HOUSEHOLD SIZE - General Target Variable Report (GVR)

1. General Information

The target variable T_HH_SIZE indicates the number of individuals, including children, who reside within the respondent's household. Its values range from 1, which denotes a one-person household, to 96 – the maximum value for household size in the source data (see Table 1.1). Source data come either directly from the respondent (e.g. in response to the question: "*How many people live in your household?")* or indirectly through sources like a Kish grid or a set of questions about household members.

T_HH_SIZE is accompanied by the quality control variable QR_HH_SIZE_OUTLIER (see Table 1.1 and Section 3.3). This methodological indicator is assigned the code 1 whenever a respondent's household comprises more than 20 members, and is coded 0 otherwise.

The target variable report for T_HH_SIZE is accompanied by the Crosswalk Table (CWT): T_HH_SIZE_CWT_SDR2.xlsx. CWT Excel files in SDR2 contain details about mapping of source values to target values.

Table 1.1. HOUSEHOLD SIZE: Description of the target, source, and control variables

	Variable description	**Variable name**	**Variable values[a]**
Target variable	Household size (number of people in the household)	T_HH_SIZE	1 = One-person household (minimum number) 96 = Maximum number of people in household
Source variables			See: T_HH_SIZE_CWT_SDR2.xlsx
Control variables	Respondent's household comprises more than 20 members	QR_HH_SIZE_OUTLIER	0 = Household comprises 20 or fewer members 1 = Household comprises over 20 members

[a] Missing values are assigned according to the SDR2 missing codes schema, provided in the Appendix.

2. Survey Projects

Source variables that we used for T_HH_SIZE appear in 18 international survey projects: ABS, AMB, CB, CDCEE, CNEP, EB, EQLS, ESS, EVS, ISJP, ISSP, LB, LITS, NBB,

NEB, PA2, PPE7N, WVS, 125 waves and 2241 national surveys. The data cover 106 countries and years from 1966 to 2017.

3. General Rules and Procedures

3.1. Source data description

Source items on household size exhibit some variation in terms of how these data are captured and who they come from. Some surveys ask respondents directly about the total number of household members, including children. Other surveys do not include a direct question on household size. Instead, they provide a Kish table or a set of questions – for example about the age of each household member – that allows us to calculate the total household size.

To construct the target variable T_HH_SIZE, we use three types of source information:

1) Source variables that capture respondents' answers to a direct question about the total number of household members. These source variables for household size are the most common in SDR2. Typical wording for these items are: "Including yourself, how many people – including children – live here regularly as members of this household?" (ESS_4) or "How many people live in your residence?" (CNEP_3_MX).
2) Source data that include a set of questions about household members, which allows us to estimate respondents' household size. For example, the source survey provides information about the number of people of different ages living in the household: aged 18 or more, children aged 13 to 17, children aged 5 to 12, and children less than 5 years old (CNEP_3_ES). The sum of numbers from all sub-questions gives an estimate of the total household size.
3) Roster-type source questions asking the respondent to provide the names of all members of the household (LITS_3).

3.2. Rules of transformation of source variables into target variable

In constructing T_HH_SIZE, we give priority to source variables that capture respondents' answers to a direct question about the size of their household. We take into account only questions about the size of the household including family members only, and dismiss source variables that include other people residing with family members, such as servants.

When the source data do not include an explicit question about the respondent's household size, we use other variables to derive household size. For example, in CNEP_3_ES we combine information from four source questions: number of people aged 18 or more living in the household + number of children aged 13 to 17 + number of children aged 5 to 12 + number of children less than 5 years old; in ISSP_1985 we use the variable on household composition (i.e. v121), while in LITS_3 we rely on roster-type questions where the respondent is asked to list all the members of their household.

We do not harmonize source questions that ask only about adult members of the household (for exceptions, see Section 4, Special cases).

If a source data file provides more than one household size measure, we always prioritize direct questions about household size over indirect questions (roster-type questions, number of adults + number of children etc.). For any deviations in question wording, see Section 4.

Some source variables topcode the maximum number of household members (e.g. NEB_5 and NEB_7 use "5 or more," while other surveys, such as ISSP_2008, use "10 and more"). In T_HH_SIZE, such instances are assigned the target value corresponding to the source cut-off value. In the examples above, the target values are 5 and 10, respectively. We do not store this transformation in a separate control indicator.

Missing values are assigned according to the SDR2 missing codes schema, provided in the Appendix.

3.3. Methodological variables that accompany T_HH_SIZE

The quality control variable QR_HH_SIZE_OUTLIER identifies national surveys where values on T_HH_SIZE are larger than 20. The indicator is coded 1 whenever a respondent's household comprises more than 20 members and 0 when the number of household members is 20 or below (see Table 1.1). QR_HH_SIZE_OUTLIER provides an easy way to identify countries and years where values on T_HH_SIZE are larger than 20, and, if needed, to filter these instances out.

