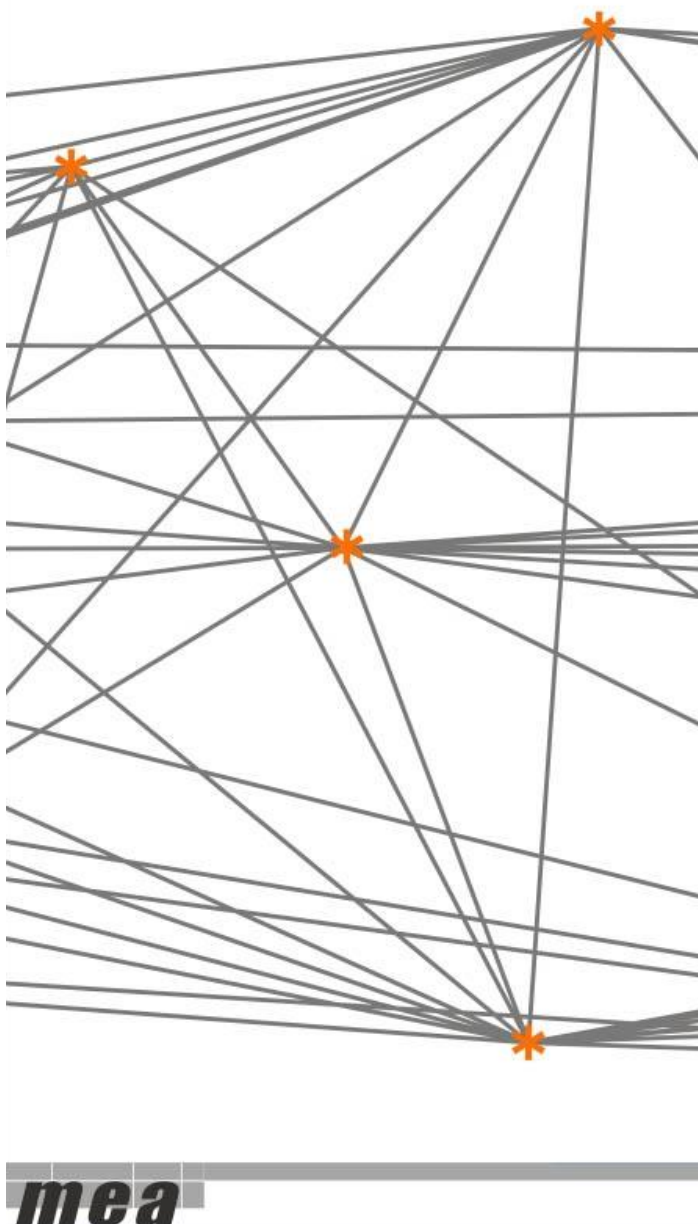




## Release Guide 6.0.0



March 31, 2017

## Table of Contents

1 Sources of Information.....	5
2 SHARE Data Releases.....	6
3 New in SHARE Wave 6.....	6
4 Additional Files .....	7
4.1 <i>easySHARE</i> .....	7
4.2 Interviewer Survey (Wave 5 & Wave 6) .....	7
5 Countries, Languages and Fieldwork times in SHARE .....	8
6 Eligibility Rules.....	9
7 Data Collection .....	9
8 Types of Respondents .....	10
9 Merging the Data.....	13
10 Missing Codes.....	14
11 Naming Conventions .....	15
11.1 General Conventions .....	15
11.2 Double Loops.....	15
11.3 Dummy Variables .....	16
11.4 Euro Conversion .....	17
11.5 Unfolding Brackets .....	17
11.6 Renames since release 5.0.0 .....	18
12 Types of Questionnaires and Modules.....	18
12.1 Regular Questionnaire Modules: Short Description .....	20
12.2 Special Questionnaire Modules .....	22
12.2.1 Coverscreen.....	22
12.2.2 Paper and Pencil Drop-Off .....	23
12.2.3 Vignettes .....	23
12.2.4 End-of-Life Interviews .....	24
12.2.5 Technical Variables.....	24
13 Special Aspects of Questionnaire Modules.....	25
13.1 DN Module: Citizenship/Country of Birth Coding .....	25
13.2 CH Module: Information on Children in SHARE .....	25
13.3 PH Module: <i>phrandom</i> and <i>ph006</i> .....	26
13.4 EX Module: Definition of Life Expectancy Target Age.....	27

