CIVIL STATUS - General Target Variable Report (GVR)

1. General Information

The target variable T_CIVIL_STATUS is a harmonized measure of respondents' civil status that prioritizes, whenever possible, the respondent's status of living with a partner over their legal marital status. For example, if the source dataset provides two questions, one about legal marital status and one about the relationship with the person the respondent is currently living with, we prioritize the latter when coding T_CIVIL_STATUS. To illustrate, in ESS/5 we code T_CIVIL_STATUS as "Cohabitating" when the respondent is widowed but reports currently living with a partner.

T_CIVIL_STATUS classifies respondents' civil status into six categories: married, cohabitating, single, separated/divorced, widowed and other.

Table 1.1 describes the target, source, and control variables. Section 3.2 outlines the harmonization rules employed to construct the target variable.

T_CIVIL_STATUS is accompanied by five harmonization controls that capture source variable characteristics that deviate from the standard definitions of the target variable categories. The control variable C_CIVIL_STATUS_COHAB records source variables' features related to the target value 2 - "Cohabitating". C_CIVIL_STATUS_SINGLE, C_CIVIL_STATUS_SEP_DIV, C_CIVIL_STATUS_WIDOWED, and C_CIVIL_STATUS_OTHER capture deviations related to, respectively, target value 3 - "Single"; 4 - "Separated or Divorced"; 5 - "Widowed"; and 6 - "Other". For details, see Table 1.1, and Section 3.3.

The target variable report for respondent's civil status is accompanied by the following Excel documents:

- The Detailed Variable Report (DVR): T_CIVIL_STATUS_DVR_SDR2.xlsx. DVR Excel files in SDR2 systemize all information about source variables that were used for harmonization into a given target variable of the SDR2 database;
- The Crosswalk Table (CWT): T_CIVIL_STATUS_CWT_SDR2.xlsx. CWT Excel files in SDR2 contain details about the mapping of source values to target values.

Table 1.1. CIVIL STATUS: Description of the target, source, and control variables

	Variable description	**Variable name**	**Variable values***
Target variable	Respondent's civil status	T_CIVIL_STATUS	1 = married 2 = cohabitating 3 = single 4 = separated or divorced 5 = widowed 6 = other
Source variables			See:T_CIVIL_STATUS_DVR_SDR2.xlsx T_CIVIL_STATUS_CWT_SDR2.xlsx

Control variables	Harmonization control for target values “married” (1) and “cohabitating” (2)	C_CIVIL_STATUS_ COHAB	0 = distinct answer option for **cohabitating;** no answer options for legal cohabitation/partnership 1 = available answer options for **civil union** or for **legalized cohabitation/partnership** 2 = **merged** answer options: **married** together with the **cohabitation** 3 = **cohabitating** not present among answer options
	Harmonization control for target value “other” (6)	C_CIVIL_STATUS_ OTHER	0 = distinct answer option for **other** 1 = **other** not present among answer options
	Harmonization control for target values “single” (3)	C_CIVIL_STATUS_ SINGLE	0 – distinct answer option for **single** 1 – **single** not present among answer options
	Harmonization control for target value “separated or divorced” (4)	C_CIVIL_STATUS_ SEP_DIV	0 – Both **separated** and **divorced** options are present 1 - **Merged** answer option: Separated and/or Divorced together with **widowed** 2 - **divorced** not present among answer options 3 - **separated** not present among answer options 4 - neither **separated** nor **divorced** present among answer options
	Harmonization control for target value “widowed” (5)	C_CIVIL_STATUS_ WIDOWED	0 – Distinct answer option for **widowed** 1 - **Merged** answer option: separated and/or divorced together with widowed 2 - **widowed** not present among answer options

[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_CIVIL_STATUS appear in 22 international survey projects: ABS, AMB, ARB, ASES, CB, CDCEE, CNEP, EB, EQLS, ESS, EVS, ISJP, ISSP, LB, LITS, NBB, NEB, PA1, PA2, PPE7N, VPCPCE, WVS, 166 waves and 3140 national surveys. The data cover 135 countries and years from 1966 to 2017.

