Global Monitoring Database (GMD) Harmonization Guidelines

Data for Goals (D4G) Team

Poverty Global Practice

The World Bank

September 2023

The Global Monitoring Database (GMD) harmonization has benefited from many World Bank Group (WBG) colleagues feedback.

The guidelines developed by GMD project in the Poverty Global Practice team address important knowledge gaps in harmonization and comparability of survey indicators measuring inequality, poverty and living conditions on a global scale.

Data for Goals (D4G) task Team leaders are appreciative to all the contributions that have involved since the inception in 2004.

Table of Contents

1 Introduction

1.1 Global Micro Database (GMD) Project

The Global Micro Database (GMD) founded with the aim of developing a method for survey standardization to construct indicators globally that comparable microdata across countries, regions and across years for global poverty monitoring and welfare measurement. GMD is based on the best multi-purpose[1] available surveys that cater for many applications and different users, have wide coverage of years, and are available, accessible, and shareable within the World Bank Group (WBG).

Household surveys are one of the top three sources of social and demographic information in many countries.[2] They collect detailed and diverse socio demo-graphic data on people's living conditions and well-being; activities they engage; and demographic characteristics and cultural factors that influence behavior, social and economic change among many.

Although generally focused on examining household consumption expenditures and income patterns and their characteristics, etc., household surveys represent a wide variety of survey instrument designs, with country unique variable coding standards. Integrating or comparing multiple surveys can therefore be very costly and time-consuming exercise. Moreover, the lack of a standardized survey structure adversely affects the comparability of survey indicators and thus severely limiting their applications and usefulness of the analyses. This presents a serious constraint not only for cross-country analysis but also within country over time and for national research.

The primary purpose of these Guidelines is to ensure that relevant statistical and research teams, both internal and external to the WBG adhere to recognized standards in producing harmonized data, document files and apply a common strategy best suited to the original source materials. It also aims to provide country, regional, and global teams, including the User community and data harmonizers, with the tools and material resources necessary for collecting, processing, harmonizing, and analyzing survey data. Consistency and comparability of indicators derived from the survey can only be ensured by using this a common framework for the harmonization of survey indicators, which envisages standardization of indicators and indicator definitions across time and space.

The guidelines are primarily intended for local statistical teams coordinating regular surveys, but they also provide useful information for analysts, policy makers, researchers, and other users of survey data. Additionally, with the growing need for data on SDG indicators, GMD will support the development of products aligned with banks' corporate objectives, in addition to those developed by the Poverty and Equity Global Practice.

---

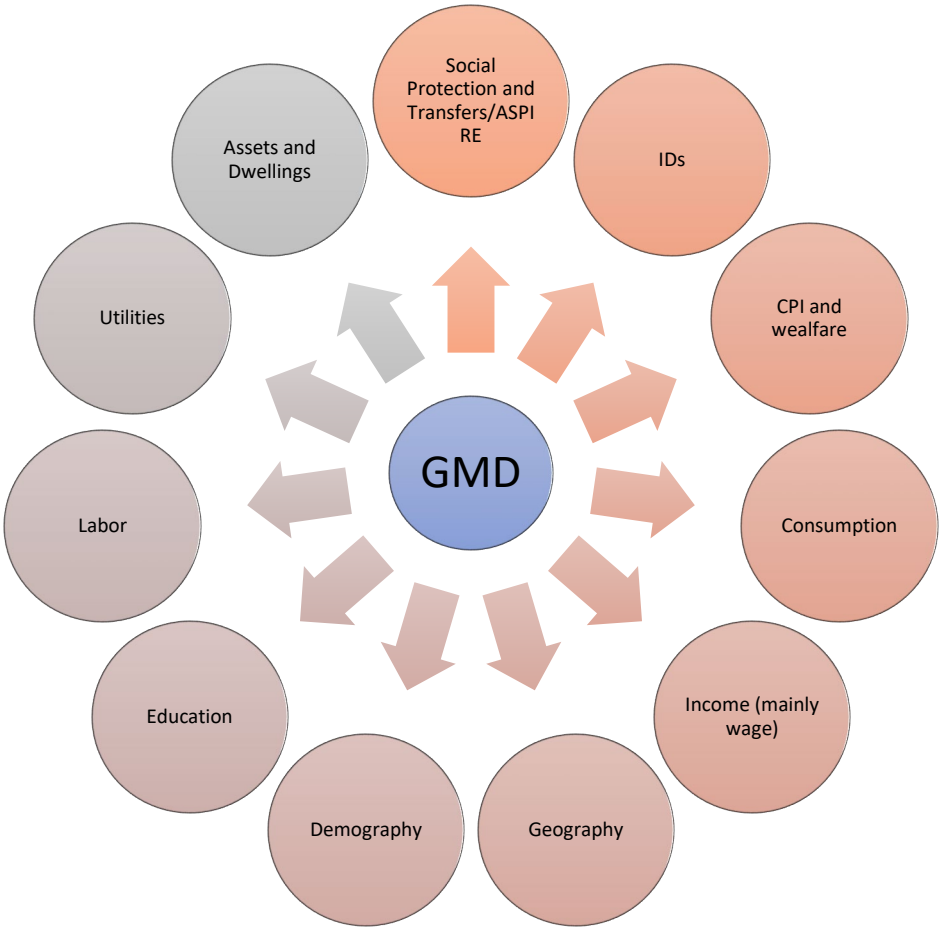
1 In most of the countries, the household budget survey is the best available type of multi-purpose household level information. However, the Living Standard Measurement Survey (LSMS), EU Statistics on Income and Living Conditions (EU-SILC), and other types of household surveys are also used.

2 The other two major data sources are population and housing census and administrative record systems.

1.2 File Structure and Content

The Global Monitoring Databases consists of 10 modules, as presented in Figure 1. The modules covered comprise 1. IDs; 2. CPI and welfare; 3. Consumption; 4. Income; 5. Geography; 6. Demography; 7. Education; 8. Labor; 9. Utilities; 10. Assets and Dwellings, and 11. Social Protection.

Figure 1. Modules of the Global Monitoring Database



Source: Based on initial GSG proposal.

Each module file contains data for all module subject areas available in a survey. The files also contain selected identification and demographic information from the core dataset, making it possible to conduct the analysis independently from the core file and full panel files. If more detailed social or demographic information is required for an analysis, users can obtain that information by merging module files with each other or with any other statistical information.

To facilitate documentation consistency, all topical module files have been assigned with their unique identification numbers. For example, the "Geography" module can be identified under the code 5. Furthermore, some modules may be broken down further into sub-modules. For example, the "Assets and Dwellings" module is comprised of two sub-sections – Assets and Dwellings.

1.3 Data Dictionary

This section discusses the names, definitions, and attributes of data elements within the GMD data system. The goal of the GMD Data Dictionary is to provide a comprehensive information catalog of data definitions, relationships, collection groupings, and sources of data.

The inventory revealed about 300 variables available in the latest version of the GMD Data Dictionary. It should be noted that as the GMD team continues updating and revising the Data Dictionary, the number of variables may differ from the number of variables covered in the current version of the Guidelines.

All topical GMD modules consistently use the same format and naming convention. As noted, all GMD modules contain a set of common identification and demographic variables, such as a country code, year variable, household identifier, individual identifier (if applicable), and household weights. The number of module specific variables typically varies across the subject areas depending on the type of module under consideration.

To facilitate common understanding, each topical GMD module contains a table summarizing all the variables covered in the module, including variable name, variable label, variable description, acceptable variable type after harmonization, as well as sources of variable information. Each variable element is also mapped to its tier, GMD module, as well as sub-module.

1.4 File organization

Since GMD is a collaborative effort where different individuals work on the harmonization of different surveys, it is essential that everyone follows the same file structure and do-file organization to allow for easy access and understanding.

1.4.1 Where to find the original raw data files

The first step for each harmonization is to obtain the original raw data files. The original data files used for the GMD harmonization should be cataloged and stored in the regional shared drive and the central Microdata Library catalog (http://microdatalib/). We assume that the depositing of the original microdata in the Microdata Library is a collective responsibility shared by all World Bank staff working in all the World Bank's regions, irrespective of global practice, since the dialogue and data acquisition through the NSO and line ministries often take place in a decentralized manner. Those original data files are stored in a secured server, to which access is limited only to the regional admins (see Azevedo and Cancho, 2013 for further details).

To avoid having multiple data files stored and saved in different location, it is recommended that the users should use and query the original files within the datalibweb system.

1.4.2 Do-file and data file naming guidelines

Data sets and do-files have identical file naming conventions. Do-files used to create a data file must have the same file name and follow the naming guidelines below.

Example: The following files are the do file and STATA file for the GMD consumption module harmonization using the Kazakhstan 2011 HBS.

KAZ_2011_HBS_v01_M_v01_A_GMD_CON.do
KAZ_2011_HBS_v01_M_v01_A_GMD_CON.dta

Table 1. GMD File Naming Convention

Generic Structure: ${CCC}_${YYYY}_SurveyName_vnn_M_vmm_A_GMD_mod		
Component	**Description**	**Values**
CCC	Country Code	ISO 3 code, for example: KAZ, GEO, TJK
YYYY	Year of the Survey	The starting year of the survey
SurveyName	Abbreviation of the survey (acronymic) – See the metadata for reference	Example: HBS, KIHS, LFS, EUSILC Character length can vary.
v*nn*	stands for the version of the master file	*nn*=01, 02, …
v*mm*	stands for the version of the harmonization, as there can be revisions after the first release	*mm*=01, 02, …
mod	denotes the specific GMD dataset (module)	Always 3 letters corresponding to the module. IDN=ID COR=Core GEO=Geographic DEM=Demographic DWL=Dwelling LBR=Labor UTL=Utilities CON=Consumption

1.4.3 How to store do files and GMD harmonized files

While the following instructions are specifically for data harmonizers at the World Bank, following a clear set of naming and file storing guidelines will save time and ensure accuracy for all researchers.

Data organization and data harmonization do-files must use global to define the root of the datalibweb folder or call the files through datalibweb. This way do-files are not hardcoded to path does not exists anymore or harmonizer temporal working root or inaccessible path. Having the do-files linked to datalibweb folder structure but not to a specific hardcoded path allow maintain the replicability of the harmonization when data is hosted in a different place. Also, user can replicate easily from inputs from

datalibweb. The file names and versions would then be saved using the global within the .do file, allowing others to run the .do files with few (or even without any) updates needed.

Do this:

<Top of the do-file>

```
*<_Program setup_>
clear all
set more off
glo rootdatalib "ADD PATH"
glo CCC         "LKA"
glo YYYY         "2019"
glo SURVEY       "HIES"
glo vnn          "01"
glo vmm          "01"
glo mod          "COR"
…
glo out "${rootdatalib}\\${CCC}\Data\Harmonized"
```

. . .

<Save harmonization>

```
cap mkdir "${rootdatalib}"
save ${out}\${CCC}_${YYYY}_${SURVEY}_v${vnn}_M_v${vmm}_A_GMD_${mod}.dta
```

Folders with raw and harmonized data, and the corresponding documentation and final do-files must follow WB folder structure, which is as follow.

Folder structure:

I. Raw data

- DATALIB
  - CCC
    - ${CCC}_${YYYY}_SURVEY
      - ${CCC}_${YYYY}_SURVEY_VNN_M
        - Data
          - Original
          - Stata
        - Doc
        - Programs

II. Harmonized data

- DATALIB
  - CCC
    - ${CCC}_${YYYY}_SURVEY
      - ${CCC}_${YYYY}_SURVEY_VNN_M_VMM_A
        - Data
          - Harmonized
        - Doc
        - Programs

Example of how to create folder for raw data:

```
*<_Program setup_>
clear all
set more off
glo rootdatalib "ADD PATH"
local ccc         "LKA"
local yyyy         "2019"
local survey       "HIES"
local vnn          "01"

*------------------------------------------------------------------
*------------------------------------------------------------------
local yearfolder   "`ccc'_`yyyy'_`survey'"
*<_Folder creation_>
cap mkdir "${rootdatalib}"
cap mkdir "${rootdatalib}\\`ccc'"
cap mkdir "${rootdatalib}\\`ccc'\\`yearfolder'"
cap mkdir "${rootdatalib}\\`ccc'\\`yearfolder'\\`yearfolder'_v`vnn'_M"
cap mkdir "${rootdatalib}\\`ccc'\\`yearfolder'\\`yearfolder'_v`vnn'_M\Data"
cap mkdir "${rootdatalib}\\`ccc'\\`yearfolder'\\`yearfolder'_v`vnn'_M\Data\Original"
cap mkdir "${rootdatalib}\\`ccc'\\`yearfolder'\\`yearfolder'_v`vnn'_M\Data\Stata"
cap mkdir "${rootdatalib}\\`ccc'\\`yearfolder'\\`yearfolder'_v`vnn'_M\Doc"
cap mkdir "${rootdatalib}\\`ccc'\\`yearfolder'\\`yearfolder'_v`vnn'_M\Programs"
*</_Folder creation_>
```

In the review process, versions will be zipped with time stamp for version control.

1.4.4 Saving *.dta base files

In addition to the clean, harmonized files created in the process described below, an intermediate version of the .dta files should also be saved in the same location. This version should include the variables used in the process of creating the harmonized data. The file should be saved using the following convention:

```
${CCC}_${YYYY}_${SURVEY}_v${vnn}_M_v${vmm}_A_GMDBASE_${mod}
```

1.4.5 Working versions of the harmonized files

Over the course of harmonizing raw data into the appropriate format, files invariably go through several versions. To keep track of these versions, the last version of the harmonization that is “live” on datalib should be used as the root name, but with a slight modification to differentiate updates that occur between datalib versions. As an example, a first version of a harmonization would start with 00wrk1:

```
${CCC}_${YYYY}_${SURVEY}_v${vnn}_M_v00wrk1_A_GMD_${mod}.dta
```

Once the file is moved to datalib for the first time, the version changes to v01. Any updates from that point would be done using the new root version: v01wrk, and so forth. For clarity and ease of navigation, each time a new file is created it should be placed in its own folder, identically named to the harmonization .do file.

1.4.6 How to use the datalibweb system to load the raw/original data

Often the original/raw data was saved with Stata format in the datalib folder with a structured folder. This well-defined structure allows programs to query the data easily, making the sharing of the work much easier.

Harmonization from original/raw data should start with datalibweb, thus ensuring that the data is available for all to use.

For example: the code below is used to query the file "NHIES_2015_16_individual_level.dta" from the NAM 2015 survey.

```
datalibweb, country(NAM) year(2015) type(SSARAW) survey(NHIES)
filename(NHIES_2015_16_individual_level.dta) clear files
```

or

```
datalibweb, country($CCC}) year(${YYYY}) type(${type})
survey(${survey}) filename(NHIES_2015_16_individual_level.dta)
clear files
```

For more examples on how to use datalibweb to query the data, please see the helpfile of datalibweb.

1.5 Do-File Organization and Guidelines

1.5.1 Do-files: header guidelines

Each module (*mod*) and each do-file must have the same file name structure.
All relevant information about the project should be included in the preamble of each do file (see Box 1). The relevant information includes:

- Description of the original survey
- Name of the researcher who created the current do-file (and of researchers who worked on previous versions)
- Complete names of the input and output dataset
- Global paths where temporary output data will be stored
- Essential variables, such as country name, country code, country management unit, year and survey name

Box 1: Do file – Preamble

```
/* ------------------------------------------------------------------------
                          GMD Harmonization
 ------------------------------------------------------------------------
 <_Program name_>                      MNE_2013_HBS_v01_M_v02_A_GMD_CON.do </_Program name_>
 <_Application_>                       STATA 13 <_Application_>
 <_Author(s)_>                         ABC, XYZ </_Author(s)_>
 <_Date created_>                      2014-03-01 </_Date created_>
 <_Date modified>                      2015-03-01 </_Date modified_>
 ------------------------------------------------------------------------
 <_Country_>                           Montenegro (MNE) </_Country_>
 <_Survey Title_>                      Household Budget Survey (HBS) - 2013 </_Survey Title_>
 <_Survey Year_>                       2013 </_Survey Year_>
 <_Reference Year_>                    2013 </_Reference Year_>
 <_Study ID_>                          1234 </_Study ID_>
 <_Data collection from (M/Y)>         MM/YYYY </_Data collection from (M/Y)_>
 <_Data collection to (M/Y)_>          MM/YYYY </_Data collection to (M/Y)_>
 <_Source of dataset_>                 Poverty and Education teams </_Source of dataset_>
 <_Sample size (HH)_>                  12240 </_Sample size (HH)_>
 <_Sample size (IND)_>                 52240 </_Sample size (IND)_>
 <_Sampling method_>                   Stratified, multi-stage </_Sampling method_>
 <_Geographic coverage_>               National </_Geographic coverage_>
 <_Geo_1_>                             division </_Geo_1_>
 <_Geo_2_>                             region </_Geo_2_>
 <_PSU variable_>                      psu </_PSU variable_>
 <_Number of food items_>              133 </_Number of food items_>
 <_Food recall period(s)_>             Last 14 days </_Food recall period(s)_>
 <_Number of non-food items_>          221 </_Number of non-food items_>
 <_Non-food recall period(s)_>         One month/3 months/12 months </_Non-food recall period(s)_>
 <_Rent imputation_>                   Actual and owner's estimate rental values, taken from existing
data </_Rent imputation_>
 <_Durables_>                          Use value of durable goods could not be calculated as value
of the items when purchased are not available </_Durables_>
 <_Regional price deflators_> Paasche price index using food consumption items </_Regional price
deflators_>
 <_Consumer price index_>              1.56 value of base year 2005 </_Consumer price index_>
 <_Currency_>                          EURO </_Currency_>
 ------------------------------------------------------------------------

 <_Version Control_>
 Date: 2015-03-01
 File: MNE_2013_HBS_v01_M_v02_A_GMD_CON.do - Changes on education variables (5 level to 3 level)
- Add more durable items to consumption aggregates (items are ...)
 Date: 2014-03-01
 File: MNE_2013_HBS_v01_M_v01_A_GMD_CON.do - First version
 </_Version Control_>
 */
```

1.5.2 Do-files: variable and variable note tagging guidelines

The location of i) variable creation and ii) survey/headings information in the .do files will be tagged using a standardized approach. All harmonized variables in the in each module in the data will be tagged according to the following convention:

- The beginning of the code relating to a harmonized variable should be proceeded by **<_var_>*
- The end of the code relating to the variable creation should read **</_var_>* where "var" is the harmonized variable being created.
- Variables that are already named (such as in the case when "HHID" is already defined) should be noted when the file when opened, using the same convention as above. Between the "open" and "close" codes, a starred outline should read:"**'var' brought in from 'source'*"
- If a variable is created more than once (for example, HHID is created from several sources, then used to merge), it should be tagged only once.

- If possible, the DDI tags should include any value labels for the variable
- For survey information (such as spatial deflation) a similar convention should be used: **<_var_s_>*, the "s" signifies that the variable is "survey" information that is included in the general description of the data in the DDI.

It would also be important to note the comments, if necessary, when creating the harmonized variables. For any variables with notes or comments on how the variable is created, that information is also needed to tag so one can pull out those variable-specific notes by using similar taggings **<_var_note_> and *</_var_note_>*.

- For example, the variable "LFSTATUS" is created only for individuals with age of 15 and above, then one can put the variable-specific note as follow: **<_ LFSTATUS_note_> Only for individuals with age of 15 and above *</_LFSTATUS_note_>.*

These tags will then be located automatically using the *strpos* command in Stata. It is incorporated in the ado called **ddi2.dta** to either generate DDI or to extract information from the dofile.

There are two useful purposes of tagging the harmonization variables: (1) tagging is useful when cross checking the definitions of harmonized variables overtime, and when comparing the comparability of such variables with different countries; (2) tagging will improve the automated updating of the DDI by adding the block of codes used for generating the harmonized variables in the variable description of the DDI. Tagging will also improve the transparency of the metadata DDI for basic users in the Microdata Library.

Some rules/comments in tagging (to be updated when testing):

- No double quotes: " in the tagging block, especially with * comment. If double quotes are included in the label, the ado will extract the double quote and put the rest in the DDI.
- Take the size of the block into account of (not too long, focus on the main idea). For example, for the spatial deflator variable, a short version of the code can be included.
- Tagging should be done for one variable at a time, not a group of variables.

1.5.3 Updates tracking/vintages

If there are any changes resulting into the new harmonization .do files, a summary of the changes in a text (.txt) document should be included. The name of the file should be the same as the harmonized Survey ID. Duplicate do-file and data file, put all changes in the .txt (same study ID and in the do-file). Cumulative and start with the latest one and goes backwards. This readme file should be stored together with the do-file (same folder). For example, the name could be:

```
"MNE_2013_HBS_v01_M_v03_A_GMD_readme.txt"
*<_Version Control_>
 Date: 2015-12-09
 File: MNE_2013_HBS_v01_M_v03_A_GMD_CON.do - File created
 Harmonization - outcome: MNE_2013_HBS_v01_M_v03_A_GMD_CON.dta
 Notes: Missing consumption in COICOP 9 and 10 added. GMD Harmonization 3.0
conducted.
*</_Version Control_>
```

If there are changes in content, we should make the new version of the data and the dofile all together, and the readme should explain the changes. For other modules where the data/dofiles are not affected, the names should also be updated to match the updated module.

For example, changes made to the labor module (LBR) only, we have "ARM_2012_ILCS_v01_M_v05_A_GMD_LBR.dta", then version v05 should be in other modules of this harmonization.

It is also important that the notes on the changes are added to the data cumulatively, so we keep the historical records of all the changes. In Stata, one can do that by:

```
note: ARM_2012_ILCS_v01_M_v05_A_GMD_LBR.dta; 11/02/2015; age is fixed, and labor
information is added.

For more detail, please see the template header and labeling dofiles for each
module.
```

1.6 Missing Value Codes

Surveys allow answers like "Do not know", "Decline" or other similar. Sometimes, in the anonymization process, monetary variables get truncated. Sometimes, the original files of those answers are coded in the format of -99, -999, and -88, among others. Part of the harmonization process include the investigation of those type of codes, and those must be replaced with missing values.

Stata allows distinguishing between missing values: ". a", ".b", …, ".z". The following missing codes are proposed, and additional missing values could be defined, in which case they must be defined in the labels and included in the "notes" of the variable.

".d" as "Decline to answer"
".k" as "Do not know"
".t" for truncated values in continue variables. For example, if wages over a threshold are coded as "999999".

See **Stata Manual Missing Values** for more details on missing values.

Harmonizers need to clearly differentiate missing values of variables from variables that were present in the survey but could not be harmonized due to reasons such as time unavailability. This will help the future harmonizers to focus on the unharmonized variables. The missing value code for these two scenarios are:

- For variables unavailable in survey = "."
- For variables available in the survey but not harmonized =" .a" .
- For variables cannot be harmonized because data does not meet definition= ".b" .

```
gen varname=".a"
```

Harmonizers should reach out to the TTL and/or regional focal points to discuss any situation in which variables are planned to be classified as ".a", "b", or other special missing.

Note:

- Please follow the definition instructions of the manual as close as possible.
- If unclear, ask Task Team Leader (TTL).

2 ID module (IDN)

2.1 Framework for Harmonization

The ID module covers identification variables that are commonly derived from survey data sets or that can be constructed using existing variables. The primary objective of the indicator harmonization is to generate a unified data source of globally comparable identification statistical indicators to support routine analytical and corporate business functions of the World Bank.

The ID variables are essential for necessary for keeping data well-organized and attributing the findings to a specific country or year. When the World Bank receives survey data from NSOs, those data sets can be in one file or spread between different files. When the data is spread between different files, household IDs are necessary to merge the data. Household IDs mat contain several identifying variables such as the region and/or PSU they come from. In the case that households don't have clear HHID, harmonizers can reconstruct new HHID to ensure that data is correctly attributed to the correct household. In this module, the primary units of analysis are the level of individuals within a household and the household.

The GMD variables is split into two groups: -

- Tier 1 variables are compulsory and must be harmonized for every survey;
- Tier 2 are optional variables to be derived depending on region needs.[3]

2.2 Creating IDs

The ID module is a file at individual level with basic household and individual identifiers. The household identifier is HHID, and the individual identifier is PID. Those variables must be present and must be the same across the different modules of the harmonized database. For efficiency of the data and merging across different files, those HHID and PID variables must be stored in the same format (numeric preferable) throughout all data files within each survey. In addition, the ID module must also have the original variables that were used to construct the HHID and PID variables. Those variables are useful when users want to merge the harmonized data with the original data files. Those original variables must be in the same format and type with those variables in the original data files. In practice, the identifier variables should be common in all the raw/original data files so that they can be merged between them. Those variables are often the good candidates for the ID variables.

Example 1: the below codes are used in creating the HHID and PID variables for Albania's LSMS 2012:

```
datalibweb, country(ALB) year(2012) filen(Modul_2B_education.dta)
type(ECARAW) surveyid(ALB_2012_LSMS)

*<_pid_>
rename idcode pid
```

[3] As noted under each topic, Tier 2 variables are comprised of previously identified variables under GMD 2.0. Regional teams with the capacity and systems ready to keep producing Tier 2 variables are very strongly encouraged to do so.

```
*</_pid_>

*<_hhid_>
gen hhid= psu * 100 + hh
*</hhid_>
```

In this example, those original variables are IDCODE, PSU, and HH. Therefore, in the ID module we have the following variables: COUNTRYCODE, YEAR, HHID, PID, IDCODE, PSU, and HH.

Example 2: For the case of HHID and PID are already in the raw/original database, either in the same numeric format or in different format, it would be still important to have the ID variables standardized and also keep the original variables. Those variables can be renamed with the suffix _ORIG. For example, in a country where HHID and PID are used as ID variables, then in the ID module, we should have the following variables: COUNTRYCODE*, YEAR, HHID, PID, HHID_ORIG,* and *PID_ORIG*; where HHID_ORIG and PID_ORIG are the same in content as HHID and PID, respectively.

Example 3: Create the HHID variable from the concat- function in Stata from two variables. Note that -*sort*- must be used before the -*group*- function, otherwise the HHID variable will point to different households in each data files. One good practice is to do that *concat()* function only one time and use the variables in the *group()* to merge across the data files.

```
*<_hhid_>
sort folio e10
egen hhid =concat(folio e10)
*</hhid_>

*<_pid_>
clonevar pid = eglin
*</_pid_>
```

2.3 Mapping and Description of Variables

The aim of this section is to provide the readers with basic information about the main sources of information and then some details about how GMD ID are produced, and the issues surrounding them.

countrycode
This is a string variable that specifies the 3-character country ISO3 code used by the World Bank to identify each country. Although there are different naming conventions, it is necessary to use those specified to ensure that the data for each country is appropriately labeled.

year
This is a numeric variable that denotes the year in which the implementation of the household survey was begun. For example, if a survey was implemented during October 2018 and September 2019, the *year* would be 2018. This is based on World Bank survey catalogue metadata guidance https://microdata.worldbank.org/index.php/home

int_year

This in a numeric variable when most of the survey data collection was conducted using the following rules:

i) if the period of reference for the survey covers multiple years, use as reference year the one with most of the survey respondents; and,

ii) if the period of reference is half in one year and half in the other year, the first year will be used.

int_month

This is a numeric variable that specifies the month when the survey questionnaire was administered to the household.

1 = January
2 = February
3 = March
4 = April
5 = May
6 = June
7 = July
8 = August
9 = September
10 = October
11 = November
12 = December

hhid_orig

This is the household identifier available in the original data. The original format of original data (string or numeric) should be preserved.

hhid

This provides a unique household identification number within the data file. This is a string variable. Make sure the HHID is unique when converted to string. If there is Household ID in the original data is a string, HHID and HHID_ORIG should be the same.

If HHID_ORIG is missing, it is constructed by "variable names in raw data" variables. **See in Section 2.2.**

pid_orig

This is the individual identifier within the household available in the original data set.

pid

This allows identification of individuals. Variable will vary in length depending on how the identification code was constructed in each country. Depending on individual countries, this variable may be a concatenation of several variables in the raw data file. Keep format (string or numeric) of original data. If there is Personal ID in the original data, PID and PID_ORIG should be the same. If PID_ORIG is missing, it is constructed by "variable names in raw data" variables.

variable names in raw data
These are the variables in the raw data used to construct HHID and/or PID, when HHID_ORIG and PID_ORIG are not available.

Table 2.1: ID Module

	Module Code	Variable name	Variable label	Description	Tier
1	ID	**countrycode**	country code	String	1
2	ID	**year**	4-digit survey start year	Numeric	1
3	ID	**int_year**	4-digit year of household survey interview	Numeric	1
4	ID	**int_month**	Month of household survey interview	Numeric	1
5	ID	**hhid_orig**	Household unique identifier in the raw data	Numeric or string	1
6	ID	**hhid**	Household unique identifier	String	1
7	ID	**pid_orig**	Personal unique identifier in the raw data	Numeric or string	1
8	ID	**pid**	Personal unique identifier	String or numeric	1
9	ID	**weight**	Household weights	Numeric	1
10	ID	***variable name in raw data***	Variables used to construct Household identifier	Numeric or string	1

2.4 Lessons Learned and Challenges

- Creating and checking IDs.
  Avoid using the sequential index of the observation as the ID (i.e., gen hhid = _n). This is dangerous as the order of each observation may be different, even across vintages of the same file sorted by to different variables.

Create `hhid` like this:

✓
```
gen hhid= psu * 100 + hh
tostring hhid,replace
```

Not like this:

✘
```
gen hhid = _n
```

- Check if HHID is missing.

```
assert missing(hhid)
```

- When creating HHID and PID, especially from string variables or from *group(varlist)* or *concat(varlist)* functions, users should try to create them from roster data files first where all information or observations are available. In addition, the order of the variables in the *varlist* option above must be the same across the files. Across the data files, the order and the sort on the variables in the varlist must be done in the same way across files.

- When the HHID and PID are in numeric format but less precision, it is recommended to bring them the accurate precision level so it can be used in the merging correctly. For example, the value of the HHID for an observation might be 100021210121 (a long number), users should format the variable by "format %15.0g hh" before converting to string.

- It is recommended to check the uniqueness level of the data files with identifier variables at the corresponding level of the data (i.e. household vs individual level data).

- HHID and PID need to be unique in the database.

```
isid hhid pid
cap destring pid, replace
duplicates report hhid pid
local n=r(unique_value)
`N'!= `n'
```

- Harmonizers should also ensure that there is a perfect match between the hhid and the pid of each household and individual, respectively, across modules. Even if some variables are missing for some individuals/households, these observations should be included in the dataset with missing values for the related information to ensure that the datasets merge.

- In case a household survey is conducted more than once per year – e.g. quarterly HH surveys – you may want to use this as panel data, in which case the household ID can remain as is. However, if you want to use the data as cross-sectional, then new HHIDs can be constructed for each HH for each quarter.

Quarter	Quarter 1	Quarter 2	Quarter 3	Quarter 4
hhid_orig	hhid=1	hhid=1	hhid=1	hhid=1
hhid	hhid=1Q1	hhid=1Q2	hhid=1Q3	hhid=1Q4

- Checks on the country code and year

Ensure that country is a three-letter country code.

```
cap confirm str3 var country _rc!=0
```

- Harmonizers should also ensure that country codes are updated according to the ISO 3 country codes (can be found in Appendix A). Some common adjustments include the following:

```
cap replace countrycode="XKX" if countrycode=="KSV"
cap replace countrycode="TLS" if countrycode=="TMP"
cap replace countrycode="PSE" if countrycode=="WBG"
cap replace countrycode="COD" if countrycode=="ZAR"
```

- Furthermore, harmonizers should check that the years used are in an appropriate range. The year needs to be a four-digit number in the range of 1980 to the current year (assumed here to be 2020).

```
(year<1980 | year>2020) & mod(year, 1) == 0
```

- Run basic tabulations for all categorical variables and check if within range. If not within the allowed range, return and fix variable.

```
tab int_month int_year,m
tab1 year urban
```

- Household weighting coefficient cannot be missing.

```
count if weight==.
```

The first step is to analyze and try to understand the reason why weights are missing for a given observation. Based on the documentation and revision of the data:

  - If you believe it is an error, one alternative is:
  - If those observations are not main members of the household (domestic workers of the household or renters), GMD does not include no main members of the household.

- Run the dofile ***MODULE_01-IDN.do*** which labels (variable and value) and order variables.
    - ✓ This is a must process with no exception.

3 Geography (GEO)

3.1 Framework for Harmonization

The geography module presents the geographic context in which households are situated. This chapter describes geographical dimensions that are commonly derived from survey data sets or that can be constructed using existing variables. The overall objective of this indicator harmonization is to generate a unified data source of globally comparable geographical indicators to support routine analytical and corporate business functions of the World Bank.

The international guidelines presented in this section contribute to the development of internationally comparable geographical statistics. Geographical indicators are instrumental in building an evidence base of the drivers and consequences of regional and local inequities and in informing policies for fostering regional convergence and other development outcomes. Conversely, the lack of adequate and internationally comparable geographical data is a major impediment to informed and effective policies. Regional breakdowns of information are necessary for several SDGs, such as SDG 1.1, which requires information for populations in different geographic areas. Moreover, it is crucial to understand uneven progress on SDGs within countries wherever possible, which requires information at the lowest subnational level available.

In this module, the primary unit of analysis is the household. Geography module variables typically cover the entire country with all its subdivisions, unless indicated otherwise. The harmonization framework consists of fifteen variables in the sub-national administrative structure of the country. For clarity and convenience, the largest administrative subdivision of a country is referred to as the "first-level administrative division" or "first administrative level". The next level is called "second-level administrative division" or "second administrative level." Each variable can then be used to identify a country subdivision in a global context uniquely. The geographical mapping is also compatible with the classification of territorial units for statistics. In ECA, the NUTS classification is used for regional classification.[4] In other regions, the remaining sub-national indicators are country-specific and follow existing naming conventions. To the extent possible, however, the administrative codes used in the survey should match an existing shapefile used by the national statistics office that describes the location of the administrative units.

3.2 Mapping and Description of Variables

This section describes all variables available in the geography module. A few variables from the IDN are included in this module (See table 3.1).

[4] NUTS areas aim to provide a single and coherent territorial breakdown for the compilation of EU regional statistics.

subnatid1
Country-specific categorical variable. This refers to a subnational identifier at the highest level within the country's administrative structure. This is typically a province or state. The variable is string and country-specific categorical. Numeric entries are coded in string format using the following naming convention: "1 – Hatay".

subnatid2
Country-specific categorical variable. This refers to a subnational identifier at which survey is representative at the second highest level within the country's administrative structure. This is typically a district. The variable is string and country-specific categorical. Numeric entries are coded in string format using the following naming convention: "1 – Hatay".

subnatid3
Country-specific categorical variable. This refers to a sub-national identifier at which survey is representative at the third level within the country's administrative structure. This is typically a sub-district. The variable is string and country-specific categorical. Numeric entries are coded in string format using the following naming convention: "1 – Hatay".

subnatid4
Country-specific categorical variable. This refers to a sub-national identifier at which survey is representative at the lowest level within the country's administrative structure. In some countries, this is effectively a village. The variable is string and country-specific categorical. Numeric entries are coded in string format using the following naming convention: "1 – Hatay".

subnatidsurvey
Country-specific categorical variable. This is a string variable that refers to the lowest level of the administrative level at which the survey is representative. In most cases this will be equal to "SUBNATID1" or "SUBNATID2".

