Harmonized Datafiles and Variables

for

High-Frequency Phone Surveys on COVID-19

Version 1 (February 2021)

Prepared by the Living Standards Measurement Study (LSMS) team

Objective

To facilitate the use of data collected through the high-frequency phone surveys on COVID-19, the Living Standards Measurement Study (LSMS) team has created the harmonized datafiles using two household surveys: 1) the country's latest face-to-face survey which has become the sample frame for the phone survey, and 2) the country's high-frequency phone survey on COVID-19.

The LSMS team has extracted and harmonized variables from these surveys, based on the harmonized definitions and ensuring the same variable names. These variables include demography as well as housing, household consumption expenditure, food security, and agriculture. Inevitably, many of the original variables are collected using questions that are asked differently. The harmonized datafiles include the best available variables with harmonized definitions.

Two harmonized datafiles are prepared for each survey. The two datafiles are:
1. HH: This datafile contains household-level variables. The information include basic household characterizes, housing, water and sanitation, asset ownership, consumption expenditure, consumption quintile, food security, livestock ownership. It also contains information on agricultural activities such as crop cultivation, use of organic and inorganic fertilizer, hired labor, use of tractor and crop sales.
2. IND: This datafile contains individual-level variables. It includes basic characteristics of individuals such as age, sex, marital status, disability status, literacy, education and work.

Harmonization Guidelines

- Household and individual IDs are not harmonized. The variable names and values are extracted and kept as they are in the original dataset.
- When the survey does not include variables necessary to create harmonized variables, the variables are still created as missing.
- Any assumptions made during the harmonization process are noted in the do-file as comments. When possible, these will also be attached as a note to the harmonized variables.
- When the data from the new round of phone surveys become available, they will be harmonized, and the harmonized variables will be added to the end of each harmonized datafile.
- Once the variables are harmonized, a quality check is performed by using programs called hh_validation and ind_validation. They test if all the variables are in the harmonized datafile, if all the variables are named as in the data dictionary, if all the variables have the correct format, if all the variables take plausible values or are in the plausible ranges, and if some of the variables are mutually consistent.
- Once the quality check is performed, the harmonized variables are labeled using programs called hh_label and ind_label. The programs label variables and value labels ensuring their consistency with the data dictionary.

Note

- The high-frequency phone survey on COVID-19 has multiple rounds of data collection. When variables are extracted from multiple rounds of the survey, the originating round of the survey is noted with "_rX" in the variable name, where X represents the number of the round. For example, a variable with "_r3" presents that the variable was extracted from Round 3 of the

high-frequency phone survey. Round 0 refers to the country's latest face-to-face survey which has become the sample frame for the high-frequency phone surveys on COVID-19. When the variables are without "_rX", they were extracted from Round 0.

- All harmonized datasets contain a variable which is a unique identifier for the household (Household ID). This variable is used as the unique key variable in the merging of the two harmonized datafiles.
- The harmonized datasets can be merged with (1) the country's latest face-to-face survey which has become the sample frame for the phone survey, and (2) the country's high-frequency phone survey on COVID-19. The household's unique identifier variable (Household ID) is used to merge the household-level datasets and the individual's unique identifier variables (both Household ID and Individual ID) are used to merge the individual-level files.
- Individuals who are added after Round 0 (during the country's high-frequency phone survey on COVID-19) are missing certain information which were collected only in Round 0 (the country's latest face-to-face survey). It is also impossible to merge these individuals with the Round 0 dataset. These individuals can be identified with the variable "member_r0" (member_r0=0) in the individual-level datafile.
- There are 2 variables available for age of household members: (1) age at Round 0, and (2) age at Round 1+. The former is not available for the individuals who are added after Round 0 (during the country's high-frequency phone survey on COVID-19). The latter is updated with the most recent round of the survey.

Data Dictionary

Household-level file

The datafile is named CCC_HH where CCC=three-letter ISO country code. In addition to the harmonized variables, the datafile also includes country-specific geographic variables.