13.5 CF Module: Ten Words List Learning .....	27
13.6 SN Module (Waves 4 and 6) .....	28
13.7 SP and FT Module: List of Relations differences due to SN .....	29
13.8 EP Module .....	31
13.8.1 EP005 & Re-Interview in Israel (Wave 1) .....	31
13.8.2 Values in EP078_3, EP078_7, and EP078_12 in Greece (Wave 6) .....	31
13.9 Social Exclusion Items (Wave 5) .....	32
14 Generated Variables .....	32
14.1 gv_allwaves_cv_r .....	34
14.2 Weights .....	34
14.2.1 Sampling design weights .....	34
14.2.2 Calibrated weights .....	36
14.2.3 Structure of the SHARE weights in release 6.0.0 .....	39
14.3 Imputations .....	42
14.3.1 Practical decisions about imputations .....	42
14.3.2 The imputation procedure .....	44
14.3.3 Additional comments on the usage of imputations .....	46
14.4 gv_iscd .....	48
14.5 gv_health .....	48
14.6 gv_housing .....	50
14.7 gv_networks .....	50
14.8 gv_exrates .....	53
14.9 gv_job_episodes_panel .....	53
14.10 gv_grossnet .....	53
14.11 gv_isco .....	53
14.12 gv_ssw .....	54
14.13 gv_deprivation .....	55
14.14 gv_children .....	56
14.15 gv_linkage .....	58
14.16 gv_dbs .....	59
15 APPENDIX .....	60

## Index of Tables

Table 1: Links to Further Documentation Material.....	5
Table 2: Countries, Language Versions & Fieldwork Times in SHARE Waves 1-6.....	8
Table 3: Types of Respondents and Variable Names .....	10
Table 4: Who Answers What in the Questionnaire? .....	12
Table 5: Adjustment of ac004_ Variable Names in Wave 2 .....	16
Table 6: Questionnaire Modules of Waves 1, 2, 4, 5 and 6 .....	19
Table 7: Variables in the Technical Variables Module .....	24
Table 8: Changes in Information on Children.....	26
Table 9: Relationship Type Coding .....	30
Table 10: Items on Social Exclusion.....	32
Table 11: Generated Variable Modules of Waves 1 to 6 .....	33
Table 12: Sampling design and calibrated cross-sectional weights .....	39
Table 13: Sampling design and calibrated longitudinal weights.....	40
Table 14: Generated Variables of gv_health.....	48
Table 15: Generated Housing Variables.....	50
Table 16: Generated Variables in Wave 4 gv_networks .....	51
Table 17: Generated Variables in Wave 6 gv_networks .....	52
Table 18: Variables Used for ISCO and NACE Coding.....	54
Table 19: Items Used for Material Deprivation Index (depmat).....	55
Table 20: Items Used for the Social Deprivation Index (depsoc).....	56
Table 21: Generated Variables in Wave 6 gv_children .....	57

## Index of Appendix Tables

Table 1: Digital Object Identifier of the SHARE Datasets .....	60
Table 2: SHARE Data Releases.....	61
Table 3: Overview of Additional Codes for Country of Birth and Citizenship.....	64
Table 4: Drop-Off Correspondence in Waves 1 & 2 .....	65
Table 6: Country-specific composition of imputed pension variables in wave 1 .....	70
Table 7: Composition of aggregated imputation variables in wave 2 .....	72
Table 8: Composition of aggregated imputation variables in wave 4 .....	74
Table 9: Composition of aggregated imputation variables in wave 5 .....	76
Table 10: Composition of aggregated imputation variables in wave 6 .....	78
Table 11: List of variables in gv_imputations.....	80
Table 12: Description of flag variables associated with imputed variables.....	84

# 1 Sources of Information

This release guide contains general information on the SHARE database like naming of variables, missing code scheme, merging modules and/or waves as well as wave-specific information like important questionnaire innovations, methodological advancements and new procedures introduced between waves. Because of its divergence from other waves, the life history data of wave 3 (SHARELIFE) is documented separately: [http://www.share-project.org/fileadmin/pdf\\_documentation/SHARELIFE\\_release\\_guide\\_6.0.0.pdf](http://www.share-project.org/fileadmin/pdf_documentation/SHARELIFE_release_guide_6.0.0.pdf)

Additional information to enable and facilitate the use of SHARE data is available on the SHARE website: [www.share-eric.eu](http://www.share-eric.eu). Since 2013, the SHARE datasets are registered with a Digital Object Identifier (DOI) in order to enable a permanent identification and citation of the SHARE data (see table 0 of the appendix). The DOIs can be found on website (<http://www.share-project.org/index.php?id=891>) as well as specific information on data access (<http://www.share-project.org/index.php?id=832>) and further documentation material (<http://www.share-project.org/index.php?id=833>). Furthermore the data resource profile published in the *International Journal of Epidemiology* as well as the methodology volumes give a compact overview on SHARE.

The following overview contains links to further important documentation material like the wave-specific questionnaires, country specifics (so far only available for waves 1 and 2), deviations between waves, methodology volumes and the Data Resource Profile.

