly, all children  $\leq$  5 years must be single.

Children between the ages of 5-12 should be single. However, depending on culture and in exceptionally cases; one may find 10-12 year olds that are married.

viii) **Literacy levels** should be coded appropriately.

If a criterion is in place (say above a minimum age), then anyone less than the minimum age should have values missing.

For example, in some countries, literacy is asked for persons 15 years and above (herein referred as adult literacy). This should be checked. All other persons must have missing values or system missing.

If literacy is asked for all persons regardless of age, then no missing value is permitted.

Check literacy versus age. A child less than 5 years is illiterate with some degree of accuracy.

ix) **School attendance** should be coded appropriately.

Usually children aged less than 3 years will not be in school. Ages 4 and 5 may be in pre-school and will be considered as school attendance. Definition of pre-school is country-specific.

This can be crosschecked with education level attained.

x) **Education level** reached or attained.

In most cases, children less than 5 years will not be in primary school. Children aged 4-5 may be in pre-school. Pre-school definition is country-specific. As a result, one may find children less than 3 in school that may not formal schooling. Therefore this is checked with country in question for correct education definition.

Few exceptions may exist with 4-5 year olds in primary school.

Children less than 3 years will not be in school.

xi) **Education years** completed

Age of member  $\neq$  number of years in school.

One can guesstimate a difference of about 3 years between age and number of years in school. Otherwise code this to missing.

xii) **Morbidity** (sickness) should be for all persons. No missing values permitted.

Only 2 responses allowed – Yes and No. If type of sickness present and morbidity missing and may be some individual-level health expenditure present, then morbidity will be YES.

Missing values not permitted.

xiii) **Household size** must be greater than zero.

xiv) **Number of rooms** must be greater than zero.

xv) Age in months can be calculated if date of interview and date of birth are present.

xvi) For welfare aggregate data, country-specific requirements will be considered and assumed to be correct, within range and reasonable. Expenditure becomes one of the most difficult file to edit. There are no hard and fast rules but each country should be observed independently of another. Consumption expenditure patterns are unique for each country and so one case fits all does not apply.

*(b) Sources of error reviews*

This provides quantitative information on specific sources of error in the data. Sources of errors can occur during field data collection and/or at the editing stage. The results of these reviews are only available after the official release of data. However, this does not imply that the results produced earlier be rendered void but helps in the improvement of data collection methodologies and techniques in the next phase. Sources of errors include: -

(i) *Sampling errors*

Occur when results are based on a sample population instead of the entire population. The sampling error is a measure of variability between all possible samples and can be evaluated statistically. These are grouped into: -

- ✓ Probability sampling – This is used where registers are available and accuracy can be estimated for variables. If estimating from a probability random sample, then a measure of accuracy of an estimate is the square root of the mean square error.
- ✓ Non-probability sampling – National Statistical Institutes may use expert samples based on a high coverage of relevant characteristics. In such circumstances, it is impossible to obtain an objective assessment of the accuracy of the estimates. However, some rough accuracy of quality can be designed using sensitivity analyses.

(ii) *Non-sampling errors*

Non-sampling errors are impossible to avoid and are difficult to evaluate statistically. These include: -