However, in some cases the lowest level is classified in terms of urban, rural or any other regional categorization cannot be mapped to SUBNATIDs. The variable would contain survey representation at lowest level irrespective of its mapping to SUBNATIDs.

strata
Country specific and a unique identifier is created for each stratum.

This refer to the division of the target population – typically the census sample frame -- into subpopulations based on auxiliary information that is known about the full population. Sampling is conducted separately for each stratum.

The strata are mutually exclusive: every element in the population must be assigned to only one stratum. The strata should also be collectively exhaustive: no population element can be excluded.

Sampling strata need to be considered when constructing the variance (or confidence intervals) of population estimates. Strata is needed for the correct calculation of standard deviation for each sample design. STRATA is numeric and country specific. A unique identifier is created for each stratum.

In STATA, users are advised to specify strata through the SVYSET command. The variable is in string format with the following naming convention “code of stratum – stratum name”, for example: “1 – Dar-es-salaam”.

psu
Country-specific and a unique identifier created for each primary sampling unit.

The primary sampling unit (PSU) refers to sampling units that are selected in the first (primary) stage of multi-stage sample design. These sampling units typically correspond to several large aggregate units (clusters), each of which contains sub-units.

For example, a primary sampling unit can represent the set of all housing units contained in a well-defined geographic area, such as a municipality or a group of contiguous municipalities. Primary sampling units are numeric and country specific. A unique identifier is created for each primary sampling unit. In Stata, users are advised to specify the primary sampling unit through the SVYSET command.

subnatid1_prev
Country-specific categorical variable. This is coded as missing unless the classification used for subnatid1 has changed since the previous survey. In that case, it refers to the subnatid1 code used in the previous survey. This provides a way of tracking splits.

For example, if province “32 – West Java” split into province “32 – West Java” and “36 – Banten” since the most recent survey, this variable would contain "32 – West Java” for both provinces 32 and 36.

subnatid2_prev
Country-specific categorical variable. This is coded as missing unless the classification used for SUBNATID2 has changed since the previous survey. In that case, it refers to the subnatid2 code used in the previous survey.

subnatid3_prev
Country-specific categorical variable. This is coded as missing unless the classification used for SUBNATID3 has changed since the previous survey. In that case, it refers to the subnatid3 code used in the previous survey.

subnatid4_prev
Country-specific categorical variable. This is coded as missing unless the classification used for subnatid4 has changed since the previous survey. In that case, it refers to the subnatid4 code used in the previous survey.

gaul_adm1_code
This is numeric and country-specific based on the GAUL database. It should be taken from the same data in the GAUL database (a copy of those codes is available at the D4G team where the geographical area can be identified in the survey based on the name of the location/area. *The number of unique values from the SUBNATID1 and the GAUL_ADM1_CODE could be different or the same.*

For example, in the case of a fictional country, if the highest-level representation is the state level (53 states) and Gaul also has 53 states, it is the same in this case. In a different example, the survey is representative at the level of statistical regions (7) while the identifiable GAUL code is at state level (53 states); with this setup, one can know how the seven statistical regions are constructed.

gaul_adm2_code
This is numeric and country-specific based on the GAUL database. It should be taken from the same data in the GAUL database where the geographical area can be identified in the survey based on the name of the location/area.

urban
This is a dummy variable that specifies the location type – urban or rural - of the household. This variable is country specific as each country uses its own criterion to distinguish urban from rural areas. In many cases there is no clear division between urban and rural areas, and areas are classified as "semi-urban" or "mixed". Harmonizers are advised to classify such categories as "urban." Urban categories:

1 = Urban
0 = Rural

Table 3.1: Geography Module

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	ID	**countrycode**	**country code**	String See Table 2.1	1
2	ID	**year**	**Year**	Numeric See Table 2.1	1
3	ID	**hhid**	**Household identifier**	String See Table 2.1	1
4	ID	**weight**	**Household weights**	Numeric See Table 2.1	1
5	Geography	**subnatid1**	**Subnational ID - highest level**	String, country-specific categorical variable; numeric entries in string format using the following naming convention: "1 – Hatay" (as string)	1

6	Geography	**subnatid2**	**Subnational ID - second highest level**	String, country-specific categorical variable; numeric entries in string format using the following naming convention: “1 – Hatay” (as string)	1
7	Geography	**subnatid3**	**Subnational ID - third highest level**	String, country-specific categorical variable; numeric entries in string format using the following naming convention: “1 – Hatay” (as string)	2
8	Geography	**subnatid4**	**Subnational ID - fourth highest level**	String, country-specific categorical variable; numeric entries in string format using the following naming convention: “1 – Hatay” (as string)	2
9	Geography	**subnatidsurvey**	**Lowest level of Subnational ID**	String	1
10	Geography	**strata**	**Strata**	String	1
11	Geography	**psu**	**PSU**	Numeric, country-specific	1
12	Geography	**subnatid1_prev**	**SUBNATID previous - highest level**	String, country-specific categorical variable;	2
13	Geography	**subnatid2_prev**	**SUBNATID previous – second highest level**	String, country-specific categorical variable;	2
14	Geography	**subnatid3_prev**	**SUBNATID previous - third highest level**	String, country-specific categorical variable;	2
15	Geography	**subnatid4_prev**	**SUBNATID previous - lowest level**	String, country-specific categorical variable;	2
16	Geography	**gaul_adm1_code**	**Gaul Code**	Numeric	1
17	Geography	**gual_adm2_code**	**Gaul Code**	Numeric	1
18	Geography	**urban**	**Urban**	1 = Urban 0 = Rural	1

3.3 Challenges and Common Mistakes

- SUBNATID codes should reflect the most recent codes that pertain to that survey. SUBNATID_PREV codes can be used to track splits and new administrative units that have been introduced since the previous survey. It is important to ensure there is consistency in geographic variables across time. Sub-nationally representative units may be added in later additions of surveys, so names of subnational units must be consistent across time. This will allow analysts to make the current administrative units “backwards-compatible” with little additional effort.

- Harmonizers should ensure the subnatid1 through subnatid4 are string variables NOT categorical.

- The urban variable cannot be different from zero or one

```
urban!= 1 & urban!= 0
```

- After creating all the variables: -
    - ✓ Run basic tabulations for all categorical variables and check those are within range. If not within the allowed range, return and fix variable.

4 Demography (DEM)

4.1 Framework for Harmonization

Multiple-topic household surveys collect data on the characteristics of both households and individuals within those households. The GMD Demography module covers demographic indicators that are commonly derived from survey data or that can be constructed using existing variables. The primary objective of the indicator harmonization is to generate a unified data source of globally comparable demographic characteristics to support routine analytical and corporate business functions of the World Bank.

Since there is no commonly agreed-upon framework for demographic and migration indicators, these guidelines draw on various major projects and internationally accepted principles for the collection and production of demographic and migration statistical indicators based on household surveys. The mentioned projects include The World Bank Living Standards Measurement Study- Integrated Surveys on Agriculture (LSMS-ISA), the Demographic Health Surveys (DHS), International Income Distribution Database (I2D2) and other WB global and regional harmonization referenced in line with UN Sustainable Development Goal indicators.

Building upon the conceptual and operational foundations of this work, and in collaboration with a wide range of various relevant partners within and outside the World Bank, this chapters proposes a set of commonly collected and used indicators to assess the demographic and migration characteristics of individuals and households.

Demographic and migration statistics are essential for planning and monitoring socio-economic development programs. They are instrumental in building an evidence base of the drivers and consequences of poverty and inequality and in informing the necessary policy approaches for fostering poverty reduction and boosting shared prosperity, as well as other relevant development outcomes. Statistics on population composition by age and sex are among the most basic data necessary to describe a population and/or a subgroup of a population. The need for data decomposed by urban/rural, gender and age are clear from the indicators for the SDGs which call for breakdowns on these lines in indicators on poverty (SDG 1), education (SDG 4), gender issues (SDG 5), decent work (SDG 8), inequality (SDG 10), and urban issues (SDG 11). Conversely, the lack of adequate and internationally comparable demographic and migration statistics is a major impediment to informed and effective policies.

4.2 Mapping and Description of Variables

The demography module contains many metadata that provides a wealth of information about the variables, including their type, description, sources, etc. To improve readability, only the most significant information has been included in this section. For a complete list of all variables captured in the module please consult Table at the end of the chapter. This section aims to provide the readers with basic information about the main sources of information and then some details about how GMD demographic indicators are produced, and the issues surrounding them.

Characteristics such as race, gender, age, and marital status, are all typical examples of demographics that are used in surveys. These are fundamental building blocks for most statistical analysis. In this module, the primary unit of analysis is the level of individuals within a household.

weight
This contains household weights, typically inversely proportional to the probability of the household being selected for the sample, that should be applied to all analysis to make the results representative of the population.

language
This is a string variable that refers either to the one the respondent normally speaks in his or her present home (usual language) or the language usually spoken in the individual's home in his or her early childhood (mother tongue), or the language that the person commands best (main language). Its classification is country specific. Information on language (including any sign language) should be harmonized for all persons. In the tabulated results, the criterion for determining the language for children not yet able to speak should be clearly indicated. Numeric entries are coded in string format using the following naming convention: "2 – language".

Include in "notes" of the variable the type of question used in the harmonization, for example, usual language or mother tongue.

age
This variable must be recorded in completed years for all persons. It refers to the interval of time between the date of birth and the date of the survey. Every effort should be made to determine the precise and accurate age of each person, particularly of children[5] and older persons. Information on age may be secured either by obtaining the date (year, month, and day) of birth or by asking directly for age at the person's last birthday. Lastly, if the information on age is not available, it should be coded as missing ".k" or ".d" rather than some other value such as "99" or "999". **See Section 1.6 on missing code**.

childyr
For children under 5 years (aged less than 60 months), age in complete years should be recorded. Harmonizer check if this is consistent with age in completed years (AGE).

childmth
For those under 5 years (aged less than 60 months), age in complete months should be recorded.

agecat
This is a string variable that refers to age groups defined in the survey if information on age is only available in age categories rather than in years. For example:

"15 years or younger"
"15-24 years old"

[5] Especially in the case of children aged less than 5 years old, age is used to interpret Anthropometrics data.

"25-54 years old"
"55-64 years old"
"65 years or older"

male

This is a dummy variable that specifies the sex – male or female – of an individual within a household. While constructing this variable, it is important to make sure that all relevant values are included. Some countries are including a third option in their gender classification (transgender, gender neutral, non-binary, agender, pangender, etc.) and this should be coded as 9. Variable values coded as '98' or other numeric characters should be excluded from the values of the `male' variable. Sex of household member, two categories after harmonization:

1 = Male
0 = Female
9 = Other gender

relationharm

This is a string variable that indicates a relationship to the reference person of household (usually the head of household). Variable values coded as '98' or other numeric characters should be excluded from the values of RELATIONHARM variable. Relationship to head of household, six categories after harmonization:

1 = Head
2 = Spouse
3 = Children
4 = Parents
5 = Other relative
6 = Non-relative

Note: In cases where head is missing or a migrant, we assign spouse as the head of the household. If spouse is also not available, then we will use oldest member of the household as the head and recode all the relations to head accordingly.

relationcs

This is a country-specific categorical variable that indicates the relationship to the head of the household. The categories for relationship to the head of the household are defined according to the region or country requirements.

marital

This is a categorical variable that refers to the personal status of each individual in relation to the marriage laws or customs of the country. In some countries not all persons are asked marital status and this depends on the person's age. Harmonizer must not guestimate the marital status of those not asked.

The categories of marital status to be identified should include at least the following: (a) single (in other words, never married); (b) married; (c) married but separated; (d) widowed and remarried; (e) divorced and not remarried. In some countries, category (b) may require a subcategory of persons who are

contractually married but not yet living as man and wife. In all countries, category (c) should comprise both the legally and the de facto separated, who may be shown as separate subcategories if desired.

The marital variable should not be imputed but rather calculated only for those to whom the question was asked (in other words, the youngest age at which information is collected may differ depending on the survey). The consistency between age and marital needs to be cross-checked. In most countries, there are also likely to be persons who were permitted to marry below the legal minimum age because of special circumstances. To permit international comparisons of data on marital status, however, any tabulations of marital status not cross-classified by exact age should at least distinguish between persons under 15 years of age and over. If it is not possible to distinguish between married and living together, then it should be assumed that the individual is married. Variable values coded as '98' or other numeric characters should be excluded from the values of the 'marital' variable. Marital status, five categories after harmonization:

1 = Married
2 = Never married
3 = Living together
4 = Divorced/separated
5 = Widowed

literacy

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 definition. Be careful while coding 1; one must be able to both read and write. If a person can either read or write, he/she will be considered illiterate (LITERACY=0). Accepted codes: -

1 = Yes
0 = No

Note: <u>It can be assumed with some degree of accuracy that if respondent has secondary level and above of education, then must be literate.</u>

everattend

Ask to respondents if they have ever attended school. The length of attendance is not important but at some point, member attended school.

1 = Yes
0 = No

mineducatage

Minimum age for which education section is applied in country. For this reason, the lower age cut-off at which information is collected will vary from country to country. The questionnaire and/or manual specifies this. If unknown leave as missing or ask Supervisor or Country Poverty Economist.

Note: <u>The subsequent education variables (EDUCAT7, EDUCAT5, EDUCAT4, PRIMARYCOMP), harmonization will be for those equal or above MINEDUCATAGE.</u>

educat7

Do not try to guestimate education levels for population but leave it as MISSING. This question can be captured in two ways (a) country categories or (b) ISCED classification.

1 = No education
2 = Primary incomplete
3 = Primary complete
4 = Secondary incomplete
5 = Secondary complete
6 = Higher than secondary but not university
7 = University incomplete or complete

- *Primary complete* implies that one completed the stipulated primary education by undertaking some form of assessment.
- *Secondary complete* implies that one completed the stipulated secondary education by undertaking some form of assessment.
- *Higher than secondary but not university* refers to any higher education after successfully completing secondary level of education such as higher professional schooling, college, etc.
- *University and higher education level* refer to undergraduate and higher.

This classification is based on the UNESCO ISCED 2011 education categories. Check this link for country ISCED mappings http://uis.unesco.org/en/isced-mappings.

See Annex II for detailed groupings "ISCED-2011 education groupings.xlsx". If a country uses ISCED classification, see below: -

- Incomplete secondary education: corresponds to incomplete grades five to twelve or:
  - ISCED-A Level 2
- General Secondary: corresponds to ISCED-11
  - ISCED-A Level 3 Category 34 (Upper secondary general education)
- Specialized Secondary: corresponds to ISCED-11:
  - ISCED-A Level 3 Category 35
  - ISCED-A Level 4 Category 44 and 45
- Tertiary: corresponds to ISCED-11:
  - ISCED-A Levels 5, 6, 7, 8 (all categories)
- Incomplete primary: corresponds to ISCED-11
  - ISCED-A Level 0 Category 03.
- Complete primary: corresponds to ISCED-11
  - ISCED-A Level 1 Category 10
- None corresponds to
  - ISCED-A Level 0 Category 01-02

Note:

- If Harmonizer cannot create all the 7 groups, leave as missing.
- Value must be missing for individuals less than the required age (MINEDUCATAGE).

- In some surveys, incomplete primary and no education will be difficult to distinguish. Please liaise with TTL for what to do in such instances and record in the dofile this issue.
- In several ECA surveys education is specified for persons 15+ probably due to compulsory education system. This means that education levels for children under 16 will all missing. Do not attempt to guestimate education levels. Consult TTL on what to do,
- If the survey has option categories not included in ISCED 2011 or country categories, use special missing option. Document in the notes the variable and with the country-specific labels.

educat5

Do not try to guestimate education levels for population but leave it as MISSING. EDUCAT5 can be derived from EDUCAT7 is derived. However, if EDUCAT7 is missing, see EDUCAT7 definitions to derive this.

1 = No education
2 = Primary incomplete
3 = Primary complete but secondary incomplete
4 = Secondary complete
5 = Some tertiary/post-secondary

Note:

- If Harmonizer cannot create all the 5 groups, leave as missing.
- Value must be missing for individuals less than the required age (MINEDUCATAGE).
- In some surveys, incomplete primary and no education will be difficult to distinguish. Please liaise with TTL for what to do in such instances.
- In several ECA surveys education is specified for persons 15+ probably due to compulsory education system. This means that education levels for children under 16 will all missing. Do not attempt to guestimate education levels. Consult TTL on what to do.

educat4

Do not try to guestimate education levels for population but leave it as MISSING. This variable must be recoded if EDUCAT7 and EDUCAT5 are missing.

1 = No education
2 = Primary (complete or incomplete)
3 = Secondary (complete or incomplete)
4 = Tertiary (complete or incomplete)

Note:

- At the bare minimum, this variable must be generated. If unclear, contact TTL.
- Value must be missing for individuals less than the required age (MINEDUCATAGE).
- In some surveys, incomplete primary and no education will be difficult to distinguish. Please liaise with TTL for what to do in such instances and record in the dofile this issue.
- In several ECA surveys education is specified for persons 15+ probably due to compulsory education system. This means that education levels for children under 16 will all missing. Do not attempt to guestimate education levels. Consult TTL on what to do.

educy

If grade level not listed, leave EDUCY as missing and do not guestimate by using age and education level. For individuals who are currently enrolled in school, their years of education completed correspond to the class currently attending minus one. 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 years 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 mappings.

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: -

- 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.
- 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.

The variable does not consider the actual number of years spent to reach a grade level. In other words, first grade repeated three times only counts as 1 year of completed education.

primarycomp

Record at least primary completion for every individual in household; from EDUCAT5 or EDUCAT7. If not code accordingly.

1 = Yes
0 = No

school

There is no age cut-off. But do not guestimate but use the country-specific age cut-off.

1 = Yes
0 = No

Note: <u>If asked for persons older than a certain age, do not guestimate enrolment for the younger population. Leave as missing.</u>

eye_dsablty

This is a numerical variable that indicates whether an individual has any difficulty in seeing, even when wearing glasses. Categories after harmonization:

1 = No – no difficulty

2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

hear_dsablty
This is a numerical variable that indicates whether an individual has any difficulty in hearing even when using a hearing aid. Categories after harmonization:

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

walk_dsablty
This is a numerical variable that indicates whether an individual has any difficulty in walking or climbing steps. Categories after harmonization:

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

conc_dsord
This is a numerical variable that indicates whether an individual has any difficulty concentrating or remembering. Categories after harmonization:

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

slfcre_dsablty
This is a numerical variable that indicates whether an individual has any difficulty with self-care such as washing all over or dressing. Categories after harmonization:

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

comm_dsablty
This is a numerical variable that indicates whether an individual has any difficulty communicating or understanding usual (customary) language. Categories after harmonization:

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

Table 4.1: Demography Module

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	ID	**countrycode**	**country code**	String See Table 2.1	1
2	ID	**year**	**Year**	Numeric See Table 2.1	1
3	ID	**hhid**	**Household identifier**	String See Table 2.1	1
4	ID	**pid**	**Personal identifier**	String or numeric	1
5	ID	**weight**	**Weight**	Numeric See Table 2.1	1
6	Demography	**language**	**Language**	String, country-specific categorical variable	2
7	Demography	**age**	**Age of individual (continuous)**	Numeric, continuous	1
8	Demography	**agecat**	**Age of individual (categorical)**	String, country-specific categorical variable	1
9	Demography	**childyr**	**Child age in years for those under 5**	Numeric, continuous	1
10	Demography	**childmth**	**Child age in months those under 5**	Numeric, continuous	1
11	Demography	**male**	**Sex of household member**	1 = Male 0 = Female	1
12	Demography	**relationcs**	**Relationship to head of household country/region specific**	String, country-specific categorical variable	1
13	Demography	**relationharm**	**Relationship to head of household harmonized across all regions**	1 = Head 2 = Spouse 3 = Children 4 = Parent 5 = Other relative 6 = Non-relative	1
14	Demography	**marital**	**Marital status**	1 = Married 2 = Never married 3 = Living together 4 = Divorced/separated 5 = Widowed	1
15	Education	**literacy**	**Can both read and write**	1 = Yes 0 = No	1
16	Education	**everattend**	**Ever attended school**	1 = Yes 0 = No	1
17	Education	**mineducatage**	**Education module application age (country specific)**	Numeric	1

18	Education	**educat7**	**Highest level of education completed (7 categories)**	1 = No education 2 = Primary incomplete 3 = Primary complete 4 = Secondary incomplete 5 = Secondary complete 6 = Higher than secondary but not university 7 = University incomplete or complete	1
19	Education	**educat5**	**Highest level of education completed (5 categories)**	1 = No education 2 = Primary incomplete 3 = Primary complete but secondary incomplete 4 = Secondary complete 5 = Some tertiary/post-secondary	1
20	Education	**educat4**	**Highest level of education completed (4 categories)**	1 = Incomplete secondary 2 = General Secondary 3 = Special Secondary 4 = Tertiary 5 = None or incomplete primary	1
21	Education	**educy**	**Years of education**	Numeric	
22	Education	**primarycomp**	**Primary school completion**	1 = Yes 0 = No	1
23	Education	**school**	**Currently enrolled in school**	1 = Yes 0 = No	1
24	Disability	**eye_dsablty**	**Eye disability**	1 = No – no difficulty 2 = Yes – some difficulty 3 = Yes – a lot of difficulty 4 = Cannot do at all	1
25	Disability	**hear_dsablty**	**Hearing disability**	1 = No – no difficulty 2 = Yes – some difficulty 3 = Yes – a lot of difficulty 4 = Cannot do at all	1
26	Disability	**walk_dsablty**	**Walk disability**	1 = No – no difficulty 2 = Yes – some difficulty 3 = Yes – a lot of difficulty 4 = Cannot do at all	1
27	Disability	**conc_dsord**	**Concentration disorder**	1 = No – no difficulty 2 = Yes – some difficulty 3 = Yes – a lot of difficulty 4 = Cannot do at all	1

28	Disability	**slfcre_dsablty**	**Selfcare disability**	1 = No – no difficulty 2 = Yes – some difficulty 3 = Yes – a lot of difficulty 4 = Cannot do at all	1
29	Disability	**comm_dsablty**	**Communication disability**	1 = No – no difficulty 2 = Yes – some difficulty 3 = Yes – a lot of difficulty 4 = Cannot do at all	1

Yellow is/are the new variables in GMD3.0.

4.3 Challenges and Common Mistakes

Data sets that are harmonized incorrectly can lead to skewed and/or incorrect data analysis. Harmonizers should run a series of checks to ensure data is harmonized properly.

- Check for duplicates of PID within HHID. If duplicates fix by renumbering but be careful if any merges will be needed to be done later.

```
duplicates tag (hhid pid),gen(dup)
tab dup
```

- WEIGHT cannot be missing.

```
weight==.
```

- Check to make sure that AGE is an integer for persons aged 5 and over.

```
age/int(age)!= 1 & !mi(age) & age > 5
```

- AGE cannot have negative or extreme values (>120)

```
(age < 0 | age>120) & age<.
```

- AGE cannot be missing.

```
mi(age)
```

- MALE variable can only take one of two values 0 or 1 (or missing).

```
!mi(male) & !inlist(male,0,1)
```

- Check if MALE is missing.

```
male==.
mi(male)
```

- Check to make sure that there is variation in MALE.

```
egen sdmale = sd(male)      sdmale==0
```

- Check if head and spouse are same sex. This will depend on country context.

- Check for if RELATIONHARM=1 is unique and not missing. Only one head is allowed per household.
  Step 1: Every household must have one head.

```
gen head=1  if ctryvarname==1   //ctryvarname that identifies head
bys hhid: egen heads=total(head)
ta heads
```

  Step 2: Missing heads (a) assign oldest as head (b) if not collected PID=1 becomes head.

```
(a)
gsort hhid -age
bys hhid: gen countdup=_n
replace relathh=1   if heads==0 & countdup==1

(b)
replace relathh6=1  if pid==1
```

- RELATIONHARM must be an integer in the range [1,6].

```
!inlist(relationharm,1,2,3,4,5,6)
```

- Check consistency on married unions.
  - Be cognizant that a blanket replacement is not ideal as the relationship to head may have been an error. That is, the head and spouse are not actually that but a different relationship. If they are few cases, please peruse case by case and fix. Please liaise with the TTL and reflect this in the dofile.

- MARITAL must be an integer in the range [1,5].

```
!inlist(marital,1,2,3,4,5)
```

- Children are "Never married" and should be coded as so even though it may be perceived as obvious. The marital status of individuals should be harmonized for all individuals. Harmonizers should check to make sure children are not systematically left with missing values for marital.

```
tab age marital, missing
```

- If there are more than 1 percent of missing values based on MINEDUCATAGE, please report to the TTL for further instructions.

```
tab educat7,m   if age>ed_mod_age
```

- Consistency checks between education and AGE should be done

```
tab age educat7,m   if age>ed_mod_age
```

- If any outlier is detected edit accordingly. For example, child aged 12 cannot be in university; or a 5-year completed secondary. Small fixes can be done based on country education structure for primary-level or secondary-level school ages but if unable leave as missing. ISCED country-level specifics can be used to guestimate for this population.

- EDUCAT5 can be programmed from EDUCAT7 if available.

```
recode educat7 (3 4=3) (5=4) (6 7=5), gen(educat5)
tab educat7 educat5,m
tab age educat5
```

- EDUCAT4 can be programmed from EDUCAT7 if available.

```
recode educat7 (2 3=2) (4 5=3) (6 7=4), gen(educat4)
tab educat7 educat4
tab age educat4
```

- PRIMARYCOMP: One can assume with a degree of certainty certain conditions qualify primary-school completion.

```
gen primarycomp=inrange(educat7,3,7) if !mi(educat7)
ta educat7
or
gen primarycomp=inrange(educat5,3,5) if !mi(educat5)
gen primarycomp=1       if (educat5>=3 & educat5<=5)
replace primarycomp=0   if primarycomp==. & !mi(educat5)
```

- Additionally, harmonizers should ensure that the household size variable is calculated correctly. Not all the individuals reported in a household that form the raw data are current household members. For example, for the EU-SILC survey, a household contains the current member, but also the members of the previous survey who have left the household for reasons such as death or migration.

Run the dofile ***MODULE_04_Demography.do*** which labels (variable and value) and order variables

- ✓ This is a must process with no exception.

5 Labor Module (LBR)

5.1 Framework of Harmonization

The GMD labor module contains a variety of variables relating to individuals' labor status and jobs. This includes their primary activity (employed, unemployed, or out of the labor force), as well as types, sectors of employment and wages of workers. This set of variables is identical to the labor variables coded in the International Income Distribution Database (I2D2). The primary objective of this module is to generate a unified data source of globally comparable labor indicators that can be easily linked to poverty and welfare indicators and support routine analytical and corporate business functions of the World Bank.

In this module, the primary unit of analysis is at the individual level. The age at which the labor module starts being applied, or the legal working age, is different for each country. Hence, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

The data harmonized in this module is critical for understanding progress on SDG indicators. SDG 8 calls for data on the proportion of informal employment in non-agriculture sectors by sex (8.3), the average hourly earnings of female and male employees by occupation and age (8.5, partially), the unemployment rate by sex and age (8.5, partially), and the proportion and number of children aged 5-17 years engaged in child labor by sex and age (8.7). The data for these indicators can be linked to poverty, welfare, and other GMD variables through this module.

The module contains variables that encompass two different time units: the last seven days and the past twelve months. Considering that labor market dynamics differ among regions and countries, the idea of having two different time units is to be able to get a better characterization of the labor market at a country level. Developing countries, for example, usually have a substantial share of their labor force dedicated to agricultural activities. This type of labor activity is characterized by an important seasonal component, meaning that labor participation and the economic status of individuals varies significantly throughout the year. Therefore, it is probable that labor variables regarding last seven days will be affected by this seasonality, in contrast with annual variables, which are less prone to be sensitive to the seasonal component but may be more prone to recall error.

The last seven days variables include the employment status (paid employee, employer, self-employed, etc.), sector (public or private), industry and occupation classification as well as other important job's characteristics. GMD also contains variables for contract, social security, health insurance, and union membership, which are proxy measures of the extent to which a worker's job is formal. The last 12 months variables include employment status and number of jobs.

Consistent with the above, in many developing regions individuals usually have more than one job. The GMD labor module also contains variables regarding the second job both in the last seven days and during the past year. The second job variables include employment status, sector, industry, and occupation classification.

ILO Resolution I

Labor variables are provided only for working-age population, which might be different for different countries. Furthermore, even though most countries have an established retiring age, GMD files do not consider an upper bounding age for the working population, which is line with ILO recommendations.[6] In this way, working age population is defined as those individuals whose age is greater than or equal to the minimum legal age to work, defined by each country. Besides this, many surveys (particularly in developing regions) also include a child labor module; this information is not provided in GMD files, which only include labor outcomes for those deemed old enough to be included in each country's labor module.

In 2013, The ILO adopted a new definition of employment, which reclassifies own-production and many forms of unpaid labor as not working.[7] As of 2019, these guidelines have yet to be widely adopted. Because most labor modules do not contain the required information to follow these guidelines, and to maintain comparability with older surveys, this module maintains the traditional pre-2013 definitions of employment.

5.2 Mapping and Description of Variables

The labour module contains a large number of metadata that provides a wealth of information about variables, including their types, descriptions, sources, etc. There are eleven sections based on a 7-day and 12-month recall.

Recall period	Section	Comment
7-day	5.2.1; 5.2.2; 5.2.3; 5.2.4; 5.2.5	• Leave missing if not asked
12-month	5.2.6; 5.2.7; 5.2.8; 5.2.9; 5.2.10	• Leave missing if not asked • These sections regardless of whether they responded to questions with a reference period of the last 7 days.

5.2.1 Labor status, 7-day reference period

All variables are numeric unless specified.

minlaborage
This is the lowest age for which the labor module is implemented in the survey or the minimum working age in the country. For this reason, the lower age cutoff at which information is collected will vary from country to country.

[6] Resolution Concerning statistics of work, employment and labor underutilization http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_230304.pdf

[7] ILO. 2013. Statistics and Databases. See here: https://www.ilo.org/global/statistics-and-databases/statistics-overview-and-topics/WCMS_470295/lang--en/index.htm

lstatus

This is an individual's ***labor status in the last 7 days***. The Value must be missing for individuals less than the required age (MINLABORAGE). Three categories are used after harmonization:

1 = Employed
2 = Unemployed
3 = Not-in-labor force

- *Employed*: 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.
- *Unemployed*: 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 for the start of a new job is considered unemployed.
- *Not-in-labor force*: A person is defined as not-in-labor force if he or she is, presently not working and it is not actively seeking a job during the last 7 days or reference week.

Note: <u>All persons are considered active in the labor force if they presently have a job (formal or informal, i.e., employed) or do not have a job but are actively seeking work (i.e., unemployed).</u>

nlfreason

This is the reason an individual was not in the labor force ***in the last 7 days***. This variable is constructed for all those who are not presently employed and are not looking for work (LSTATUS=3) and missing otherwise. Five categories after harmonization:

1 = Student
2 = Housewife
3 = Retired
4 = Disabled
5 = Other

- *Student*: A person currently studying.
- *Housewife*: A person who takes care of the house, older people, or children.
- *Disabled*: A person who cannot work due to physical conditions.
- *Other*: A person does not work for any other reason.

unempldur_l

This is a continuous variable specifying the ***duration of unemployment in months*** (***lower bracket***).

The variable is constructed for all unemployed persons (LSTATUS=2, otherwise missing). If it is specified as continuous in the survey, it records the numbers of months in unemployment. If the variable is categorical, it records the lower boundary of the bracket.

Missing values are allowed for everyone who is not unemployed. Other missing values are also allowed.

unempldur_u
This is a continuous variable specifying the ***duration of unemployment in months (upper bracket)***.

The variable is constructed for all unemployed persons (LSTATUS=2, otherwise missing). If it is specified as continuous in the survey, it 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.

Missing values are allowed for everyone who is not unemployed. Other missing values are also allowed.
If the duration of unemployment is not reported as a range, but as continuous variables, the UNEMPLDUR_L and UNEMPLDUR_U variables will have the same value. If the high range is open-ended the UNEMPLDUR_U variable will be missing.

Table 5.1: Labor status, 7-day reference period

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	ID	**countrycode**	**country code**	String See Table 2.1	1
2	ID	**year**	**Year**	Numeric See Table 2.1	1
3	ID	**hhid**	**Household identifier**	String See Table 2.1	1
4	ID	**pid**	**Personal identifier**	String or numeric See Table 2.1	1
5	ID	**weight**	**Weight**	Numeric See Table 2.1	1
6	Labor	**minlaborage**	**Labor module application age (country-specific)**	Numeric	1
7	Labor	**lstatus**	**Labor status (7-day ref period)**	1 = Employed 2 = Unemployed 3 = Not-in-labor force	1
9	Labor	**nlfreason**	**Reason not in the labor force (7-day ref period)**	1 = Student 2 = Housewife 3 = Retired 4 = Disabled 5 = Other	1
10	Labor	**unempldur_l**	**Unemployment duration (months) lower bracket (7-day ref period)**	Continuous variable	1
11	Labor	**unempldur_u**	**Unemployment duration (months) upper bracket (7-day ref period)**	Continuous variable	1

Note: <u>All the subsequent harmonization will be for those equal or above MINLABORAGE.</u>

5.2.2 Primary Employment, 7-day reference period

empstat

This is a categorical variable that specifies the ***main employment status in the last 7 days*** of any individual with a job (LSTATUS=1) and is missing otherwise. The variable is constructed for all individuals. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

The definitions are taken from the International Labor Organization's Classification of Status in Employment with some revisions to consider the data available. Five categories after harmonization:

1 = Paid Employee
2 = Non-Paid Employee
3 = Employer
4 = Self-employed
5 = Other, workers not classifiable by status

- *Paid Employee*: 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 due to country specificity.
- *Non-Paid Employee*: Non-paid employee includes contributing family 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.
    - All apprentices should be mapped as 'non-paid employee'
- *Employer*: An 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 self-employed.
- *Self-employed*: Own account or self-employment includes jobs where remuneration is directly dependent from the goods and service produced (where home consumption is part of the profits) and where one has not engaged any permanent employees to work for them on a continuous basis during the reference period.
    - 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.
- *Other, workers not classifiable by status*: This include those for whom insufficient relevant information is available and/or who cannot be included in any of the above categories.

ocusec

This is a categorical variable that specifies the ***sector of activity in the last 7 days***. It classifies the main job's sector of activity of any individual with a job (LSTATUS=1) and is missing otherwise. The variable is constructed for all persons administered this module in each questionnaire.

Four categories after harmonization:

1 = Public sector, Central Government, Army (including armed forces)
2 = Private, NGO
3 = State-owned
4 = Public or State-owned, but cannot distinguish

- *Public Sector, Central Government, Army (including armed forces)*: Public sector is the part of economy run by the government.
- *Private, NGO*: Private sector 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
  - NGO is a nonprofit organization that operates independently of any government whose purpose is to address a social or political issues.
- *State-owned enterprises*: State-owned includes para-state firms and all others in which the government has control (participation over 50%).
- *Public or State-owned but cannot distinguish*: Select this option is the questionnaire does not ask for State-owned enterprises, and only for public sector.