**Table 1: Links to Further Documentation Material**

	Wave 1	Wave 2	Wave 3 (SHARELIFE)	Wave 4	Wave 5	Wave 6
Questionnaires	<a href="#"><u>X</u></a>	<a href="#"><u>X</u></a>	<a href="#"><u>X</u></a>	<a href="#"><u>X</u></a>	<a href="#"><u>X</u></a>	<a href="#"><u>X</u></a>
Deviations between countries ...	<a href="#"><u>X</u></a>	<a href="#"><u>X</u></a>				
... and waves	<a href="#"><u>Deviation between waves 1, 2...</u></a>			<a href="#"><u>... 4 ...</u></a>	<a href="#"><u>...5...</u></a>	<a href="#"><u>...and 6</u></a>
Methodology Volumes	<a href="#"><u>X</u></a>	<a href="#"><u>Chapter 8 of W2 FRB</u></a>	<a href="#"><u>X</u></a>	<a href="#"><u>X</u></a>	<a href="#"><u>X</u></a>	
Scales Manual	<a href="#"><u>Scales and Multi-Item Indicators</u></a>					
Data Resource Profile	<a href="#"><u>Börsch-Supan A. et al. (2013): Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE), Int J of Epidemiology</u></a>					

In case you have questions that are neither answered by the content of these documents nor by the [FAQ](#) section of the website, you can contact the SHARE Central-Team: [info\[at\]share-project.org](mailto:info[at]share-project.org)

## 2 SHARE Data Releases

After being registered as a SHARE user you can easily download the data free of charge from the SHARE Research Data Center. Please find the link as well as more specific information on the data access procedure here: <http://www.share-project.org/index.php?id=832>

Releases are indicated as follows:

- Minor changes will be indicated by the third digit, e.g. release 2.2.1. Please check the website for updates regularly.
- Major changes will be announced to users via e-mail and are indicated by the second digit, e.g. release 2.5.0.
- Combined releases of all waves are indicated by the first digit and have the number of the latest wave as first digit, e.g. release 6.0.0 with wave 6 being the latest wave.

The new release 6.0.0 of waves 1 to 6 comes with some major improvements. It comprises the latest state of data cleaning, harmonisation across waves as well as a considerable range of updates and innovations (see chapter 2). Moreover, the sample has been improved by conducting thorough consistency checks based on information from all six waves. An overview on previous SHARE release updates is provided in the appendix.

## 3 New in SHARE Wave 6

- A new country joined SHARE: Croatia
- Longitudinal information on social networks: SN module (see chapter 12.1) and gv\_networks (see chapter 14.7)
- New generated variables module for children: gv\_children (see chapter 14.14)
- Collection of Dried Blood Spots: BS module (see chapter 12.1) and gv\_dbs (see chapter 14.16)
- Updated imputations (see chapter 14.3)
- Updated weights (see chapter 14.2)
- **New since Release 5.0.0**
- New modules: gv\_allwaves\_cv\_r (see chapter 14.1), gv\_exrates (see chapter 14.8) and gv\_ssw (see chapter 14.12)
- Improved coverscreen module (see chapter 12.2.1)
- Cause of death (xt012) now coded (xt012c) (see chapter 12.2.4)
- Information from the modules gv\_dol and cv\_h now stored in the cv\_r or gv\_allwaves\_cv\_r
- mergeid for all household members
- All monetary amounts in Euro
- Interviewer IDs for waves 1, 2 and 4
- Ireland wave 3 data included
- Labels in wave 4 drop-off questionnaire now translated into English

## 4 Additional Files

### 4.1 *easySHARE*

*easySHARE* is a simplified HRS-adapted dataset for student training, and for researchers who have little experience in quantitative analyses of complex survey data. *easySHARE* stores information on all respondents and of all currently released data collection waves in one single dataset. For the subset of variables covered, the complexity was considerably reduced. *easySHARE* is stored as long format panel dataset. In addition to the data and the release guide, the download zip files include the Stata program that was used to extract *easySHARE* from the regular SHARE data. For further information see <http://www.share-project.org/special-data-sets/easyshare.html>.

### 4.2 Interviewer Survey (Wave 5 & Wave 6)

The SHARE wave 5 and wave 6 Interviewer Survey was implemented as a web survey. Besides basic demographics, the questionnaire contains questions about interviewers' attitudes towards surveys in general, their expectations and experiences towards some specific SHARE modules as well as some hypothetical questions of how they would behave as a respondent.

Five of the wave 5 SHARE countries participated in the SHARE Interviewer Survey: Austria, Belgium, Germany, Spain and Sweden. In SHARE's sixth wave 12 countries implemented the SHARE Interviewer Survey (Austria, Belgium, Estonia, Germany, Greece, Italy, Luxembourg, Poland, Portugal, Spain, Slovenia and Sweden). The participation of interviewers was voluntary and confidential, i.e. responses were not shared with the survey agencies. To link the interviewer survey data with the SHARE survey data, a wave-specific interviewer ID variable (*intid*) is provided in the IV module.