Name	Labels and codes	Instructions / notes
hhid	**Household ID**	Household Unique Identifier Production note: The variable name and the values are kept as they are in the Round 0 original dataset.
year	**Survey year (Round 0)**	The year when data collection started for Round 0 (format YYYY).
rural	**Rural/Urban** *1=Rural* *2=Urban*	Urban/rural jurisdiction (based on country-specific definition). *Production note:* The 'semi-urban' category is assimilated to 'urban'.
ea_latitude	**EA Latitude (Modified)**	Average of household GPS coordinates in each EA taken at Round 0 (a random offset within a specified range is applied following the MeasureDHS methodology).

ea_longitude	EA Longitude (Modified)	Average of household GPS coordinates in each EA taken at Round 0 (a random offset within a specified range is applied following the MeasureDHS methodology).
dwelling	**Ownership of dwelling** *0=No* *1=Yes*	Ownership of dwelling that the household resides in. • *Yes* includes ownership whether or not full-payment has yet been made. • *No* includes free (authorized and not authorized) and rented.
roof	**Modern roof** *0=No* *1=Yes*	The dwelling that the household resides in has modern roof. • *Yes* includes corrugated iron sheets, clay tiles, concrete, cement, plastic sheet, asbestos sheet, step tiles, long/short span sheets, zinc and aluminum. • *No* includes thatch and mud.
floor	**Modern floor** *0=No* *1=Yes*	The dwelling that the household resides in has modern floor. • *Yes* includes cement, concrete, wood, and tile. • *No* includes sand, dirt, straw, and smoothed mud.
walls	**Modern exterior walls** *0=No* *1=Yes*	The dwelling that the household resides in has modern exterior walls. • *Yes* includes burnt bricks, cement, concrete, and iron sheets. • *No* includes mud, stone, unburnt bricks, wood, bamboo, and cardboard.
toilet	**Access to improved toilet** *0=No* *1=Yes*	The household has access to improved toilet, based on JMP standard definition. • *Yes* includes flush/pour-flush toilet to sewers, septic tanks or pit latrines, pit latrine with slab, composting toilet, twin pit latrine with slab, and container based sanitation. This also includes flush/pour-flush toilet to don't know where. • *No* includes flush/pour-flush to open drain, flush/pour-flush to elsewhere, pit latrine without slab/open-pit, bucket, hanging toilet/hanging latrine, and no facility/bush/field.
water	**Access to improved drinking water source** *0=No* *1=Yes*	The household has access to improved water source, based on JMP standard definition. • *Yes* includes piped, public tap or standpipe, borehole or tube well, protected well, protected spring, rainwater collection, tanker-truck, cart with

		small tank/drum, water kiosk, bottled water, and sachet water.
		• *No* includes unprotected well, unprotected spring, and surface water.
		If the main source of water differs between the wet and dry season, refers to the water source during dry season.
rooms	**Number of rooms**	Number of habitable rooms occupied by the household. It excludes storerooms, toilets, bathrooms, kitchens and garage.
elect	**Connection to electricity** *0=No* *1=Yes*	The household has connection to electricity, irrespective of its source or its use.
tv	**Ownership of television** *0=No* *1=Yes*	Ownership of a television, irrespective of who owns it in the household.
radio	**Ownership of radio** *0=No* *1=Yes*	Ownership of a radio, irrespective of who owns it in the household.
refrigerator	**Ownership of refrigerator** *0=No* *1=Yes*	Ownership of a refrigerator, irrespective of who owns it in the household.
bicycle	**Ownership of bicycle** *0=No* *1=Yes*	Ownership of a bicycle, irrespective of who owns it in the household.
mcycle	**Ownership of motorcycle** *0=No* *1=Yes*	Ownership of a motorcycle, irrespective of who owns it in the household.
car	**Ownership of car or other vehicle** *0=No* *1=Yes*	Ownership of a car or any other vehicle, irrespective of who owns it in the household.
mphone	**Ownership of mobile phone** *0=No* *1=Yes*	Ownership of a mobile phone, irrespective of who owns it in the household.
computer	**Ownership of computer** *0=No* *1=Yes*	Ownership of a computer, irrespective of who owns it in the household.
internet	**Access to Internet** *0=No* *1=Yes*	Access to Internet by a device owned by the household, irrespective of who owns it in the household.