Note: Do not code basis of occupation (ISCO) or industry (ISIC) codes.

industry_orig
This is a string variable that specifies the ***original industry codes in the last 7 days for the main job*** provided in the survey (the actual question) and should correspond to whatever is in the original file with no recoding.

The variable is constructed for all persons administered this module in each questionnaire. It will contain missing values for people below the working age. Other missing values are allowed. It classifies the main job of any individual with a job (LSTATUS=1) and is missing otherwise.

Code and name: Example: "1 - Agriculture"; "2 - Fishing"; "3 - Construction"; etc.

If classification not ISIC, labels must be translated to English. Make sure translation is correct from a language expert.

If the variables are coded exactly as the ISIC classification, depending on the ISIC digit level (2-, 3-. 4-level) or NACE (2-, 3-, 4-level), one can label the variables by running the dofiles provided that are already in English. **See Section 5.4 labeling dofiles.**

Note: This will only apply if the country uses the same ISIC or NACE classifications.

industrycat10
This is a categorical variable that specifies the 1-digit industry classification ***in the last 7 days for the main job*** of any individual with a job (LSTATUS=1) and is missing otherwise.

The variable is constructed for all persons administered this module in each questionnaire. The codes for the main job are given here based on the UN International Standard Industrial Classification (ISIC) (revision 3.1/4.0)[8]. Ten categories after harmonization:

1 = Agriculture, Hunting, Fishing, etc.
2 = Mining
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

Notes:

- Can be recoded from INDUSTRY_ORIG.
- When creating the industry variable, one needs to carefully check the type of economic activity codes and its revision which the household survey uses. Some surveys follow ISIC (International Standard Industrial Classification), while others follow NACE (Statistical Classification of Economic Activities in the European Community).
- In the case of different classifications (former Soviet Union republics, for example), recoding has been done to best match the ISIC codes.
- Category 10 is also assigned for unspecified categories or items.
- See Annex III.1 for ISIC details and how to map the classifications.
- See Annex III.2 for NACE details and how to map the classifications.

If all 10 categories cannot be identified in the questionnaire create this variable as missing and proceed to create INDUSTRYCAT4.

industrycat4

This is a categorical variable that specifies the 1-digit ***industry classification in the last 7 days for the main job*** for Broad Economic Activities. This variable is either created directly from the data (if industry classification does not exist for ten categories) or created from industrycat10. Four categories after harmonization:

1 = Agriculture
2= Industry
3 = Services
4 = Other

This variable is either created directly from the data (if industry classification does not exist for ten categories) or created from INDUSTRYCAT10.

[8] https://unstats.un.org/unsd/classifications/Econ/ISIC#ISIC3

occup_orig
This is a string variable that specifies the ***original occupation code in the last 7 days for the main job***. This variable corresponds to whatever is in the original file with no recoding.

For each value label, there should be a space between the hyphen (before and after). If value label is truncated, make sure it is written in full. For example, pharm., law., etc are not allowed.

Code and name: Example: "1 - Pharmacist"; "2 - Engineer"; "3 - Lawyer"; etc.

If classification is not ISCO, labels must be translated to English. Make sure translation is correct from a language expert.

> Depending on the ISCO digit level (2-, 3-. 4-level), one can label the variables by running the dofiles provided that are already in English. **See Section 5.4 labeling dofiles.**

Note: This will only apply if the country uses the same ISCO classifications.

occup
This is a categorical variable that specifies the 1-digit ***occupation classification for the main job in the last 7 days*** of any individual with a job (LSTATUS=1) and is missing otherwise. This 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.

Most surveys collect detailed information and then code it, without keeping the original data, no attempt has been made to correct or check the original coding.

The classification is based on the International Standard Classification of Occupations (ISCO-08). The main categories subsume the following codes. Eleven categories after harmonization:

1 = Managers
2 = Professionals
3 = Technicians and associate professionals
4 = Clerical support workers
5 = Service and sales workers
6 = Skilled agricultural, forestry and fishery workers
7 = Craft and related trades workers
8 = Plant and machine operators, and assemblers
9 = Elementary occupations
10 = Armed forces occupations
99 = Other/unspecified

Note: See Annex III.3 for ISCO details and how to map the country classifications.

wage_nc
This is a continuous variable that specifies the ***last wage payment in local currency*** of any individual (LSTATUS=1 & EMPSTAT=1) in its primary occupation at the reference period reported in the survey and it is missing

otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the week preceding the survey.

This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments. The 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.

Notes:

- 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.
- Non-paid employees (EMPSTAT=2) should have WAGE=0.
- The reference period of the WAGE_NC will be recorded in the UNITWAGE variable.

unitwage

This is a categorical variable that specifies the time reference for the WAGE_NC variable. It specifies the time unit measurement for the wages of any individual (LSTATUS=1 & EMPSTAT=1) and it is missing otherwise. Acceptable values include:

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

whours

This is a continuous variable that specifies the hours of work last week for the main job of any individual with a job (LSTATUS=1) and is missing otherwise. The main job defined as that occupation that the person dedicated more time to over the past week. The variable is constructed for all persons administered this module in each questionnaire.

Note:

- If the respondent was absent from the job in the week preceding the survey due to holidays, vacation, or sick leave, then record the time worked in the previous 7 days that the person worked.
- Sometimes the questions are phrased as, “on average, how many hours 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 4.3 to get weekly hours.

wmonths

This is a continuous variable that specifies the number of months worked in the last 12 months for the main job of any individual with a job (LSTATUS=1) and is missing otherwise. The main job is defined as that occupation that the person dedicated more time to over the past week. The variable is constructed for all persons administered this module in each questionnaire.

wage_total

This is a continuous variable that specifies the ***annualized wage payment*** (regular wage plus bonuses, in-kind, compensation, etc.) for the primary occupation in local currency of any individual (LSTATUS=1 & EMPSTAT=1) and is missing otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the week preceding the survey. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments.

WAGE_TOTAL should be equal to WAGE_NC in case there are no bonuses, tips etc. offered as part of the job.

The variable is constructed for all persons administered this module in each questionnaire. The annualization of the WAGE_TOTAL should consider the number of months/weeks the persons have been working and receiving this income. Harmonizer should not assume the person has been working the whole year.

```
gen double wage_total=.
replace wage_total=(wage_nc*5*4.3)*wmonths if unitwage==1
  //Wage in daily unit
replace wage_total=(wage_nc*4.3)*wmonths if unitwage==2
  //Wage in weekly unit
replace xx=(wage_nc*2.15)*wmonths if unitwage==3
  //Wage in every two weeks unit
replace wage_total=(wage_nc)/2*wmonths if unitwage==4
  //Wage in every two months unit
replace wage_total=(wage_nc)*wmonths if unitwage==5
  //Wage in monthly unit
replace wage_total=(wage_nc)/3*wmonths if unitwage==6
 //Wage in every quarterly unit
replace wage_total=(wage_nc)/6*wmonths if unitwage==7
  //Wage in every six months unit
replace wage_total= wage_nc/12*wmonths if unitwage==8
 //Wage in annual unit
replace wage_total=(wage_nc*whours*4.3)*wmonths if unitwage==9
  //Wage in hourly unit
```

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors**.**

contract

This is a dummy variable that classifies the contract status (YES/NO) of any individual with a job (LSTATUS=1) and is missing otherwise. It indicates whether a person has a signed (formal) contract, regardless of duration. Two categories after harmonization:

0 = No
1 = Yes

Note: This variable is only constructed if there is an explicit question about contract between employee and employer.

healthins
This is a dummy variable that classifies the health insurance status (YES/NO) of any individual with a job (LSTATUS=1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. Two categories after harmonization:

0 = No
1 = Yes

Note: This variable is only constructed if there is an explicit question about health insurance provided by the job.

socialsec
This is a dummy variable that classifies the social security status (YES/NO) of any individual with a job (LSTATUS=1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. Two categories after harmonization:

0 = No
1 = Yes

Note: This variable is only constructed if there is an explicit question about pension plans or social security.

union
This is a dummy variable that classifies the union membership status (YES/NO) of any individual with a job (LSTATUS=1) and is missing otherwise. 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. Two categories after harmonization:

0 = No
1 = Yes

Note: This variable is only constructed if there is an explicit question about trade unions.

firmsize_l
This specifies the lower bracket of the firm size. The variable is constructed for all persons who are employed ***in the last 7 days for the main job***. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the lower boundary of the bracket.

firmsize_u
This specifies the upper bracket of the firm size. The variable is constructed for all persons who are employed ***in the last 7 days for the main job***. If it is continuous, it 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, this variable should be missing.

Table 5.2: Primary Employment, 7-day reference period

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Labor	**empstat**	**Employment status, primary job (7-day ref period)**	1 = Paid Employee 2 = Non-Paid Employee 3 = Employer 4 = Self-employed 5 = Other, workers not classifiable by status	1
2	Labor	**ocusec**	**Sector of activity, primary job (7-day ref period)**	1 = Public sector, Central Government, Army 2 = Private, NGO 3 = State owned 4 = Public or State-owned, but cannot distinguish	1
3	Labor	**industry_orig**	**Original industry code, primary job (7-day ref period)**	Country specific	1
4	Labor	**industrycat10**	**1 digit industry classification, primary job (7-day ref period)**	1 = Agriculture, Hunting, Fishing, etc. 2 = Mining 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	1
5	Labor	**industrycat4**	**4-category industry classification, primary job (7-day ref period)**	1 = Agriculture 2 = Industry 3 = Services 4 = Other	1
6	Labor	**occup_orig**	**Original occupation classification, primary job (7-day ref period)**	Country specific	1
7	Labor	**occup**	**1-digit occupation classification, primary job (7-day ref period)**	1 = Managers 2 = Professionals 3 = Technicians and associate professionals 4 = Clerical support workers 5 = Service and sales workers 6 = Skilled agricultural, forestry and fishery workers 7 = Craft and related trades workers 8 = Plant and machine operators, and assemblers	1

				9 = Elementary occupations 10 = Armed forces occupations 99 = Other/unspecified	
8	Labor	**wage_no_compen**	**Last wage payment, primary job, excl. bonuses, etc. (7-day ref period)**	Continuous variable	1
9	Labor	**unitwage**	**Time unit of last wages payment, primary job (7-day ref period)**	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	1
10	Labor	**whours**	**Hours of work in last week, primary job (7-day ref period)**	Continuous variable	1
11	Labor	**wmonths**	**Months worked in the last 12 months, primary job (7-day ref period)**	Continuous variable	1
12	Labor	**wage_total**	**Annualized total wage, primary job (7-day ref period)**	Continuous variable	1
13	Labor	**contract**	**Contract (7-day ref period)**	0 = No 1 = Yes	1
14	Labor	**healthins**	**Health insurance (7-day ref period)**	0 = No 1 = Yes	1
15	Labor	**socialsec**	**Social security (7-day ref period)**	0 = No 1 = Yes	1
16	Labor	**union**	**Union membership (7-day ref period)**	0 = No 1 = Yes	1
17	Labor	**firmsize_l**	**Firm size (lower bracket), primary job (7-day ref period)**	Continuous variable	1
18	Labor	**firmsize_u**	**Firm size (upper bracket), primary job (7-day ref period)**	Continuous variable	1

5.2.3 Secondary Employment, 7-day reference period

empstat_2
This is a categorical variable that specifies employment status of the secondary job ***with reference period of last 7 days*** of any individual with a job (LSTATUSs=1) and is missing otherwise. The variable is constructed for all individuals. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

The definitions are taken from the International Labor Organization's Classification of Status in Employment with some revisions to consider the data available.
Five categories after harmonization:

1 = Paid Employee
2 = Non-Paid Employee
3 = Employer
4 = Self-employed
5 = Other workers not classifiable by status

See EMPSTAT in Section 5.2.2 for definitions.

ocusec_2
This is a categorical variable that specifies the **sector of activity in the last 7 days**. It classifies the secondary job's sector of activity of any individual with a job (LSTATUS=1) and is missing otherwise. The variable is constructed for all persons administered this module in each questionnaire. Four categories after harmonization:

1 = Public sector, Central Government, Army (including armed forces)
2 = Private, NGO
3 = State-owned
4 = Public or State-owned, but cannot distinguish

Note: Do not code on the basis of occupation (ISCO) or industry (ISIC) codes.

See OCUSEC in Section 5.2.2 for definitions.

industry_orig_2
This is a string variable that specifies the original industry codes for the second job ***with reference period of the last 7 days*** and should correspond to whatever is in the original file with no recoding. Do not put missing values for people below the working age. Other missing values are allowed. It classifies the main job of any individual with a job (LSTATUS=1) and is missing otherwise.

See INDUSTRY_ORIG in Section 5.2.2 for how to label files.

industrycat10_2
This is a categorical variable that specifies the 1-digit industry classification that classifies the second job ***with reference period of the last 7 days*** of any individual with a job (LSTATUS=1) and is missing otherwise. The variable is constructed for all persons administered this module in each questionnaire. The codes for the second job are given here based on the UN International Standard Industrial Classification. Ten categories after harmonization:

1 = Agriculture, Hunting, Fishing, etc.
2 = Mining
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

Note:

- In the case of different classifications (former Soviet Union republics, for example), recoding has been done to best match the ISIC codes.
- Category 10 is also assigned for unspecified categories or items.

See INDUSTRYCAT10 in Section 5.2.2 for definitions.

If all 10 categories cannot be identified in the questionnaire create this variable as missing and proceed to create INDUSTRYCAT4.

industrycat4_2
This is a categorical variable that specifies the 1-digit industry classification for Broad Economic Activities for the second job ***with reference period of the last 7 days***. This variable is either created directly from the data (if industry classification does not exist for 10 categories) or created from INDUSTRYCAT10_2. Four categories after harmonization:

1 = Agriculture
2 = Industry
3 = Services
4 = Other

This variable is either created directly from the data (if industry classification does not exist for 10 categories) or created from INDUSTRYCAT10.

occup_orig_2
This is a string variable that specifies the ***original occupation code in the last 7 days for the secondary job***. This variable corresponds to whatever is in the original file with no recoding.

occup_2
This is a categorical variable that specifies the 1-digit occupation classification. It classifies the second job of any individual with a job (LSTATUS=1) and is missing otherwise. This variable is constructed for all persons administered this module in each questionnaire.

Most surveys collect detailed information and then code it, without keeping the original data. No attempt has been made to correct or check the original coding.

The classification is based on the International Standard Classification of Occupations (ISCO). In the case of different classifications, re-coding has been done to best match the ISCO. Eleven categories after harmonization:

1 = Managers
2 = Professionals
3 = Technicians and associate professionals
4 = Clerical support workers
5 = Service and sales workers
6 = Skilled agricultural, forestry and fishery workers
7 = Craft and related trades workers
8 = Plant and machine operators, and assemblers
9 = Elementary occupations
10 = Armed forces occupations
99 = Other/unspecified

See OCCUP in Section 5.2.2 for definitions.

wage_nc_2

This is a continuous variable that specifies ***the last wage payment in local currency of any individual*** (LSTATUS=1 & EMPSTAT=1) ***in its secondary occupation*** and is missing otherwise. The wage should come from the second job, in other words, the job that the person dedicated the second most amount of time in the week preceding the survey.

This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments. The 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.

Note:

- 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.
- By definition, non-paid employees (EMPSTAT_2=2) should have WAGE=0.
- The reference period of the WAGE_NC_2 will be recorded in the UNITWAGE_2 variable.

Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

unitwage_2

This is a categorical variable that specifies the time reference for the wage_no_compen_2 variable. It specifies the time unit measurement for the wages for the secondary job of any individual (LSTATUS=1 & EMPSTAT=1) and is missing otherwise. Ten categories after harmonization:

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

whours_2
This is a continuous variable that specifies the hours of work in ***last week for the second job with reference period of the last 7 days*** of any individual with a job (LSTATUS=1) and is missing otherwise. The second job defined as that occupation that the person dedicated the second most amount of time to over the past week. The variable is constructed for all persons administered this module in each questionnaire. The lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

Notes:

- If the respondent was absent from the job in the week preceding the survey due to holidays, vacation, or sick leave, then record the time worked in the previous 7 days that the person worked.
- Sometimes the questions are phrased as, "on average, how many hours 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 4.3 to get weekly hours.

wmonths_2
This is a continuous variable that specifies the number of months worked in the last 12 months for the secondary job of any individual with a job (LSTATUS=1) and is missing otherwise. The secondary job is defined as that occupation in which the person dedicated less time than the primary job over the past week. The variable is constructed for all persons administered this module in each questionnaire.

wage_total_2
This is a continuous variable that specifies the *annualized wage payment* (regular wage plus bonuses, in-kind, compensation, etc.) in local currency of any individual (LSTATUS=1 & EMPSTAT=1) in its secondary occupation and is missing otherwise. The wage should come from the secondary job, in other words, the job that the person dedicated the second most amount of time in the week preceding the survey. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments.

WAGE_TOTAL_2 should be equal to WAGE_NC_2 in case there are no bonuses, tips etc. offered as part of the job. The 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.

Notes:

- The annualization of the WAGE_TOTAL_2 should consider the number of months/weeks the persons have been working and receiving this income. Harmonizer should not assume the person has been working the whole year.

See WAGE_TOTAL in Section 5.2.2 for derivation formulae.

firmsize_l_2
This specifies the lower bracket of the firm size. The variable is constructed for all persons who are employed. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the lower boundary of the bracket.

firmsize_u_2
This specifies the upper bracket of the firm size. The variable is constructed for all persons who are employed. If it is continuous, it 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.

Table 5.3: Secondary Employment, 7-day reference period

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Labor	**empstat_2**	**Employment status, secondary job (7-day ref period)**	1 = Paid Employee 2 = Non-Paid Employee 3 = Employer 4 = Self-employed 5 = Other, workers not classifiable by status	2
2	Labor	**ocusec_2**	**Sector of activity, secondary job (7-day ref period)**	1 = Public sector, Central Government, Army 2 = Private, NGO 3 = State owned 4 = Public or State-owned, but cannot distinguish	2
3	Labor	**industry_orig_2**	**Original industry code, secondary job (7-day ref period)**	Country specific	2
4	Labor	**industrycat10_2**	**1 digit industry classification, secondary job (7-day ref period)**	1 = Agriculture, Hunting, Fishing, etc. 2 = Mining 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	2
5	Labor	**industrycat4_2**	**4-category industry classification, secondary job (7-day ref period)**	1=Agriculture 2=Industry 3=Services 4=Other	2

6	Labor	**occup_orig_2**	**Original occupational classification, secondary job (7-day ref period)**	Country specific	2
7	Labor	**occup_2**	**1-digit occupation classification, secondary job (7-day ref period)**	1 = Managers 2 = Professionals 3 = Technicians and associate professionals 4 = Clerical support workers 5 = Service and sales workers 6 = Skilled agricultural, forestry and fishery workers 7 = Craft and related trades workers 8 = Plant and machine operators, and assemblers 9 = Elementary occupations 10 = Armed forces occupations 99 = Other/unspecified	2
8	Labor	**wage_nc_2**	**Last wage payment, secondary job, excl. bonuses, etc. (7-day ref period)**	Continuous variable	2
9	Labor	**unitwage_2**	**Time unit of last wages payment, secondary job (7-day ref period)**	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	2
10	Labor	**whours_2**	**Hours of work in last week, secondary job (7-day ref period)**	Continuous variable	1
11	Labor	**wmonths_2**	**Months worked in the last 12 months, secondary job (7-day ref period)**	Continuous variable	1
12	Labor	**wage_total_2**	**Annualized total wage, secondary job (7-day ref period)**	Continuous variable	2
13	Labor	**firmsize_l_2**	**Firm size (lower bracket), secondary job (7-day ref period)**	Continuous variable	2

14	Labor	**firmsize_u_2**	**Firm size (upper bracket), secondary job (7-day ref period)**	Continuous variable	2

5.2.4 Other Employment Earnings, 7-day reference period

This refers to all other employment earnings excluding main and secondary jobs.

t_hours_others
Annualized hours worked in all but primary and secondary jobs (7-day ref period)
This is a continuous variable that specifies the hours of work in last 12 months in all jobs excluding the primary and secondary ones.

t_wage_nc_others
Annualized wage in all but primary & secondary jobs excl. bonuses, etc. (7-day ref period)
This is a continuous variable that specifies the annualized wage in all jobs excluding the primary and secondary ones. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

This will depend on the data available to derive annualized value but see WAGE_TOTAL in Section 5.2.2 for derivation formulae.

t_wage_others
Annualized wage in all but primary and secondary jobs (7-day ref period)
This is a continuous variable that specifies the annualized wage in all jobs excluding the primary and secondary ones.

This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. WAGE_OTHERS should be equal to WAGE_NC others in case there are no bonuses, tips etc. offered as part of any of the jobs.

This will depend on the data available to derive annualized value but see WAGE_TOTAL in Section 5.2.2 for derivation formulae.

Table 5.4: Other employment earnings, 7-day reference period

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Labor	**t_hours_others**	**Total hours of work in the last 7-day in other jobs excluding the primary and secondary ones**	Continuous variable	1

2	Labor	**t_wage_nc_others**	**Annualized wage in all jobs excluding the primary and secondary ones (excluding tips, bonuses, etc.)**	Continuous variable	1
3	Labor	**t_wage_others**	**Annualized wage (including tips, bonuses, etc.) in all other jobs excluding the primary and secondary ones**	Continuous variable	1

5.2.5 Total Employment Earnings, 7-day reference period

t_hours_total
This is a continuous variable that specifies the hours of work in last 12 months in **all jobs including primary, secondary and others**.

t_wage_nc_total
This is a continuous variable that specifies the total annualized wage income in **all jobs including primary, secondary and others**. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

This will depend on the data available to derive annualized value but see WAGE_TOTAL in Section 5.2.2 for derivation formulae.

t_wage_total
This is a continuous variable that specifies the total annualized wage income in **all jobs including primary, secondary and others**. This income includes tips, compensations such as bonuses, dwellings or clothes, and other payments. T_WAGE_TOTAL should be equal to T_WAGE_NC_TOTAL in case there are no bonuses, tips etc. offered as part of any of the jobs. If the number of months worked in this job is missing the harmonizer could assumed that the person worked the whole year in this job.

This will depend on the data available to derive annualized value but see WAGE_TOTAL in Section 5.2.2 for derivation formulae.

Table 5.5: Total employment earnings, 7-day reference period

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Labor	**t_hours _total**	Annualized hours worked in all jobs (7-day ref period)	Continuous variable	1
2	Labor	**t_wage_nc_total**	Annualized wage in all jobs excl. bonuses, etc. (7-day ref period)	Continuous variable	1

3	Labor	**t_wage_total**	Annualized total wage for all jobs (7-day ref period)	Continuous variable	1

5.2.6 Labor status, 12-month reference period

This section must be filled only for those individuals who responded to labor questions with a reference period of 12 months, regardless of whether they responded to questions with a reference period of the last 7 days.

All variables are numeric unless specified.

minlaborage_year
This is the lowest age for which the labor module is implemented in the survey or the minimum working age in the country. For this reason, the lower age cutoff at which information is collected will vary from country to country.

lstatus_year
This is an individual's ***labor status in the last 12 months***. The Value must be missing for individuals less than the required age (MINLABORAGE_YEAR).
Three categories are used after harmonization:

1 = Employed
2 = Unemployed
3 = Not-in-labor force

See LSTATUS in Section 5.2.2 for definitions.

Note: All persons are considered active in the labor force if they presently have a job (formal or informal, i.e., employed) or do not have a job but are actively seeking work (i.e., unemployed).

nlfreason_year
This is the reason an individual was not in the labor force ***in the last 12 months***. This variable is constructed for all those who are not presently employed and are not looking for work (LSTATUS_YEAR=3) and missing otherwise. Five categories after harmonization:

1 = Student
2 = Housewife
3 = Retired
4 = Disabled
5 = Other

See NLFREASON in Section 5.2.2 for definitions.

unempldur_l_year
This is a continuous variable specifying the ***duration of unemployment in months*** (***lower bracket***). The variable is constructed for all unemployed persons (LSTATUS_YEAR=2, otherwise missing). If it is specified as continuous in the survey, it records the numbers of months in unemployment. If the variable is categorical, it records the lower boundary of the bracket.

Missing values are allowed for everyone who is not unemployed. Other missing values are also allowed.

unempldur_u_year
This is a continuous variable specifying the ***duration of unemployment in months (upper bracket)***.

The variable is constructed for all unemployed persons (LSTATUS_YEAR=2, otherwise missing). If it is specified as continuous in the survey, it 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.

Missing values are allowed for everyone who is not unemployed. Other missing values are also allowed.
If the duration of unemployment is not reported as a range, but as continuous variables, the UNEMPLDUR_L_YEAR and UNEMPLDUR_U_YEAR variables will have the same value. If the high range is open-ended the UNEMPLDUR_U_YEAR variable will be missing.

Table 5.6: Labour force status, 12-month reference period

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Labor	**minlaborage_year**	**Labor module application age**	numeric	1
2	Labor	**lstatus_year**	**Labor status (12-month ref period)**	1 = Employed 2 = Unemployed 3 = Not-in-labor force	1
3	Labor	**nlfreason_year**	**Reason not in the labor force (12-month ref period)**	1 = Student 2 = Housewife 3 = Retired 4 = Disabled 5 = Other	1
4	Labor	**unempldur_l_year**	**Unemployment duration (months) lower bracket (12-month ref period)**	Continuous variable	1
5	Labor	**unempldur_u_year**	**Unemployment duration (months) upper bracket (12-month ref period)**	Continuous variable	1

5.2.7 Primary Employment, 12-month reference period

All variables are numeric unless specified.

empstat_year
This is a categorical variable that specifies the ***main employment status in the last 12 months*** of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. The variable is constructed for all individuals. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

The definitions are taken from the International Labor Organization's Classification of Status in Employment with some revisions to consider the data available. Five categories after harmonization:

1 = Paid Employee
2 = Non-Paid Employee
3 = Employer
4 = Self-employed
5 = Other, workers not classifiable by status

See EMPSTAT in Section 5.2.2 for definitions.

ocusec_year
This is a categorical variable that specifies the ***sector of activity in the last 12 months***. It classifies the main job's sector of activity of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. The variable is constructed for all persons administered this module in each questionnaire.
Four categories after harmonization:

1 = Public sector, Central Government, Army (including armed forces)
2 = Private, NGO
3 = State-owned
4 = Public or State-owned, but cannot distinguish

Note: Do not code basis of occupation (ISCO) or industry (ISIC) codes.

See OCUSEC in Section 5.2.2 for definitions.

industry_orig_year
This is a string variable that specifies the ***original industry codes in the last 12 months for the main job*** provided in the survey (the actual question) and should correspond to whatever is in the original file with no recoding.

The variable is constructed for all persons administered this module in each questionnaire. It will contain missing values for people below the working age. Other missing values are allowed. It classifies the main job of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise.

Code and name: Example: "1 - Agriculture"; "2 - Fishing"; "3 - Construction"; etc.

If classification not ISIC, labels must be translated to English. Make sure translation is correct from a language expert.

Depending on the ISIC digit level (2-, 3-. 4-level) or NACE (2-, 3-, 4-level), one can label the variables by running the dofiles provided that are already in English. **See Section 5.4 labeling dofiles.**

industrycat10_year

industrycat10 is a categorical variable that specifies the 1-digit industry classification ***in the last 12 months for the main job*** of any individual with a job (LSTATUS=1) and is missing otherwise.

The variable is constructed for all persons administered this module in each questionnaire. The codes for the main job are given here based on the UN International Standard Industrial Classification (ISIC) (revision 3.1/4.0)[9]. Ten categories after harmonization:

1 = Agriculture, Hunting, Fishing, etc.
2 = Mining
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

Note:

- When creating the industry variable, one needs to carefully check the type of economic activity codes and its revision which the household survey uses. Some surveys follow ISIC (International Standard Industrial Classification), while others follow NACE (Statistical Classification of Economic Activities in the European Community).
- See Annex III.1 for ISIC details and how to map the classifications.
- See Annex III.2 for NACE details and how to map the classifications.
- In the case of different classifications (former Soviet Union republics, for example), recoding has been done to best match the ISIC codes.
- Category 10 is also assigned for unspecified categories or items.

Note: If all 10 categories cannot be identified in the questionnaire create this variable as missing and proceed to create INDUSTRYCAT4.

See INDUSTRY_ORIG in Section 5.2.2 for how to label files.

industrycat4_year

industrycat4 is a categorical variable that specifies the 1-digit ***industry classification in the last 12 months for the main job*** for Broad Economic Activities. This variable is either created directly from the data (if industry

[9] https://unstats.un.org/unsd/classifications/Econ/ISIC#ISIC3

classification does not exist for ten categories) or created from industrycat10. Four categories after harmonization:

1 = Agriculture
2= Industry
3 = Services
4 = Other

This variable is either created directly from the data (if industry classification does not exist for ten categories) or created from industrycat10.

occup_orig_year

This is a string variable that specifies the ***original occupation code in the last 12 months for the main job***. This variable corresponds to whatever is in the original file with no recoding.

For each value label, there should be a space between the hyphen (before and after). If value label is truncated, make sure it is written in full. For example, pharm., law., etc are not allowed.

Code and name: Example: "1 - Pharmacist"; "2 - Engineer"; "3 - Lawyer"; etc.

If classification is not ISCO, labels must be translated to English. Make sure translation is correct from a language expert.

Depending on the ISCO digit level (2-, 3-. 4-level), one can label the variables by running the dofiles provided that are already in English. **See Section 5.4 labeling dofiles.**

Note: <u>This will only apply if the country uses the same ISCO classifications.</u>

occup_year

This is a categorical variable that specifies the 1-digit ***occupation classification for the main job in the last 12 months*** of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. This 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. Most surveys collect detailed information and then code it, without keeping the original data, no attempt has been made to correct or check the original coding. The classification is based on the International Standard Classification of Occupations (ISCO-08). The main categories subsume the following codes (see Annex III.3 for details). Eleven categories after harmonization:

1 = Managers
2 = Professionals
3 = Technicians and associate professionals
4 = Clerical support workers
5 = Service and sales workers
6 = Skilled agricultural, forestry and fishery workers
7 = Craft and related trades workers
8 = Plant and machine operators, and assemblers
9 = Elementary occupations
10 = Armed forces occupations

99 = Other/unspecified

wage_nc_year

This is a continuous variable that specifies the ***last wage payment in local currency*** of any individual (LSTATUS_YEAR=1 & EMPSTAT_YEAR=1) in its primary occupation at the reference period reported in the survey and it is missing otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the week preceding the survey.

This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments. The 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.

Note:

- 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.
- By definition, non-paid employees (EMPSTAT_YEAR=2) should have WAGE_YEAR=0.
- The reference period of the WAGE_NC_YEAR will be recorded in the UNITWAGE_YEAR.

unitwage_year

This is a categorical variable that specifies the time reference for the WAGE_NC_YEAR variable. It specifies the time unit measurement for the wages of any individual (LSTATUS_YEAR=1 & EMPSTAT_YEAR=1) and it is missing otherwise. Acceptable values include:

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

whours_year

This s a continuous variable that specifies the hours of work last week for the main job of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. The main job defined as that occupation that the person dedicated more time to over the past week. The variable is constructed for all persons administered this module in each questionnaire.

Note:

- If the respondent was absent from the job in the week preceding the survey due to holidays, vacation, or sick leave, then record the time worked in the previous 7 days that the person worked.
- Sometimes the questions are phrased as, “on average, how many hours 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 4.3 to get weekly hours.

wmonths_year
This is a continuous variable that specifies the number of months worked in the last 12 months for the main job of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. The main job is defined as that occupation that the person dedicated more time to over the past week. The variable is constructed for all persons administered this module in each questionnaire.

wage_total_year
This is a continuous variable that specifies the ***annualized wage payment*** (regular wage plus bonuses, in-kind, compensation, etc.) for the primary occupation in local currency of any individual (LSTATUS_YEAR=1 & EMPSTAT_YEAR=1) and is missing otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the week preceding the survey. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. WAGE_TOTAL_YEAR should be equal to WAGE_NC_YEAR in case there are no bonuses, tips etc. offered as part of the job.

The variable is constructed for all persons administered this module in each questionnaire. The annualization of the WAGE_TOTAL_YEAR should consider the number of months/weeks the persons have been working and receiving this income. Harmonizer should not assume the person has been working the whole year.

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors**.**

See WAGE_TOTAL in Section 5.2.2 for derivation formulae.

contract
This is a dummy variable that classifies the contract status (YES/NO) of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. It indicates whether a person has a signed (formal) contract, regardless of duration. Two categories after harmonization:

0 = No
1 = Yes

Note: This variable is only constructed if there is an explicit question about contract between employee and employer.

healthins_year
This is a dummy variable that classifies the health insurance status (YES/NO) of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. Two categories after harmonization:

0 = No
1 = Yes

Note: This variable is only constructed if there is an explicit question about health insurance provided by the job.

socialsec_year

This is a dummy variable that classifies the social security status (YES/NO) of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. Two categories after harmonization:

0 = No
1 = Yes

Note: This variable is only constructed if there is an explicit question about pension plans or social security.

union_year

This is a dummy variable that classifies the union membership status (YES/NO) of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. 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. Two categories after harmonization:

0 = No
1 = Yes

Note: This variable is only constructed if there is an explicit question about trade unions.

firmsize_l_year

This specifies the lower bracket of the firm size. The variable is constructed for all persons who are employed ***in the last 12 months for the main job***. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the lower boundary of the bracket.

firmsize_u_year

This specifies the upper bracket of the firm size. The variable is constructed for all persons who are employed ***in the last 12 months for the main job***. If it is continuous, it 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, this variable should be missing.