For further information about the interviewer survey and the questionnaire see: <http://www.share-project.org/special-data-sets/interviewer-survey.html>

## 5 Countries, Languages and Fieldwork times in SHARE

**Table 2: Countries, Language Versions & Fieldwork Times in SHARE Waves 1-6**

Country ID	Language ID	Country & language	Wave 1	Wave 2	Wave 3 SHARELIFE	Wave 4	Wave 5	Wave 6
11	11	Austria	2004	2006/07	2008/09	2011	2013	2015
12	12	Germany	2004	2006/07	2008/09	2011/12	2013	2015
13	13	Sweden	2004	2006/07	2008/09	2011	2013	2015
14	14	Netherlands	2004	2007	2008/09	2011	2013	-
15	15	Spain (Castilian)	2004	2006/07	2008/09	2011	2013	2015
15	38	Spain/Girona (Castilian or Catalan)					2013	2015
15	39	Spain/Girona (Catalan)	-	-	-	-	2013	2015
15	40	Spain/Girona (Castilian)	-	-	-	-	2013	2015
16	16	Italy	2004	2006/07	2008/09	2011	2013	2015
17	17	France	2004/05	2006/07	2009	2011	2013	2015
18	18	Denmark	2004	2006/07	2008/09	2011	2013	2015
19	19	Greece	2004/05	2007	2008/09	-	-	2015
20	20	Switzerland (German)	2004	2006/07	2008/09	2011	2013	2015
20	21	Switzerland (French)	2004	2006/07	2008/09	2011	2013	2015
20	22	Switzerland (Italian)	2004	2006/07	2008/09	2011	2013	2015
23	23	Belgium (French)	2004/05	2006/07	2008/09	2011	2013	2015
23	24	Belgium (Flemish)	2004/05	2006/07	2008/09	2011	2013	2015
25	25	Israel (Hebrew)	2005/06	2009/10	-	-	2013	2015
25	26	Israel (Arabic)	2005/06	2009/10	-	-	2013	2015
25	27	Israel (Russian)	2005/06	2009/10	-	-	2013	2015
28	28	Czech Republic	-	2006/07	2008/09	2011	2013	2015
29	29	Poland	-	2006/07	2008/09	2011/12	-	2015
30	30	Ireland	-	2007	2009/10/11	-	-	-
31	41	Luxembourg (French)	-	-	-	-	2013	2015
31	42	Luxembourg (German)	-	-	-	-	2013	2015
32	32	Hungary	-	-	-	2011	-	-
33	33	Portugal	-	-	-	2011	-	2015
34	34	Slovenia	-	-	-	2011	2013	2015



35	35	Estonia (Estonian or Russian)	-	-	-	2010/11	2013 (XT only)	2015
35	36	Estonia (Estonian)	-	-	-	-	2013	2015
35	37	Estonia (Russian)	-	-	-	-	2013	2015
47	47	Croatia						2015

## 6 Eligibility Rules

The SHARE target population consists of all persons aged 50 years and over at the time of sampling who have their regular domicile in the respective SHARE country. Persons are excluded if they are incarcerated, hospitalised or out of the country during the entire survey period, unable to speak the country's language(s) or have moved to an unknown address. In wave 1 all household members born 1954 or earlier are eligible for an interview. Starting in wave 2, for new countries or refreshment samples, there is only one selected respondent per household who has to be born 1956 or earlier in wave 2, 1960 or earlier in wave 4, 1962 or earlier in wave 5 and 1964 or earlier in wave 6. In addition – in all waves – current partners living in the same household are interviewed regardless of their age.

All SHARE respondents who were interviewed in any previous wave are part of the longitudinal sample. If they have a new partner living in the household, the new partner is eligible for an interview as well, regardless of age. Age eligible respondents who participated are traced and re-interviewed if they move within the country and end-of-life interviews are conducted if they decease. Younger partners, new partners and partners who never participated in SHARE will not be traced if they move and are not eligible for an end-of-life interview.

## 7 Data Collection

SHARE data collection is based on computer-assisted personal interviewing (CAPI). The interviewers conduct face-to-face interviews using a laptop computer on which the CAPI instrument is installed. Personal interviews are necessary for SHARE because they make the execution of physical tests possible.

SHARE applies a concept of ex-ante harmonisation: there is one common generic questionnaire that is translated into the national languages (in some countries more than one language) using an internet based translation tool and processed automatically in a common CAPI instrument. However, some internationally highly diverse variables require country-specific measurements and ex-post harmonisation, for example in the areas of education (ISCED) or occupation (ISCO, NACE).

## 8 Types of Respondents

Selected household members serve as family, financial or household respondents. They answer specific questions on behalf of the couple or the whole household. Therefore, the answers to financial, housing and family questions in the modules FT, AS, HO, HH, CO, CH and parts of SP are only available for the financial, family or household respondents, respectively. However, for generated variables (see next chapter) the information is stored for all respondents, regardless of their status as regular or financial/household/family respondent.

- A **family respondent** answers the questions of the CH module and parts of the SP module on behalf of the couple. They are indicated by the dummy variable *fam\_resp* that can be found in both the *cv\_r* and the technical variables module. Family respondents are selected by the chronological order of interviews per couple (married or not): The couple's first person interviewed is the family respondent. Note that the naming of the *mergeid* variable does not indicate the chronology of interviews within a household. A respondent living without a partner in the household is automatically defined as family respondent.
- The **financial respondent** answers the modules FT and AS on behalf of the couple and is indicated by the dummy variable *fin\_resp* in the *cv\_r* and the technical variables module. In case of a one-person household or a respondent living as single, the respondent is always the financial respondent. In multi-person households, the number of financial respondents may vary: respondents living without a partner in multi-person households are always financial respondents. In wave 1 eligible couples could decide to answer questions about their finances separately. Otherwise, one partner can answer on behalf of the couple.
- Questions about household features (HO, HH, CO) are asked only to one **household respondent** on behalf of the whole household. The household respondent indicated by the dummy *hou\_resp* and stored both in the *cv\_r* and the technical variables module.