generator	**Ownership of generator** *0=No* *1=Yes*	Ownership of a generator, irrespective of who owns it in the household.
land	**Ownership of land** *0=No* *1=Yes*	Ownership of any type of land by the household irrespective of its use. This includes residential land, agricultural land (cultivated, fallow, rented out), pastureland, forest and business/commercial plots.
land_tot	**Total land size owned (ha)**	Includes both residential and agricultural land. Land size should be in hectares. By convention 1 ha = 2.471 acres. *Production note:* land_tot equals to 0 if the household does not own any land (land=0).
land_cultivated	**Total land size cultivated (ha)**	Total land area cultivated by the household in hectares regardless of ownership. By convention 1 ha = 2.471 acres. *Production note:* land_cultivate equals to 0 if the household does not cultivate any crops (crop=0).
cons_quint	**Consumption quintile** *1=Poorest* *2=Poorer* *3=Middle* *4=Richer* *5=Richest*	The household's quintile ranking identified by the household's annual normalized per-capita consumption expenditure with price (spatial and temporal) adjustments. This includes food and non-food items, and includes purchased, own produced and received (gifts and other sources). *Production note:* Consumption aggregates used are retrieved from the Round 0 original dataset and are not recalculated for the harmonized dataset.
totcons	**Total annual per capita consumption**	Total annual per capita consumption in local currency, without price (spatial and temporal) adjustments. *Production note:* Consumption aggregates used are retrieved from the Round 0 original dataset and are not recalculated for the harmonized dataset.
foodcons	**Total annual per capita food consumption**	Total annual per capita food consumption in local currency, without price (spatial and temporal) adjustments. *Production note:* Consumption aggregates used are retrieved from the Round 0 original dataset and are not recalculated for the harmonized dataset.
nonfoodcons	**Total annual per capita non-food consumption**	Total annual per capita non-food consumption in local currency, without price (spatial and temporal) adjustments. *Production note:* Consumption aggregates used are retrieved from the Round 0 original dataset and are not recalculated for the harmonized dataset.

totcons_adj	**Total per capita consumption, spatially and temporally adjusted**	Total annual per capita consumption in local currency, with price (spatial and temporal) adjustments. *Production note:* Consumption aggregates used are retrieved from the Round 0 original dataset and are not recalculated for the harmonized dataset.
foodcons_adj	**Total annual per capita food consumption, spatially and temporally adjusted**	Total annual per capita food consumption in local currency, with price (spatial and temporal) adjustments. *Production note:* Consumption aggregates used are retrieved from the Round 0 original dataset and are not recalculated for the harmonized dataset.
nonfoodcons_ adj	**Total annual per capita non-food consumption, spatially and temporally adjusted**	Total annual per capita non-food consumption in local currency, with price (spatial and temporal) adjustments. *Production note:* Consumption aggregates used are retrieved from the Round 0 original dataset and are not recalculated for the harmonized dataset.
rent	**Rental income** *0=No* *1=Yes*	The household has income from rent.
remit	**Received remittance** *0=No* *1=Yes*	The household reported to have received remittances (international and/or domestic).
assist	**Received assistance** *0=No* *1=Yes*	The household reported to have received assistance from institutions or government.
finance	**Account from financial institutions** *0=No* *1=Yes*	Ownership of an account from financial institutions, irrespective of who owns it in the household.
any_work	**% of working adults working**	Percentage of the working age (15-64) members of the household who worked in any income-generating activities in the last 7 days.
ag_work	**% of working adults working in agriculture**	Percentage of the working age (15-64) members of the household who worked in agricultural activities in the last 7 days.
nfe_work	**% of working adults working in non-farm family enterprise**	Percentage of the working age (15-64) members of the household who worked in non-farm family enterprise in the last 7 days.
ext_work	**% of working adults working in wage work**	Percentage of the working age (15-64) members of the household who worked in wage work in the last 7 days.