Table 5.7: Primary Employment, 12-month reference period

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Labor	**empstat_year**	**Employment status, primary job (12-month ref period)**	1 = Paid Employee 2 = Non-Paid Employee 3 = Employer 4 = Self-employed 5 = Other workers not classifiable by status	1
2	Labor	**ocusec_year**	**Sector of activity, primary job (12-month ref period)**	1 = Public sector, Central Government, Army 2 = Private, NGO 3 = State owned	1

				4 = Public or State-owned, but cannot distinguish	
3	Labor	**industry_orig_year**	**Original industry code, primary job (12-month ref period)**	Country specific	1
4	Labor	**industrycat10_year**	**1 digit industry classification, primary job (12-month ref period)**	1 = Agriculture, Hunting, Fishing, etc. 2 = Mining 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	1
5	Labor	**industrycat4_year**	**4-category industry classification, primary job (12-month ref period)**	1 = Agriculture 2 = Industry 3 = Services 4 = Other	1
6	Labor	**occup_orig_year**	**Original occupation classification, primary job (12-month ref period)**	Country specific	1
7	Labor	**occup_year**	**1-digit occupation classification, primary job (12-month ref period)**	1 = Managers 2 = Professionals 3 = Technicians and associate professionals 4 = Clerical support workers 5 = Service and sales workers 6 = Skilled agricultural, forestry and fishery workers 7 = Craft and related trades workers 8 = Plant and machine operators, and assemblers 9 = Elementary occupations 10 = Armed forces occupations 99 = Other/unspecified	1
8	Labor	**wage_nc_year**	**Last wage payment, primary job, excl. bonuses, etc. (12-month ref period)**	Continuous variable	1
8	Labor	**unitwage_year**	**Time unit of last wages payment, primary job (12-month ref period)**	1 = Daily 2 = Weekly 3 = Every two weeks 4 = Every two months 5 = Monthly 6 = Quarterly 7 = Every six months 8 = Annually	1

				9 = Hourly 10 = Other	
10	Labor	**whours_year**	**Hours of work in last week, primary job (12-month ref period)**	Continuous variable	1
11	Labor	**wmonths_year**	**Months worked in the last 12 months, primary job (12-month ref period)**	Continuous variable	1
12	Labor	**wage_total_year**	**Annualized total wage, primary job (12-month ref period)**	Continuous variable	1
13	Labor	**contract_year**	**Contract (12-month ref period)**	0 = No 1 = Yes	1
14	Labor	**healthins_year**	**Health insurance (12-month ref period)**	0 = No 1 = Yes	1
15	Labor	**socialsec_year**	**Social security (12-month ref period)**	0 = No 1 = Yes	1
16	Labor	**union_year**	**Union membership (12-month ref period)**	0 = No 1 = Yes	1
17	Labor	**firmsize_l_year**	**Firm size (lower bracket), primary job (12-month ref period)**	Continuous variable	1
18	Labor	**firmsize_u_year**	**Firm size (upper bracket), primary job (12-month ref period)**	Continuous variable	1

5.2.8 Secondary Employment, 12-month reference period

empstat_2_year

This is a categorical variable that specifies the ***main employment status in the last 12 months*** of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. The variable is constructed for all individuals. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

The definitions are taken from the International Labor Organization's Classification of Status in Employment with some revisions to consider the data available. Five categories after harmonization:

1 = Paid Employee
2 = Non-Paid Employee
3 = Employer
4 = Self-employed
5 = Other, workers not classifiable by status

See EMPSTAT in Section 5.2.2 for definitions.

ocusec_2_year

This is a categorical variable that specifies the ***sector of activity in the last 12 months***. It classifies the main job's sector of activity of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. The variable is constructed for all persons administered this module in each questionnaire.

Four categories after harmonization:

1 = Public sector, Central Government, Army (including armed forces)
2 = Private, NGO
3 = State-owned
4 = Public or State-owned, but cannot distinguish

Note: Do not code basis of occupation (ISCO) or industry (ISIC) codes.

See OCUSEC in Section 5.2.2 for definitions.

industry_orig_2_year

This is a string variable that specifies the ***original industry codes in the last 12 months for the main job*** provided in the survey (the actual question) and should correspond to whatever is in the original file with no recoding.

The variable is constructed for all persons administered this module in each questionnaire. It will contain missing values for people below the working age. Other missing values are allowed. It classifies the main job of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise.

Code and name: Example: "1 - Agriculture"; "2 - Fishing"; "3 - Construction"; etc.

If classification not ISIC, labels must be translated to English. Make sure translation is correct from a language expert.

See INDUSTRY_ORIG in Section 5.2.2 for how to label files.

industrycat10_2_year

industrycat10 is a categorical variable that specifies the 1-digit industry classification ***in the last 12 months for the main job*** of any individual with a job (LSTATUS=1) and is missing otherwise.

The variable is constructed for all persons administered this module in each questionnaire. The codes for the main job are given here based on the UN International Standard Industrial Classification (ISIC) (revision 3.1/4.0)[10].

Ten categories after harmonization:

1 = Agriculture, Hunting, Fishing, etc.
2 = Mining
3 = Manufacturing
4 = Public Utility Services
5 = Construction
6 = Commerce
7 = Transport and Communications

[10] https://unstats.un.org/unsd/classifications/Econ/ISIC#ISIC3

8 = Financial and Business Services
9 = Public Administration
10 = Other Services, Unspecified

Note:

- When creating the industry variable, one needs to carefully check the type of economic activity codes and its revision which the household survey uses. Some surveys follow ISIC (International Standard Industrial Classification), while others follow NACE (Statistical Classification of Economic Activities in the European Community).
- In the case of different classifications (former Soviet Union republics, for example), recoding has been done to best match the ISIC codes.
- Category 10 is also assigned for unspecified categories or items.

Note: If all 10 categories cannot be identified in the questionnaire create this variable as missing and proceed to create industrycat4.

See INDUSTRYCAT10 in Section 5.2.2 for definitions.

industrycat4_2_year
industrycat4 is a categorical variable that specifies the 1-digit ***industry classification in the last 12 months for the main job*** for Broad Economic Activities. This variable is either created directly from the data (if industry classification does not exist for ten categories) or created from industrycat10. Four categories after harmonization:

1 = Agriculture
2= Industry
3 = Services
4 = Other

This variable is either created directly from the data (if industry classification does not exist for ten categories) or created from industrycat10.

occup_orig_2_year
This is a string variable that specifies the ***original occupation code in the last 12 months for the main job***. This variable corresponds to whatever is in the original file with no recoding.

For each value label, there should be a space between the hyphen (before and after). If value label is truncated, make sure it is written in full. For example, pharm., law., etc are not allowed.

Code and name: Example: “1 - Pharmacist”; “2 - Engineer”; “3 - Lawyer”; etc.

If classification is not ISCO, labels must be translated to English. Make sure translation is correct from a language expert.

See OCCUP_ORIG in Section 5.2.2 for how to label files.

occup_2_year

This is a categorical variable that specifies the 1-digit ***occupation classification for the main job in the last 12 months*** of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. This 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.

Most surveys collect detailed information and then code it, without keeping the original data, no attempt has been made to correct or check the original coding. The classification is based on the International Standard Classification of Occupations (ISCO-08). The main categories subsume the following codes (see Annex III.4 for details). Eleven categories after harmonization:

1 = Managers
2 = Professionals
3 = Technicians and associate professionals
4 = Clerical support workers
5 = Service and sales workers
6 = Skilled agricultural, forestry and fishery workers
7 = Craft and related trades workers
8 = Plant and machine operators, and assemblers
9 = Elementary occupations
10 = Armed forces occupations
99 = Other/unspecified

See OCCUP in Section 5.2.2 for definitions.

wage_nc_2_year

This is a continuous variable that specifies the ***last wage payment in local currency*** of any individual (LSTATUS_YEAR=1 & EMPSTAT_YEAR=1) in its secondary occupation at the reference period reported in the survey and it is missing otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the week preceding the survey. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments. The 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.

Notes:

- 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.
- By definition, non-paid employees (EMPSTAT_YEAR=2) should have WAGE_YEAR=0.
- The reference period of the WAGE_NC_YEAR will be recorded in the UNITWAGE_YEAR.

unitwage_2_year

This is a categorical variable that specifies the time reference for the WAGE_NC_YEAR variable. It specifies the time unit measurement for the wages of any individual (LSTATUS_YEAR=1 & EMPSTAT_YEAR=1) and it is missing otherwise. Acceptable values include:

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

whours_2_year
This s a continuous variable that specifies the hours of work last week for the main job of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. The main job defined as that occupation that the person dedicated more time to over the past week. The variable is constructed for all persons administered this module in each questionnaire. Notes:

- If the respondent was absent from the job in the week preceding the survey due to holidays, vacation, or sick leave, then record the time worked in the previous 7 days that the person worked.
- Sometimes the questions are phrased as, "on average, how many hours 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 4.3 to get weekly hours.

wmonths_2_year
This is a continuous variable that specifies the number of months worked in the last 12 months for the main job of any individual with a job (LSTATUS_YEAR=1) and is missing otherwise. The main job is defined as that occupation that the person dedicated more time to over the past week. The variable is constructed for all persons administered this module in each questionnaire.

wage_total_2_year
This is a continuous variable that specifies the ***annualized wage payment*** (regular wage plus bonuses, in-kind, compensation, etc.) for the secondary occupation in local currency of any individual (LSTATUS_YEAR=1 & EMPSTAT_YEAR=1) and is missing otherwise.

The wage should come from the main job, in other words, the job that the person dedicated most time in the week preceding the survey. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments.

WAGE_TOTAL_YEAR should be equal to WAGE_NC_YEAR in case there are no bonuses, tips etc. offered as part of the job. The variable is constructed for all persons administered this module in each questionnaire. The annualization of the WAGE_TOTAL_YEAR should consider the number of months/weeks the persons have been working and receiving this income. Harmonizer should not assume the person has been working the whole year.

See WAGE_TOTAL in Section 5.2.2 for derivation formulae.

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors**.**

firmsize_l_2_year
This specifies the lower bracket of the firm size. The variable is constructed for all persons who are employed ***in the last 12 months for the main job***. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the lower boundary of the bracket.

firmsize_u_2_year
This specifies the upper bracket of the firm size. The variable is constructed for all persons who are employed ***in the last 12 months for the main job***. If it is continuous, it 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, this variable should be missing.

Table 5.8: Secondary Employment, 12-month reference period

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Labor	**empstat_2_year**	**Employment status, primary job (12-month ref period)**	1 = Paid Employee 2 = Non-Paid Employee 3 = Employer 4 = Self-employed 5 = Other workers not classifiable by status	2
2	Labor	**ocusec_2_year**	**Sector of activity, secondary job (12-month ref period)**	1 = Public sector, Central Government, Army 2 = Private, NGO 3 = State owned 4 = Public or State-owned, but cannot distinguish	2
3	Labor	**industry_orig_2_year**	**Original industry code, secondary job (12-month ref period)**	Country specific	2
4	Labor	**industrycat10_2_year**	**1 digit industry classification, secondary job (12-month ref period)**	1 = Agriculture, Hunting, Fishing, etc. 2 = Mining 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	2

5	Labor	**industrycat4_2_year**	**4-category industry classification, secondary job (12-month ref period)**	1=Agriculture 2=Industry 3=Services 4=Other	2
6	Labor	**occup_orig_2_year**	**Original occupation classification, secondary job (12-month ref period)**	Country specific	2
7	Labor	**occup_2_year**	**1-digit occupation classification, secondary job (12-month ref period)**	1 = Managers 2 = Professionals 3 = Technicians and associate professionals 4 = Clerical support workers 5 = Service and sales workers 6 = Skilled agricultural, forestry and fishery workers 7 = Craft and related trades workers 8 = Plant and machine operators, and assemblers 9 = Elementary occupations 10 = Armed forces occupations 99 = Other/unspecified	2
8	Labor	**wage_nc_2_year**	**Last wage payment, secondary job, excl. bonuses, etc. (12-month ref period)**	Continuous variable	1
9	Labor	**unitwage_2_year**	**Time unit of last wages payment, secondary job (12-month ref period)**	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	1
10	Labor	**whours_2_year**	**Hours of work in last week, secondary job (12-month ref period)**	Continuous variable	1
11	Labor	**wmonths_2_year**	**Months worked in the last 12 months, secondary job (12-month ref period)**	Continuous variable	1
12	Labor	**wage_total_2_year**	**Annualized total wage, secondary job (12-month ref period)**		2

13	Labor	**firmsize_l_2_year**	**Firm size (lower bracket), secondary job (12-month ref period)**	Continuous variable	2
14	Labor	**firmsize_u_2_year**	**Firm size (upper bracket), secondary job (12-month ref period)**	Continuous variable	2

5.2.9 Other Employment, 12-month reference period

t_hours_others_year
This is a continuous variable that specifies the hours of work in last 12 months **in all jobs excluding the primary and secondary ones**.

t_wage_nc_others_year
This is a continuous variable that specifies annual wage **in all jobs excluding the primary and secondary ones**. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

This will depend on the data available to derive annualized value but see WAGE_TOTAL in Section 5.2.2 for derivation formulae.

t_wage_others_year
This is a continuous variable that specifies the annual wage **in all jobs excluding the primary and secondary ones**. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. T_WAGE_OTHERS should be equal to T_WAGE_NC_OTHERS in case there are no bonuses, tips etc. offered as part of any of the jobs.

This will depend on the data available to derive annualized value but see WAGE_TOTAL in Section 5.2.2 for derivation formulae.

Table 5.9: Other employment earnings, 12-month reference period

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Labor	**t_hours_others_year**	**Total hours of work in the last 12 months in other jobs excluding the primary and secondary ones**	Continuous variable	1
2	Labor	**t_wage_nc_others_year**	**Annualized wage in all jobs excluding the primary and secondary ones (excluding tips, bonuses, etc.).**	Continuous variable	1

3	Labor	**t_wage_others_year**	**Annualized wage (including tips, bonuses, etc.) in all other jobs excluding the primary and secondary ones.**	Continuous variable	1

5.2.10 Total Employment Earnings, 12-month reference period

t_hours_total_year
This is a continuous variable that specifies the hours of work in last 12 months in **all jobs including primary, secondary and others**. Note:

t_wage_nc_total_year
This is a continuous variable that specifies the total annualized wage income in **all jobs including primary, secondary and others**. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

> **This will depend on the data available to derive annualized value but see WAGE_TOTAL in Section 5.2.2 for derivation formulae.**

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

t_wage_total_year
This is a continuous variable that specifies the total annualized wage income in **all jobs including primary, secondary and others**. This income includes tips, compensations such as bonuses, dwellings or clothes, and other payments. T_WAGE_TOTAL should be equal to T_WAGE_NOCOMPEN_TOTAL in case there are no bonuses, tips etc. offered as part of any of the jobs.

> **This will depend on the data available to derive annualized value but see WAGE_TOTAL in Section 5.2.2 for derivation formulae.**

Table 5.10: Total employment earnings, 12-month reference period

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Labor	**t_hours _total_year**	Annualized hours worked in all jobs (12-month ref period)	Continuous variable	1
2	Labor	**t_wage_nc_total_year**	Annualized wage in all jobs excl. bonuses, etc. (12-month ref period)	Continuous variable	1
3	Labor	**t_wage_total_year**	Annualized total wage for all jobs (12-month ref period)	Continuous variable	1

5.2.11 Total Labor Income

Total Labor income will be created based on either the 7 days or 12 months reference period variables or a combination of both.
Harmonizers should make sure that all jobs are included and none of them are double counted.

njobs
This is a numeric variable that specifies the total number of jobs. Do not put missing value for people below working age, unemployed and people out of the labor force. Other missing values allowed.

t_hours_annual
This is a continuous variable that specifies the annual numbers of hours worked in all the jobs including primary, secondary and others regardless of their period of reference.

linc_nc
This is a continuous variable that specifies the total annualized wage income in all the jobs including primary, secondary and others regardless of their period of reference. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

This will depend on the data available to derive annualized value but see WAGE_TOTAL in Section 5.2.2 for derivation formulae.

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

laborincome
This is a continuous variable that specifies the total annualized individual labor income in all jobs including primary, secondary and others regardless of their period of reference. This income includes tips, compensations such as bonuses, dwellings or clothes, and other payments. This variable should be used as the total annual labor income of an individual.

This will depend on the data available to derive annualized value but see WAGE_TOTAL in Section 5.2.2 for derivation formulae.

Table 5.11: Total Labor Income

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Labor	**njobs**	Total number of jobs	Continuous variable	1
2	Labor	**t_hours_annual**	Total hours worked in all jobs in the previous 12 months	Continuous variable	1

3	Labor	**linc_nc**	Total annual wage income in all jobs, excl. bonuses, etc.	Continuous variable	1
4	Labor	**laborincome**	Total annual individual labor income in all jobs, incl. bonuses, etc.	Continuous variable	1

5.3 Challenges and Lessons Learned

Notes:

- For any variable not collected in a country, the 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.
- Labor status during last 12 months only reflects if the person has worked during last year or not, as many of the surveys do not provide enough information to distinguish those individuals that are unemployed from those that are out of the labor force.
- Individuals working in cooperatives are considered as "paid employee" in the employment status variable.

Several checks should be conducted to ensure that the data is harmonized correctly.

- LSTATUS should be an integer in the range [1,3].

```
lstatus<0 & lstatus>3 & mod(lstatus, 1) == 1
```

- If LSTATUS is classified as employed, then the employment type needs to be defined.

```
lstatus==1 & empstat>5
```

- MINILABORAGE should be an integer.

```
round(minlaborage)==minlaborage
```

- The minimum age for employment should not be higher than 20.

```
minlaborage >20 & !mi(minlaborage)
```

- EMPSTAT should be an integer in the range [1,5].

```
inlist(empstat,1,2,3,4,5)
```

- If employment type is defined then labor force status should be employed.

```
empstat<=5 & lstatus!=1
```

- Labelling the INDUSTRY_ORIG, INDUSTRY_ORIG2, INDUSTRY_ORIG_YEAR, INDUSTRY_ORIG_YEAR2 depeding on ISIC revisions (Rev 3.1., Rev 4) or NACE revisions (Rev 1 or Rev 2). This will only apply if the country uses exactly the same ISIC or NACE classifications.

    - ISIC 2-digit or 3-digit or 4-digit

```
clonevar industry_orig*=ctryvarname  //ctryname is the variable name provided by country
format industry_orig* %02.0f
run "X:\ILO labor\ISIC Rev* 2-digit labelling.do"

   or

clonevar industry_orig*=ctryvarname
format industry_orig* %03.0f
run "X:\ILO labor\ISIC Rev* 3-digit labelling.do"

   or

clonevar industry_orig*=ctryvarname
format industry_orig* %02.0f
run "X:\ILO labor\ISIC Rev* 4-digit labelling.do"
```

    - NACE 2-digit or 3-digit or 4-digit

```
clonevar industry_orig*=ctryvarname  //ctryname is the variable name provided by country
format industry_orig* %02.0f
run "X:\ILO labor\NACE Rev* 2-digit labelling.do"

   or

clonevar industry_orig=ctryvarname
format industry_orig* %03.0f
run "X:\ILO labor\NACE Rev* 3-digit labelling.do"

   or

clonevar industry_orig*=ctryvarname
format industry_orig* %02.0f
     run "X:\ILO labor\NACE Rev* 4-digit labelling.do"
```

- INDUSTRY10 should be an integer in the range [1,10].

```
!inlist(industrycat10,1,2,3,4,5,6,7,8,9,10)
```

- INDUSTRY4 should be an integer in the range [1,4].

```
!inlist(industrycat4,1,2,3,4)
```

- There should not be a mismatch between industry and industrycat4.

```
((industrycat4==1 & industrycat10!=1 ) | (industrycat4==2 &
(industrycat10 <2 | industrycat10 >5)) |  (industrycat4==3 &
(industrycat10 <6 | industrycat10 >9)) | (industrycat4==1 & industrycat10
!=1 ) ) &  industrycat10 !=.
```

- Labelling the OCCUP_ORIG, OCCUP_ORIG2, OCCUP_ORIG_YEAR, OCCUP_ORIG_YEAR2 depending on ISCO revisions (Rev 3.1., Rev 4). This will only apply if the country uses exactly the same ISCO classifications.
    - ISCO 1-digit, 2-digit or 3-digit or 4-digit

```
clonevar industry_orig*=ctryvarname  //ctryname is the variable name provided by country
format industry_orig* %02.0f
run "X:\ILO labor\ISCO-08 1-digit labelling.do"

   or

clonevar industry_orig*=ctryvarname  //ctryname is the variable name
provided by country
format industry_orig* %02.0f
run "X:\ ILO labor\ISCO-08 2-digit labelling.do"

   or

clonevar industry_orig*=ctryvarname
format industry_orig* %03.0f
run "X:\ ILO labor\ISCO-08 3-digit labelling.do"

   or

clonevar industry_orig*=ctryvarname
format industry_orig* %02.0f
run "X:\ ILO labor\ISCO-08 4-digit labelling.do"
```

- FIRMSIZE_U should not be lower than FIRMSIZE_L

```
firmsize_u<firmsize_l
```

Combine all Sections 5 and run the dofile ***MODULE_05_Labor.do*** which labels (variable and value) and order variables

- ✓ This is a must process with no exception.

6 Utilities (UTL)

6.1 Framework of Harmonization

Multiple-topic household surveys collect data on the characteristics of both households and individuals within those households. This module covers affordability and access related indicators that are commonly derived from survey data sets, or that can be constructed using existing variables. The primary objective of this indicator harmonization is to generate a unified data source of globally comparable indicators for utilities affordability and access to support routine analytical and corporate business functions of the World Bank.

In this module, the primary unit of analysis is the level of a household. The units of classification are consumption expenditures made by households for satisfying their needs or wants for various goods and services.

6.2 Mapping and description of variables

The Utilities module contains a large number of metadata that provides a wealth of information about variables, including their types, descriptions, sources, etc. To improve readability, only the most significant information has been included in this section. For a complete list of all variables captured in the module, please consult table at the end of the module.

The Utilities module consists of affordability variables and access to services variables, which follow COICOP and MTF and Access Plus frameworks, respectively. The utilities affordability variables are monetary variables expressed at current prices in the local currency unit (LCU) and non-deflated either temporal or spatial. It should include not only monetary expenses, but also value of in-kind acquisitions.

6.2.1 Access to Services

Non-energy or non-WASH variables that are broadly classified as utilities are included as additional variables: Maintenance and Repair of the Dwelling (4.3); Part of Fuels and lubricants for personal transport equipment (transport fuels only, but not lubricants; 7.2.2); Telephone and telefax services (8.3.0); Part of Cultural Services (TV broadcasting services only; 9.4.2).

The harmonization framework for GMD utilities access to services variables is based on three independent frameworks: Human Opportunity Index (HOI)[11], developed by the World Bank's Latin America and Caribbean (LAC) department in 2008; WASH (Water, sanitation, and hygiene) Access Plus Framework developed by the Water Global Project in 2015 (draft); and the Multi-tier access Tracking Framework (MTF)[12] introduced by the World Bank and SE4ALL Knowledge Hub in 2015. According to the HOI typology, all WASH and energy access indicators are defined as binary indices, while the other two frameworks define WASH and energy indicators as

[11] For further reference, see http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/LACEXT/EXTLACREGTOPPOVANA/0,,contentMDK:21881102~pagePK:34004173~piPK:34003707~theSitePK:841175,00.html

[12] For further reference, see http://www.worldbank.org/en/topic/energy/publication/energy-access-redefined

multi-tier indices. Due to the limitation of ex-post harmonization, GMD variables may not match exactly multi-tier frameworks, but the tiers are embedded into the categories whenever possible.
Access to and affordability of WASH and energy are important nonmonetary dimensions of welfare. The SDGs call for universal and equitable access to safe and affordable drinking water and access to adequate and equitable sanitation and hygiene for all. The indicators require an understanding of the proportion of the population which has access to safely managed WASH services or facilities. In the same vein, understanding the proportion of the population with access to electricity, and more specifically, the types of fuels used is necessary to understand the progress made on SDG 7- Affordable and Clean Energy.

The utilities access to services indicators include access to water, sanitation, and hygiene (WASH) and access to energy.

6.2.1.1 Access to Water, Sanitation and Hygiene (WASH)

watertype_quest
This is a categorical variable that specifies the type of water questions in the survey. The variable records the type of question(s) asked about the type of water source. For example, the survey had a specific question on the water source on drinking water, or on water source on general water or both. Type of water question, four categories after harmonization:

1 = Drinking water
2 = General water
3 = Both
4 = Other

If unknown code as “Other”.

Subsequent questions on water will depend on WATERTYPE_QUEST response.

water_original
This is a string variable that specifies the original survey response for the main water sources.

If the main source of water differs between the wet and dry season, water source during dry season is referred. The best possible match is sought, but in many cases the correspondence between country-specific values and these standardized codes is imperfect.

It is a country-specific variable. It must follow the naming convention: “1 – Piped Water” (as string).

water_source
This is a categorical variable that indicates the main source of drinking water for the household. If the main source of water differs between the wet and dry season, water source during dry season is referred. The best possible match is sought, but in many cases the correspondence between country-specific values and these standardized codes is imperfect.

If several types of water are used by the household, only main source required.

Unless otherwise, must be coded from WATER_ORIGINAL. Harmonizers should refer to the survey questionnaire to assess the best matches. Main source of drinking water, fourteen categories after harmonization:

1 = Piped water into dwelling
2 = Piped water to yard/plot
3 = Public tap or standpipe
4 = Tubewell or borehole
5 = Protected dug well
6 = Protected spring
7 = Bottled water
8 = Rainwater
9 = Unprotected spring
10 = Unprotected dug well
11 = Cart with small tank/drum
12 = Tanker-truck
13 = Surface water
14 = Other

- *Piped into dwelling,* 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. When tubewell or borehole is delivered via a pipe system this is considered as piped (see tubewell/borehole definition).
- *Piped water to yard/plot*, 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.
- *Public standpipe* 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.
- *Tubewell or borehole* 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. However, boreholes delivering water to an overhead tank which supplies multiple compounds through a reticulated piped system should be classified as one of the types of 'piped water', depending on where the household collects the water.
- *Protected dug well* 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 is also covered to prevent any infiltration.
- *Protected spring* 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.
- *Surface water* is water located above the ground and includes lakes, rivers, ponds, streams, canals, and irrigation canals.
- *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.
- *Tanker-truck* is water trucked into a community and sold from a water truck. The water source unknown.

- *Other* includes other water sources not mentioned above.

imp_wat_rec

This is a categorical variable that estimates the "recommended" categorization for access to improved water sources in each country, or how evidence suggests that the expected error might be minimized. If the relevant survey was on file in the SDG calculations, this would be considered 1 if most of the problematic category was estimated therein to be of an improved type at the rural level, and otherwise considered 0. If the survey was not already in the SDG calculations, recommendations are based on the standard international classifications plus any relevant insights from other surveys on file for the specific country. In the few instances where there was no evidence, 0 is used. For an example of this, see the Main challenges/lessons learned section. The recommended access, two categories after harmonization:

0 = No
1 = Yes

Improved drinking water sources should, but do not always, provide safe drinking water, and include:

- Piped household water connection
- Public standpipe
- Borehole
- Protected dug well
- Protected spring
- Rainwater collection
- Bottled water previously treated as unimproved due to lack of data on accessibility, availability and quality.

But for SDG monitoring the WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP)[13] will treat them as 'improved'.

Unimproved drinking water sources include:

- Unprotected dug well
- Unprotected spring
- Surface water (river, dam, lake, pond, stream, canal, irrigation channel)
- Vendor-provided water (cart with small tank/drum, tanker truck)
- Tanker truck water

The JMP data excel file is a good source of cross-validation on this variable harmonization (https://washdata.org/data#!/) but one must be cognizant there could be differences and that does not make one correct than the other as the differences could be due to use of different survey data source.

piped

This is a categorical variable that indicates whether the household has access to piped water. There are two major types of water supply – within premises and outside premises. 'Within premises' refers to water service piped connection to own tap. It includes both household connection (in-house plumbing) and yard connection (yard or plot outside the house plumbing).

[13] https://washdata.org/how-we-work/about-jmp

Conversely, outside premise refers to a public water point from which people can collect water, shared among houses. It includes public tap and standpipe or a public fountain. Access to piped water, two categories after harmonization:

0 = No

1 = Yes

piped_to_prem

This is a categorical variable that specifies whether a household has access to piped water on premises. There are two major types of water supply – within premises and outside premises. 'Within premises' refers to water service piped connection to own tap. It includes both household connection (in-house plumbing) and yard connection (yard or plot outside the house plumbing).

Conversely, outside premise refers to a public water point from which people can collect water, shared among houses. It includes public tap and standpipe or a public fountain. Access to piped water on premises, two categories after harmonization:

0 = No

1 = Yes

Only if WATER_SOURCE<=2.

w_30m

This is a categorical variable that specifies whether a household has access to water within 30 minutes. This includes time taken for a *round trip and waiting time* in case of queues. This variable can be used in conjunction with the IMP_WAT_REC dummy to identify whether the improved water source is available within 30 minutes. Two categories after harmonization:

1 = Time to WATER_SOURCE less than or equal to 30 mins

0 = Time to WATER_SOURCE more than 30 mins

w_avail

This is a categorical variable that specifies whether water is available when needed. This variable can be used in conjunction with the IMP_WAT_REC dummy to identify where the improved water source is available reliably 24/7. Availability of water when needed, two categories after harmonization:

1 = Water is available continuously, reliable source

0 = Water source is unreliable

sanitation_original

This is a string variable that specifies the original survey response for the SANITATION_SOURCE variable. It is a country-specific variable. It must follow the naming convention: "1 – Flush toilet" (as string).

sanitation_source

This is a categorical variable that specifies the source of sanitation facilities. The best possible match is sought, but in many cases the correspondence between country-specific values and these standardized codes is imperfect.

If several types of toilets are used by the household, only main source required.

Unless otherwise, must be coded from SANITATION_ORIGINAL. Harmonizers should refer to the survey questionnaire to assess the best matches. Main sanitation source, fourteen categories after harmonization:

1 = A flush toilet
2 = A piped sewer system
3 = A septic tank
4 = Pit latrine
5 = Ventilated improved pit latrine (VIP)
6 = Pit latrine with slab
7 = Composting toilet
8 = Special case
9 = A flush/pour flush to elsewhere
10 = A pit latrine without slab
11 = Bucket
12 = Hanging toilet or hanging latrine
13 = No facilities or bush or field
14 = Other

- *Flush toilet* 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: -
    - pour flush uses a water seal, but unlike a flush toilet, it uses water poured by hand for flushing (no cistern is used)
    - flush toilet to "somewhere else" such a flushed to a river, hanging toilet or some place
- *Ventilated Improved Pit latrine* (VIP): 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.
- A *composting toilet* is a type of dry toilet that uses a predominantly aerobic processing system to treat human excreta, by composting or managed aerobic decomposition. These toilets generally use little to no water and may be used as an alternative to flush toilets.
- *Pit latrine* is a simple pit latrine but covered or with a slab.
- *No facility* includes, open fields, bush.
- *Other* includes bucket, pan, and open/uncovered pit latrines among others.

toilet_acc

This is a categorical variable that indicates type of access to a flush toilet. Access to flush toilet, four categories after harmonization:

0 = No
1 = Yes, in premise
2 = Yes, but not in premise including public toilet
3 = Yes, unstated whether in or outside the premise

sewer
This is a categorical variable that specifies whether a household has access to a toilet connected to a piped sewer system. Access to sewer, two categories after harmonization

0 = No
1 = Flush/pour flush to piped sewer system

open_def
This is a categorical variable that specifies whether a household has access to any sanitation facility. Access to any sanitation facility, two categories after harmonization:

0 = Availability of any facility
1 = No facility, or bush, or field

imp_san_rec
This is a categorical variable that estimates the categorization for access to improved sanitation facilities, or how evidence suggests that the expected error might be minimized.

If the relevant survey was on file in the SDG computations, this would be considered 1 if most of the problematic category was estimated therein to be of an improved type at the rural level, and otherwise considered 0.

If the survey was not already in the SDG calculations, recommendations are based on the standard international classifications plus any relevant insights from other surveys on file for the specific country. In the few instances where there was no evidence, 0 is used. The recommended access, two categories after harmonization:

0 = No
1 = Yes

Improved sanitation facilities include:	Unimproved sanitation includes:
• Flush or pour-flush to piped sewer system, septic tank or pit latrine • Ventilated improved pit latrine • Pit latrine with slab • Composting toilet	• Flush or pour-flush to elsewhere • Pit latrine without slab or open pit • Bucket, hanging toilet or hanging latrine and • No facilities or bush or field (open defecation)

Source: (WHO & UNICEF, 2010) http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=9

Note: Sanitation facilities are not considered improved when shared with other households, or open to public use.

waste
This is a categorical variable that indicates the type of solid waste disposal. This variable contains information on the usual manner of collection and disposal of solid waste or garbage generated by occupants of the housing unit. It is categorized by the manner of disposal, such as collection, disposal, burial, or compost and by the administrator of the waste disposal, such as authorized collectors, self-appointed collectors, and dump supervised by authorities.

1 = Solid waste collected on a regular basis by authorized collectors
2 = Solid waste collected on an irregular basis by authorized collectors
3 = Solid waste collected by self-appointed collectors

4 = Occupants dispose of solid waste in a local dump supervised by authorities
5 = Occupants dispose of solid waste in a local dump not supervised by authorities
6 = Occupants burn solid waste
7 = Occupants bury solid waste
8 = Occupants dispose solid waste into river, sea, creek, pond
9 = Occupants compost solid waste
10 = Other arrangement

Table 6.1: Utilities-Access to Water, Sanitation and Hygiene (WASH)

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	ID	**countrycode**	**Country code**	String See Table 2.1	1
2	ID	**year**	**Year**	Numeric discrete See Table 2.1	1
3	ID	**hhid**	**Household identifier**	String See Table 2.1	1
4	ID	**weight**	**Weight**	Numeric See Table 2.1	1
5	Utilities	**watertype_quest**	**Type of water questions used in the survey**	Numeric categorical 1 = Drinking water 2 = General water 3 = Both 4 = Other (undefined)	2
6	Utilities	**water_original**	Main source of water (country specific)	String	1
7	Utilities	**water_source**	**Main source of drinking water**	1 = Piped water into dwelling 2 = Piped water to yard/plot 3 = Public tap or standpipe 4 = Tubewell or borehole 5 = Protected dug well 6 = Protected spring 7 = Bottled water 8 = Rainwater 9 = Unprotected spring 10 = Unprotected dug well 11 = Cart with small tank/drum 12 = Tanker-truck 13 = Surface water 14 = Other	1
8	Utilities	**imp_wat_rec**	**Improved water source**	0 = No 1 = Yes	1
9	Utilities	**piped**	**Access to piped water**	0 = No	1

				1 = Yes	
10	Utilities	**piped_to_prem**	**Access to piped water on premises**	0 = No 1 = Yes	1
11	Utilities	**w_30m**	**Access to water within 30 minutes**	1 = Collection time less than or equal to 30 mins 0 = Collection time more than 30 mins	2
12	Utilities	**w_avail**	**Water is available when needed**	1= Water is available continuously, reliable source 0 = Water source is unreliable	2
13	Utilities	**sanitation_original**	**Main toilet facility (country specific)**	String	1
14	Utilities	**sanitation_source**	**Main toilet facility**	1 = A flush toilet 2 = A piped sewer system 3 = A septic tank 4 = Pit latrine 5 = Ventilated improved pit latrine (VIP) 6 = Pit latrine with slab 7 = Composting toilet 8 = Special case 9 = A flush/pour flush to elsewhere 10 = A pit latrine without slab 11 = Bucket 12 = Hanging toilet or hanging latrine 13 = No facilities or bush or field 14 = Other	1
15	Utilities	**toilet_acc**	**Access to a flush toilet**	0 = No 1 = Yes, in premise 2 = Yes, but not in premise including public toilet 3 = Yes, unstated whether in or outside premise	2
16	Utilities	**sewer**	**Access to sewer**	0 = No 1 = Flush/pour flush to piped sewer system	2
17	Utilities	**open_def**	**Access to any sanitation facility**	0 = Availability of any facility 1 = No facility	2
18	Utilities	**imp_san_rec**	**Improved sanitation facility**	0 = No 1 = Yes	1
19	Utilities	**waste**	**Main types of solid waste disposal**	1 = Solid waste collected on a regular basis by authorized collectors 2 = Solid waste collected on an irregular basis by authorized collectors 3 = Solid waste collected by self-appointed collectors 4 = Occupants dispose of solid waste in a	2

				local dump supervised by authorities 5 = Occupants dispose of solid waste in a local dump not supervised by authorities 6 = Occupants burn solid waste 7 = Occupants bury solid waste 8 = Occupants dispose solid waste into river, sea, creek, pond 9 = Occupants compost solid waste 10 = Other arrangement	

6.2.1.2 Access to Energy

central_acc

This is a dummy variable that indicates the access to central heating in the dwelling. Categories after harmonization:

0 = No

1 = Yes

heatsource

This is a categorical variable that indicates the main source of heating. Main source of heating refers to the type of system used to provide heating for most of the space. It may be central heating covering all or parts of living quarters, or it may not be central, in which case the heating will be provided separately within the living quarters by a stove, fireplace or some other heating body.