**Table 3: Types of Respondents and Variable Names**

Type of respondent	Variable name in <i>cv_r</i> and technical variables module	Name of Filter in w1 & w2 questionnaire	Name of Filter in w4, w5 and w6 questionnaire
Family respondent	<i>fam_resp</i>	MN006_ (FAMILY RESPONDENT)	MN006_NumFamR
Financial respondent	<i>fin_resp</i>	MN007_ (FINANCIAL RESPONDENT)	MN007_NumFinR
Household respondent	<i>hou_resp</i>	MN008_ (HOUSEHOLD RESPONDENT)	MN008_NumHHR

Only if physical and/or cognitive limitations make it too difficult for a respondent to complete the interview her-/himself, it is possible that the respondent is assisted by a so-called **proxy respondent** to complete the interview (“partly proxy” interview). If the proxy respondent answers the entire questionnaire in lieu of the respondent, the interview is referred to as a “fully proxy” interview. Examples of conditions under which proxy interviewing is allowed are hearing loss, speaking problems, Alzheimer’s disease and difficulties in concentrating for the whole interview time period. Proxy respondents are also asked for end-of-life interviews in case of a respondent’s decease (see chapter 12.2.4). Some questionnaire modules are defined as non-proxy sections because those cannot be answered by other persons.

Usually at the end of each module interviewers are asked if the previous section was answered by the respondent only, by respondent and proxy or the proxy only (e.g. variable *ch023\_* in the CH module). Please have a look at the questionnaires for the respective variables in other modules. Information on the relationship between proxy and respondent is stored in variable *iv020\_* in the IV module.

**Table 4: Who Answers What in the Questionnaire?**

Module	Name	All respondents <sup>1</sup>	Household Resp.	Financial Resp.	Family Resp.	non-proxy
CV_R	Coverscreen	cv-respondent only				
DN	Demographics	x				
SN	Social Networks	x				x
CH	Children				x	
PH	Physical Health	x				
BR	Behavioural Risks	x				
CF	Cognitive Function	x				x
MH	Mental Health	x				x (partly)
HC	Health Care	x				
EP	Employment and Pensions	x				
IT	IT module	x				
MC	Mini Childhood	x				
GS	Grip Strength	x				x
WS	Walking Speed	x				x
BS	Blood Sample	x				x
PF	Peak Flow	x				x
CS	Chair Stand	x				x
SP	Social Support	x (partly)			x (partly)	
FT	Financial Transfers			x		
HO	Housing		x			
HH	Household Income		x			
CO	Consumption		x			
AS	Assets			x		
AC	Activities	x				x
EX	Expectations	x				x
IV	Interviewer Observations	Interviewer only				
XT	End-of-Life Interview	proxy only				
dropoff	Drop off	x				x
vignettes	Vignettes	x				x

<sup>1</sup> "All respondents" only refers to types of respondents here. Due to other filters/routing not all modules are answered by all respondents.

## 9 Merging the Data

To merge different modules and/or waves on individual level *mergeid* is the key variable. For each individual the variable *mergeid* is a unique and non-changing person identifier for all waves. It has the format “CC-hhhhhh-rr” (e.g. “AT-070759-01”), where CC refers to the short country code (here: “AT” for Austria), “hhhhhh” are digits to identify the household, and “rr” is the respondent identifier within each household. The variable *mergeid* is included in all modules that contain individuals’ answers and thus can be used to combine these modules on the individual level.

If the data are to be merged on household level one of the *hhid`w`* (where `w` stands for the respective wave) variables should be used as key identifier. Thus, *hhid2* refers to the household in which the individual resided in wave 2. *hhid`w`* has the following format “CC-hhhhhh-S” (e.g. “AT-070759-A”), where “CC” refers to the short country code, “hhhhhh” is the household identifier, and “S” identifies possible split households, i.e. the household of a panel member who moved out of a previous household. In case of a household split there is not only an “A”-suffix but also “B”, “C”, etc. Note that *mergeid* does NOT change when a respondent moves from the original household.

As in previous releases, researchers interested in identifying all household members in a current wave – note that this includes also ineligible household members – can use the variable *hhid`w`*. New since release 5.0.0 is the *gv\_allwaves\_cv\_r* module which provides an overview of all respondents and non-participating household members across all currently released waves (see chapter 14.1 for further information).

For merging data collected in the interviewer survey the variable *intid* has to be used. To merge *gv\_exrates* to other wave-specific modules, the *country* identifier is to be used.

## 10 Missing Codes

Apart from so called system missing values due to filters/routing or an abandoned interview, SHARE uses various labelled missing codes.