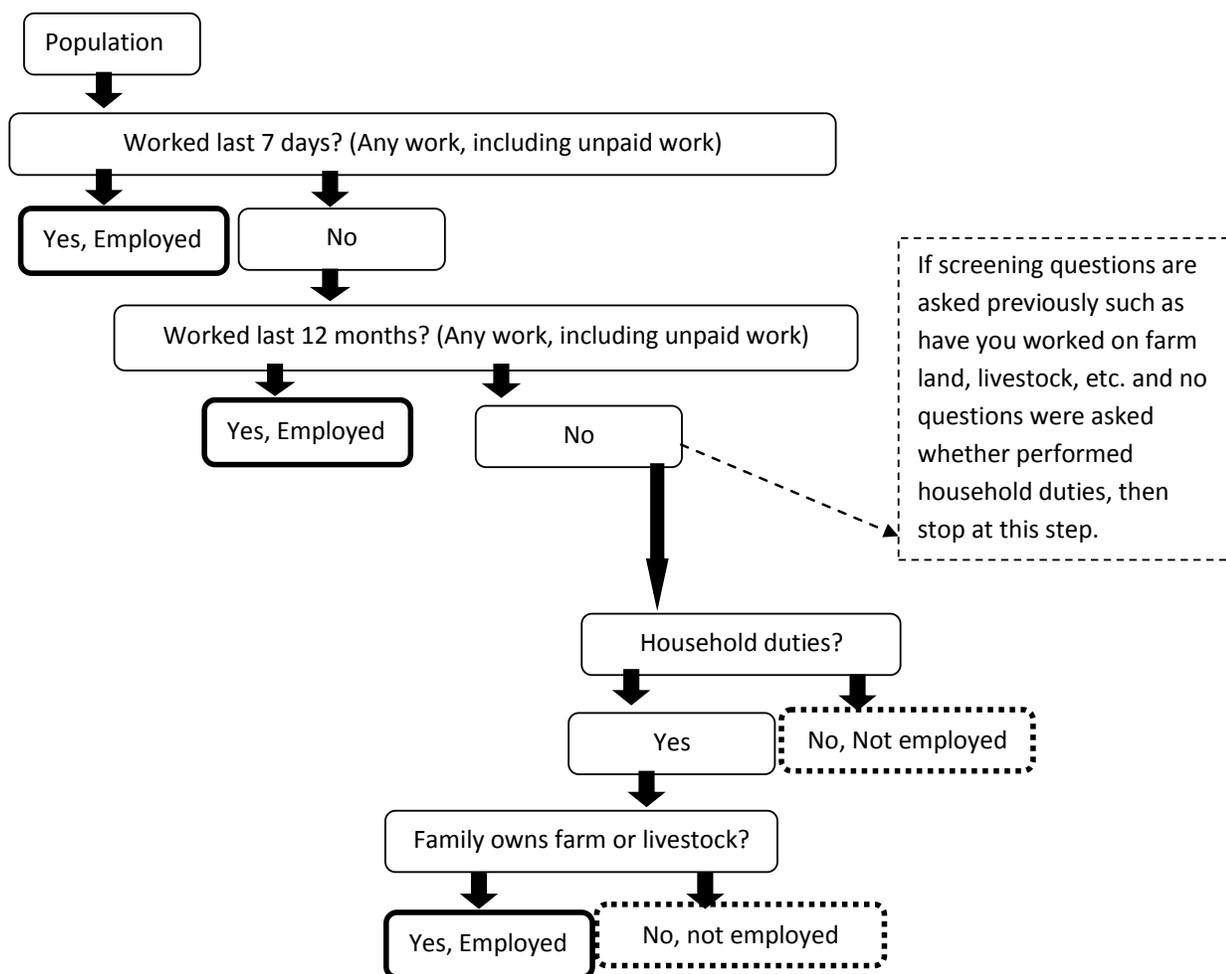
- ✓ Coverage errors – consist of omissions, erroneous inclusions and frame duplication while conducting the survey. This affects every estimate produced and may cause a bias in the results and the effect can vary among sub-groups.
- ✓ Non-response errors – this occurs when an effective sample size is not attained. This may increase the variance due to the decrease in sample size. This can be corrected either through imputation or adjusting weights of the responding units.
- ✓ Measurement errors – occurs when measures differ from true values. This occurs at the time of data collection. This may be caused by the: -
  - *Interviewer* - The interviewer may influence respondent in such way that measurement errors arise.
  - *Respondent* (e.g. lack of understanding of survey question, respondent fatigue, long recall period)
  - *Information system* (e.g. reference period requested may be different from the period i.e. calendar year and accounting year).
  - *Survey instrument* although this is difficult to assess in an objective way. However, a description of the pilot survey and the conclusion of the analysis are necessary in order to assess the questionnaire. Errors also arise due to large questionnaire, vague questions.
  - *Mode of data collection* (interviewing technique – face-to-face, telephone, self-administered, etc.)
- ✓ Processing errors – occurs during post-data collection processes such as validation of data editing, coding, imputation, capture and tabulation.

**ANNEX IV: ISO 3166-1 ALPHA-3 COUNTRY CODES (SUB-SAHARAN AFRICA)**

<b>Country code</b>	<b>Country name</b>
AGO	Angola
BEN	Benin
BWA	Botswana
BFA	Burkina Faso
BDI	Burundi
CMR	Cameroon
CPV	Cape Verde
CAF	Central African Republic
TCD	Chad
COM	Comoros
COD	Congo, Dem. Rep.
COG	Congo, Rep.
CIV	Cote d'Ivoire
GNQ	Equatorial Guinea
ERI	Eritrea
SWZ	Eswatini
ETH	Ethiopia
GAB	Gabon
GMB	Gambia, The
GHA	Ghana
GIN	Guinea
GNB	Guinea-Bissau
KEN	Kenya
LSO	Lesotho

<b>Country code</b>	<b>Country name</b>
LBR	Liberia
MDG	Madagascar
MWI	Malawi
MLI	Mali
MRT	Mauritania
MUS	Mauritius
MOZ	Mozambique
NAM	Namibia
NER	Niger
NGA	Nigeria
RWA	Rwanda
STP	Sao Tome and Principe
SEN	Senegal
SYC	Seychelles
SLE	Sierra Leone
SOM	Somalia
ZAF	South Africa
SSD	South Sudan
SDN	Sudan
TZA	Tanzania
TGO	Togo
UGA	Uganda
ZMB	Zambia
ZWE	Zimbabwe

## Definition of Employment in Africa Region



Note: This diagram tries to capture unpaid economic activities in Africa, based on a research about under reporting labor participations in Tanzania. It found that many unpaid family workers, especially women, under report their economic activities due to poor questionnaire design. For details see Bardasi, Beegle, Dillon and Serneels "Do Labor Statistics Depend on How and to Whom the Questions Are Asked", World Bank Policy Research Working Paper 5192.