2. General Rules and Procedures

3.1. Source data description

For harmonization into T_CIVIL_STATUS, we select source questions that contain the information on the marital status of the respondent. We rely on the English language questionnaires and codebooks describing the source survey data.

When a given dataset contains more questions than one related to marital status, we generally take the question that best matches our target variable categories. This procedure leads to, for example, selecting for harmonization a question with more answer options over a question with fewer answer options. Among all the source datafiles considered for harmonization of T_CIVIL_STATUS, 38% contained more than one question related to civil status. These ranged from project waves like ABS/1 that included two questions on the marital status that differed only in the number of available response options (`se004a`: 1|Yes, 2|No, 98|Don't Know, 99|No Answer; or se004: 1|Married, 2|Living-in as married, 3|Widowed, 4|Separated, 5|Divorced, 6|Single_Never married, 98|Don't know, 99|No Answer) to multiple waves of CB in which the question about marital status was asked about each household member.

In some cases, we had to combine two source questions to assign a target value on T_CIVIL_STATUS. For example, in ASES `V0230`, instead of a general question about marital status, respondents were asked multiple questions about the relationship with the person they live with if they do not live alone. In this project, we combine the question about living alone with the question specifying whether the respondent lives with their spouse or partner.

A different example of combining two source questions for each project is when we combine a question about legal marital status with a question about who the respondent is currently living with: our purpose is to prioritize the relationship with the person the respondent is currently living with over their legal marital status. Such questions appear in EB, ESS, ISSP, and LITS projects. Ultimately, we combine the question about legal marital status and the relationship with the person the respondent is currently living with for the following datasets: EB/40 `v552/v553`, EB/47.1 `v434/v444`, ESS/5-8 `rshpsts/marsts`, LITS/2 `q701/q702`, LITS/3 `q107_1/q108_1`.

While ESS/1-4 and ISSP provide more than one source item on marital status, due to too many unclear situations regarding respondent's ultimate civil status it was not possible to combine their information into a unified target value. Hence, we code only the marital status source variable and do not include source variables about cohabitation in harmonization of T_CIVIL_STATUS.

3.2. Rules of the transformation of source variables into the target variable

Married (1) refers to the source response options "married," in a "civil union" and "remarried".

If source variables permit differentiation, the target category "Married" does not include cohabitation. However, if the source variable joins "married" and "cohabitating" into a single answer category (e.g., married/living as married, or similar formulations), we code it as "Married", T_CIVIL_STATUS = 1, and control for the harmonization process. Specifically, since this combination of source categories is a deviation from the target categorization standard, we capture it by assigning the value 2 to the harmonization control C_CIVIL_STATUS_COHAB (described in Table 1.1).

Married but not living together with a spouse, is coded as target category **Separated or Divorced**, T_CIVIL_STATUS = 4 (e.g., response option "married, not living with partner" in source variable ISJP/1-2 `v68`).

The source option “married but living with a different partner than married to” is coded as **Cohabiting**, T_CIVIL_STATUS = 2 (e.g., response option “Living with my partner - not legally recognised / Legally married” from the combined source variable ESS/5 `rshpsts/marsts`).

Cohabitating (2) refers to living with a partner, legally or otherwise (living together, living as married, common law marriage, free union). If “legal partnership”, “civil partnership”, “legal cohabitation” or “pacte de solididarité” (ESS for France) appears among source answers, we store this information in the harmonization control, C_CIVIL_STATUS_COHAB = 1. Cohabitating does not include source options for civil unions (we code them as Married).

Single (3) refers to the source answer option “single” (or its equivalents, like “unmarried”, “never married” or “bachelor”). Having a partner but not living in the same household is also coded as “single”, T_CIVIL_STATUS = 3.

Separated or Divorced (4). Source answer options “separated” or “divorced” are both coded as T_CIVIL_STATUS = 4. **Separated or Divorced** also includes respondents who indicate that they are married but not living together.