### **A: general missing codes**

- 1: "Don't know"
- 2: "Refusal"
- 3: "Implausible value/suspected wrong"
- 4: "Not codable"
- 5: "Not answered"
- 7: "Not yet coded"
- 9: "Not applicable"

### **B: not applicable – specified**

- 91: "Not applicable (not yet part of sample)"
- 92: "Not applicable (no participation in this wave)"
- 93: "Not applicable (respondent was not listed as household member in this wave)"
- 94: "Not applicable (respondent has an End-of-Life interview in previous waves)"
- 95: "Not applicable (no main interview done)"
- 98: "Not applicable (other reason)"
- 99: "Not applicable (missing by design)"

### **C: missing codes - financial variables**

- 9999991: "Don't know"
- 9999992: "Refusal"

### **D: missing codes – gv\_networks variables in Wave 6**

- 20: "Missing information"
- 21: "W4 only"
- 22: "W6 only"
- 23: "W4 SN size of 0"
- 24: "W6 SN size of 0"
- 25: "No SN in w6 & w4"

Both Stata and SPSS users should define missing values as all values smaller than 0 for all variables except financial amounts. Missing values for financial amounts should be defined smaller than -9999990.

### **Treating SHARE missing codes with sharetom (only available for Stata)**

Stata users can download an ado file (sharetom.ado & sharetom.hlp) from the data download website: <https://share-project.centerdata.nl/sharedatadissemination/users/login>.

The program recodes missing values to Stata's "extended missing values" and labels them appropriately. Stata's extended missing values, e.g. ".a" or ".b", are treated the same way as system missing values ".". This means, in a tabulate var-command they are not tabulated, unless you add the missing option (tabulate var, missing). sharetom should run before you do any other changes to the data.

The recent version *sharetom5* builds on its predecessors and is compatible with all release versions previously covered by *sharetom4*. In addition, the new missing codes were integrated into the program which also covers the new missing codes from *gv\_allwaves\_cv\_r* as well as those of *easySHARE*. Please note that the new missing codes for *gv\_networks* are not yet integrated in *sharetom5*.

## 11 Naming Conventions

### 11.1 General Conventions

The naming of variables is harmonised across waves. Variable names in the CAPI instrument data use the following format: mmXXXyyy\_LL. "mm" is the module identifier, e.g. DN for the demographics module, "XXX" refers to the question number, e.g. 001, and "yyy" are optional digits for dummy variables (indicated by "d"), euro conversion (indicated by "e") or unfolding brackets (indicated by "ub"). The separation character "\_" is followed by "LL" optional digits for category or loop indication ("outer loop").

Examples:

*ho045\_* stores "The main reason to move", hence does not allow multiple responses and is not asked within a loop.

*ft003\_1*, *ft003\_2* and *ft003\_3* store the relationship to whom the respondent provided financial gifts for up to three people ("outer loop" over three persons)

### 11.2 Double Loops

Some variables like *ep111* – *ep116* are embedded in a double loop: They are looped both over  $c:=\{1,2,...,6\}$  and  $p:=\{1,2,...,20\}$ , where *c* refers to the categories of dummy variable *ep110* (received public benefits) and *p* refers to up to 20 periods of incidence the respondent may report. In order to provide an intuitive naming of the double-looped variables, we allow for multiple loop indicators. In the case of *ep111* – *ep116*, mmXXX\_LL is changed to mmXXX\_c\_p whereby "c" refers to the type of public benefit (1 = old age pension, 2 = early retirement pension, 3 = unemployment benefits, 4 = sickness benefits, 5 = disability insurance and = social assistance) and "p" to the corresponding period. Example: *ep111\_3\_5* refers to the receipt of unemployment benefits in period 5.

### 11.3 Dummy Variables

Answers to all questions that allow for multiple responses appear in the data as several dummy variables. E.g. in wave 1 and 2 question *br005* ("What do or did you smoke") has three answer categories: 1 = Cigarettes, 2 = Pipe and 3 = Cigars or cigarillos. The dataset thus contains three dummy variables: *br005d1*, *br005d2*, and *br005d3* corresponding to the three answer categories. Note that in general we omit leading zeros for the first nine answers in a response set with more than ten answer categories to allow for easier looping in commands.

Value "1" in any of these variables means that the respondent selected the particular option as an answer and in case of value "0" the respondent did not select the option.

- In case the respondent answers with a "none of these" or "other", the naming of the dummy variables has the following structure:  
mmXXXdno "None of these"  
mmXXXdot "Other"
- In case the respondent answers with "Don't know" or "Refusal", all corresponding dummy variables of that question are set to the respective missing code, i.e. -1 for "Don't know" and -2 for "Refusal".
- In case the question requires loop indication, the digit(s) right after the "d" correspond(s) to the multiple response categories. The loop indication is added as the last part of the variable name separated by a "\_" as usual. Hence, all multiple response variables can be identified by the "d"-separator, all loops are identified by the presence of digit(s) after a separation indicator "\_".