As for the energy used for heating purposes, it is closely related to the type of heating and refers to the predominant source of energy, such as solid fuels (coal, lignite, and products of coal and lignite, wood), oils, gaseous fuels (natural or liquefied gas), or electricity. Main sources of heating, seven categories after harmonization:

1 = Firewood

2 = Kerosene

3 = Charcoal

4 = Electricity

5 = Gas

6 = Central

9 = Other

10= No heating

gas

This is a categorical variable that identifies type of gas usage. The categories after harmonization are:

0 = No

1 = Yes, piped gas (LNG)

2 = Yes, bottled gas (LPG)

3 = Yes, but don't know

cooksource

This is a categorical variable that identifies the source of cooking. If several fuels asked in survey, only main source required. The categories after harmonization are:

1 = Firewood
2 = Kerosene
3 = Charcoal
4 = Electricity
5 = Gas
9 = Other
10=No cook source

- *Firewood* includes both purchased and collected.
- *Electricity* refers to mains, generator and solar energy provided by the government or private entity.
- Other includes fuel derived from coffee waste, saw dust, crop residue, cow dung among others.

lightsource

This is a categorical variable that identifies the source of light. The categories after harmonization are:

1 = Electricity
2 = Kerosene
3 = Candles
4 = Gas
9 = Other
10 = No light source

elec_acc

This is a categorical variable that identifies type of connection to electricity. For instance, access to electricity ('Yes') may be public/quasi-public referring to mains electricity (i.e. the term used to refer to the electricity supply from power stations to households) or private referring to electricity from generator or solar or private company. The quality of electricity is assessed by other Tier 2 variables, such as number of electricity hours per day (ELECHR_ACC). Access to electricity, categories after harmonization

1 = Yes, public/quasi-public
2 = Yes, private
3 = Yes, source unstated
4 = No

Note: Having an electrical connection says nothing about the actual electrical service received by the household in each country or area.

electricity

This is a dummy variable that specifies the access to electricity in the household irrespective of source. Categories after harmonization:

0 = No
1 = Yes

elechr_acc

This is a numeric continuous variable that specifies the access to electricity in hours per day.

electyp

This is a categorical variable that specifies the source of energy when COOKSOURCE and LIGHTSOURCE variables are not available and there is only one question about the type of energy source in the household. When COOKSOURCE and LIGHTSOURCE are available this variable must be created prioritizing electricity, then gas, then lamp. Categories after harmonization:

1 = Electricity
2 = Gas
3 = Lamp
4 = Others
10 = No cook and light source

Table 6.2: Utilities-Access to Energy

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Utilities	**central_acc**	**Access to central heating**	0 = No 1 = Yes	2
2	Utilities	**heatsource**	**Main source of heating**	1 = Firewood 2 = Kerosene 3 = Charcoal 4 = Electricity 5 = Gas 6 = Central 9 = Other 10 = No heating	2
3	Utilities	**gas**	**Connection to gas/Usage of gas**	0 = No 1 = Yes, piped gas (LNG) 2 = Yes, bottled gas (LPG) 3 = Yes, but don't know	2
4	Utilities	**cooksource**	**Main source of cooking fuel**	1 = Firewood 2 = Kerosene 3 = Charcoal 4 = Electricity 5 = Gas 9 = Other 10 = No cook source	1
5	Utilities	**lightsource**	**Main source of lighting**	1 = Electricity 2 = Kerosene 3 = Candles 4 = Gas 9 = Other 10= No light source	1

6	Utilities	**elec_acc**	**Access to electricity**	1 = Yes, public/quasi-public 2 = Yes, private 3 = Yes, source unstated 4 = No	1
7	Utilities	**electricity**	**Access to electricity in dwelling**	0 = No 1 = Yes	1
8	Utilities	**elechr_acc**	**Electricity availability (hr/day)**	Numeric continuous	2
9	Utilities	**electyp**	**Lighting and/or electricity type**	1 = Electricity 2 = Gas 3 = Lamp 4 = Others 10 = No cook and no light source	1

6.2.2 Affordability

6.2.2.1 Essential variables – Water, Sanitation and Hygiene (WASH)

The utilities affordability variables are monetary variables expressed at current prices in the local currency unit (LCU) and non-deflated.

The typology of utilities affordability variables follows the Classification of Individual Consumption According to Purpose (COICOP[14]) developed by the United Nations Statistics Division (UNSD). The objective of COICOP is to provide a framework for comparable classification of homogeneous categories of goods and services, which are considered a function or purpose of household consumption expenditure.

COICOP has 14 consumption categories, of which utility related expenditures are classified under the "4 Housing, water, electricity, gas and other fuels" category. While more detailed information on water and urban services consumption is provided in the "4.4 Water Supply and Miscellaneous Services related to the dwelling" category, expenditure on energy is sub-categorized into "4.5 Electricity, Gas, and other Fuels".

The "Water Supply and Miscellaneous Services (4.4)" group is further sub-divided into the "Water supply (4.4.1)", "Refuse collection (4.4.2)", "Sewage collection (4.4.3)" and "Other services relating to the dwelling (4.4.4)" categories. Similarly, the "Electricity, gas and other fuels (4.5)" group are sub-categorized into "Electricity (4.5.1)", 'Gas (4.5.2)", "Liquid fuels (4.5.3)", "Solid fuels (4.5.4)" and "Heat energy (4.5.5)". According to COICOP, energy-related household consumption is not sub-categories beyond the 3-digit level, but GMD divides them into commonly used categories except electricity expenditures.

[14] The classification of individual consumption by purpose, abbreviated as COICOP, is a classification developed by the United Nations Statistics Division to classify and analyze individual consumption expenditures incurred by households, non-profit institutions serving households and general government according to their purpose. It includes categories such as clothing and footwear, housing, water, electricity, and gas and other fuels.

pwater_exp
This is a continuous variable that refers to total annual household expenditures on water supply/piped water. It includes associated expenditure such as hire of meters, reading of meters, standing charges, etc. Water consumption variables include an aggregate water variable comprising water supply (PWATER_EXP) and hot water (HWATER_EXP) and defined as WATER_EXP.

As in the case of the COICOP classification, the variable excludes household expenditures on hot water. Drinking water sold in bottles or containers is also excluded from water supply.

hwater_exp
This is a continuous variable that refers to total annual household expenditure on hot water supply.

water_exp
This is a continuous variable that refers to total annual household expenditure on water supply and hot water supply. This variable specifies the sum of expenditure of water supply (PWATER_EXP) and hot water supply (HWATER_EXP).

```
egen water_exp=rsum(pwater_exp hwater_exp)
```

garbage_exp
This is a continuous variable that refers to total annual household expenditures on collection and disposal of garbage or refuse.

sewage_exp
This is a continuous variable that refers to total annual household expenditures on collection and disposal of wastewater.

waste_exp
This is a continuous variable that refers to the total annual household expenditure on garbage (GARBAGE_EXP) and sewage (SEWAGE_EXP) collection.

```
egen waste_exp=rsum(garbage_exp sewage_exp)
```

dwelothsvc_exp
This is a continuous variable that refers to total annual household expenditures on other services relating to the dwelling.

These expenditures typically include co-proprietor charges in multi-occupied buildings, security services, and other miscellaneous services. Co-proprietor charges include charges for caretaking, gardening, stairwell cleaning, heating, and lighting, maintenance of lifts and refuse disposal chutes, etc.

This variable does not include household services such as window cleaning, disinfecting, fumigation, and pest extermination[15]; bodyguards[16]. Maintenance and repair of the dwelling[17] is also excluded from other services relating to the dwelling (DWELOTHSVC_EXP) but included as additional variables defined as DWELMAT_EXP and DWELSVC_EXP.

Table 6.3: Utilities Module-Wash, Sanitation and Hygiene Expenditure Variables

	Module code	Variable name	Variable name	Tier
1	Utilities	pwater_exp	**Total annual consumption of water supply**	2
2	Utilities	hwater_exp	**Total annual household consumption of hot water supply**	2
3	Utilities	water_exp*	**Total annual consumption of water supply and hot water**	2
4	Utilities	garbage_exp	**Total annual consumption of garbage collection**	2
5	Utilities	sewage_exp	**Total annual consumption of sewage collection**	2
6	Utilities	waste_exp*	**Total annual consumption of garbage and sewage collection**	2
7	Utilities	dwelothsvc_exp	**Total annual consumption of other services relating to the dwelling**	2

* These are secondary variables that are aggregated using primary variables. However, there might be surveys that report expenditures on secondary level only.

6.2.2.2 Essential variables - Energy[18]

elec_exp
This is a continuous variable that refers to total annual household expenditures on electricity and other associated expenditures such as hire of meters, reading of meters and standing charges.

ngas_exp
This is a continuous variable that refers to total annual household expenditure on town gas and natural gas.

LPG_exp
This is a continuous variable that refers to total annual household expenditure on LPG that includes butane, propane, “bottled gas” etc.

gas_exp
This is a continuous aggregate variable comprised of total annual household expenditures on network/natural gas and liquefied gas (LPG).

Due to differences in characteristics and price patterns, two types of gas are recorded as separate variables under GAS_EXP:

15 Also known as COICOP 5.6.2.
16 Also known as COICOP 12.7.0.
17 Also known as COICOP 4.3
18 It also serves as indices for 5. Affordability for “Access to Cooking Solutions” and “Access to Space Heating” in MTF.

a) Town gas and natural gas (NGAS_EXP); and
b) LPG (liquefied petroleum gas (LPG_EXP): includes butane, propane, “bottled gas”, etc.

Associated expenditure such as hire of meters, reading of meters, storage containers, standing charges, etc. are included in the construction of the variable.

```
egen gas_exp=rsum(ngas_exp LPG_exp)
```

gasoline_exp
This is a continuous variable that refers to total annual household expenditure on gasolines. Use mostly in vehicles and motorcycles.

diesel_exp
This is a continuous variable that refers to total household expenditure on diesel or gasoil. Mostly use on electricity generators, SUV, Trucks, buses, very few sedan cars use this type of fuel.

kerosene_exp
This is a continuous variable that refers to total annual household expenditure on kerosene.

othliq_exp
This is a continuous variable that refers to total annual household expenditure on other liquid fuels such as heating oil, black oil and lighting oil.

liquid_exp
This is a continuous aggregate variable comprised of total annual household expenditures on all liquid fuels.

Liquid fuels are subcategorized into:

- gasoline/petrol (GASOLINE_EXP)
- diesel (DIESEL_EXP_EXP)
- kerosene (KEROSENE_EXP)
- other liquid fuels (OTHLIQ_EXP). Other liquid fuels category includes all other liquid fuels other than diesel and kerosene. Examples include “heating oil”, “black oil” and “lighting oil”.

```
egen liquid_exp=rsum(gasoline_exp diesel_exp kerosene_exp othliq_exp)
```

wood_exp
This is a continuous variable that refers to total annual household expenditure on firewood.

In some countries, this includes imputed cost for collected firewood. Harmonizer check accordingly.

coal_exp
This is a continuous variable that refers to total annual household expenditure on coal.

peat_exp
This is a continuous variable that refers to total annual household expenditure on peat.

othsol_exp
This is a continuous variable that refers to total annual household expenditure on other solid fuels such as agricultural residue and charcoal.

solid_exp
This is a continuous aggregate variable comprised of total annual household expenditures on all solid fuels.

Solid energy is subcategorized into expenditures on

- coal (COAL_EXP)
- firewood (WOOD_EXP)
- peat (PEAT_EXP)
- other solid fuels (OTHSOL_EXP). Other solid fuels category includes all other solid fuels not included in the above three categories. Examples include "pressed dung, corn brans, brushwood", and "other solid".

```
egen solid_exp=rsum(coal_exp wood_exp peat_exp)
```

othfuel_exp
This is a continuous variable that refers to total annual household expenditure on other fuels that are not captured under OTHLIQ_EXP and OTHSOL_EXP.

central_exp
This is a continuous variable that refers to total annual household expenditure on central heating.

heating_exp
This is a continuous aggregate variable comprised of total annual household expenditures on heating. These expenditures can be subcategorized into expenditures on central heating (CENTRAL_EXP) and hot water (HWATER_EXP).

It is worth to note that COICOP narrowly defines heat energy to purchase from district heating plant only, but GMD includes all heat energy from building or other sources.

Note: The expenditure for central heating is frequently combined either with expenditures under hot water or rent. Hot water is also often combined with cold water. Also note that COICOP categorizes hot water under 4.5.5 Heat energy, while cold water is reflected under 4.4.1 Water supply.
Be careful when aggregating this variable to avoid double counting.

utl_exp
This is a continuous aggregate variable comprised of total annual household expenditure on all utilities excluding telecom and other housing expenses. Utilities expenditure in this case is sum of the following variables:

- electricity (ELEC_EXP)
- gas (GAS_EXP)
- liquid fuels (LIQUID_EXP)
- solid fuels (SOLID_EXP)
- central heating (CENTRAL_EXP)
- water (WATER_EXP)

- waste (WASTE_EXP and
- other fuels (OTHFUEL_EXP).

Excludes expenditures for other housing (OTHHOUSING_EXP), fuel for transportation (TRANSFUEL_EXP), telecommunication services (COMM_EXP) and television services (TV_EXP).

```
egen utl_exp=rsum(water_exp waste_exp elec_exp gas_exp liquid_exp
solid_exp central_exp othfuel_exp)
```

Table 6.4: Utilities Module-Energy Expenditure Variables

	Module code	Variable name	Variable label	Tier
1	Utilities	**elec_exp**	**Total annual consumption of electricity**	1
2	Utilities	**ngas_exp**	**Total annual consumption of network/natural gas**	1
3	Utilities	**LPG_exp**	**Total annual consumption of liquefied gas**	1
4	Utilities	**gas_exp***	**Total annual consumption of network/natural and liquefied gas**	1
5	Utilities	**gasoline_exp**	**Total annual consumption of gasoline**	1
6	Utilities	**diesel_exp**	**Total annual consumption of diesel**	1
7	Utilities	**kerosene_exp**	**Total annual consumption of kerosene**	1
8	Utilities	**othliq_exp**	**Total annual consumption of other liquid fuels**	1
9	Utilities	**liquid_exp***	**Total annual consumption of all liquid fuels**	1
10	Utilities	**wood_exp**	**Total annual consumption of firewood**	1
11	Utilities	**coal_exp**	**Total annual consumption of coal**	1
12	Utilities	**peat_exp**	**Total annual consumption of peat**	1
13	Utilities	**othsol_exp**	**Total annual consumption of other solid fuels**	1
14	Utilities	**solid_exp***	**Total annual consumption of all solid fuels**	1
15	Utilities	**othfuel_exp**	**Total annual consumption of all other fuels**	1
16	Utilities	**central_exp**	**Total annual consumption of central heating**	1
17	Utilities	**heating_exp**	**Total annual consumption of hot and cold water**	1
18	Utilities	**utl_exp***	**Total annual consumption of all utilities excluding telecom and other housing, current year prices**	2

* These are aggregated derived secondary variables that are aggregated using primary variables. However, there might be surveys that report expenditures on secondary level only.

6.2.2.3 Additional variables

dwelmat_exp
This is a continuous variable that refers to total annual household expenditures on product and materials for maintenance and repair of the dwelling.

Products and materials for minor maintenance and repair typically include expenditures on paints and varnishes, renderings, wallpapers, fabric wall coverings, windowpanes, plaster, cement, putty, wallpaper pastes. Fitted carpets and linoleum (5.1.2); hand tools, door fittings, power sockets, wiring flex and lamp bulbs (5.5.2); brooms, scrubbing brushes, dusting brushes and cleaning products (5.6.1); products, materials and fixtures used for major maintenance and repair (intermediate consumption) or for extension and conversion of the dwelling (capital formation) are excluded.

dwelsvc_exp
This is a continuous variable that refers to total annual household expenditures on services for minor maintenance and repair of the dwelling.

This variable generally includes expenditures on services of plumbers, electricians, carpenters, glaziers, painters, decorators, floor polishers, etc as well as total value of the service (that is, both the cost of labor and the cost of materials are covered). It excludes separate purchases of materials made by the household with the intention of undertaking the maintenance or repair by themselves (4.3.1); services engaged for major maintenance and repair (intermediate consumption) or for the extension and conversion of the dwelling (capital formation).

othhousing_exp
This is a continuous variable that refers to total annual household expenditures on other materials and services for minor maintenance and repair of the dwelling.

```
egen othhousing_exp=rsum(dwelmat_exp  dwelsvc_exp)
```

transfuel_exp
This is a continuous variable that refers to total annual household expenditures on fuels for personal transportation.

According to COICOP, fuels use for transportation purposes are classified under Fuels and lubricants for personal transport equipment (COICOP 7.2.2). COICOP 7.2.2 also includes lubricants, which are excluded from this GMD indicator.

landphone_exp
This is a continuous variable that refers to total annual household expenditures on land phone.

This includes installation, subscription, and service usage fees. Expenditure on equipment is not included.

cellphone_exp
This is a continuous variable that refers to total annual household expenditures on cellphone.

This includes installation, subscription, and service usage fees. Expenditure on equipment is not included.

tel_exp

This is a continuous aggregate variable comprised of total annual household expenditures on landline phone (LANDPHONE_EXP) and cell phone (CELLPHONE_EXP) which may include

(i) Installation and subscription costs of personal telephone equipment

(ii) telephone calls from a private line or from a public line (public telephone box, post office cabin, etc.); telephone calls from hotels, cafés, restaurants and the like,

(iii) hire of telephones, telefax machines, telephone-answering machines and telephone loudspeakers. Expenditures on relevant equipment are not included.

Telephone and telefax services (COICOP 8.3.0) are subcategorized into 4 categories: landline phone, cell phone, internet and telefax services.

```
egen tel_exp=rsum(landphone cellphone)
```

internet_exp

This is a continuous variable that refers to total annual household expenditures on information transmission and Internet connection services. This variable also includes installation, subscription, and service usage fees and costs, but excludes consumption for equipment. Telefax services (TELEFAX_EXP) includes telegraphy, telex and telefax services, as well as radiotelephony, radiotelegraphy and radio-telex services. Expenditures on relevant equipment are not included[19].

telefax_exp

This is a continuous variable that refers to total annual household expenditures on telegraphy, telex and telefax services. This includes radiotelephony, radiotelegraphy and radio telex services.

comm_exp

This is a continuous variable comprised of total annual household expenditures on all telephone and telefax services, including expenditures on landline phone (LANPHONE_EXP), cell phone (CELLPHONE_EXP), internet (INTERNET_EXP) and telefax services (TELEFAX_EXP).

```
egen comm_exp=rsum(tel_exp internet_exp telefax_exp)
```

tv_exp

This is a continuous variable that refers to total annual household expenditures on television broadcasting services, license fees for television equipment and subscriptions to television networks. This variable is compatible with COICOP 9.4.2 Cultural services but does not include spending on such services as theatres, museums, and historic monuments.

[19] A radiotelephone (or radiotelephone) is a radio communication system for conducting conversations. It is different from radiotelegraphy, which a radio that transmits telegrams (messages), or television, which transmits video and audio.

tvintph_exp

This is a continuous aggregate variable comprised of total annual household expenditures on internet (INTERNET_EXP), telephone (TEL_EXP) and television broadcasting services (TV_EXP).

Table 6.5 below provides the summary of all the utilities expense variables. The variables highlighted in yellow are secondary variables that are aggregated using primary variables. However, there might be surveys that report expenditures on secondary level only. For example: waste expenditure (WASTE_EXP) is sum of garbage expenditure (GARBAGE_EXP) and sewage expenditure (SEWAGE_EXP). In surveys where expenditures are reported on disaggregated level will include values for garbage expenditure and sewage expenditure and then WASTE_EXP is created by adding garbage and sewage expenditures. However, some surveys will report expenditure only for total waste i.e. WASTE_EXP, leading to missing values for GARBAGE_EXP and SEWAGE_EXP.

```
egen tvintph_exp=rsum(tel_exp internet_exp tv_exp)
```

Table 6.5: Utilities Module-Additional Expenditure Variables

	Module code	Variable name	Variabel label	Tier
1	Utilities	**dwelmat_exp**	**Total annual consumption of materials for the maintenance and repair of the dwelling**	2
2	Utilities	**dwelsvc_exp**	**Total annual consumption of services for the maintenance and repair of the dwelling**	2
3	Utilities	**othhousing_exp***	**Total annual consumption of maintenance and repair of the dwelling**	2
4	Utilities	**transfuel_exp**	**Total annual consumption of fuels for personal transportation**	2
5	Utilities	**landphone_exp**	**Total annual consumption of landline phones**	2
6	Utilities	**cellphone_exp**	**Total annual consumption of cell phones**	1
7	Utilities	**tel_exp***	**Total annual consumption of all phones**	2
8	Utilities	**internet_exp**	**Total annual consumption of internet**	1
9	Utilities	**telefax_exp**	**Total annual consumption of other telefax services**	2
10	Utilities	**comm_exp***	**Total annual consumption of telecommunication services**	2
11	Utilities	**tv_exp**	**Total annual consumption of television broadcasting services**	2
12	Utilities	**tvintph_exp***	**Total annual consumption of television, internet and telephone services**	2

* These are aggregated derived secondary variables that are aggregated using primary variables. However, there might be surveys that report expenditures on secondary level only.

6.3 Challenges and Common Mistakes

6.3.1 Data harmonization

The main challenge in this module is inconsistency between what is available in the survey and variable and allowed codes of the harmonization. This needs to be considered carefully in the harmonization process and will differ between the affordability and access submodules.

Distinguishing missing, value 0 and skip patterns also needs careful attention. Some variables or questionnaires may contain 0 in the raw data when the value should be missing. y. This could lead to underestimation of the variable for the survey, distorting the snapshot of energy-related information. Skip patterns should also be distinguished from both missing and value 0.

Water

This variable is created with the help of the WASH team (Luis Alberto Andres landres@worldbank.org), the poverty GP is in charge of proposing the classification of the categories, the WASH team will be in charge of reviewing the proposal, then approved it or recommend another classification following the next work flow.

1. The Regional TSD will fill out the excel template including all categories for water access and the actual question in the questionnaire as it is shown below, each country year must be submitted in a separate worksheet.

Improved water

Harmonization

Category	Share of population (%)			Improved	
	National	Rural	Urban	Recommended estimate	WaterGP review
Centralize water supply	97.39	93.20	99.76	Yes	Y
Spring Water, wells	0.75	1.78	0.16	Yes	
Own system of water supply	1.40	3.81	0.04	Yes	
River, lake	0.00	0.00	0.00	No	
Delivered (imported) water	0.10	0.25	0.02	Yes	
Bought water	0.01	0.04	0.00	Yes	
Rainwater	0.00	0.00	0.00	Yes	
Other	0.35	0.93	0.02	No	
Missing	0.00	0.00	0.0	Missing	Missing

Share improved	Share of population (%)			JMP National	JMP Rural	JMP Urban
	National	Rural	Urban			
Recommended estimate	99.65	99.07	99.98			

Snippet from questionnaire

C.13. What are the sources of water you use?

1. centralized water supply
2. spring water, wells ⇨ C. 16.
3. own system of water supply
4. river, lake ⇨C.16.
5. delivered (imported) water ⇨ C. 16.
6. bought water (Noy, Byuregh, etc) ⇨ C. 16.
7. rainwater ⇨ C. 16.
8. other _____(specify) ⇨ C. 16

C.14. Where is your water tap located?
(respondents are points 1 and 3 of item C. 13)

1. in door
2. in the yard
3. in the street

2. The WASH team reviews and proposes a classification in the section “Water GP review”

Improved water

Harmonization

Category	Share of population (%)			Improved	
	National	Rural	Urban	Recommended estimate	WaterGP review
Centralize water supply	97.39	93.20	99.76	Yes	Yes
Spring Water, wells	0.75	1.78	0.16	Yes	Yes
Own system of water supply	1.40	3.81	0.04	Yes	Yes
River, lake	0.00	0.00	0.00	No	No
Delivered (imported) water	0.10	0.25	0.02	Yes	Yes
Bought water	0.01	0.04	0.00	Yes	Yes
Rainwater	0.00	0.00	0.00	Yes	Yes
Other	0.35	0.93	0.02	No	No
Missing	0.00	0.00	0.0	Missing	Missing

Share improved	Share of population (%)			JMP National	JMP Rural	JMP Urban
	National	Rural	Urban			
Recommended estimate	99.65	99.07	99.98	100%	100%	100%

Snippet from questionnaire

C.13. What are the sources of water you use?

1. centralized water supply
2. spring water, wells ⇨ C. 16.
3. own system of water supply
4. river, lake ⇨C.16.
5. delivered (imported) water ⇨ C. 16.
6. bought water (Noy, Byuregh, etc) ⇨ C. 16.
7. rainwater ⇨ C. 16.
8. other _____(specify) ⇨ C. 16

C.14. Where is your water tap located?
(respondents are points 1 and 3 of item C. 13)

1. in door
2. in the yard
3. in the street

3. Regional TSD will create the harmonized variables based on the recommendation of WASH team (based on “Water GP review”).

Sanitation

This variable is created with the help of the WASH team (Luis Alberto Andres landres@worldbank.org), the poverty GP is in charge of proposing the classification of the categories, the WASH team will be in charge of reviewing the proposal, then approved it or recommend another classification following the next work flow.

1. The Regional TSD will fill out the excel template including all categories for access to sanitation and the actual question in the questionnaire as it is shown below, each country year must be submitted in a separate worksheet.

Improved sanitation

Harmonization

Category	Share of population (%)			Improved	
	National	Rural	Urban	Recommended	WaterGP review
Flush Toilet	2.79	0.53	7.49	Y	
VIP Latrine	6.85	3.50	13.81	Y	
Covered Pit Latrine with a slab	27.73	20.01	43.79	Y	
Covered Pit Latrine without a slab	39.29	47.15	22.95	N	
Uncovered Pit Latrine with a slab	3.02	3.01	3.03	Y	
Uncovered Pit Latrine without a slab	11.22	13.97	5.52	N	
Ecosan (compost toilet)	0.27	0.26	0.30	Y	
No facility/bush/ polythene bags/ bucket/ et	8.49	11.21	2.82	N	
Other (specify)	0.34	0.36	0.30	N	
Missing	0.00	0.00	0.00	Missing	

Share improved	Share of population (%)			JMP National	JMP Rural	JMP Urban
	National	Rural	Urban			
Recommended estimate	40.7	27.3	68.4			
- unshared	18.9	16.3	24.4			

Snippet from questionnaire

What type of toilet facility does this household mainly use?

01= Flush Toilet
02= VIP Latrine
03= Covered Pit Latrine with a slab
04= Covered Pit Latrine without a slab
05= Uncovered Pit Latrine with a slab
06= Uncovered Pit Latrine without a slab
07= Ecosan (compost toilet)
08= No facility/bush/ polythene bags/ bucket/ etc.
96= Other (specify)

Does the household share this toilet facility with other households?

1= Yes
2= No (>> HC17)
97= N/A (For “No facility/bush/ polythene bags/bucket/ etc). (>> HC18)

2. The WASH team reviews and proposes a classification in the section “Water GP review”

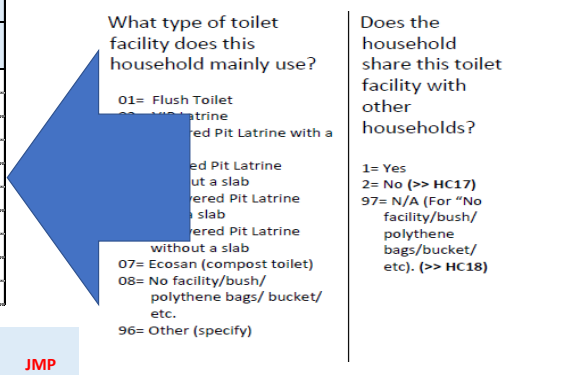
Improved sanitation

Harmonization

Category	Share of population (%)			Improved	
	National	Rural	Urban	Recommended	WaterGP review
Flush Toilet	2.79	0.53	7.49	Y	Y
VIP Latrine	6.85	3.50	13.81	Y	Y
Covered Pit Latrine with a slab	27.73	20.01	43.79	Y	Y
Covered Pit Latrine without a slab	39.29	47.15	22.95	N	N
Uncovered Pit Latrine with a slab	3.02	3.01	3.03	Y	Y
Uncovered Pit Latrine without a slab	11.22	13.97	5.52	N	N
Ecosan (compost toilet)	0.27	0.26	0.30	Y	Y
No facility/bush/ polythene bags/ bucket/ et	8.49	11.21	2.82	N	N
Other (specify)	0.34	0.36	0.30	N	N
Missing	0.00	0.00	0.00	Missing	Missing

Share improved	Share of population (%)			JMP National	JMP Rural	JMP Urban
	National	Rural	Urban			
Recommended estimate	40.7	27.3	68.4			
- unshared	18.9	16.3	24.4	19%	16%	26%

Snippet from questionnaire



3. Regional TSD will create the harmonized variables based on the recommendation of WASH team (based on “Water GP review”).

PIPED

```
recode water_source (1/3=1) (4/14=0), gen(piped)
```

PIPED_TO_PERM

```
recode water_source (1/2=1) (3/14=0), gen(piped_to_perm)
```

IMP_WAT_REC

```
recode water_source (1/8=1) (9/14=0),gen(imp_wat_rec)
```

OPEN_DEF

```
recode toilet6 (1/4 9=0) (5=1), gen(open_def)
```

When COOKSOURCE and LIGHTSOURCE are available, ELECTYP can be created using the following code:

```
gen electyp=.
replace electyp=1 if cooksource==4 | lightsource==1
replace electyp=2 if (cooksource==5 | lightsource==4) & mi(electyp)
replace electyp=3 if (cooksource==2 | inlist(lightsource,2,3)) &
mi(electyp)
replace electyp=4 if inlist(cooksource,1,3,9) | lightsource==9) &
mi(electyp)
```

```
ta electricity elecsource
```

6.3.2 Affordability

The harmonization process must carefully consider Inconsistencies between what is available in the survey and variable and the codes specified in the harmonization. To solve this problem, the affordability submodule includes several variables that are defined as a sum of other variables. For instance, some surveys provide separate expenditures for solid fuels such as wood or coal, while others provide total expenditure on solid fuels only. The GMD provides both levels of expenditures variables such as solid fuel, wood and coal, so that surveys with more detailed information retains their information in the harmonization and also can be easily compared with other surveys without the information.

How to annualize monthly or quarterly expenditures while taking seasonality into account is also a point of discussion. GMD annualizes variables by annualizing expenditures. For example, if the value is from 3 months recall, it will be annualized by multiplying 4. Though this is unlikely to cause when the surveyed month is evenly distributed across the year. But it could cause biases when considering households interviewed at a particular time. When analyzing these variables, it is typically useful to interpret them in the context of the interview month.

Valuing items that were not purchased – such as collected, received as gift or in-kind – is also challenging. GMD adds the value of purchased items and non-purchased items for the expenditure for the item when the value of non-purchased item is given. However, when only other measures, such as quantity, weight, or volume is given for non-purchased items, it is not calculated as a part of the expenditure.

6.3.2.1 Access to Services

Recommendations or framework for utilities access to variables distinguishes access to services within and outside the premise, though many existing surveys do not provide the item. GMD allowed codes include a third category that “unstated” the raw data do not provide the location of the access point for the utility.

Box 2: Countries with extreme seasonality in utilities consumption

Notice that the consumption aggregate should reflect the *normal* consumption of a household over the period considered (in our case, one year). In some countries, and particularly in the ECA region, there is an extreme fluctuation of energy consumption between winter and summer. Beware that households interviewed over the summer months might have a very different consumption than households interviewed during winter months (for example, in Tajikistan). Nevertheless, under the assumption that households are interviewed throughout the year and assuming that there is no sample bias, this should not be an issue.

Box 3: Inclusion of self-collected or received in-kind items as a part of consumption

Energy expenditures include both purchased items and also self-collected or received in-kind items when they are available and are reported in monetary terms. Non-purchased items that are only reported in quantity, volume or any other non-monetary terms are not valued in monetary terms to be added as a part of energy consumption. Estimating value per unit for monetary conversion can be biased, even if value per unit purchased for the same item for the same household is given, due to differences in quality, inaccuracy in unit measurement, and other factors not listed here.

Box 4: Note on access to services inferred from expenditures

When there is no direct question on access to services, expenditure on the item if often used as a proxy for the access. However, it should be noted that access inferred from expenditure generally underestimates actual access, because it leaves out cases such as when the household do not pay for the item due to the subsidy, or the expense is charged along with other expenditure such as rent.

Not all surveys have detailed information that would allow for better understanding of access to water and sanitation. For example, round trip collection time and location for water and whether facilities are shared and if there is safe disposal of excreta for sanitation faculties, would allow for better understanding of the service levels.

Combine all Sections 6 and run the dofile ***MODULE_06_Utilities.do*** which labels (variable and value) and order variables.

- ✓ This is a must process with no exception.

7 Assets and Dwellings (DWL)

7.1 Framework for Harmonization

The Assets and Dwellings module contains information on housing conditions and asset ownership. The chapter introduces the concepts, definitions, and data requirements for examining asset/dwelling ownership and condition and provide guidance for producing statistics on housing indicators that are commonly derived from household survey data sets or that can be constructed using existing variables. The overall objective of this indicator harmonization is to generate a unified data source of globally comparable indicators for assets and dwellings to support routine analytical and corporate business functions of the World Bank.

Since there is no commonly agreed-upon framework for assets and dwellings survey indicators, these guidelines draw on various major projects and internationally accepted principles for the collection and production of assets and dwellings statistical indicators based on household survey data. The mentioned harmonization projects include United Nations Principles and Recommendations for Population and Housing Censuses Revision 3 (2015), World Bank Living Standards Measurement Study (LSMS), International Income Distribution Database (I2D2) and other relevant regional and global harmonization efforts of the World Bank are referenced as appropriate.

Note that the GMD Dwellings and Assets module adopts narrow definition of Housing/Dwelling, and only include dwelling description and ownership variables in the dwellings module among the following five categories commonly used by UN, WB and Eurostat: 1) Description of Dwellings 2) Dwelling services 3) Dwelling expenditures 4) Assets in their Housing/Dwellings module and 5) Others.

This module consists of Dwellings and Assets subsections. The Dwellings section covers variables that describe dwelling condition and ownership, which can be categorized into three groups covering (i) materials, (ii) facilities and characteristics and (iii) ownership.[20] The Assets section is sub-divided into three parts: (i) household appliances, (ii) means of transportation and (iii) household animals.

Assets serve multiple functions. In their productive capacity, they generate income and facilitate access to capital and credit. They also strengthen a household's capacity to cope with and respond to shocks by enhancing its ability to diversify income and ease liquidity constraints. Moreover, assets comprise a store of wealth that can be sold to generate income or passed on to future generations. Finally, assets may provide status and security to individuals or households. Assets are therefore an important indicator of economic welfare that is complementary to consumption or income.