If the source variable merges “separated” and/or “divorced” with “widowed” into one response option, we code it as “separated or divorced”, T_CIVIL_STATUS = 4, and capture this deviation from the standard target option definition via the harmonization control C_CIVIL_STATUS_SEP_DIV = 1.

Widowed (5) refers to the source answer option “widowed” either from a marriage or civil partnership. If the respondent is both widowed and currently living with a partner, we code T_CIVIL_STATUS = 2, **Cohabitating**. If the respondent is widowed and has a partner but does not live together, we code it as “**Widowed**”, T_CIVIL_STATUS = 5.

Other (6) refers to the source answer option “other” or “neither”. In some EB waves, the source response option “other” is marked as “spontaneous.” Such cases are also coded as “**Other**”, T_CIVIL_STATUS = 6, and the control C_CIVIL_STATUS_OTHER is set to 0 - meaning that option “other” was available.

Missing values and different situations that warrant to be treated as missing data are coded according to the SDR2 missing codes schema, provided in Table A.1 in the Appendix.

3.3. Methodological variables that accompany T_CIVIL_STATUS

The target variable T_CIVIL_STATUS is accompanied by five harmonization controls that capture specific features of the “civil status” source variables response options (see Table 1.1), namely: the source variable provides merged response options (e.g., married/living as married or separated/widowed); the response option was not available in the source variable or, specifically for C_CIVIL_STATUS_COHAB, when the option for legal cohabitation was available.

As a rule, if a given target option is available in the data dictionary (except for “married” as it was always available), the respective control variable takes the value 0.

1) **C_CIVIL_STATUS_COHAB**

   This harmonization control informs whether the source variable provides a distinct answer option for cohabitation, either non-formal, like living together (C_CIVIL_STATUS_COHAB = 0) or

formalized, like a legal partnership or civil union (C_CIVIL_STATUS_COHAB = 1). Second, it informs whether "cohabitation" and "married" appear as a single (merged) source response option explicitly referencing both statuses (C_CIVIL_STATUS_COHAB = 2). Third, this control records instances when the source variable provides no 'cohabitation' answer option: neither formal, non-formal nor merged (C_CIVIL_STATUS_COHAB = 3). If the answer option "married" is merged with "cohabitation" and there is also a separate question about cohabitation, we code C_CIVIL_STATUS_COHAB = 0 (distinct answer option). Thus, we prioritize distinct answer options over merged answer options.

It is important to note that we determine the presence of the cohabitation answer option and the type of cohabitation (formal or non-formal) based on data files only, and not by analyzing the codebook or the questionnaire. Therefore, we cannot rule out that specific countries did have the cohabitation option available even if it was not present in the general labels derived from data files.

2) C_CIVIL_STATUS_SINGLE

This harmonization control informs whether the source variable provides a distinct answer option for "single", C_CIVIL_STATUS_SINGLE = 0, or not, C_CIVIL_STATUS_SINGLE = 1.

3) C_CIVIL_STATUS_SEP_DIV

This harmonization control informs whether the source variable provides distinct answer options for "separated" or "divorced". We code C_CIVIL_STATUS_SEP_DIV = 0 if both options appear (they could be presented in the source variable as either separate options or merged into one, e.g., ISSP/2010-2015, ARB/4). If the source variable merges either "separated" or "divorced" with the answer option "widowed", we code C_CIVIL_STATUS_SEP_DIV = 1. The next two categories indicate if either the option "divorced" or the option "separated" are missing from the available source response options. If C_CIVIL_STATUS_SEP_DIV = 2, the option "Divorced" was not realized in the data, which additionally conveys the message that option "Separated" was present. Analogically, when C_CIVIL_STATUS_SEP_DIV = 3, the option "Separated" was not realized in the data. If the source variable provided neither option for "Divorced" nor for "Separated", we code C_CIVIL_STATUS_SEP_DIV = 4.