In general, the numbering of answer categories in the generic questionnaire determines the optional digits in the dummy variables. We only deviate from this rule, if otherwise misleading variable names across waves emerge. *ac004\_* is one example where this is the case: Wave 2 includes fewer answer categories than wave 1, however, the reduced set is comparable. Variable names in wave 2 were adjusted to match the respective wave 1 answer categories in the following way:

**Table 5: Adjustment of *ac004\_* Variable Names in Wave 2**

wave 2: <i>ac004_</i> questionnaire categories	variable names
1. To meet other people	<i>ac004d1_#</i>
2. To contribute something useful	<i>ac004d2_#</i>
3. Because I am needed	<i>ac004d4_#</i>
4. To earn money	<i>ac004d5_#</i>
5. To use my skills or to keep fit	<i>ac004d7_#</i>
96. None of these	<i>ac004dno_#</i>



## 11.4 Euro Conversion

All answers referring to amounts of money are converted into Euro values. The variable *exrate* containing the corresponding exchange rate used for Euro conversion is stored in all modules containing financial questions as well as in the *gv\_exrates* module. If the amount as denominated by the local non-Euro currency is of interest, a simple multiplication with the variable *exrate* is needed to undo the Euro-conversion. The fixed exchange rate for Euro countries refers to the official exchange rate used when Euro was implemented in a specific country and is used mainly in wave 1 when pre-Euro currency answers were possible in financial questions.

The format of the Euro converted variables is *mmXXXe\_LL* with “mm” representing the module identifier, “XXX” the question number, “e” being the indication of Euro conversion and “LL” referring to optional digits for loop indication. When the respondent answers with “Don’t know” (DK) or “Refusal” (RF) to a question indicating a financial amount, the following values are included in the dataset: -9999991 for DK and -9999992 for RF.

Please see chapter 14.8 for information about ppp-adjustment.

## 11.5 Unfolding Brackets

Financial variables are very important, but often subject to non-response. SHARE therefore uses unfolding brackets in order to reduce non-response. It is found that non-respondents tend to be more willing to answer, if they have to classify e.g. their earnings in rough categories rather than a specific amount. Thus, if a respondent does not provide information for a financial variable she/he is forwarded to a set of bracket values (usually a salient amount such as 2,000 €). In each country there are three different entry points for an unfolding bracket sequence. The interviewer then asks: Do you earn a) more than this amount, b) less than this amount or c) approximately this amount. A respondent can enter up to 3 unfolding bracket sequences. The scientific release includes the country-specific bracket values (amounts in Euros) and the final category in which the respondent ended. When a DK or RF is given during the unfolding bracket sequence, the value for the final category is set to the respective missing code.

The country-specific bracket values are indicated as *mmXXXv1*, *mmXXXv2*, and *mmXXXv3*. In case of a loop, there mostly exists only one set of bracket values as the country-specific thresholds are constant over loop numbers. Thus, *mmXXXv1* contains the lowest threshold for unfolding bracket variable *mmXXXub\_1*, as well as for *mmXXXub\_2*, etc.

The format of the summarising unfolding bracket variable is as follows: *mmXXXub\_LL* with “mm” representing the module identifier, “XXX” question number, “ub” indicating an unfolding bracket sequence and “LL” being optional digits for loop indication.

In releases prior to release 5.0.0, the naming of looped financial variables was not consistent across waves. This is why some of the variables have been renamed. For instance, *ep078e1*

and *ep078ub1* have been renamed *ep078e\_1* and *ep078ub\_1* in waves where this was not already the case. These changes facilitate the use of loops across waves.

Moreover, the bracket values have been converted to Euro in waves 4 and 5 to be in line with waves 1 and 2. If the researcher is interested in the bracket value as denominated by the local non-Euro currency, a simple multiplication with the variable *exrate* suffices that is included in all modules with financial variables as well as in the *gv\_exrates* module.

## 11.6 Renames since release 5.0.0

The names of the drop-off and vignettes datasets have been changed into “do” and “vi”. Additionally, some variables have been renamed in the course of cross-wave harmonisation when this was in line with the questionnaire. This mainly holds for financial variables and looped variables which were previously named differently across waves. The harmonisation is intended to facilitate looping across waves.

## 12 Types of Questionnaires and Modules

The questionnaires can be downloaded from the SHARE website: <http://www.share-project.org/data-documentation/questionnaires.html>

The SHARE questionnaire is designed in such a way that not every respondent gets every question. Some modules or questions are restricted to certain subgroups of respondents indicated by “if”-statements in the questionnaire. This leads to missing values in the respective variables for those respondents who are not part of the “if”-condition. Basically, two types of questionnaires can be distinguished: the baseline questionnaire for respondents who participate in a SHARE interview for the first time and the longitudinal questionnaire for respondents who participated in SHARE before. The variable *mn101\_* indicates the questionnaire version (see also chapter 12.2.5). Please note that for some items the question text can differ between longitudinal and baseline questionnaire. One example in this context is *ph006\_*: For baseline respondents the question text is “*Has a doctor ever told you that you had any condition ...*” whereas for longitudinal respondents we already collected the information on past diagnoses. Therefore the question text is changed to “*Do you currently have any condition ...*”.