Understanding assets and dwellings is also critical to forming indicators under the SDGs. SDG 1 and 5, poverty and gender respectively, both require data on land access and ownership. For example, SDG 1.4 is requires knowing the proportion of total adult population with secure tenure rights to land, with legally recognized documentation and who perceive their rights to land as secure, by sex and by type of tenure.

[20] Note that access to and expenditure on utilities services, such as energy, water and urban services are included in GDM Utilities module. Rent and/or imputed rent is discussed in GDM Welfare.

Indicator 5.a.1 requires further information, including the proportion of total agricultural population with ownership or secure rights over agricultural land, disaggregated by sex. Proxies for these indicators can be created using harmonized data in this module.

7.2 Mapping and Description of Variables

GMD Assets and Dwellings module contains a large number of metadata that provides a wealth of information about variables, including their types, descriptions, and sources. To improve readability, only the most significant information has been included in this section.

In this module, the primary unit of analysis is the level of household.

7.2.1 Assets

Assets consist of household appliances and means of transportation assessed at the household level. Household animals are tier 3 indicators and are not included in this version.

7.2.1.1 Appliances

This section refers to the presence of a functioning appliance irrespective of who owns it within the household. It also does not take the number of items into account within the household. Do not guestimate that appliance exists within household based on other information such as expenditure.

landphone
This is a dummy variable indicating whether the household owns a landline phone. It is generally defined as landline phone, home telephone, or fixed phone. Two categories after harmonization:

0 = No
1 = Yes

cellphone
This is a dummy variable indicating whether anyone in the household owns a cell phone. Two categories after harmonization:

0 = No
1 = Yes

phone
This is a dummy variable indicating whether the household owns *either a land phone or a cell phone*. It should only be coded in cases where the survey does not distinguish between ownership of landline and cell phones. In other cases, it may be coded as missing. Two categories after harmonization:

0 = No
1 = Yes

computer
This is a dummy variable indicating whether the household owns a computer, including desktop and laptop computer. Two categories after harmonization:

0 = No
1 = Yes

etablet
This is a dummy variable indicating the ownership of an electronic tablet. Two categories after harmonization:

0 = No
1 = Yes

internet
This is a categorical variable indicating whether anyone in the household can use a device that is connected to the internet within the home or have access to internet outside the house. Connection to the Internet can be both wired and wireless and does not depend on who manages it within the household. Four categories after harmonization:

1 = Subscribed in the house
2 = Accessible outside the house (includes internet cafes and smartphones with internet access)
3 = Either (Use this category when the questionnaire does not specify whether the access is in the house or outside the house)
4 = No internet

radio
This is a dummy variable indicating whether the household owns a radio (i.e. radio, radio cassette, and 3-in-1 radio cassette player). Two categories after harmonization:

0 = No
1 = Yes

tv
This is a dummy variable indicating whether the household owns a TV set. This includes both color and black and white TVs. Two categories after harmonization:

0 = No
1 = Yes

tv_cable
This is a dummy variable indicating whether the household owns a cable or dish antenna services. Only for households that reported having a television (TV=1). Two categories after harmonization:

0 = No
1 = Yes

video
This is a dummy variable indicating whether the household owns a video cassette player and/or video cassette recorder (VCR). It also digital video disc (DVD) player. Two categories after harmonization:

0 = No
1 = Yes

fridge
This is a dummy variable indicating whether the household owns a refrigerator (i.e. refrigerator or freezer). It does not include cooler, icebox, or ice chest. Two categories after harmonization:

0 = No
1 = Yes

sewmach
This is a dummy variable indicating whether the household owns a sewing machine. Two categories after harmonization:

0 = No
1 = Yes

washmach
This is a dummy variable indicating whether the household owns a machine for washing clothes and household linen; but does not include non-electric washing machine. Two categories after harmonization:

0 = No
1 = Yes

stove
This is a dummy variable indicating whether the household owns a stove. Stove generally refers to a portable or fixed apparatus that burns fuel or uses electricity to provide heat for cooking or heating purposes and includes a cooker (stove). Two categories after harmonization:

0 = No
1 = Yes

ricecook
This is a dummy variable indicating whether the household owns a rice cooker. Two categories after harmonization:

0 = No
1 = Yes

fan
This is a dummy variable indicating whether the household owns a fan operated by electricity. Two categories after harmonization:

0 = No
1 = Yes

ac
This is a dummy variable indicating whether the household owns a central or wall air conditioner. It includes both air conditioners (both wall and central ACs) and air coolers. Air coolers use simple water to

cool the air while air conditioners use a chemical coolant to cool the air as it is drawn into the device Two categories after harmonization:

0 = No

1 = Yes

ewpump

This is a dummy variable indicating the ownership of an electric water pump. Two categories after harmonization:

0 = No

1 = Yes

Table 7.1: Asset ownership--appliances

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	ID	**countrycode**	**country code**	String See Table 2.1	1
2	ID	**year**	**Year**	Numeric See Table 2.1	1
3	ID	**hhid**	**Household identifier**	String See Table 2.1	1
4	ID	**weight**	**Weight**	Numeric See Table 2.1	1
5	Assets	**landphone**	**Ownership of a land phone**	0 = No 1 = Yes	1
6	Assets	**cellphone**	**Ownership of a cell phone**	0 = No 1 = Yes	1
7	Assets	**phone**	**Ownership of a telephone**	0 = No 1 = Yes	1
8	Assets	**computer**	**Ownership of a computer**	0 = No 1 = Yes	1
9	Assets	**etablet**	**Ownership of a electronic tablet**	0 = No 1 = Yes	1
10	Assets	**internet**	**internet connection**	1 = Subscribed in the house 2 = Accessible outside the house 3 = Either 4 = No internet	1
11	Assets	**radio**	**Ownership of a radio**	0 = No 1 = Yes	1
12	Assets	**tv**	**Ownership of a tv**	0 = No 1 = Yes	1

13	Assets	**tv_cable**	**Ownership of a cable tv**	0 = No 1 = Yes	1
14	Assets	**video**	**Ownership of a video**	0 = No 1 = Yes	1
15	Assets	**fridge**	**Ownership of a refrigerator**	0 = No 1 = Yes	1
16	Assets	**sewmach**	**Ownership of a sewing machine**	0 = No 1 = Yes	1
17	Assets	**washmach**	**Ownership of a washing machine**	0 = No 1 = Yes	1
18	Assets	**stove**	**Ownership of a stove**	0 = No 1 = Yes	1
19	Assets	**ricecook**	**Ownership of a rice cooker**	0 = No 1 = Yes	2
20	Assets	**fan**	**Ownership of an electric fan**	0 = No 1 = Yes	2
21	Assets	**ac**	**Ownership of an air conditioner**	0 = No 1 = Yes	1
22	Assets	**ewpump**	**Ownership of a electric water pump**	0 = No 1 = Yes	2

7.2.1.2 Means of transportation

This section refers to the presence of a transport equipment that is for household use. The transport equipment must be functioning irrespective of who owns it within the household. It also does not take the number of items into account within the household. Do not guestimate that appliance exists within household based on other information such as expenditure.

Note: Any equipment used for commercial must be excluded.

car

This is a dummy variable indicating whether the household owns a car or truck for household use. This refers to car for household use and NOT a commercial vehicle. Two categories after harmonization:

0 = No

1 = Yes

mcycle

This is a dummy variable indicating whether the household owns a motorcycle. Motorcycle refers to an automotive vehicle with two in-line wheels, including motorbike or moped. Two categories after harmonization:

0 = No
1 = Yes

bcycle
This is a dummy variable indicating whether the household owns a bicycle. Note that motored bicycles are classified as motorcycle regardless of motor type. Two categories after harmonization:

0 = No
1 = Yes

oxcart
oxcart is a dummy variable indicating whether the household owns an animal cart, which is used as a means of transport or a farm tool. Two categories after harmonization:

0 = No
1 = Yes

boat
This is a dummy variable indicating whether the household owns a boat. This refers to boat for household use and NOT for commercial use. Two categories after harmonization:

0 = No
1 = Yes

canoe
This is a dummy variable indicating the ownership of a canoe. This refers to canoe for household use and NOT for commercial use. Two categories after harmonization:

0 = No
1 = Yes

Table 7.2: Asset ownership—means of transport

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Assets	**car**	**Ownership of a Car**	0 = No 1 = Yes	1
2	Assets	**bcycle**	**Ownership of a bicycle**	0 = No 1 = Yes	1
3	Assets	**mcycle**	**Ownership of a motorcycle**	0 = No 1 = Yes	1
4	Assets	**oxcart**	**Ownership of an oxcart**	0 = No 1 = Yes	1
5	Assets	**boat**	**Ownership of a boat**	0 = No 1 = Yes	1
6	Assets	**canoe**	**Ownership of a canoes**	0 = No 1 = Yes	1

7.2.2 Main Dwelling

Dwellings variables are mapped into three sections: Materials; Facilities and Characteristics; and Ownership. Materials, characteristics, ownership variables are categorical variables, and most facility variables are YES/NO dummies. All variables are assessed at household level.

7.2.2.1 Materials

The best possible match is sought, but in many cases the correspondence between country-specific values and these standardized codes is imperfect. If a perfect match is not found, users could classify the type of roofs as Natural-Other, Rudimentary-Other, Finished-Other or ultimately as Other – "Specific".

All variables are assessed at household level.

roofcs
This is a string variable. If more than one material is used for structure, the dominant material is the information required. Variables must be translated into English. 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 - Stone"; "2 - Mud"; etc.

roof
This is a categorical variable that indicates ***type of material used for roof***, such as adobe, thatch, iron, and tiles. The roof material is categorized into 3 broad categories namely: Natural, rudimentary, and finished. For cases that cannot be covered in the above three categories, please use code 96 = Other – "Specific".
Main source of material used for roof:

11 = Natural - No roof
12 = Natural - Thatch/palm leaf
13 = Natural - Sod
14 = Natural - Other
21 = Rudimentary - Rustic mat
22 = Rudimentary - Palm/bamboo
23 = Rudimentary - Wood planks
24 = Rudimentary - Other
31 = Finished - Wood
32 = Finished - Asbestos
33 = Finished - Ceramic tile
34 = Finished - Cement
35 = Finished - Metal tile
36 = Finished - Roofing shingles
37 = Finished - Other
96 = Other

- *Adobe, wattle, mud* includes all building techniques that rely on earth or mud put over a frame or mixed with other materials for strength.

- *Thatch* includes grass or any form of natural vegetation for roofing.
- *Iron or metal sheets* are processed tin, zinc, and the like
- *Cement* includes concrete and stone and cement blocks.
- *Tiles/bricks* include baked bricks.
- *Other* includes tin, cardboard among others.

wallcs

This is a string variable. If more than one material is used for structure, the dominant material is the information required. Variables must be translated into English. 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 - Stone"; "2 - Mud"; etc.

wall

This is a categorical variable that indicates ***type of material used for walls***. The wall material is categorized into 3 broad categories namely: Natural, rudimentary, and finished. For cases that cannot be covered in the above three categories, please use code 96 = Other – "Specific". Main source of material used for walls, 21 categories after harmonization:

11 = Natural - No wall
12 = Natural - Cane/palm/trunks
13 = Natural - Dirt
14 = Natural - Other
21 = Rudimentary - Bamboo with mud
22 = Rudimentary - Stone with mud
23 = Rudimentary - Uncovered adobe
24 = Rudimentary - Plywood
25 = Rudimentary - Cardboard
26 = Rudimentary - Reused wood
27 = Rudimentary - Other
31 = Finished - Woven Bamboo
32 = Finished - Stone with lime/cement
33 = Finished - Cement blocks
34 = Finished - Covered adobe
35 = Finished - Wood planks/shingles
36 = Finished - Plaster wire
37 = Finished - GRC/Gypsum/Asbestos
38 = Finished - Other
96 = Other

- *Adobe, wattle, mud* includes all building techniques that rely on earth or mud put over a frame or mixed with other materials for strength.
- *Bricks* include baked bricks.
- *Wood* includes timber and wood planks, unfinished.
- *Iron /metal sheets* are processed tin, zinc and the like.

- *Cement* includes concrete and stone and cement block.
- *Other* includes tin, cardboard among others.

floorcs

this is a string variable. If more than one material is used for structure, the dominant material is the information required. Variables must be translated into English. 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 - Stone"; "2 - Mud"; etc.

floor

This is a categorical variable that indicates ***type of material used for floors***. The floor material is categorized into 3 broad categories namely: Natural, rudimentary, and finished. For cases that cannot be covered in the above three categories, please use code 96 = Other – "Specific". Main source of material used for floors, 14 categories after harmonization as shown below.

11 = Natural – Earth/sand

12 = Natural – Dung
13 = Natural – Other
21 = Rudimentary – Wood planks
22 = Rudimentary – Palm/bamboo
23 = Rudimentary – Other
31 = Finished – Parquet or polished wood
32 = Finished – Vinyl or asphalt strips
33 = Finished – Ceramic/marble/granite
34 = Finished – Floor tiles/terrazzo
35 = Finished – Cement/red bricks
36 = Finished – Carpet
37 = Finished – Other
96 = Other

- *Earth* implies dirt or mud floors.
- *Bricks* include baked bricks.
- *Polished wood/tiles* include finished wood floors, parquet floors, as well as ceramic tiles.
- *Terrazzo* is a composite material, poured in place or precast. It is used for floor and wall treatments and consists of chips of marble, quartz, granite, glass, or other suitable material, poured with a cementitious binder, polymeric, or a combination of both.

Table 7.3: Main Dwelling-Materials

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Dwelling	**roofcs**	**Main material used for roof (country-specific)**	String variable	1
2	Dwelling	**roof**	**Main material used for roof**	11 = Natural - No roof 12 = Natural - Thatch/palm leaf	1

				13 = Natural - Sod 14 = Natural - Other 21 = Rudimentary - Rustic mat 22 = Rudimentary - Palm/bamboo 23 = Rudimentary - Wood planks 24 = Rudimentary - Other 31 = Finished - Wood 32 = Finished - Asbestos 33 = Finished - Ceramic tile 34 = Finished - Cement 35 = Finished - Metal tile 36 = Finished - Roofing shingles 37 = Finished - Other 96 = Other	
3	Dwelling	**wallcs**	**Main material used for wall (country-specific)**	String variable	1
4	Dwelling	**wall**	**Main material used for external walls**	11 = Natural - No wall 12 = Natural - Cane/palm/trunks 13 = Natural - Dirt 14 = Natural - Other 21 = Rudimentary - Bamboo with mud 22 = Rudimentary - Stone with mud 23 = Rudimentary - Uncovered adobe 24 = Rudimentary - Plywood 25 = Rudimentary - Cardboard 26 = Rudimentary - Reused wood 27 = Rudimentary - Other 31 = Finished - Woven Bamboo 32 = Finished - Stone with lime/cement 33 = Finished - Cement blocks 34 = Finished - Covered adobe 35 = Finished - Wood planks/shingles 36 = Finished - Plaster wire 37 = Finished - GRC/Gypsum/Asbestos 38 = Finished - Other 96 = Other	1
5	Dwelling	**floorcs**	**Main material used for floor (country-specific)**	String variable	1
6	Dwelling	**floor**	**Main material used for floor**	11 = Natural - Earth/sand 12 = Natural - Dung 13 = Natural - Other 21 = Rudimentary - Wood planks 22 = Rudimentary - Palm/bamboo 23 = Rudimentary - Other 31 = Finished - Parquet or polished wood 32 = Finished - Vinyl or asphalt strips 33 = Finished - Ceramic/marble/granite	1

				34 = Finished - Terrazzo 35 = Finished - Cement/red bricks 36 = Finished - Carpet 37 = Finished - Other 96 = Other	

7.2.2.2 Facilities characteristics

All variables are assessed at household level. This refers to the Section 7.2.2.1 dwelling characteristics.

dweltyp

This is a categorical variable that specifies the type of dwelling unit household lives in.

1 = *Detached house*
2 = Multi-family house
3 = Separate apartment
4 = Communal apartment
5 = Room in a larger dwelling
6 = Several buildings connected
7 = Several separate buildings
8 = Improvised housing unit
9 = Other

- *A Separate apartment* is a self-contained apartment and is not shared among families.
- A *Communal apartment* is shared by two or more families. Each family has its own room, which serves as a living room, dining room, and bedroom for the entire family. The hallways, kitchen, bathroom, and telephone are shared among all the residents.

typlivqrt

This is a categorical variable that specifies the type of living quarters. Categories after harmonization are:

1 = Housing units, conventional dwelling with basic facilities
2 = Housing units, conventional dwelling without basic facilities
3 = Other housing units

kitchen

This is a dummy variable indicating whether the household has a separate kitchen in the dwelling*, implying an independent space is set aside for cooking inside the dwelling* (kitchen). Any other space reserved for cooking, such as kitchenette or an outer space for kitchen, is not considered as a kitchen. The unit of enumeration for this topic is the housing unit. However, some countries may find it useful to collect information on the availability of kitchen facilities for the use of occupants in collective living quarters, such as hotels, lodging houses, institutions camps and workers' quarters, though people living in these places are generally not captured in a household survey. Two categories after harmonization:

0 = No
1 = Yes

bath

This is a dummy variable indicating whether the household has a separate bathing facility such a shower or bathroom in the dwelling. Fixed bath or shower outside housing unit is not considered. Two categories after harmonization:

0 = No
1 = Yes

rooms

This is an integer variable that refers to the number of habitable rooms in the whole household dwelling unit. It may consist of one or more structure(s) (rooms), including all rooms used for living, sleeping, and eating. It excludes storerooms, bathrooms, kitchens, and rooms used for business or professional purposes. In the case of a one-room dwelling this variable will have the value of one. Must be >=1. Zero rooms are an outlier.

areaspace

This is a continues variable that refers to the total floor area (in square meters) of all rooms and auxiliary premises (kitchen, vestibule, cloakroom, hallway, toilet room, sauna that is within the dwelling, pantry, interstice, bathroom, storeroom, porch, integrated wall closets) in the whole household dwelling unit.

The area of the dwelling does not include cellars, garages (incl. in private houses), boiler rooms, attics (if they are not suitable for permanent habitation) and common rooms (such as stairways, corridors, saunas, etc.) in buildings with multiple dwellings. Open areas (loggias, balconies and terraces) are not included in the area of the dwelling. However, if such areas have been closed in and insulated, they should be added to the total area of the dwelling. If a household lives in an uncompleted residential building, enter the area of the finished part of the house.

ybuilt

This is an integer variable that indicates the year when the dwelling was built. This information should be gathered, when available, for all HHs, irrespective of ownership status.

Table 7.4: Main Dwelling-Facilities Characteristics

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Dwelling	**dweltyp**	**Type of dwelling**	1 = Detached house 2 = Multi-family house 3 = Separate apartment 4 = Communal apartment 5 = Room in a larger dwelling 6 = Several buildings connected 7 = Several separate buildings 8 = Improvised housing unit 9 = Other	1

2	Dwelling	**typlivqrt**	**Types of living quarters**	1 = Housing units, conventional dwelling with basic facilities 2 = Housing units, conventional dwelling without basic facilities 3 = Other housing units	1
3	Dwelling	**kitchen**	**Separate kitchen in dwelling**	0 = No 1 = Yes	1
4	Dwelling	**bath**	**Bathing facility in the dwelling**	0 = No 1 = Yes	1
5	Dwelling	**rooms**	**Number of habitable rooms**	Numeric, continuous	1
6	Dwelling	**areaspace**	**Area of main household space (sq. meters)**	Numeric, continuous	1
7	Dwelling	**ybuilt**	**4-digit year the dwelling built**	Numeric, continuous	1

7.2.2.3 Ownership

ownhouse

This is a categorical variable that specifies whether a household owns, rents, is provided for free, or squats in their house.

This variable has four categories after harmonization:

1 = Ownership/ secure rights
2 = Renting
3 = Provided for free
4 = Without permission

- Ownership includes ownership or other equivalent of secure tenure, whether or not full payment has been made yet.
- Rental denotes that regular payment is made to the owner (which could be private, corporate, or government) with or without formal agreement.

acqui_house

This is a categorical variable that specifies the mode of acquisition for their dwellings. Only for household owners (OWNHOUSE=1). Acquisition of house, three categories after harmonization:

1 = Purchased
2 = Inherited
3 = Other

dwelownlti

This is a dummy variable specifying whether a household has legal evidence for ownership (YES/NO). See Lessons learned/Challenges in the next section for more information on what can be considered legal evidence. Two categories after harmonization:

0 = No

1 = Yes

fem_dwelownlti

This is a dummy variable that specifies whether the names of female household members are listed on the legal document specifying ownership of the dwelling (YES/NO). This will be derived from questions asking about the roster ID of the household member(s) whose name(s) are on the legal document for the dwelling. Two categories after harmonization:

0 = No

1 = Yes

dwelownti

This is a categorical variable that specifies the type of legal document the household has as evidence for ownership of their dwelling. Type of legal document, six categories after harmonization:

1 = Title, deed, freehold

2 = Government issued leasehold

3 = Occupancy certificate – govt issued

4 = Legal document in the name of group (community; cooperative)

5 = Condominium (apartment)

6 = Other

selldwel

This is a dummy variable that specifies whether the respondent has alienation rights (i.e. the right to sell) for their dwelling (YES/NO). Two categories after harmonization:

0 = No

1 = Yes

transdwel

This is a dummy variable that specifies whether the respondent has the right to bequeath the dwelling to the next generation of their family (YES/NO). Two categories after harmonization:

0 = No

1 = Yes

Table 7.5: Main Dwelling-Ownership

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Dwelling	**ownhouse**	**Ownership of house**	1 = Ownership/secure rights 2 = Renting 3 = Provided for free 4 = Without permission	1

2	Dwelling	**acqui_house**	**Acquisition of house**	1 = Purchased 2 = Inherited 3 = Other	2
3	Dwelling	**dwelownlti**	**Legal title for Ownership**	0 = No 1 = Yes	1
4	Dwelling	**fem_dwelownlti**	**Legal title for Ownership - Female**	0 = No 1 = Yes	1
5	Dwelling	**dwelownti**	**Type of Legal document**	1 = Title, deed, freehold 2 = Government issued leasehold 3 = Occupancy certificate – govt issued 4 = Legal document in the name of group (community; cooperative) 5 = Condominium (apartment) 6 = Other	2
6	Dwelling	**selldwel**	**Right to sell dwelling**	0 = No 1 = Yes	2
7	Dwelling	**transdwel**	**Right to transfer dwelling**	0 = No 1 = Yes	2

7.2.3 Land ownership

7.2.3.1 Residential

This section refers to residential land ownership and characteristics. Do not confuse this with other types of land ownership.

ownland
This is a dummy variable that specifies whether a household owns residential land (YES/NO). Ownership for property versus residential land on which property is constructed can be different in certain jurisdictions (land vested in a state or municipality). Two categories after harmonization:

0 = No
1 = Yes

acqui_land
This is a categorical variable that specifies the mode of acquisition for any residential land that the household uses. Only for the main residence. Only for landowners (OWNLAND==1). Acquisition of residential land, categories after harmonization:

1 = Purchased
2 = Inherited
3 = Other

doculand

This is the dummy variable specifying whether the household has a legal document for their residential land (YES/NO). See the main challenges/lessons learned in the next section for more information on what can be considered legal evidence. Only for landowners (OWNLAND=1). Two categories after harmonization:

0 = No

1 = Yes

fem_doculand

This is the dummy variable specifying whether the household has the name of female household members listed on a legal document for their residential land (YES/NO). This will be derived from questions asking about the roster ID of the household member(s) whose name(s) are on the legal document for residential land. Only for landowners (OWNLAND=1). Two categories after harmonization:

0 = No

1 = Yes

landownti

This is a categorical variable that specifies the type of document that a household has to prove residential land ownership. The two customary rights categories (3 and 4) differentiate whether issued by plot or as a joined group title. Customary groups and cooperatives are differentiated, as well. Customary groups not required to have formal membership declared, while cooperative members have formalized status.

Land ownership type of document. Only for landowners (OWNLAND=1). If household have pieces of land under several type of ownership, the harmonizer should collapse the plots area by title type and then pick the type of ownership for the largest area. Six categories after harmonization:

1 = Title; deed

2 = leasehold (govt issued)

3 = Customary land certificate/plot level

4 = Customary based / group right

5 = Cooperative group right

6 = Other

sellland

This is a dummy variable that specifies whether the respondent has alienation rights (i.e. the right to sell) for their residential land (YES/NO). Only for landowners (OWNLAND=1). Two categories after harmonization:

0 = No

1 = Yes

transland

This is a dummy variable that specifies whether the respondent has the right to bequeath residential land to the next generation of their family (YES/NO). Only for landowners (OWNLAND=1). Two categories after harmonization:

0 = No

1 = Yes

Table 7.6: Land ownership

	Module Code	Variable name	Variable label	Allowed codes after standardization	Tier
1	Land	**ownland**	**Ownership of residential land**	0 = No 1 = Yes	1
2	Land	**acqui_land**	**Acquisition of residential land**	1 = Purchased 2 = Inherited 4 = Other	2
3	Land	**doculand**	**Legal document for residential land**	0 = No 1 = Yes	2
4	Land	**fem_doculand**	**Legal document for residential land - female**	0 = No 1 = Yes	2
5	Land	**landownti**	**Residential land ownership**	1 = Title; deed 2 = Leasehold (govt issued) 3 = Customary land certificate/plot level 4 = Customary based / group right 5 = Cooperative group right 6 = Other	2
6	Land	**sellland**	**Right to sell residential land**	0 = No 1 = Yes	1
7	Land	**transland**	**Right to transfer residential land**	0 = No 1 = Yes	1

7.2.3.2 Agricultural

This section refers to agricultural land ownership and characteristics. Do not confuse this with other types of land ownership.

agriland

This is a dummy variable that specifies whether a household is using agricultural land according to the classification of the World Census of Agriculture 2020 (YES/NO)[21]. This variable Is not about ownership but use of agricultural land. Two categories after harmonization:

[21] FAO (2015). "WORLD PROGRAMME FOR THE CENSUS OF AGRICULTURE 2020". Paragraph (8.2.35) FAO's recommended land use classification in the Figure 1 includes the following aggregate classes:

- **Arable land** is land that is used in most years for growing temporary crops. It includes land used for growing temporary crops during a twelve-month reference period, as well as land that would normally be so used but is lying fallow or has not been sown due to unforeseen circumstances. Arable land does not include land under permanent crops or land that is potentially cultivable but is not normally cultivated. Such land should be classified as "permanent meadows and pastures" if used for grazing or haying, "forest and other wooded land" if overgrown with trees and not used for grazing or haying, or "other area not elsewhere classified" if it becomes wasteland.

0 = No
1 = Yes

area_agriland
This is a numeric, continuous variable that specifies the ***total area of agricultural land used*** by household members in hectares. This could be land that is owned, rented, or sharecropped, or some combination.

2.471 acres = 1 hectare
10,000 sq metres = 1 hectare

ownagriland
This is a dummy variable that specifies whether a household owns agricultural land (YES/NO). Owned land can be by freehold, deed, customary, or government leasehold. Only those households that declared using agricultural land (AGRILAND=1). Two categories after harmonization:

0 = No
1 = Yes

area_ownagriland
This is a numeric, continuous variable that specifies the ***total area of agricultural land owned*** in hectares. Only if OWNAGRILAND=1.

2.471 acres = 1 hectare
10,000 sq metres = 1 hectare

purch_agriland
purch_agriland is a dummy variable specifies whether a household has purchased the agricultural land they own (YES/NO). Only if OWNAGRILAND=1. Two categories after harmonization:

0 = No
1 = Yes

areapurch_agriland
This is a numeric, continuous variable that specifies the total area of agricultural land purchased in hectares. Only if PURCH_AGRILAND=1.

2.471 acres = 1 hectare
10,000 sq metres = 1 hectare

- **Cropland** is the total of arable land and land under permanent crops.
- **Agricultural land** is the total of cropland and permanent meadows and pastures.
- **Land used for agriculture** is the total of "agricultural land" and "land under farm buildings and farmyards".

0203 Area of holding according to land tenure types

- Legal ownership or legal owner-like possession
- Non-legal ownership or non-legal owner-like possession
- Rented from someone else
- Other types of land tenure

inher_agriland

This is a dummy variable specifying whether a household has inherited the agricultural land they own (YES/NO). Only if OWNAGRILAND=1. Two categories after harmonization:

0 = No

1 = Yes

areainher_agriland

This is a numeric, continuous variable that specifies the total area of agricultural land inherited in hectares. Only if INHER_ARILAND=1.

2.471 acres = 1 hectare

10,000 sq metres = 1 hectare

rentout_agriland

This is a dummy variable that specifies whether any of the agricultural land a household uses is rented–out land or sharecropped (YES/NO). This refers to land (or use rights) owned by the household but cultivated or utilized by someone else irrespective of the type of the tenant (individual, household, legal entity, etc.) and contractual arrangements (fixed rental, sharecropping, etc.). Only if OWNAGRILAND=1. Two categories after harmonization:

0 = No

1 = Yes

arearentout_agriland

This is a numeric, continuous variable that specifies the total area of agricultural land rented out or share cropped in hectares. Only if RENTOUT_AGRILAND=1.

2.471 acres = 1 hectare

10,000 sq metres = 1 hectare

rentin_agriland

This is a dummy variable that specifies whether any of the agricultural land a household uses is rented–in land or sharecropped (YES/NO). This refers land owned by others (not members of the household) but cultivated or used by the household under fixed rental, sharecropped or similar arrangements. Only if OWNAGRILAND=1. Two categories after harmonization:

0 = No

1 = Yes

arearentin_agriland

This is a numeric, continuous variable that specifies the total area of agricultural land rented in or share cropped in hectares. Only if RENTIN_AGRILAND=1.

2.471 acres = 1 hectare

10,000 sq metres = 1 hectare

docuagriland

This is the dummy variable specifying whether the household has a legal document for their agricultural land (YES/NO). See main challenges/lessons learned in the next section for more information on what can be considered legal evidence. Only if OWNAGRILAND=1. Two categories after harmonization:

0 = No

1 = Yes

area_docuagriland

This is a numeric, continuous variable that specifies the total area of agricultural land owned with legal documentation in hectares. Only if DOCUAGRILAND=1.

2.471 acres = 1 hectare

10,000 sq metres = 1 hectare

fem_agrilandownti

This is the dummy variable specifying whether the household has the name of female household members listed on a legal document for their agricultural land (YES/NO). This will be derived from questions asking about the roster ID of the household member(s) whose name(s) are on the legal document for agricultural land. Only if DOCUAGRILAND=1. Two categories after harmonization:

0 = No

1 = Yes

agrilandownti

This is a categorical variable that specifies the type of document that a household has to prove agricultural land ownership. The two customary rights categories (3 and 4) differentiate whether issued by plot or as a joined group title. Customary groups and cooperatives are differentiated, as well. Customary groups not required to have formal membership declared, while cooperative members have formalized status. Agricultural land ownership type of document. Only if DOCUAGRILAND=1.

If household have several plots under several type of ownership, the harmonizer should collapse the plots by area and then pick the type of ownership for the largest area. Categories after harmonization:

1 = Title; deed

2 = leasehold (govt issued)

3 = Customary land certificate/plot level

4 = Customary based / group right

5 = Cooperative

6 = Other

See Section 7.2.2.3 DWELOWNIT for definitions.

sellagriland

This is a dummy variable that specifies whether the respondent has alienation rights (i.e. the right to sell) for their agricultural land (YES/NO). Only if OWNAGRILAND=1. Two categories after harmonization:

0 = No

1 = Yes

transagriland

This is a dummy variable that specifies whether the respondent has the right to bequeath agricultural land to the next generation of their family (YES/NO). Only if OWNAGRILAND=1. Two categories after harmonization:

0 = No

1 = Yes

Table 7.7: Agricultural Land ownership

	Module code	Variable	Variable and value label	Variable description	Tier
1	Land	**agriland**	**Use agricultural land**	1 = Yes 0 = No	2
2	Land	**area_agriland**	**Area of agricultural land used (Ha)**	Numeric, continuous	2
3	Land	**ownagriland**	**Own agricultural land**	1 = Yes 0 = No	2
4	Land	**area_ownagriland**	**Area of agricultural land owned (Ha)**	Numeric, continuous	2
5	Land	**purch_agriland**	**Purchased agricultural land**	1 = Yes 0 = No	2
6	Land	**areapurch_agriland**	**Area of purchased agriculture land (Ha)**	Numeric, continuous	2
7	Land	**inher_agriland**	**Inherit agriculture land**	1 = Yes 0 = No	2
8	Land	**areainher_agriland**	**Area of inherited agriculture land (Ha)**	Numeric, continuous	2
9	Land	**rentout_agriland**	**Rent out land**	1 = Yes 0 = No	2
10	Land	**arearentout_agriland**	**Area of rent out agricultural land (Ha)**	Numeric, continuous	2
11	Land	**rentin_agriland**	**Rent in Land**	1 = Yes 0 = No	2
12	Land	**arearentin_agriland**	**Area of rent in agricultural land (Ha)**	Numeric, continuous	2
13	Land	**docuagriland**	**Documented agricultural land**	1 = Yes 0 = No	2
14	Land	**fem_agrilandownti**	**Ownership Agri Land – Female**	1 = Yes 0 = No	2
15	Land	**area_docuagriland**	**Area of documented agricultural land (ha)**	Numeric, continuous	2

16	Land	**agrilandownti**	**Type agricultural land ownership document**	1 = Title; deed 2 = leasehold (govt issued) 3 = Customary land certificate/plot level 4 = Customary based/ group right 5 = Cooperative 6 = Other	2
17	Land	**sellagriland**	**Right to sell agricultural land**	1 = Yes 0 = No	2
18	Land	**transagriland**	**Right to transfer agricultural land**	1 = Yes 0 = No	2

7.3 Challenges and Lessons learned

The dwelling and land variables will differ in different country surveys. Documentation of legal ownership and types of dwelling/land ownership will be different depending on the country context and will often be missing in household budget surveys. To help harmonizers to understand what can be considered documentation across countries, see Table 7.7.

Table 7.7: Legal documentation of property and land ownership and use

Examples of title or deed – full ownership	Title deed
	Registered title
	Land hold title
	Real property title based on court decision
	Donation deed
	Registered conveyance of property/land in country with deeds system
Long term leasehold with government	Registered leasehold (Rwanda, Vietnam, Ethiopia etc.)
	Right of use and benefit of land (DUAT in Mozambique)
Occupancy agreement/semi-formal title for dwelling	Certificate of localization
	Occupancy permit
	Certificate of occupancy
	Grant of admission
	Bails (French system)
	Rural/ urban concession
Occupancy agreement/semi-formal title for land (including registration of customary land)	Land use certificate
	Certificate of customary ownership
	Certificate of hereditary acquisition listed in registry
	Provisional concession
	Official “petits papier”
	Collective land certificate
	Community DUAT (Right of use and benefit of land)

Collective agreement of occupancy / semi-formal	Tribal/Land certificate
Examples of **NOT** possessing a legal document as evidence of ownership for dwelling or land	Council tax letter
	Minutes of palaver
	Mortgage agreement
	Mortgage bond
	Private land sale contract
	Promise of purchase document
	Property tax receipt
	Simple petits papier
	Survey plan
	Utility bills
	Private Rental contract (with property owner)

- Furthermore, harmonizers should checks to ensure data has been harmonized correctly. Dummy variables should have a value of 0 or 1 (or be missing). They should also have variation as not all households will have certain items. Using landphone as an example, the Stata code to check that the value is 0 or 1:

```
landphone!=. & landphone!= 1 & landphone!= 0
```

- code to check variation:

```
su landphone if sdlandphone==0, d
```

- After translating the value labels to English generate variable. Check variation:

```
decode (varname), gen(varname_str)
egen roofcs = concat(varname varname_str), punct(" – ")
```

- Combine all Sections 7 and run the dofile ***MODULE_07_Dwelling.do*** which labels (variable and value) and order variables
    - ✓ This is a must process with no exception.