nfe	**Ownership of non-farm family enterprise** *0=No* *1=Yes*	Ownership of non-farm family enterprise, irrespective of who owns it in the household.
crop	**Crop cultivation** *0=No* *1=Yes*	The household cultivates any crops, irrespective of the final destination of the output.
crop_number	**Number of crops cultivated**	Number of crops cultivated by the household. *Production note:* Calculate only if the household cultivates any crops (crop=1).
cash_crop	**Cash crop cultivation** *0=No* *1=Yes*	The household cultivates at least one of the main cash crops in the country. *Production note:* Calculate only if the household cultivates any crops (crop=1).
org_fert	**Use of organic fertilizer** *0=No* *1=Yes*	The household uses organic fertilizer in at least one plot. *Production note:* Calculate only if the household cultivates any crops (crop=1).
inorg_fert	**Use of inorganic fertilizer** *0=No* *1=Yes*	The household uses inorganic fertilizer in at least one plot. *Production note:* Calculate only if the household cultivates any crops (crop=1).
pest_fung_herb	**Use of pesticides, fungicides or herbicides** *0=No* *1=Yes*	The household uses pesticides, fungicides or herbicides in at least one plot. *Production note:* Calculate only if the household cultivates any crops (crop=1).
hired_lab	**Use of hired labor** *0=No* *1=Yes*	The household uses hired labor in at least one plot. This excludes post-harvest activities. *Production note:* Calculate only if the household cultivates any crops (crop=1).
ex_fr_lab	**Use of exchange and/or free labor** *0=No* *1=Yes*	The household uses exchange and/or free labor in at least one plot. This excludes post-harvest activities. *Production note:* Calculate only if the household cultivates any crops (crop=1).
hired_lab_ph	**Use of hired labor for post-harvest activities** *0=No* *1=Yes*	The household uses hired labor in at least one of the harvested crops. *Production note:* Calculate only if the household cultivates any crops (crop=1).

ex_fr_lab_ph	**Use of exchange and/or free labor for post-harvest activities** *0=No* *1=Yes*	The household uses exchange and/or free labor in at least one of the harvested crops. *Production note:* Calculate only if the household cultivates any crops (crop=1).
tractor	**Use of tractor** *0=No* *1=Yes*	The household uses a tractor in at least one plot. *Production note:* Calculate only if the household cultivates any crops (crop=1).
ph_loss	**Post-harvest crop loss** *0=No* *1=Yes*	The household lost a portion of any crop after harvest. *Production note:* Calculate only if the household cultivates any crops (crop=1).
sell_crop	**Sale of crop** *0=No* *1=Yes*	The household sells a portion of the harvest of at least one crop in any state. • *Yes* includes sales of unprocessed and/or processed crops. *Production note:* Calculate only if the household cultivates any crops (crop=1).
sell_process	**Sale of processed crop** *0=No* *1=Yes*	The household sells a portion of processed crops. *Production note:* Calculate only if the household cultivates any crops (crop=1).
sell_unprocess	**Sale of unprocessed crop** *0=No* *1=Yes*	The household sells a portion of unprocessed crops. *Production note:* Calculate only if the household cultivates any crops (crop=1).
livestock	**Ownership of livestock** *0=No* *1=Yes*	Ownership of any livestock by the household.
lruminant	**Ownership of large ruminant** *0=No* *1=Yes*	Ownership of any large ruminants (cattle) by the household. *Production note:* lruminant equals to 0 (No) if the household does not own any livestock (livestock=0).
sruminant	**Ownership of small ruminant** *0=No* *1=Yes*	Ownership of any small ruminants (sheep/goat) by the household. *Production note:* sruminant equals to 0 (No) if the household does not own any livestock (livestock=0).
poultry	**Ownership of poultry** *0=No* *1=Yes*	Ownership of any chicken or other poultry by the household. It refers to any type of birds including geese, and doves.