4. Special Cases

- ISSP_1985 Australia, the source variable measures household composition rather than number of people in the household. Hence, in T_HH_SIZE we recoded the maximum to the target value 12 that corresponds to the largest available source option ("Three adults, nine children") for this country.
- We exclude certain ISSP national surveys from harmonization of household size, because the source items do not allow calculating the number of persons in respondents' households, for example "Couple with little children (the oldest child less than 15 years)" (ISSP_1985 Italy) or "Couple with child(ren)" (ISSP_1987 Netherland). As a result, we did not harmonize source variables from: ISSP_1985 for Great Britain and Italy, ISSP_1987 and ISSP_1988 for Netherlands, and ISSP_1989 for Netherlands and Hungary.
- PPE7N_IN V340 provides source variables where the number of household members is measured via intervals. We recoded the source values so that "(2) TWO TO THREE MEMBERS" takes the the target value 2 (i.e. two people), "(3) FOUR TO SIX MEMBERS" is assigned the target value 5, "(4) SEVEN TO NINE MEMBERS" is recoded into 8, "(5) TEN TO TWELVE MEMBERS" is recoded to

11, "(6) THIRTEEN TO FIFTEEN MEMBERS" takes the value 14, and "(7) SIXTEEN OR MORE" is recoded to 16.

- In the exceptional cases listed below, we decided to harmonize source variables although they measure only the number of adults in the household. We did so given that these situations are rare and confined to ISSP (hence also the decision not to construct a control indicator).
  - Australia in ISSP_1991 v116 and ISSP_1992 v119 ("How many adults (over 18) live in this household?"),
  - Switzerland in ISSP_2000 v272 ("Total number of persons living in household CH: Respondents are considered as adults with 15 and older")
  - Israel in ISSP_2011 HOMPOP, ISSP_2012 HOMPOP, ISSP_2013 HOMPOP, ISSP_2014 HOMPOP, ISSP_2015 HOMPOP ("Adults of 18 years and older living in your household")
  - Norway in ISSP_2011 HOMPOP, ISSP_2012 HOMPOP, ISSP_2013 HOMPOP ("Including yourself, how many people live normally in your household, and how are the household members distributed on age groups? Number of people 18 years or more.")
  - Georgia in ISSP_2013 HOMPOP, ISSP_2014 HOMPOP, ISSP_2015 HOMPOP ("Including yourself, how many people – including children – usually live in your household? Adults 18 years and older")
- We do not harmonize household-size related variables that ask also about other persons than family members, for instance: ISSP_1992 v119 Philippines – "Actual number of household member including servants, transients, boarders", or ISSP_2004 v256 Spain – "How many persons live in your household, incl. respondent, maids and domestic helpers?".
- ISSP_2003 hompop Venezuela was omitted because of its very low maximum value (answer options for the household size were only 1 and 2).
- ISSP_2004 v256 South Africa is worded differently from other ISSP surveys: "Please list all persons in the household who eat from the same cooking pot and were resident 15 out of the past 30 days"). We included this source variable for harmonization.
- LITS_3: In the absence of a variable measuring the number of persons for this wave, we harmonized roster-type source questions where the respondent was asked to list all the members of their family (Q. 1.02 "What is [NAME]'s gender").

Appendix: Codes for missing values in SDR2

In the SDR database v.2 we identify different situations that warrant to be treated as missing data. Table A.1 lists all SDR2 missing value codes:

Table A.1. Codes for missing values in SDR2

SDR tag [a]	SPSS (STATA) codes	Label
		Standardized source codes for missing values
DK	-1 (.a)	Don't know
NA	-2 (.b)	No answer
REF	-3 (.c)	Refusal
DU	-4 (.d)	Don't understand the question
DNR	-5 (.e)	Any combination of DK, NA, REF, DU
INAP	-6 (.f)	Inapplicable
NEC	-7 (.g)	Not elsewhere classified
		SDR created codes for missing values
UNFIT	-8 (.h)	Source value does not fit to target
ERR	-9 (.i)	Errors in source data and undocumented source values
COMBI	-10 (.j)	Different missing codes on multiple sources taken for a target
CINAP	-11 (.k)	For control variables only: inapplicable
INSUF	-12 (.l)	For survey: Insufficiently defined response categories
QNA	-13 (.m)	For survey: Question not available

[a] Abbreviations for the labels corresponding to the SDR2 codes for missing values. These tags are used in the Crosswalk Table (CWT) files (Excel) that accompany documentation of SDR2 target variables.

In exceptional situations when codes for missing data listed in Table A.1 cannot be used, we apply a system missing <null> value.