8 Consumption (CONS)

8.1 Framework for Harmonization

One of the most important information from the household survey data is household expenditure aggregates, upon which poverty and inequality measures are based. The Consumption module contains information on consumption and expenditures of households. The chapter introduces the concepts, definitions and data requirements and provide guidance for producing statistics. In many of the countries, these variables are already provided, and the GMD takes these at face value without adjustment. Only a few countries will the Harmonizer derive these variables with the help of the Regional Focal point.

In this module, the primary unit of analysis is the level of a household. The units of classification are consumption expenditures made by households for satisfying their needs or wants for various goods and services.

8.2 Mapping and description of variables

The Consumption module contains a small number of metadata that provides a wealth of information especially on poverty. To improve readability, only the most significant information has been included in this section. These variables are the estimates used for international poverty estimation in pip.worldbank.org[22].

The consumption variables are monetary variables expressed at current prices in the local currency unit (LCU).
Because of the different time references by country, the Harmonizer must annualize all variables.

hhsize
This is the total number of residents (regular members) in the household, excluding maids and servants. The definition of regular member is country specific. Compare this variable with number of persons in Chapter 4 (DEM). Missing values are not allowed.

Note: Values cannot be missing.

ctry_adq
To measure poverty accurately requires taking into consideration the household composition, in particular the size and demographic structure of household. This varies from country to country, because different adult scales exist worldwide.

Can be derived from the individual-level file Chapter 4 (DEM) if the scales are provided or use the ones provided by NSO.

[22] https://pip.worldbank.org/home

Note: Values cannot be zero if applicable but missing if a country does not use equivalent scales for poverty estimation.

fdtexp_own
This is a continuous variable that refers to total nominal annual household own food consumption.

This variable is important for estimating welfare impact of price changes due to shocks that affects supply of goods as well as changes in the demand.

```
CTRY_VARNAME refers to Country variable.
gen fdtexp_owm=ctry_varname*365        //daily
  or
gen fdtexp_own=ctry_varname/7*364     //7-day or one week
  or
gen fdtexp_own=ctry_varname*12         //Monthly
  or
gen fdtexp_won=ctry_varname/28*365    //reference is 28 days
  or
gen fdtexp_won=ctry_varname/30*365     //30 days
```

fdtexp_buy
This is a continuous variable that refers to total nominal annual household food actual purchases.

See FDTEXP_OWN for derivation formulae.

This variable is important for estimating welfare impact of price changes due to shocks that affects supply of goods as well as changes in the demand.

fdtexp
This is a continuous variable that refers to total nominal annual household food expenditures. This refers to both own food consumption, gifts, and food purchases. Food share is an important secondary welfare proxy.

```
egen fdtexp=rsum(fdtexp_own fdtexp_buy)
count if fdtexp==0 | fdtexp==.
```

nfdtexp
This is a continuous variable that refers to total nominal annual household non-food expenditures. In some countries this may include own non-food consumption.

Zero expenditures are acceptable.

totexp
This is a continuous variable that refers to total nominal annual household non-food expenditures. This is the welfare aggregate used for international poverty estimation by https://pip.worldbank.org/home. Depending on a country, this can be similar to the country welfare aggregate (CTRY_EXP) or different.

Missing values not allowed but there could be exceptions. Harmonizer must run a frequency/count of zero or missing and confer with TTL who will advise accordingly. GMD wants to keep the number of observations the same to avoid changing national measures.

```
egen totexp=rsum(fdtexp nfdtexp)
count if totexp==0 | totexp==.
```

fdpindex
This is the provided country-specific food price index and varies by geography or other dimensions. The Harmonizer should try to enter as much information as reference into the dofile.

Note: Values cannot be zero or missing if applicable.

nfdpindex
This is the provided country-specific non-food price index and varies by geography or other dimensions. The Harmonizer should try to enter as much information as reference into the dofile.

Note: Values cannot be zero or missing if applicable.

pindex
This is the provided country-specific where one cannot differentiate between food or non-food price index and varies by geography or other dimensions. The Harmonizer should try to enter as much information as reference into the dofile.

Note: Values cannot be zero or missing if applicable.

ctry_totexp
This aggregate refers to the welfare measure used to measure poverty (head count, poverty gap, severity of poverty). This variable may also be used to measure inequality and other consumption measures.

pl_ext
Country-specific extreme poverty line. This will vary by country as in some countries this could be a single line or region-specific lines. On the other hand, a country may have several extreme poverty lines that measure different things, but this should refer to the line used to measure extreme poverty line. In some countries this line is equivalent to the food poverty line.

pl_abs
Absolute or overall poverty line. The Harmonizer should try to enter as much information as reference into the dofile.

Table 8.1: Consumption variables

	Module code	Variable name	Variable label	Tier
1	ID	**countrycode**	**Country code**	1
2	ID	**year**	**Year (survey start year)**	1
3	ID	**hhid**	**Household unique identifier**	1
4	ID	**weight**	**Household weights**	1
5	ID	**hhsize**	**Household size**	1
6	ID	**ctry_adq**	**Sum total of adult equivalent scales (country-specific scales)**	1
7	Consumption	**fdtexp_own**	**Total annual household own food consumption**	1
8	Consumption	**fdtexp_buy**	**Total annual household purchased food consumption**	1
9	Consumption	**fdtexp**	**Total annual household food expenditure**	1
10	Consumption	**nfdtexp**	**Total annual non-food expenditure**	1
11	Consumption	**totexp**	**Total annual consumption of food and nonfood**	1
12	Consumption	**fdpindex**	**Food price index (spatial and/or temporal)**	1
13	Consumption	**nfdpindex**	**Non-food price index (spatial and/or temporal)**	1
14	Consumption	**pindex**	**Price index (spatial and/or temporal)**	1
15	Consumption	**ctry_totexp**	**Total annual consumption of food and nonfood**	1
16	Consumption	**pl_ext**	**Extreme national poverty line**	1
17	Consumption	**pl_abs**	**Absolute/overall national poverty line**	1

8.3 Challenges and common mistakes

Data sets that are harmonized incorrectly can lead to skewed and/or incorrect data analysis. Harmonizers should run a series of checks to ensure data is harmonized properly.

- Check for duplicates of HHID. If duplicates fix by renumbering but be careful if any merges will be needed to be done later. Secondly, in some countries, same household interviewed over 2-4 cycles and this may look like a duplicate but not. In such cases, renumbering allowed to include cycle to void duplicates.

```
duplicates tag (hhid),gen(dup)
tab dup
```

```
duplicates tag (hhid totexp),gen(dup1)
ta dup1
```

- Check HHSIZE. Check for large values to make sure if there is an error on the HHID.

```
ta hhsize
sum hhsize
```

- Check CTRY_ADQ

  Total number of adult equivalent people in household: - (a) Must be greater than 0. (b) Must be greater, less than or equal to HHSIZE (Household Size).

```
ta ctry_adq
sum ctry_adq
```

- Check consumption expenditure variables.
  See Section 6.4.2 for further details.

- Run the dofile ***MODULE_08_Consumption.do*** which labels (variable and value) and order variables
  - ✓ This is a must process with no exception.

Annex I: ISO3 country codes and World Bank region classification

Country code	Country name	Region
ASM	American Samoa	East Asia & Pacific
AUS	Commonwealth of Australia	East Asia & Pacific
BRN	Brunei Darussalam	East Asia & Pacific
KHM	Kingdom of Cambodia	East Asia & Pacific
CHN	People's Republic of China	East Asia & Pacific
FJI	Republic of Fiji	East Asia & Pacific
PYF	French Polynesia	East Asia & Pacific
GUM	Guam	East Asia & Pacific
HKG	Hong Kong Special Administrative Region of the People's Republic of China	East Asia & Pacific
IDN	Republic of Indonesia	East Asia & Pacific
JPN	Japan	East Asia & Pacific
KIR	Republic of Kiribati	East Asia & Pacific
PRK	Democratic People's Republic of Korea	East Asia & Pacific
KOR	Republic of Korea	East Asia & Pacific
LAO	Lao People's Democratic Republic	East Asia & Pacific
MAC	Macao Special Administrative Region of the People's Republic of China	East Asia & Pacific
MYS	Malaysia	East Asia & Pacific
MHL	Republic of the Marshall Islands	East Asia & Pacific
FSM	Federated States of Micronesia	East Asia & Pacific
MNG	Mongolia	East Asia & Pacific
MMR	Republic of the Union of Myanmar	East Asia & Pacific
NRU	Republic of Nauru	East Asia & Pacific
NCL	New Caledonia	East Asia & Pacific
NZL	New Zealand	East Asia & Pacific
MNP	Commonwealth of the Northern Mariana Islands	East Asia & Pacific
PLW	Republic of Palau	East Asia & Pacific
PNG	The Independent State of Papua New Guinea	East Asia & Pacific
PHL	Republic of the Philippines	East Asia & Pacific
WSM	Samoa	East Asia & Pacific
SGP	Republic of Singapore	East Asia & Pacific
SLB	Solomon Islands	East Asia & Pacific
THA	Kingdom of Thailand	East Asia & Pacific
TLS	Democratic Republic of Timor-Leste	East Asia & Pacific
TON	Kingdom of Tonga	East Asia & Pacific
TUV	Tuvalu	East Asia & Pacific
VUT	Republic of Vanuatu	East Asia & Pacific
VNM	Socialist Republic of Vietnam	East Asia & Pacific
ALB	Republic of Albania	Europe & Central Asia

Country code	Country name	Region
AND	Principality of Andorra	Europe & Central Asia
ARM	Republic of Armenia	Europe & Central Asia
AUT	Republic of Austria	Europe & Central Asia
AZE	Republic of Azerbaijan	Europe & Central Asia
BLR	Republic of Belarus	Europe & Central Asia
BEL	Kingdom of Belgium	Europe & Central Asia
BIH	Bosnia and Herzegovina	Europe & Central Asia
BGR	Republic of Bulgaria	Europe & Central Asia
CHI	Channel Islands	Europe & Central Asia
HRV	Republic of Croatia	Europe & Central Asia
CYP	Republic of Cyprus	Europe & Central Asia
CZE	Czech Republic	Europe & Central Asia
DNK	Kingdom of Denmark	Europe & Central Asia
EST	Republic of Estonia	Europe & Central Asia
FRO	Faroe Islands	Europe & Central Asia
FIN	Republic of Finland	Europe & Central Asia
FRA	French Republic	Europe & Central Asia
GEO	Georgia	Europe & Central Asia
DEU	Federal Republic of Germany	Europe & Central Asia
GIB	Gibraltar	Europe & Central Asia
GRC	Hellenic Republic	Europe & Central Asia
GRL	Greenland	Europe & Central Asia
HUN	Hungary	Europe & Central Asia
ISL	Republic of Iceland	Europe & Central Asia
IRL	Ireland	Europe & Central Asia
IMN	Isle of Man	Europe & Central Asia
ITA	Italian Republic	Europe & Central Asia
KAZ	Republic of Kazakhstan	Europe & Central Asia
XKX	Republic of Kosovo	Europe & Central Asia
KGZ	Kyrgyz Republic	Europe & Central Asia
LVA	Republic of Latvia	Europe & Central Asia
LIE	Principality of Liechtenstein	Europe & Central Asia
LTU	Republic of Lithuania	Europe & Central Asia
LUX	Grand Duchy of Luxembourg	Europe & Central Asia
MDA	Republic of Moldova	Europe & Central Asia
MCO	Principality of Monaco	Europe & Central Asia
MNE	Montenegro	Europe & Central Asia
NLD	Kingdom of the Netherlands	Europe & Central Asia
MKD	Republic of North Macedonia	Europe & Central Asia
NOR	Kingdom of Norway	Europe & Central Asia
POL	Republic of Poland	Europe & Central Asia
PRT	Portuguese Republic	Europe & Central Asia

Country code	Country name	Region
ROU	Romania	Europe & Central Asia
RUS	Russian Federation	Europe & Central Asia
SMR	Republic of San Marino	Europe & Central Asia
SRB	Republic of Serbia	Europe & Central Asia
SVK	Slovak Republic	Europe & Central Asia
SVN	Republic of Slovenia	Europe & Central Asia
ESP	Kingdom of Spain	Europe & Central Asia
SWE	Kingdom of Sweden	Europe & Central Asia
CHE	Switzerland	Europe & Central Asia
TJK	Republic of Tajikistan	Europe & Central Asia
TUR	Republic of Turkey	Europe & Central Asia
TKM	Turkmenistan	Europe & Central Asia
UKR	Ukraine	Europe & Central Asia
GBR	United Kingdom of Great Britain and Northern Ireland	Europe & Central Asia
UZB	Republic of Uzbekistan	Europe & Central Asia
ATG	Antigua and Barbuda	Latin America & Caribbean
ARG	Argentine Republic	Latin America & Caribbean
ABW	Aruba	Latin America & Caribbean
BHS	Commonwealth of The Bahamas	Latin America & Caribbean
BRB	Barbados	Latin America & Caribbean
BLZ	Belize	Latin America & Caribbean
BOL	Plurinational State of Bolivia	Latin America & Caribbean
BRA	Federative Republic of Brazil	Latin America & Caribbean
VGB	British Virgin Islands	Latin America & Caribbean
CYM	Cayman Islands	Latin America & Caribbean
CHL	Republic of Chile	Latin America & Caribbean
COL	Republic of Colombia	Latin America & Caribbean
CRI	Republic of Costa Rica	Latin America & Caribbean
CUB	Republic of Cuba	Latin America & Caribbean
CUW	Curaçao	Latin America & Caribbean
DMA	Commonwealth of Dominica	Latin America & Caribbean
DOM	Dominican Republic	Latin America & Caribbean
ECU	Republic of Ecuador	Latin America & Caribbean
SLV	Republic of El Salvador	Latin America & Caribbean
GRD	Grenada	Latin America & Caribbean
GTM	Republic of Guatemala	Latin America & Caribbean
GUY	Co-operative Republic of Guyana	Latin America & Caribbean
HTI	Republic of Haiti	Latin America & Caribbean
HND	Republic of Honduras	Latin America & Caribbean
JAM	Jamaica	Latin America & Caribbean
MEX	United Mexican States	Latin America & Caribbean
NIC	Republic of Nicaragua	Latin America & Caribbean

Country code	Country name	Region
PAN	Republic of Panama	Latin America & Caribbean
PRY	Republic of Paraguay	Latin America & Caribbean
PER	Republic of Peru	Latin America & Caribbean
PRI	Puerto Rico	Latin America & Caribbean
SXM	Sint Maarten (Dutch part)	Latin America & Caribbean
KNA	St. Kitts and Nevis	Latin America & Caribbean
LCA	St. Lucia	Latin America & Caribbean
MAF	St. Martin (French part)	Latin America & Caribbean
VCT	St. Vincent and the Grenadines	Latin America & Caribbean
SUR	Republic of Suriname	Latin America & Caribbean
TTO	Republic of Trinidad and Tobago	Latin America & Caribbean
TCA	Turks and Caicos Islands	Latin America & Caribbean
URY	Oriental Republic of Uruguay	Latin America & Caribbean
VEN	República Bolivariana de Venezuela	Latin America & Caribbean
VIR	Virgin Islands of the United States	Latin America & Caribbean
DZA	People's Democratic Republic of Algeria	Middle East & North Africa
BHR	Kingdom of Bahrain	Middle East & North Africa
DJI	Republic of Djibouti	Middle East & North Africa
EGY	Arab Republic of Egypt	Middle East & North Africa
IRN	Islamic Republic of Iran	Middle East & North Africa
IRQ	Republic of Iraq	Middle East & North Africa
ISR	State of Israel	Middle East & North Africa
JOR	Hashemite Kingdom of Jordan	Middle East & North Africa
KWT	State of Kuwait	Middle East & North Africa
LBN	Lebanese Republic	Middle East & North Africa
LBY	Socialist People's Libyan Arab Jamahiriya	Middle East & North Africa
MLT	Republic of Malta	Middle East & North Africa
MAR	Kingdom of Morocco	Middle East & North Africa
OMN	Sultanate of Oman	Middle East & North Africa
QAT	State of Qatar	Middle East & North Africa
SAU	Kingdom of Saudi Arabia	Middle East & North Africa
SYR	Syrian Arab Republic	Middle East & North Africa
TUN	Republic of Tunisia	Middle East & North Africa
ARE	United Arab Emirates	Middle East & North Africa
PSE	West Bank and Gaza	Middle East & North Africa
YEM	Republic of Yemen	Middle East & North Africa
BMU	The Bermudas	North America
CAN	Canada	North America
USA	United States of America	North America
AFG	Islamic State of Afghanistan	South Asia
BGD	People's Republic of Bangladesh	South Asia
BTN	Kingdom of Bhutan	South Asia

Country code	Country name	Region
IND	Republic of India	South Asia
MDV	Republic of Maldives	South Asia
NPL	Nepal	South Asia
PAK	Islamic Republic of Pakistan	South Asia
LKA	Democratic Socialist Republic of Sri Lanka	South Asia
AGO	People's Republic of Angola	Sub-Saharan Africa
BEN	Republic of Benin	Sub-Saharan Africa
BWA	Republic of Botswana	Sub-Saharan Africa
BFA	Burkina Faso	Sub-Saharan Africa
BDI	Republic of Burundi	Sub-Saharan Africa
CPV	Republic of Cabo Verde	Sub-Saharan Africa
CMR	Republic of Cameroon	Sub-Saharan Africa
CAF	Central African Republic	Sub-Saharan Africa
TCD	Republic of Chad	Sub-Saharan Africa
COM	Union of the Comoros	Sub-Saharan Africa
COD	Democratic Republic of the Congo	Sub-Saharan Africa
COG	Republic of Congo	Sub-Saharan Africa
CIV	Republic of Côte d'Ivoire	Sub-Saharan Africa
GNQ	Republic of Equatorial Guinea	Sub-Saharan Africa
ERI	State of Eritrea	Sub-Saharan Africa
SWZ	Kingdom of Eswatini	Sub-Saharan Africa
ETH	Federal Democratic Republic of Ethiopia	Sub-Saharan Africa
GAB	Gabonese Republic	Sub-Saharan Africa
GMB	Republic of The Gambia	Sub-Saharan Africa
GHA	Republic of Ghana	Sub-Saharan Africa
GIN	Republic of Guinea	Sub-Saharan Africa
GNB	Republic of Guinea-Bissau	Sub-Saharan Africa
KEN	Republic of Kenya	Sub-Saharan Africa
LSO	Kingdom of Lesotho	Sub-Saharan Africa
LBR	Republic of Liberia	Sub-Saharan Africa
MDG	Republic of Madagascar	Sub-Saharan Africa
MWI	Republic of Malawi	Sub-Saharan Africa
MLI	Republic of Mali	Sub-Saharan Africa
MRT	Islamic Republic of Mauritania	Sub-Saharan Africa
MUS	Republic of Mauritius	Sub-Saharan Africa
MOZ	Republic of Mozambique	Sub-Saharan Africa
NAM	Republic of Namibia	Sub-Saharan Africa
NER	Republic of Niger	Sub-Saharan Africa
NGA	Federal Republic of Nigeria	Sub-Saharan Africa
RWA	Republic of Rwanda	Sub-Saharan Africa
STP	Democratic Republic of São Tomé and Principe	Sub-Saharan Africa
SEN	Republic of Senegal	Sub-Saharan Africa

Country code	Country name	Region
SYC	Republic of Seychelles	Sub-Saharan Africa
SLE	Republic of Sierra Leone	Sub-Saharan Africa
SOM	Somali Democratic Republic	Sub-Saharan Africa
ZAF	Republic of South Africa	Sub-Saharan Africa
SSD	Republic of South Sudan	Sub-Saharan Africa
SDN	Republic of the Sudan	Sub-Saharan Africa
TZA	United Republic of Tanzania	Sub-Saharan Africa
TGO	Republic of Togo	Sub-Saharan Africa
UGA	Republic of Uganda	Sub-Saharan Africa
ZMB	Republic of Zambia	Sub-Saharan Africa
ZWE	Republic of Zimbabwe	Sub-Saharan Africa

Annex II: ISCED Education groups

For details the coding of education programmes, see ISCED 2012) [23].

ISCED-A level		Category		Sub-category	
0	Early childhood education	01	Early childhood educational development	010	Early childhood educational development
		O2	Pre-primary education	020	Pre-primary education
		03	Some primary education (without level completion)	030	Some primary education (without level completion)
1	Primary education	10	Primary education	100	
2	Lower	24	General education	241	Insufficient for level completion or partial level completion, without direct access to upper secondary education
				242	Partial level completion, without direct access to upper secondary education
				243	Level completion, without direct access to upper secondary education
				244	Level completion, with direct access to upper secondary education
		25	Vocational education	251	Insufficient for level completion or partial level completion, without direct access to upper secondary education
				252	Partial level completion, without direct access to upper secondary education
				253	Level completion, without direct access to upper secondary education
				254	Level completion, with direct access to upper secondary education
3	Upper secondary education	35	General education	341	Insufficient for level completion or partial level completion, without direct access to tertiary education

[23] http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf

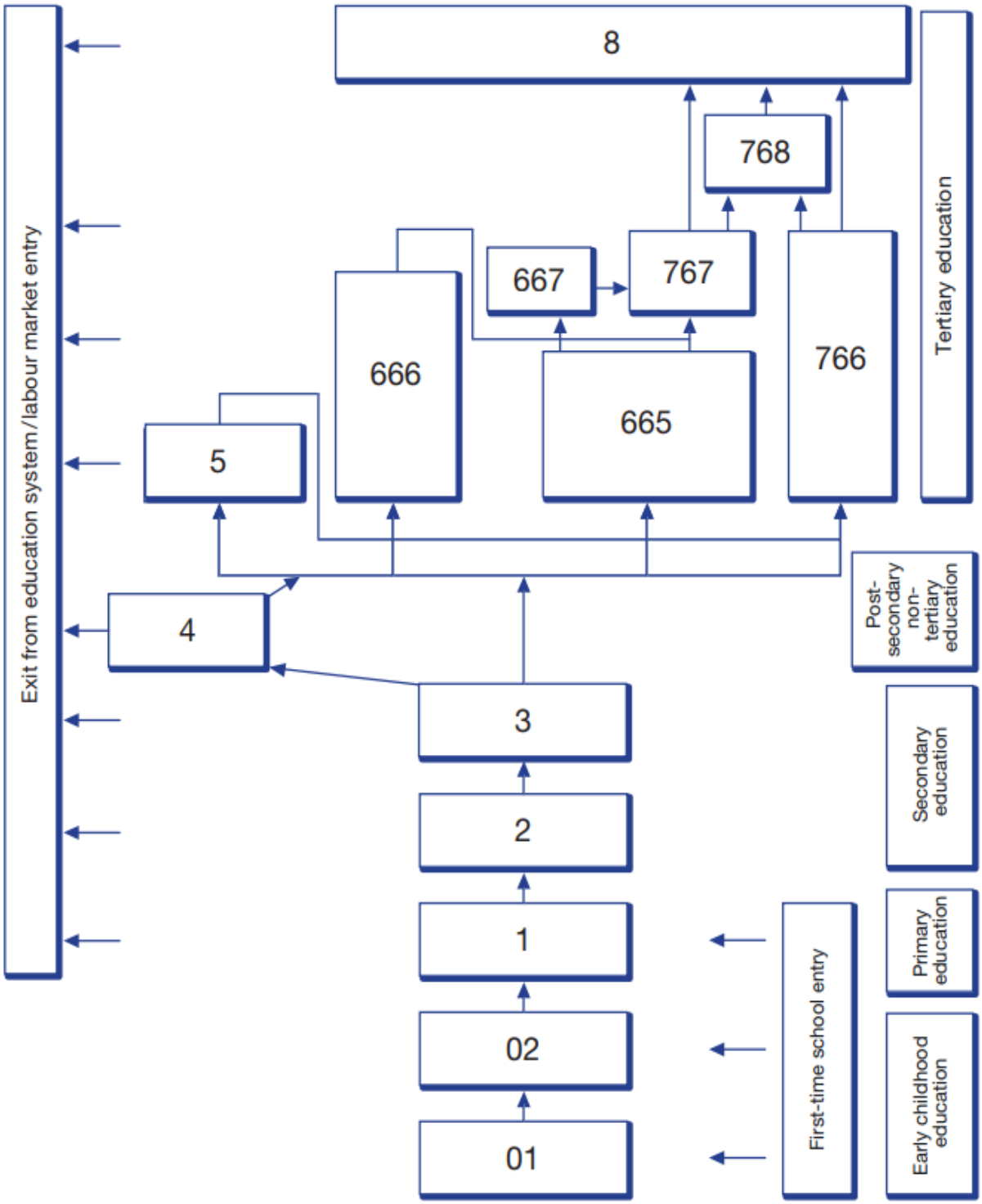
ISCED-A level		Category		Sub-category	
				342	Partial level completion, without direct access to tertiary education
				343	Level completion, without direct access to first tertiary programmes (but may give direct access to post-secondary non-tertiary education)
				344	Level completion, with direct access to first tertiary programmes (may also give direct access to post-secondary non-tertiary education
			Vocational education	351	Insufficient for level completion or partial level completion, without direct access to tertiary education
				352	Partial level completion, without direct access to tertiary education
				353	Level completion, without direct access to first tertiary programmes (but may give direct access to post-secondary non-tertiary education)
				354	Level completion, with direct access to first tertiary programmes (may also give direct access to post-secondary non-tertiary education)
4	Post-secondary non-tertiary education	44	General education	441	Insufficient for level completion, without direct access to tertiary education
				443	Sufficient for level completion, without direct access to tertiary education
				444	Sufficient for level completion, with direct access to tertiary education
		45	Vocational education	451	Insufficient for level completion, without direct access to tertiary education
				453	Sufficient for level completion, without direct access to tertiary education
				454	Sufficient for level completion, with direct access to tertiary education

ISCED-A level		Category		Sub-category	
5	Short-cycle tertiary education	54	General education[1]	541	Insufficient for level completion
				544	Sufficient for level completion
		55	Vocational education[1]	551	Insufficient for level completion
				554	Sufficient for level completion
6	Bachelor's or equivalent level	64	Academic	641	Insufficient for level completion
				645	First degree (3-4 years)
				646	Long first degree (more than 4 years)
				647	Second or further degree (following a Bachelor's or equivalent programme)
		65	Professional	651	Insufficient for level completion
				655	First degree (3-4 years)
				656	Long first degree (more than 4 years)
				657	Second or further degree (following a Bachelor's or equivalent programme)
		66	Orientation unspecified[1]	661	Insufficient for level completion
				665	First degree (3-4 years)
				666	Long first degree (more than 4 years)
				667	Second or further degree (following a Bachelor's or equivalent programme)
7	Master's or equivalent level	74	Academic	741	Insufficient for level completion
				746	Long first degree (at least 5 years)
				747	Second or further degree (following a Bachelor's or equivalent programme)
				748	Second or further degree (following a Master's or equivalent programme)
		75	Professional	751	Insufficient for level completion
				756	Long first degree (at least 5 years)
				757	Second or further degree (following a Bachelor's or equivalent programme)
				758	Second or further degree (following a Master's or equivalent programme)
		76	Orientation unspecified[1]	761	Insufficient for level completion
				766	Long first degree (at least 5 years)

ISCED-A level		Category		Sub-category	
				767	Second or further degree (following a Bachelor's or equivalent programme)
				768	Second or further degree (following a Master's or equivalent programme)
8	Doctoral or equivalent level	84	Academic	841	841 Insufficient for level completion
				844	Sufficient for completion of level
			Professional	841	841 Insufficient for level completion
				844	Sufficient for completion of level
			Orientation unspecified[1]	841	841 Insufficient for level completion
				844	Sufficient for completion of level
9	Not elsewhere classified	99	Not elsewhere classified	999	Not elsewhere classified

[1] To be used in the absence of internationally agreed definitions for academic and professional orientations at the tertiary level.

ISCED 2011 potential educational pathways



Source: ISCED 2011

Annex III: ILO Classification of Labor

Annex III.1: International Standard Industrial Classification of All Economic Activities (ISIC) Revision 4.0

ISIC classifies entities by activity. This is the fifth iteration of the International Standard Industrial Classification of Economic Activities[24] (ISIC) Revision 4.0. It is part of the international family of economic and social classifications of the United Nations. The current version, known as ISIC Revision 4.0 was published in 2008, following ISIC Revision 3.1 (2002), ISIC Revision 3.0 (1989), ISIC Revision 2.0 (1968) and ISIC Revision 1 (1958).

Section A		**Agriculture, forestry and fishing**
Division 01		**Crop and animal production, hunting and related service activities**
011		Growing of non-perennial crops
	0111	Growing of cereals (except rice), leguminous crops and oil seeds
	0112	Growing of rice
	0113	Growing of vegetables and melons, roots and tubers
	0114	Growing of sugar cane
	0115	Growing of tobacco
	0116	Growing of fibre crops
	0119	Growing of other non-perennial crops
012		Growing of perennial crops
	0121	Growing of grapes
	0122	Growing of tropical and subtropical fruits
	0123	Growing of citrus fruits
	0124	Growing of pome fruits and stone fruits
	0125	Growing of other tree and bush fruits and nuts
	0126	Growing of oleaginous fruits
	0127	Growing of beverage crops
	0128	Growing of spices, aromatic, drug and pharmaceutical crops
	0129	Growing of other perennial crop
013	0130	Plant propagation
014		Animal production
	0141	Raising of cattle and buffaloes
	0142	Raising of horses and other equines
	0143	Raising of camels and camelids
	0144	Raising of sheep and goats
	0145	Raising of swine/pigs
	0146	Raising of poultry

[24] https://unstats.un.org/unsd/publication/SeriesM/seriesm_4rev3_1e.pdf

	0149	Raising of other animals
015	0150	Mixed farming
016		Support activities to agriculture and post-harvest crop activities
	0161	Support activities for crop production
	0162	Support activities for animal production
	0163	Post-harvest crop activities
	0164	Seed processing for propagation
017	0170	Hunting, trapping and related service activities
Division 02		**Forestry and logging**
021	0210	Silviculture and other forestry activities
022	0220	Logging
023	0230	Gathering of non-wood forest products
024	0240	Support services to forestry
Division 03		**Fishing and aquaculture**
031		Fishing
	0311	Marine fishing
	0312	Freshwater fishing
032		Aquaculture
	0321	Marine aquaculture
	0322	Freshwater aquaculture
Section B		**Mining and quarrying**
Division 05		**Mining of coal and lignite**
051	0510	Mining of hard coal
052	0520	Mining of lignite
Division 06		**Extraction of crude petroleum and natural gas**
061	0610	Extraction of crude petroleum
062	0620	Extraction of natural gas
Division 07		**Mining of metal ores**
071	0710	Mining of iron ores
072		Mining of non-ferrous metal ores
	0721	Mining of uranium and thorium ores
	0729	Mining of other non-ferrous metal ores

Division 08		**Other mining and quarrying**
081	0810	Quarrying of stone, sand and clay
089		Mining and quarrying n.e.c.
	0891	Mining of chemical and fertilizer minerals
	0892	Extraction of peat
	0893	Extraction of salt
	0899	Other mining and quarrying n.e.c
Division 09		**Mining support service activities**
091	0910	Support activities for petroleum and natural gas extraction
099	0990	Support activities for other mining and quarrying

Section C Manufacturing

Division 10		**Manufacture of food products**
101	1010	Processing and preserving of meat
102	1020	Processing and preserving of fish, crustaceans and molluscs
103	1030	Processing and preserving of fruit and vegetables
104	1040	Manufacture of vegetable and animal oils and fats
105	1050	Manufacture of dairy products
106		Manufacture of grain mill products, starches and starch products
	1061	Manufacture of grain mill products
	1062	Manufacture of starches and starch products
107		Manufacture of other food products
	1071	Manufacture of bakery products
	1072	Manufacture of sugar
	1073	Manufacture of cocoa, chocolate and sugar confectionery
	1074	Manufacture of macaroni, noodles, couscous and similar farinaceous products
	1075	Manufacture of prepared meals and dishes
	1079	Manufacture of other food products n.e.c.
108	1080	Manufacture of prepared animal feeds
Division 11		**Manufacture of beverages**
	1101	Distilling, rectifying and blending of spirits
	1102	Manufacture of wines
	1103	Manufacture of malt liquors and malt
	1104	Manufacture of soft drinks; production of mineral waters and other bottled waters
Division 12		**Manufacture of tobacco products**

120	1200	Manufacture of tobacco product
Division 13		**Manufacture of textiles**
131		Spinning, weaving and finishing of textiles
	1311	Preparation and spinning of textile fibres
	1312	Weaving of textiles
	1313	Finishing of textiles
139		Manufacture of other textiles
	1391	Manufacture of knitted and crocheted fabrics
	1392	Manufacture of made-up textile articles, except apparel
	1393	Manufacture of carpets and rugs
	1394	Manufacture of cordage, rope, twine and netting
	1399	Manufacture of other textiles n.e.c
Division 14		**Manufacture of wearing apparel**
141	1410	Manufacture of wearing apparel, except fur apparel
142	1420	Manufacture of articles of fur
143	1430	Manufacture of knitted and crocheted apparel
Division 15		**Manufacture of leather and related products**
151		Tanning and dressing of leather; manufacture of luggage, handbags, saddlery and harness; dressing and dyeing of fur
	1511	Tanning and dressing of leather; dressing and dyeing of fur
	1512	Manufacture of luggage, handbags and the like, saddlery and harness
152	1520	Manufacture of footwear
Division 16		**Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials**
161	1610	Sawmilling and planning of wood
162		Manufacture of products of wood, cork, straw and plaiting materials
	1621	Manufacture of veneer sheets and wood-based panels
	1622	Manufacture of builders' carpentry and joinery
	1623	Manufacture of wooden containers
	1629	Manufacture of other products of wood; manufacture of articles of cork, straw and plaiting materials
Division 17		**Manufacture of paper and paper products**
	1701	Manufacture of pulp, paper and paperboard
	1702	Manufacture of corrugated paper and paperboard and of containers of paper and paperboard
	1709	Manufacture of other articles of paper and paperboard