4) C_CIVIL_STATUS_WIDOWED

This harmonization control informs whether the source variable provides a distinct answer option for 'widowed", C_CIVIL_STATUS_WIDOWED = 0, and if not, C_CIVIL_STATUS_WIDOWED = 2. If the source variable merges "widowed" with the option "separated" and/or "divorced, we code C_CIVIL_STATUS_WIDOWED = 1.

5) C_CIVIL_STATUS_OTHER

This harmonization control informs whether the source variable provides a distinct answer option for 'other" or "neither", C_CIVIL_STATUS_OTHER = 0, and if not, we code C_CIVIL_STATUS_OTHER = 1.

Automated coding of Harmonization controls

We noticed that, often, there is within-wave variation in how detailed answer options on civil status are. Put differently, some countries provide their respondents with a less detailed answer option

schema than other countries within the same project (e.g., "living together" appears in a subset of national surveys of a given project wave). To preserve this variability while coding control variables, we introduce automated coding on the level of a country. This country-level coding is triggered when at least one country within a wave has at least one less response option, as reflected in the data dictionary, than defined for the project wave.

Rules for automatic coding of all control variables, except C_CIVIL_STATUS_SEP_DIV:

If **C_CIVIL_STATUS_COHAB = 0** (response option for cohabitation available in the data dictionary) or **1** (response option for legal cohabitation available in the data dictionary) and the target option 2 (**Cohabitation**) is not realized in the sample, then reset the control for this sample to **C_CIVIL_STATUS_COHAB = 3** (no answer option for 'cohabitation').

If **C_CIVIL_STATUS_OTHER = 0** (response option "other" available in the data dictionary) and the target option 6 (**Other**) is not realized in the sample, then reset the control for this sample to **C_CIVIL_STATUS_OTHER = 1.**[1]

If **C_CIVIL_STATUS_SINGLE = 0** (response option for "single" was available in the data dictionary) and the target option 3 (**Single**) is not realized in the sample, then reset the control for this sample to C**_CIVIL_STATUS_SINGLE = 1.**

If **C_CIVIL_STATUS_WIDOWED = 0** (response option for "widowed" available in the data dictionary) and the target option 5 (**Widowed**) is not realized in the sample, then reset the control for this sample to **C_CIVIL_STATUS_WIDOWED = 2**

If **C_CIVIL_STATUS_WIDOWED = 1** (response option for "widowed" in the data dictionary merged with the option for "separated" or "divorced") and the target option 4 (widowed merged with separated/divorced) is not realized in the sample, then reset the control to **C_CIVIL_STATUS_WIDOWED = 2.**[2]

The control variable **C_CIVIL_STATUS_SEP_DIV** could not be automated because it is built on two response options, "separated" and "divorced". For this reason, we hand-coded directly in the CWT file (for CWT, see Section 1). The coding rules areas follows:

If **C_CIVIL_STATUS_SEP_DIV = 0** (response options for both "separated" and "divorced" available in the data dictionary) and the response option for "divorced" not realized in the sample, then reset the control for that sample to **C_CIVIL_STATUS_SEP_DIV = 2**. If the response option for "separated" was not realized in the sample, then reset the control to **C_CIVIL_STATUS_SEP_DIV = 3**. If neither response option for "divorced", nor for "separated" was realized in the sample, then reset the control for that sample to **C_CIVIL_STATUS_SEP_DIV = 4.**

In CWT we code the controls on the level of variables; these codes correspond to what we see in the data dictionary. Therefore, we cannot rule out that some answer options were available in a given country (like "cohabitation" or "widowed"), but were simply not realized in the data.

The control variable **C_CIVIL_STATUS_COHAB** for AMB/2014, ARB/4 and ISSP/2007 was hand-coded on the country-level in the CWT file. In AMB/2014 and ISSP/2007, we have country-specific response options which specify the legal status of cohabitation for some countries. In

[1] SeeSection 4. Special Cases for an exception from this rule (EB/47.1).
[2] There were no such cases.

ARB/4, the response option for cohabitation was accompanied by a note to not read it for some countries, therefore we hand-coded the C_CIVIL_STATUS_COHAB variable based on whether the option was realized for each country.