		Production note: poultry equals to 0 (No) if the household does not own any livestock (livestock=0).
equines	**Ownership of equine** *0=No* *1=Yes*	Ownership of any horses/donkeys by the household. *Production note:* equines equals to 0 (No) if the household does not own any livestock (livestock=0).
camelids	**Ownership of camelid** *0=No* *1=Yes*	Ownership of any camelids by the household. *Production note:* camelids equals to 0 (No) if the household does not own any livestock (livestock=0).
pig	**Ownership of pig** *0=No* *1=Yes*	Ownership of any pigs by the household. *Production note:* pig equals to 0 (No) if the household does not own any livestock (livestock=0).
bee	**Ownership of bee colonies** *0=No* *1=Yes*	Ownership of any bee colonies by the household. *Production note:* bee equals to 0 (No) if the household does not own any livestock (livestock=0).
phone_sample	**Phone sample** *0=No* *1=Yes*	The household is selected for phone survey.
contact_rX	**Successfully contacted (Round X)** *0=No* *1=Yes*	The household was successfully contacted (Round X). *Production note:* Calculate only if the household is selected for phone survey (phone_sample=1).
interview_rX	**Interviewed (Round X)** *0=No* 1=Yes	The household was interviewed (Round X). • *Yes* includes interview completed and partially completed. *Production note:* Calculate only if the household is successfully contacted (contact_rX=1).
complete_rX	**Interview completed (Round X)** *0=No* *1=Yes*	The interview was completed for the household (Round X). *Production note:* Calculate only if the household is interviewed (interview_rX=1).
hhsize_rX	**Household size (Round X)**	Number of household members (based on country-specific definition of a household) from Round X. *Production note:* hhsize_rX should equal to m0_14_rX+ m15_64_rX+ m65_rX+ f0_14_rX+ f15_64_rX+ f65_rX. Calculate only if the household completed the interview (complete_rX=1).
m0_14_rX	**Number of males aged 0 to 14 (Round X)**	Number of male household members aged 0 to 14 years from Round X.

		Production note: Undefined age is counted as aged 15 to 64 years. Calculate only if the household completed the interview (complete_rX=1).
m15_64_rX	**Number of males aged 15 to 64 (Round X)**	Number of male household members aged 15 to 64 years from Round X.
		Production note: Undefined age is counted as aged 15 to 64 years. Calculate only if the household completed the interview (complete_rX=1).
m65_rX	**Number of males aged 65 and above (Round X)**	Number of male household members aged 65 years and above from Round X.
		Production note: Undefined age is counted as aged 15 to 64 years. Calculate only if the household completed the interview (complete_rX=1).
f0_14_rX	**Number of females aged 0 to 14 (Round X)**	Number of female household members aged 0 to 14 years from Round X.
		Production note: Undefined age is counted as aged 15 to 64 years. Calculate only if the household completed the interview (complete_rX=1).
f15_64_rX	**Number of females aged 15 to 64 (Round X)**	Number of female household members aged 15 to 64 years from Round X.
		Production note: Undefined age is counted as aged 15 to 64 years. Calculate only if the household completed the interview (complete_rX=1).
f65_rX	**Number of females aged 65 and above (Round X)**	Number of female household members aged 65 years and above from Round X. Calculate only if the household completed the interview.
		Production note: Undefined age is counted as aged 15 to 64 years. Calculate only if the household completed the interview (complete_rX=1).
adulteq_rX	**Adult equivalence (Round X)**	Number of adult equivalents in the household, computed based on the standard FAO scale.
		Production note: Calculate only if the household completed the interview (complete_rX=1). Calculate for each household by summing up the following adult equivalent factor given to each member according to his/her age and sex:

	Male	Female
<1 yr	0.27	0.27
1-3 yrs	0.45	0.45
4-6 yrs	0.61	0.61
7-9 yrs	0.73	0.73
10-12 yrs	0.86	0.78

Name	Labels and codes	Instructions / notes
		13-15 yrs 0.96 0.83 16-19 yrs 1.02 0.77 20 and above 1.00 0.73
fies_mod_rX	**Probability of being moderately/ severely food insecure >= 50% (Round X)**	Probability of the household being moderately or severely food insecure is higher than 50% (p>=0.5) (Round X). This variable is computed based on the Food Insecurity Experience Scale (FIES) methodology.
fies_sev_rX	**Probability of being severely food insecure >= 50% (Round X)**	Probability of the household being severely food insecure is higher than 50% (p>=0.5) (Round X). This variable is computed based on the Food Insecurity Experience Scale (FIES) methodology.
head_chg_rX	**Household head changed (Round X)** *0=No* *1=Yes*	The household member identified as head of household in Round X is different from the head identified in the previous round of the survey. *Production note:* Calculate only if the household is successfully interviewed (complete_rX=1).
respond_chg_ rX	**Respondent changed (Round X)** *0=No* *1=Yes*	Respondent in Round X differs from the respondent in the previous round of the survey in which the household was successfully contacted. *Production note:* Calculate only if the household is interviewed (interview_rX=1).
wt_rX	**Cross section household weight (Round X)**	Cross section weighting coefficient to be used in all calculations referring to household level data.
wt_panel_rX	**Panel household weight (Round X)**	Panel weighting coefficient that is applicable only to the household that was successfully interviewed in all X rounds.