Division 18		**Printing and reproduction of recorded media**
181		Printing and service activities related to printing
	1811	Printing 1812 Service activities related to printing
182	1820	Reproduction of recorded media
Division 19		**Manufacture of coke and refined petroleum products**
191	1910	Manufacture of coke oven products
192	1920	Manufacture of refined petroleum products
Division 20		**Manufacture of chemicals and chemical products**
201		Manufacture of basic chemicals, fertilizers and nitrogen compounds, plastics and synthetic rubber in primary forms
	2011	Manufacture of basic chemicals
	2012	Manufacture of fertilizers and nitrogen compounds
	2013	Manufacture of plastics and synthetic rubber in primary forms
202		Manufacture of other chemical products
	2021	Manufacture of pesticides and other agrochemical products
	2022	Manufacture of paints, varnishes and similar coatings, printing ink and mastics
	2023	Manufacture of soap and detergents, cleaning and polishing preparations, perfumes and toilet preparations
	2029	Manufacture of other chemical products n.e.c. 203 2030 Manufacture of man-made fibres
Division 21		**Manufacture of pharmaceuticals, medicinal chemical and botanical products**
210	2100	Manufacture of pharmaceuticals, medicinal chemical and botanical products
Division 22		**Manufacture of rubber and plastics products**
221		Manufacture of rubber products
	2211	Manufacture of rubber tyres and tubes; retreading and rebuilding of rubber tyres
	2219	Manufacture of other rubber products
222	2220	Manufacture of plastics products
Division 23		**Manufacture of other non-metallic mineral products**
231	2310	Manufacture of glass and glass products
239		Manufacture of non-metallic mineral products n.e.c.
	2391	Manufacture of refractory products
	2392	Manufacture of clay building materials
	2393	Manufacture of other porcelain and ceramic products
	2394	Manufacture of cement, lime and plaster
	2395	Manufacture of articles of concrete, cement and plaster
	2396	Cutting, shaping and finishing of stone 2399 Manufacture of other non-metallic

		mineral products n.e.c.
Division 24		**Manufacture of basic metals**
241	2410	Manufacture of basic iron and steel
242	2420	Manufacture of basic precious and other non-ferrous metals
243		Casting of metals
	2431	Casting of iron and steel
	2432	Casting of non-ferrous metals
Division 25		**Manufacture of fabricated metal products, except machinery and equipment**
251		Manufacture of structural metal products, tanks, reservoirs and steam generators
	2511	Manufacture of structural metal products
	2512	Manufacture of tanks, reservoirs and containers of metal
	2513	Manufacture of steam generators, except central heating hot water boilers
252	2520	Manufacture of weapons and ammunition
259		Manufacture of other fabricated metal products; metalworking service activities
	2591	Forging, pressing, stamping and roll-forming of metal; powder metallurgy
	2592	Treatment and coating of metals; machining
	2593	Manufacture of cutlery, hand tools and general hardware
	2599	Manufacture of other fabricated metal products n.e.c
Division 26		**Manufacture of computer, electronic and optical products**
261	2610	Manufacture of electronic components and boards
262	2620	Manufacture of computers and peripheral equipment
263	2630	Manufacture of communication equipment
264	2640	Manufacture of consumer electronics
265		Manufacture of measuring, testing, navigating and control equipment; watches and clocks
	2651	Manufacture of measuring, testing, navigating and control equipment
	2652	Manufacture of watches and clocks
266	2660	Manufacture of irradiation, electromedical and electrotherapeutic equipment
267	2670	Manufacture of optical instruments and photographic equipment
268	2680	Manufacture of magnetic and optical media
Division 27		**Manufacture of electrical equipment**
271	2710	Manufacture of electric motors, generators, transformers and electricity

		distribution and control apparatus
272	2720	Manufacture of batteries and accumulators
273		Manufacture of wiring and wiring devices
	2731	Manufacture of fibre optic cables
	2732	Manufacture of other electronic and electric wires and cables
	2733	Manufacture of wiring devices
274	2740	Manufacture of electric lighting equipment
275	2750	Manufacture of domestic appliances
279	2790	Manufacture of other electrical equipment
Division 28		**Manufacture of machinery and equipment n.e.c.**
281		Manufacture of general-purpose machinery
	2811	Manufacture of engines and turbines, except aircraft, vehicle and cycle engines
	2812	Manufacture of fluid power equipment
	2813	Manufacture of other pumps, compressors, taps and valves
	2814	Manufacture of bearings, gears, gearing and driving elements
	2815	Manufacture of ovens, furnaces and furnace burners
	2816	Manufacture of lifting and handling equipment
	2817	Manufacture of office machinery and equipment (except computers and peripheral equipment)
	2818	Manufacture of power-driven hand tools
	2819	Manufacture of other general-purpose machinery
282		Manufacture of special-purpose machinery
	2821	Manufacture of agricultural and forestry machinery
	2822	Manufacture of metal-forming machinery and machine tools
	2823	Manufacture of machinery for metallurgy
	2824	Manufacture of machinery for mining, quarrying and construction
	2825	Manufacture of machinery for food, beverage and tobacco processing
	2826	Manufacture of machinery for textile, apparel and leather production
	2829	Manufacture of other special-purpose machinery
Division 29		**Manufacture of motor vehicles, trailers and semi-trailers**
291	2910	Manufacture of motor vehicles
292	2920	Manufacture of bodies (coachwork) for motor vehicles; manufacture of trailers and semi-trailers
293	2930	Manufacture of parts and accessories for motor vehicles
Division 30		**Manufacture of other transport equipment**
301		Building of ships and boats
	3011	Building of ships and floating structures

	3012	Building of pleasure and sporting boats
302	3020	Manufacture of railway locomotives and rolling stock Detailed structure
303	3030	Manufacture of air and spacecraft and related machinery
304	3040	Manufacture of military fighting vehicles
309		Manufacture of transport equipment n.e.c.
	3091	Manufacture of motorcycles
	3092	Manufacture of bicycles and invalid carriages
	3099	Manufacture of other transport equipment n.e.c
Division 31		**Manufacture of furniture**
310	3100	Manufacture of furniture
Division 32		**Other manufacturing**
321		Manufacture of jewellery, bijouterie and related articles
	3211	Manufacture of jewellery and related articles
	3212	Manufacture of imitation jewellery and related articles
322		Manufacture of musical instruments
	3220	Manufacture of musical instruments
323		Manufacture of sports goods
	3230	Manufacture of sports goods
324		Manufacture of games and toys
	3240	Manufacture of games and toys
325		Manufacture of medical and dental instruments and supplies
	3250	Manufacture of medical and dental instruments and supplies
329		Other manufacturing n.e.c."
	3290	Other manufacturing n.e.c."
Division 33		**Repair and installation of machinery and equipment**
331		Repair of fabricated metal products, machinery and equipment
	3311	Repair of fabricated metal products
	3312	Repair of machinery
	3313	Repair of electronic and optical equipment
	3314	Repair of electrical equipment
	3315	Repair of transport equipment, except motor vehicles
	3319	Repair of other equipment
332	3320	Installation of industrial machinery and equipment

Section D		**Electricity, gas, steam and air conditioning supply**
Division 35		**Electricity, gas, steam and air conditioning supply**
351	3510	Electric power generation, transmission and distribution
352	3520	Manufacture of gas; distribution of gaseous fuels through mains
353	3530	Steam and air conditioning supply
Division 36		**Water collection, treatment and supply**
360	3600	Water collection, treatment and supply
Division 37		**Sewerage**
370	3700	Sewerage
Division 38		**Waste collection, treatment and disposal activities; materials recovery**
381		Waste collection
	3811	Collection of non-hazardous waste
	3812	Collection of hazardous waste
382		Waste treatment and disposal
	3821	Treatment and disposal of non-hazardous waste
	3822	Treatment and disposal of hazardous waste
383	3830	Materials recovery
Division 39		**Remediation activities and other waste management services**
390	3900	Remediation activities and other waste management services
Section F		**Construction**
Division 41		**Construction of buildings**
410	4100	Construction of buildings
Division 42		**Civil engineering**
421	4210	Construction of roads and railways
422	4220	Construction of utility projects
429	4290	Construction of other civil engineering projects
Division 43		**Specialized construction activities**
431		Demolition and site preparation
	4311	Demolition
	4312	Site preparation
432		Electrical, plumbing and other construction installation activities
	4321	Electrical installation

	4322	Plumbing, heat and air-conditioning installation
	4329	Other construction installation
433	4330	Building completion and finishing
439	4390	Other specialized construction activities

Section G Wholesale and retail trade; repair of motor vehicles and motorcycles

Division 45		**Wholesale and retail trade and repair of motor vehicles and motorcycles**
451	4510	Sale of motor vehicles
452	4520	Maintenance and repair of motor vehicles
453	4530	Sale of motor vehicle parts and accessories
454	4540	Sale, maintenance and repair of motorcycles and related parts and accessories
Division 46		**Wholesale trade, except of motor vehicles and motorcycles**
461	4610	Wholesale on a fee or contract basis
462	4620	Wholesale of agricultural raw materials and live animals
463	4630	Wholesale of food, beverages and tobacco
464		Wholesale of household goods
	4641	Wholesale of textiles, clothing and footwear
	4649	Wholesale of other household goods
465		Wholesale of machinery, equipment and supplies
	4651	Wholesale of computers, computer peripheral equipment and software
	4652	Wholesale of electronic and telecommunications equipment and parts
	4653	Wholesale of agricultural machinery, equipment and supplies
	4659	Wholesale of other machinery and equipment
466		Other specialized wholesale
	4661	Wholesale of solid, liquid and gaseous fuels and related products
	4662	Wholesale of metals and metal ores
	4663	Wholesale of construction materials, hardware, plumbing and heating equipment and supplies
	4669	Wholesale of waste and scrap and other products n.e.c.
469	4690	Non-specialized wholesale trade
Division 47		**Retail trade, except of motor vehicles and motorcycles**
471		Retail sale in non-specialized stores
	4711	Retail sale in non-specialized stores with food, beverages or tobacco predominating
	4719	Other retail sale in non-specialized stores
472		Retail sale of food, beverages and tobacco in specialized stores

	4721	Retail sale of food in specialized stores
	4722	Retail sale of beverages in specialized stores
	4723	Retail sale of tobacco products in specialized stores
473	4730	Retail sale of automotive fuel in specialized stores
474		Retail sale of information and communications equipment in specialized stores
	4741	Retail sale of computers, peripheral units, software and telecommunications equipment in specialized stores
	4742	Retail sale of audio and video equipment in specialized stores
475		Retail sale of other household equipment in specialized stores 4751 Retail sale of textiles in specialized stores
	4752	Retail sale of hardware, paints and glass in specialized stores
	4753	Retail sale of carpets, rugs, wall and floor coverings in specialized stores
	4759	Retail sale of electrical household appliances, furniture, lighting equipment and other household articles in specialized stores
476		Retail sale of cultural and recreation goods in specialized stores
	4761	Retail sale of books, newspapers and stationary in specialized stores
	4762	Retail sale of music and video recordings in specialized stores
	4763	Retail sale of sporting equipment in specialized stores
	4764	Retail sale of games and toys in specialized stores
477		Retail sale of other goods in specialized stores
	4771	Retail sale of clothing, footwear and leather articles in specialized stores
	4772	Retail sale of pharmaceutical and medical goods, cosmetic and toilet articles in specialized stores
	4773	Other retail sale of new goods in specialized stores
	4774	Retail sale of second-hand goods
478		Retail sale via stalls and markets
	4781	Retail sale via stalls and markets of food, beverages and tobacco products
	4782	Retail sale via stalls and markets of textiles, clothing and footwear
	4789	Retail sale via stalls and markets of other goods
479		Retail trade not in stores, stalls or markets
	4791	Retail sale via mail order houses or via Internet
	4799	Other retail sale not in stores, stalls or markets

Section H Transportation and Storage

Division 49		**Land transport and transport via pipelines**
491		Transport via railways
	4911	Passenger rail transport, interurban
	4912	Freight rail transport

492		Other land transport
	4921	Urban and suburban passenger land transport
	4922	Other passenger land transport
	4923	Freight transport by road
493	4930	Transport via pipeline
Division 50		**Water transport**
501		Sea and coastal water transport
	5011	Sea and coastal passenger water transport
	5012	Sea and coastal freight water transport
502		Inland water transport
	5021	Inland passenger water transport
	5022	Inland freight water transport
Division 51		**Air transport**
511	5110	Passenger air transport
512	5120	Freight air transport
Division 52		**Warehousing and support activities for transportation**
521	5210	Warehousing and storage
522		Support activities for transportation
	5221	Service activities incidental to land transportation
	5222	Service activities incidental to water transportation
	5223	Service activities incidental to air transportation
	5224	Cargo handling
	5229	Other transportation support activities
Division 53		**Postal and courier activities**
531	5310	Postal activities
532	5320	Courier activities

Section I Accommodation and food service activities

Division 55		**Accommodation**
551	5510	Short term accommodation activities
552	5520	Camping grounds, recreational vehicle parks and trailer parks
559	5590	Other accommodation
Division 56		**Food and beverage service activities**
561	5610	Restaurants and mobile food service activities

562		Event catering and other food service activities
	5621	Event catering 5629 Other food service activities
563	5630	Beverage serving activities
Division 58		**Publishing activities**
581		Publishing of books, periodicals and other publishing activities
	5811	Book publishing
	5812	Publishing of directories and mailing lists
	5813	Publishing of newspapers, journals and periodicals
	5819	Other publishing activities
582	5820	Software publishing
Division 59		**Motion picture, video and television programme production, sound recording and music publishing activities**
581		Publishing of books, periodicals and other publishing activities
	5811	Book publishing
	5812	Publishing of directories and mailing lists
	5813	Publishing of newspapers, journals and periodicals
	5819	Other publishing activities
582	5820	Software publishing
Division 60		**Programming and broadcasting activities**
601	6010	Radio broadcasting
602	6020	Television programming and broadcasting activities
Division 61		**Telecommunications**
611	6110	Wired telecommunications activities
612	6120	Wireless telecommunications activities
613	6130	Satellite telecommunications activities
619	6190	Other telecommunications activities
Division 62		**Computer programming, consultancy and related activities**
	6201	Computer programming activities
	6202	Computer consultancy and computer facilities management activities
	6209	Other information technology and computer service activities
Division 63		**Information service activities**
631		Data processing, hosting and related activities; web portals
	6311	Data processing, hosting and related activities
	6312	Web portals

639		Other information service activities
	6391	News agency activities
	6399	Other information service activities n.e.c

Section K Financial and insurance activities

Division 64		**Financial service activities, except insurance and pension funding**
641		Monetary intermediation
	6411	Central banking
	6419	Other monetary intermediation
642	6420	Activities of holding companies
643	6430	Trusts, funds and similar financial entities
649		Other financial service activities, except insurance and pension funding activities
	6491	Financial leasing
	6492	Other credit granting
	6499	Other financial service activities, except insurance and pension funding activities, n.e.c
Division 65		**Insurance, reinsurance and pension funding, except compulsory social security**
651		Insurance
	6511	Life insurance
	6512	Non-life insurance
652	6520	Reinsurance
653	6530	Pension funding
Division 66		**Activities auxiliary to financial service and insurance activities**
661		Activities auxiliary to financial service activities, except insurance and pension funding
	6611	Administration of financial markets
	6612	Security and commodity contracts brokerage
	6619	Other activities auxiliary to financial service activities
662		Activities auxiliary to insurance and pension funding
	6621	Risk and damage evaluation 6622 Activities of insurance agents and brokers
	6629	Other activities auxiliary to insurance and pension funding
663	6630	Fund management activities

Section L Real estate activities

Division 68		**Real estate activities**
681	6810	Real estate activities with own or leased property

682	6820	Real estate activities on a fee or contract basis

Section M Professional, scientific and technical activities

Division 69		**Legal and accounting activities**
691	6910	Legal activities
692	6920	Accounting, bookkeeping and auditing activities; tax consultancy
Division 70		**Activities of head offices; management consultancy activities**
701	7010	Activities of head offices
702	7020	Management consultancy activities
Division 71		**Architectural and engineering activities; technical testing and analysis**
711	7110	Architectural and engineering activities and related technical consultancy
712	7120	Technical testing and analysis
Division 72		**Scientific research and development**
721	7210	Research and experimental development on natural sciences and engineering
722	7220	Research and experimental development on social sciences and humanities
Division 73		**Advertising and market research**
731	7310	Advertising
732	7320	Market research and public opinion polling
Division 74		**Other professional, scientific and technical activities**
741	7410	Specialized design activities
742	7420	Photographic activities
749	7490	Other professional, scientific and technical activities n.e.c.
Division 75		**Veterinary activities**
750	7500	Veterinary activities

Section N Administrative and support service activities

Division 77		**Rental and leasing activities**
771	7710	Renting and leasing of motor vehicles
772		Renting and leasing of personal and household goods
	7721	Renting and leasing of recreational and sports goods
	7722	Renting of video tapes and disks

	7729	Renting and leasing of other personal and household goods
773	7730	Renting and leasing of other machinery, equipment and tangible goods
774	7740	Leasing of intellectual property and similar products, except copyrighted works
Division 78		**Employment activities**
781	7810	Activities of employment placement agencies
782	7820	Temporary employment agency activities
783	7830	Other human resources provision
Division 79		**Travel agency, tour operator, reservation service and related activities**
791		Travel agency and tour operator activities
	7911	Travel agency activities
	7912	Tour operator activities
799	7990	Other reservation service and related activities
Division 80		**Security and investigation activities**
801	8010	Private security activities
802	8020	Security systems service activities
803	8030	Investigation activities
Division 81		**Services to buildings and landscape activities**
811	8110	Combined facilities support activities
812		Cleaning activities
	8121	General cleaning of buildings
	8129	Other building and industrial cleaning activities
813	8130	Landscape care and maintenance service activities
Division 82		**Office administrative, office support and other business support activities**
821		Office administrative and support activities
	8211	Combined office administrative service activities
	8219	Photocopying, document preparation and other specialized office support activities
822	8220	Activities of call centres
823	8230	Organization of conventions and trade shows
829		Business support service activities n.e.c.
	8291	Activities of collection agencies and credit bureaus

8292 Packaging activities 8299 Other business support service activities n.e.c.

Section O Public administration and defence; compulsory social security

Division 84 Public administration and defence; compulsory social security

841 Administration of the State and the economic and social policy of the community
- 8411 General public administration activities
- 8412 Regulation of the activities of providing health care, education, cultural services and other social services, excluding social security
- 8413 Regulation of and contribution to more efficient operation of businesses

842 Provision of services to the community as a whole
- 8421 Foreign affairs
- 8422 Defence activities
- 8423 Public order and safety activities

843 8430 Compulsory social security activities

Section P Education

Division 85 Education

851 8510 Pre-primary and primary education

852 Secondary education
- 8521 General secondary education
- 8522 Technical and vocational secondary education

853 8530 Higher education

854 Other education
- 8541 Sports and recreation education
- 8542 Cultural education
- 8549 Other education n.e.c.

855 8550 Educational support activities

Section Q Human health and social work activities

Division 86 Human health activities

- 861 8610 Hospital activities
- 862 8620 Medical and dental practice activities
- 869 8690 Other human health activities

Division 87 Residential care activities

871 8710 Residential nursing care facilities

872	8720	Residential care activities for mental retardation, mental health and substance abuse
873	8730	Residential care activities for the elderly and disabled
879	8790	Other residential care activities

Division 88 Social work activities without accommodation

881	8810	Social work activities without accommodation for the elderly and disabled
889	8890	Other social work activities without accommodation

Section R Arts, entertainment and recreation

Division 90 Creative, arts and entertainment activities

900	9000	Creative, arts and entertainment activities

Division 91 Libraries, archives, museums and other cultural activities

	9101	Library and archives activities
	9102	Museums activities and operation of historical sites and buildings
	9103	Botanical and zoological gardens and nature reserves activities

Division 92 Gambling and betting activities

920	9200	Gambling and betting activities

Division 93 Sports activities and amusement and recreation activities

931		Sports activities
	9311	Operation of sports facilities
	9312	Activities of sports clubs
	9319	Other sports activities
932		Other amusement and recreation activities
	9321	Activities of amusement parks and theme parks
	9329	Other amusement and recreation activities n.e.c.

Section S Other activities

Division 94 Activities of membership organizations

941		Activities of business, employers and professional membership organizations
	9411	Activities of business and employers membership organizations
	9412	Activities of professional membership organizations
942	9420	Activities of trade unions
949		Activities of other membership organizations
	9491	Activities of religious organizations
	9492	Activities of political organizations
	9499	Activities of other membership organizations n.e.c

Division 95		**Repair of computers and personal and household goods**
951		Repair of computers and communication equipment
	9511	Repair of computers and peripheral equipment
	9512	Repair of communication equipment
952		Repair of personal and household goods
	9521	Repair of consumer electronics
	9522	Repair of household appliances and home and garden equipment
	9523	Repair of footwear and leather goods
	9524	Repair of furniture and home furnishings
	9529	Repair of other personal and household goods
Division 96		**Other personal service activities**
	9601	Washing and (dry-) cleaning of textile and fur products
	9602	Hairdressing and other beauty treatment
	9603	Funeral and related activities
	9609	Other personal service activities n.e.c.
Section T		**Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use**
Division 97		**Activities of households as employers of domestic personnel**
970	9700	Activities of households as employers of domestic personnel
Division 98		**Undifferentiated goods- and services-producing activities of private households for own use**
981	9810	Undifferentiated goods-producing activities of private households for own use
982	9820	Undifferentiated service-producing activities of private households for own use
Section U		**Activities of extraterritorial organizations and bodies**
Division 99		**Activities of extraterritorial organizations and bodies**
990	9900	Activities of extraterritorial organizations and bodies

The individual categories of ISIC can be aggregated into 21 classes as shown below.

Section	Divisions	Description
A	01-03	Agriculture, forestry and fishing
B	05-09	Mining and quarrying
C	10-33	Manufacturing
D	35	Electricity, gas, steam and air conditioning supply
E	36-39	Water supply; sewerage, waste management and remediation activities
F	41-43	Construction
G	45.47	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	49-53	Transportation and storage
I	55-56	Accommodation and food service activities
J	58-63	Information and communication
K	64-66	Financial and insurance activities
L	68	Real estate activities
M	69-75	Professional, scientific and technical activities
N	77-82	Administrative and support service activities
O	84	Public administration and defence; compulsory social security
P	85	Education
Q	86-88	Human health and social work activities
R	90-93	Arts, entertainment and recreation
S	94-96	Other service activities
T	97-98	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
U	99	Activities of extraterritorial organizations and bodies

For older versions of the ISIC, kindly refer to
https://unstats.un.org/unsd/classifications/Family/Detail/27

Annex III.2: Broad structure of European Classification of Economic Activities (NACE)

The NACE Revision 2 classification is very similar to ISIC Rev 4.0 Division (see Annex III.1). NACE 2.0 classification contains 21 levels (A to U); Level 2 contains 88 divisions identified by 2-digit codes (01 to 99); Level 3 contains 272 groups identified by 3-digit codes (01.1 to 99.0) and Level 4 contains 615 classes identified by 4-digit codes (01.11 to 99.00).[25]

NACE Rev. 1.1		NACE Rev. 2.0		
Section	**Description**	**Section**	**Description**	**Divisions**
A	Agriculture, hunting and forestry	A	Agriculture, forestry and fishing	01-03
B	Fishing			
C	Mining and quarrying	B	Mining and quarrying	05-09
D	Manufacturing	C	Manufacturing	10-33
E	Electricity, gas and water supply	D	Electricity, gas, steam and air conditioning supply	35
		E	Water supply, sewerage, waste management and remediation activities	36-39
F	Construction	F	Construction	41-43
G	Wholesale and retail trade: repair of motor vehicles, motorcycles and personal and household goods	G	Wholesale and retail trade; repair of motor vehicles and motorcycles	45.47
H	Hotels and restaurants	I	Accommodation and food service activities	49-53
I	Transport, storage and communications	H	Transportation and storage	55-56
		J	Information and communication	58-63
J	Financial intermediation	K	Financial and insurance activities	64-66
K	Real estate, renting and business activities	L	Real estate activities	68
		M	Professional, scientific and technical activities	69-75
		N	Administrative and support service activities	77-82
L	Public administration and defence; compulsory social security	O	Public administration and defence; compulsory social security	84
M	Education	P	Education	85
N	Health and social work	Q	Human health and social work activities	86-88
O	Other community, social and personal services activities	R	Arts, entertainment and recreation	90-93
		S	Other service activities	94-96

[25] See Chapter 4 for NACE details and linkage to ISIC Rev 4
https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF

P	Activities of private households as employers and undifferentiated production activities of private households	T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	97-98
Q	Extraterritorial organisations and bodies	U	Activities of extraterritorial organisations and bodies	99

Source: EUROSTAT (2008).

The first four digits of the code, which is the first four levels of the classification system, are the same in all European countries. National implementations may introduce additional levels. The fifth digit might vary from country to country and further digits are sometimes placed by suppliers of databases.

For a detailed structure of NACE Revision 2.0 see Part III (Eurostat (2008) NACE Revision 2 Statistical classification of economic activities in the European Community).

Mapping ISIC and NACE codes to INDUSTRYCAT10 codes

When creating labor sector variables, one needs to carefully check the type of economic activity codes and its revision which the household survey uses. Some surveys follow ISIC (International Standard Industrial Classification), while others follow NACE (Statistical Classification of Economic Activities in the European Community). See below the relationship of all the economic activity by Section and Division.

INUSTRYCAT10	ISIC Rev 3.1		ISIC Rev 4.0		NACE Rev 1.1		NACE Rev 2.0	
	Section	Division	Section	Division	Section	Division	Section	Division
1 - Agriculture, Hunting, Fishing	A B	01-05	A	01-03	A B	01-05	A	01-03
2 - Mining	C	10-14	B	05-09	C	10-14	B	05-09
3 - Manufacturing	D	15-37	C	10-33	D	15-37	C	10-33
4 - Public Utility Services	E	40-41	D E	35; 36-39	E	40-41	D E	35-39
5 - Construction	F	45	F	41-43	F	45	F	41-43
6 - Commerce	G	50-52	G	45-47	G	50-52	G	45-47
7 - Transport and Communications	I	60-64	H J	49-53; 58-63	I	60-64	H J	49-53; 58-63
8 - Financial and Business Services	J K	65-74	K L M N	64-66; 68-82	J K	65-74	K L M N	64-82
9 - Public Administration	L	75	O	84	L	75	O	84
10 - Other services, Unspecified	H M N O P Q	55; 80-99	I P Q R S T U	55-56; 85-99	H	55; 80-99	I P Q R S T U	55-56; 86-99

Mapping INDUSTRYCAT10 codes to INDUSTRYCAT4 codes

✓ Note that construction in *INDUSTRYCAT10* is mapped to industry in *INDSUTRYCAT4.* See STATA code in variable description.

INDUSTRYCAT10	INDUSTRYCAT4			
	1 Agriculture	2 Industry	3 Services	4 Other
1 - Agriculture, Hunting, Fishing				
2 - Mining				
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				

Annex III.3: International Standard Industrial Classification of Occupations (ISCO)

The International Standard Classification of Occupations (ISCO) is the International Labour Organization (ILO) classification structure for organizing information on labor and jobs. It is part of the international family of economic and social classifications of the United Nations. The current version, known as ISCO-08, was published in 2008 and is the fourth iteration, following ISCO-58, ISCO-68 and ISCO-88. See for a detailed ISCO structure[26].

1 Managers

11 Chief Executives, Senior Officials and Legislators
- 111 Legislators and Senior Officials
- 112 Managing Directors and Chief Executives

12 Administrative and Commercial Managers
- 121 Business Services and Administration Managers
- 122 Sales, Marketing and Development Managers

13 Production and Specialized Services Managers
- 131 Production Managers in Agriculture, Forestry and Fisheries
- 132 Manufacturing, Mining, Construction and Distribution Managers
- 133 Information and Communications Technology Services Managers
- 134 Professional Services Managers

14 Hospitality, Retail and Other Services Managers
- 141 Hotel and Restaurant Managers
- 142 Retail and Wholesale Trade Managers
- 143 Other Services Managers

2 Professionals

21 Science and Engineering Professionals
- 211 Physical and Earth Science Professionals
- 212 Mathematicians, Actuaries and Statisticians
- 213 Life Science Professionals
- 214 Engineering Professionals (excluding Electrotechnology)
- 215 Electrotechnology Engineers
- 216 Architects, Planners, Surveyors and Designers

22 Health Professionals
- 221 Medical Doctors
- 222 Nursing and Midwifery Professionals
- 223 Traditional and Complementary Medicine Professionals
- 224 Paramedical Practitioners
- 225 Veterinarians

[26] https://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm

226 Other Health Professionals

23 Teaching Professionals

231 University and Higher Education Teachers
232 Vocational Education Teachers
233 Secondary Education Teachers
234 Primary School and Early Childhood Teachers
235 Other Teaching Professionals

24 Business and Administration Professionals

241 Finance Professionals
242 Administration Professionals
243 Sales, Marketing and Public Relations Professionals

25 Information and Communications Technology Professionals

251 Software and Applications Developers and Analysts
252 Database and Network Professionals

26 Legal, Social and Cultural Professionals

261 Legal Professionals
262 Librarians, Archivists and Curators
263 Social and Religious Professionals
264 Authors, Journalists and Linguists
265 Creative and Performing Artists

3 Technicians and Associate Professionals

31 Science and Engineering Associate Professionals

311 Physical and Engineering Science Technicians
312 Mining, Manufacturing and Construction Supervisors
313 Process Control Technicians
314 Life Science Technicians and Related Associate Professionals
315 Ship and Aircraft Controllers and Technicians

32 Health Associate Professionals

321 Medical and Pharmaceutical Technicians
322 Nursing and Midwifery Associate Professionals
323 Traditional and Complementary Medicine Associate Professionals
324 Veterinary Technicians and Assistants
325 Other Health Associate Professionals

33 Business and Administration Associate Professionals

331 Financial and Mathematical Associate Professionals
332 Sales and Purchasing Agents and Brokers
333 Business Services Agents
334 Administrative and Specialized Secretaries
335 Government regulatory associate professionals

34 Legal, Social, Cultural and Related Associate Professionals
341 Legal, Social and Religious Associate Professionals
342 Sports and Fitness Workers
343 Artistic, Cultural and Culinary Associate Professionals

35 Information and Communications Technicians
351 Information and Communications Technology Operations and User Support Technicians
352 Telecommunications and Broadcasting Technicians

4 Clerical Support Workers

41 General and Keyboard Clerks
411 General Office Clerks
412 Secretaries (general)
413 Keyboard Operators

42 Customer Services Clerks
421 Tellers, Money Collectors and Related Clerks
422 Client Information Workers

43 Numerical and Material Recording Clerks
431 Numerical Clerks
432 Material recording and Transport Clerks

44 Other Clerical Support Workers
441 Other Clerical Support Workers

5 Services and Sales Workers

51 Personal Services Workers
511 Travel Attendants, Conductors and Guides
512 Cooks
513 Waiters and Bartenders
514 Hairdressers, Beauticians and Related Workers
515 Building and Housekeeping Supervisors
516 Other Personal Services Workers

52 Sales Workers
521 Street and Market Salespersons
522 Shop Salespersons
523 Cashiers and Ticket Clerks
524 Other Sales Workers

53 Personal Care Workers
531 Child Care Workers and Teachers' Aides
532 Personal Care Workers in Health Services

54 Protective Services Workers

541 Protective Services Workers

6 Skilled Agricultural, Forestry and Fishery Workers

61 Market-oriented Skilled Agricultural Workers

611 Market Gardeners and Crop Growers
612 Animal Producers
613 Mixed Crop and Animal Producers

62 Market-oriented Skilled Forestry, Fishery and Hunting Workers

621 Forestry and Related Workers
622 Fishery Workers, Hunters and Trappers

63 Subsistence Farmers, Fishers, Hunters and Gatherers

631 Subsistence Crop Farmers
632 Subsistence Livestock Farmers
633 Subsistence Mixed Crop and Livestock Farmers
634 Subsistence Fishers, Hunters, Trappers and Gatherers

7 Craft and Related Trades Workers

71 Building and Related Trades Workers (excluding Electricians)

711 Building Frame and Related Trades Workers
712 Building Finishers and Related Trades Workers
713 Painters, Building Structure Cleaners and Related Trades Workers

72 Metal, Machinery and Related Trades Workers

721 Sheet and Structural Metal Workers, Moulders and Welders, and Related Workers
722 Blacksmiths, Toolmakers and Related Trades Workers
723 Machinery Mechanics and Repairers

73 Handicraft and Printing Workers

731 Handicraft Workers
732 Printing Trades Workers

74 Electrical and Electronic Trades Workers

741 Electrical Equipment Installers and Repairers
742 Electronics and Telecommunications Installers and Repairers

75 Food Processing, Woodworking, Garment and Other Craft and Related Trades Workers

751 Food Processing and Related Trades Workers
752 Wood Treaters, Cabinet-makers and Related Trades Workers
753 Garment and Related Trades Workers
754 Other Craft and Related Workers

8 Plant and Machine Operators and Assemblers

81 Stationary Plant and Machine Operators
- 811 Mining and Mineral Processing Plant Operators
- 812 Metal Processing and Finishing Plant Operators
- 813 Chemical and Photographic Products Plant and Machine Operators
- 814 Rubber, Plastic and Paper Products Machine Operators
- 815 Textile, Fur and Leather Products Machine Operators
- 816 Food and Related Products Machine Operators
- 817 Wood Processing and Papermaking Plant Operators
- 818 Other Stationary Plant and Machine Operators

82 Assemblers
- 821 Assemblers

83 Drivers and Mobile Plant Operators
- 831 Locomotive Engine Drivers and Related Workers
- 832 Car, Van and Motorcycle Drivers
- 833 Heavy Truck and Bus Drivers
- 834 Mobile Plant Operators
- 835 Ships' Deck Crews and Related Workers

9 Elementary Occupations

91 Cleaners and Helpers
- 911 Domestic, Hotel and Office Cleaners and Helpers
- 912 Vehicle, Window, Laundry and Other Hand Cleaning Workers

92 Agricultural, Forestry and Fishery Labourers
- 921 Agricultural, Forestry and Fishery Labourers

93 Labourers in Mining, Construction, Manufacturing and Transport
- 931 Mining and Construction Labourers
- 932 Manufacturing Labourers
- 933 Transport and Storage Labourers

94 Food Preparation Assistants
- 941 Food Preparation Assistants

95 Street and Related Sales and Services Workers
- 951 Street and Related Services Workers
- 952 Street Vendors (excluding Food)

96 Refuse Workers and Other Elementary Workers
- 961 Refuse Workers
- 962 Other Elementary Workers

0 Armed Forces Occupations

01 Commissioned Armed Forces Officers
011 Commissioned Armed Forces Officers

02 Non-commissioned Armed Forces Officers
021 Non-commissioned Armed Forces Officers

03 Armed Forces Occupations, Other Ranks
031 Armed Forces Occupations, Other Ranks

Mapping ISCO codes to 1-digit occupation codes

- ✓ As creating occupation variable, we need carefully check the type of occupation codes and its revision which the household survey uses. Some surveys follow ISCO-88 (International Standard Classification of Occupations being approved in 1988), while others follow ISCO-08 (International Standard Classification of Occupations being approved in 2008)
- ✓ Below is a table and STATA code as an example of mapping sections and divisions of ISCO-88 occupation codes to categories of *OCCUP_** variable.

Occupation	ISCO-88 Major Groups	ISCO-08 Major Groups
0 - Army	01	01-03
1 - Managers	11-13	11-14
2 - Professionals	21-24	21-26
3 - Technicians and associate professionals	31-34	31-35
4 - Clerical support workers	41-42	41-44
5 - Service and sales workers	51-52	51-54
6 - Skilled agricultural, forestry and fishery workers	61-62	61-63
7 - Craft and related trades workers	71-74	71-75
8 - Plant and machine operators, and assemblers	81-83	81-83
9 - Elementary occupations	91-93	91-96