3. **Special Cases**

1. In CNEP/3-4, whenever multiple questions/variables were available per country, we chose the one with the fewest missing values, more answer options to choose from and/or better-defined errors.

2. In ARB/4 `q1010` the option "living with a partner" was accompanied by the label "do not read". It was still recorded in data for some countries (Algeria, Jordania).

3. In ARB, we decided to mark the option "engaged" as with the SDR2 missing code "unfit" (see Appendix).

4. In NBB/1-6 the merged option for separated/divorced/widowed appeared only for Estonia, hence we did not retain this deviation on control variables C_CIVIL_STATUS_SEP_DIV and C_CIVIL_STATUS_WIDOWED. The same applies for WVS/1-6, where separated/divorced/widowed was merged only for Israel.

5. When source variables provide a different word (never married, not married, bachelor, unmarried, not living with partner/unmarried never couple) for "single", we coded these answers as "single", T_CIVIL_STATUS = 3:

   ARB/2 q1010, ARB/3 q1010, CB/2008 HHMMARR, CB/2009 RESPMAR, CB/2010 RESPMAR, CB/2011 RESPMAR, CB/2012 RESPMAR, CB/2013 RESPMAR, CB/2015 RESPMAR, CNEP/4 United States PPMARIT, EB/40 v552/v553, EB/52.1 v584, EB/54.1 v347, EB/55.1 v355, EB/56.1 v428, EB/57.1 v372, EB/58.2 v352, EB/60.1 v594, EB/62.2 v467, EB/63.4 v407, EB/64.4 v1921, EB/66.1 v459, EB/68.1 v417, EQLS/1-3 Y11_Q31, ESS/1 marital/martlfr, ESS/2 marital/martlfr, ESS/3 maritala, ESS/4 maritala, ESS/5 rshpsts/marsts, ESS/6 rshpsts/marsts, ESS/7 rshpsts/marsts, ESS/8 rshpsts/marsts, ISSP/1985 v120, ISSP/1986 v83, ISSP/1987 v85, ISSP/1988 v67, ISSP/1989 v87, ISSP/1990 v61, ISSP/1991 v79, ISSP/1992 v101, ISSP/1995 v202, ISSP/1996 v202, ISSP/1998 v202, ISSP/2001 marital, ISSP/2007 MARITAL, ISSP/2008 MARITAL, LITS/2 q701/q702, NEB/1 s9, NEB/6 D12CZE, NEB/6 D12POL, PA1 v209, PA2 v1197, PPE7N Netherlands V418, PPE7N United States V18

6. **Target constructed from 2 source variables** (see section 3.1. for a detailed explanation)**:**

   ASES V0230_1/V0230_2, EB/40 v552/v553, EB/47.1 v434/v444, ESS/5 rshpsts/marsts, ESS/6 rshpsts/marsts, ESS/7 rshpsts/marsts, ESS/8 rshpsts/marsts, LITS/2 q701/q702, LITS/3 q107_1/q108_1

7. **ABS/4 se4 - '**Widowed' merged with 'separated', distinct option for 'divorced', 'widowed/separated' coded as T_CIVIL_STATUS = 4, C_CIVIL_STATUS_SEP_DIV = 2, C_CIVIL_STATUS_WIDOWED = 2

8. No data realized: NEB/6 D12LAT, D12LIT, and D12EST

9. EB/47.1 - for this project we combined two variables to acquire the target variable. Even though in one of these variables the response option “other” was present, it was no longer interpretable as “other” after combining the two source variables. Therefore it was coded in C_CIVIL_STATUS_OTHER as the response option “other” was not realized.

10. During automatic coding of the control variables, some source variables were problematic because the target values were not realized on the whole dataset level. These source variables required special flags: ABS/4 Malaysia se4, CNEP/3 Spain L.Married.F, CNEP/4 (Spain, Indonesia, Mexico) L.Married, EVS/1-4 X007, VPCPCE (Czechia, Hungary) D14, WVS/1-6 X007

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	-20 (.t)	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.