Individual-level file

The datafile is named CCC_IND where CCC=three-letter ISO country code. In addition to the harmonized variables, the datafile also includes country-specific geographic variables.

Name	Labels and codes	Instructions / notes
hhid	**Household ID**	Household unique identifier. Production note: The variable name and the values are kept as they are in the Round 0 original dataset.
indiv	**Individual ID**	Individual identifier. It uniquely identifies the individual when combined with the household unique identifier. *Production note:* The variable name and the values are kept as they are in the Round 0+ original dataset.

sex	**Sex** *1=Male* *2=Female*	Sex of the household member.
age	**Age (Round 0)**	Age (in years) of the household member at Round 0. *Production note:* The values are taken from the Round 0 original dataset. Unknown values are saved as missing.
age_p	**Age (Round 1+)**	Age (in years) of the household member at Round 1+. *Production note:* The values are taken from the Round 1+ (phone survey) dataset. Unknown values are saved as missing. The values are updated with the latest data available.
married	**Currently married** *0=No* *1=Yes*	The household member is currently married. • *Yes* includes married (monogamous and polygamous) and informal/loose/civil union.
form_married	**Formerly married** *0=No* *1=Yes*	The household member is formerly married. • *Yes* includes divorced, separated and widowed.
nev_married	**Never married** *0=No* *1=Yes*	The household member is never married. • *Yes* includes single and never married.
disability	**With disability** *0=No* *1=Yes*	This household member is with disability. This variable is calculated based on the Washington Group on Disability Statistics methodology. *Production note:* Calculate it based on the cut-off recommended by the Washington Group for the Washington Group Short Set on Functioning (WG-SS): the level of inclusion is any 1 domain/question is coded A LOT OF DIFFICULTY or CANNOT DO AT ALL.
religion	**Religion** *1=Christianity* *2=Islam* *3=Other*	Religion of the household member. This variable is obtained by recoding country-specific information.
literacy	**Literacy** *0=No* *1=Yes*	This household member can read and write in any language.
educ	**Highest level of education completed** *0=None* *1=Primary* *2=Secondary* *3=Tertiary*	Highest level of education completed of the household member. This variable is obtained by recoding country-specific information. The best possible match is sought, but the correspondence between country-specific values and the standardized codes may be imperfect.

work	**Working status** *0=No* *1=Yes*	The household member has worked in the last 7 days.
member_rX	**Member of household (Round X)** *0=No* *1=Yes*	The person is identified as a member of the household (Round X). *Production note:* Calculate only if the household is successfully interviewed (complete_rX=1).
head_rX	**Head of household (Round X)** *0=No* *1=Yes*	The household member is identified as head of the household (Round X). *Production note:* For this dataset, there should be one and only one head per household. If there are more than one household head, the older one is considered as head. Calculate only if the household is successfully interviewed (complete_rX=1).
respond_rX	**Respondent (Round X)** *0=No* *1=Yes*	The household member is a respondent of the survey (Round X). *Production note:* Calculate only if the household is interviewed (interview_rX=1).
relation_rX	**Relationship to head of household (Round X)** *Country-specific codes*	Relationship of the household member to head of the household (Round X). The country-specific information is kept as is. *Production note:* Calculate only if the household is successfully interviewed (complete_rX=1).

Reference

Dupriez, Olivier (2005). Poverty PPPs: Building a Database on Household Consumption Profiles by ICP Basic Heading, Description of Work. Unpublished document, World Bank Development Data Group.