The SHARE interview consists of various thematic blocks or modules and starts with the coverscreen module (*cv\_r*). To pick up contemporary matters and due to time constraints not every module was part of every wave. The overview below lists the different questionnaire modules for the regular panel waves of SHARE and shows in which wave the corresponding module was included. Containing information on retrospective life histories wave 3 (SHARELIFE) differs from the regular SHARE waves. Please find a list of the wave 3 questionnaire modules in the [SHARELIFE Release Guide](#).

**Table 6: Questionnaire Modules of Waves 1, 2, 4, 5 and 6**

	Questionnaire Modules	Wave 1	Wave 2	Wave 4	Wave 5	Wave 6
CV_R	Coverscreen on individual level	X	X	X	X	X
DN	Demographics and Networks	X	X	X	X	X
SN	Social Networks			X		X
CH	Children	X	X	X	X	X
PH	Physical Health	X	X	X	X	X
BR	Behavioral Risks	X	X	X	X	X
CF	Cognitive Function	X	X	X	X	X
MH	Mental Health	X	X	X	X	X
HC	Health Care	X	X	X	X	X
EP	Employment and Pensions	X	X	X	X	X
IT	Computer Use				X	X
MC	Mini Childhood				X	
GS	Grip Strength	X	X	X	X	X
WS	Walking Speed	X	X			
CS	Chair Stand		X		X	
BS	Blood Sample					X
PF	Peak Flow		X	X		X
SP	Social Support	X	X	X	X	X
FT	Financial Transfers	X	X	X	X	X
HO	Housing	X	X	X	X	X
HH	Household Income	X	X	X	X	X
CO	Consumption	X	X	X	X	X
AS	Assets	X	X	X	X	X
AC	Activities	X	X	X	X	X
EX	Expectations	X	X	X	X	X
IV	Interviewer Observations	X	X	X	X	X
	<b>Special Questionnaire Modules</b>					
XT	End-of-Life Interview		X	X	X	X
DO	Drop-off	X	X	X	X	X
VI	Vignettes	X	X			
TC	Technical Variables	X	X	X	X	X

## 12.1 Regular Questionnaire Modules: Short Description

**Demographics (DN):** Basic demographic information about each respondent's marital status, country of birth and education. The module also contains questions on the respondent's parents and (ex-)spouses.

**Social Networks (SN):** Contains a detailed description of respondents' personal social networks. The respondents can name a maximum of seven persons who they consider confidants. The module collects information on the relationship to the respondent as well as additional characteristics of each social network member. The longitudinal dimension of the SN module in wave 6 provides information on whether and why social networks of respondents changed in comparison to wave 4.

**Children (CH):** Collects information about the respondent's children with respect to e.g. marital status, frequency of contact, residential proximity to the respondent, employment status, educational level and whether they have children themselves.

**Physical Health (PH):** Different aspects of people's health: self-reported general health, longstanding illness or disability, eyesight and hearing, specific diagnoses and symptoms, pain, as well as difficulties with a range of (instrumental) activities of daily living. In addition, this module contains questions about work disability and assistive devices, i.e. special aids and appliances used to assist people with disabilities.

**Behavioural Risks (BR):** Covers health behaviours such as smoking, alcohol use, and physical activities.

**Cognitive function (CF):** Contains subjective and objective measures of four aspects of the respondent's cognitive functioning: memory, concentration, numeracy, and verbal fluency.

**Mental Health (MH):** In this module information regarding the mental and emotional health of the respondents is collected.

**Health Care (HC):** Information about recent doctor visits and hospital stays. The module also contains questions about the respondent's level of health insurance.

**Mini Childhood (MC):** Introduced in wave 5, this module contains questions about the respondent's childhood, like the family's housing conditions, number of books in the residence or diseases the respondent may have suffered from. It is a complement of wave 3 (SHARELIFE) and therefore only respondents who did not participate in the SHARELIFE interview of wave 3 received these questions.

**Employment & Pensions (EP):** Information about the respondent's current work activities, income from work as well as other income sources, and any current or past pensions that he/she may be entitled to. For respondents who have retired and are receiving a pension, we ask about the number and kind of pensions and how much they receive.

**Computer Use (IT):** Information on frequency and skills in computer use.

**Grip Strength (GS):** Performance measurement that quantifies the respondent's maximum handgrip strength with the aid of a dynamometer.

**Walking Speed (WS):** In SHARE, walking speed is assessed by measuring the time (in seconds) it takes for a respondent to walk a distance of two and a half meters.

**Chair Stand (CS):** The chair stand test is measured one time per respondent and is performed only by respondents who are 74 years old or younger. The respondents fold their arms across their chest and stand up from a sitting position on a chair and sit down again for five times.

**Blood Sample (BS):** SHARE collected dried blood spots (DBS) samples in 12 countries in wave 6 (BE, CH, DK, EE, FR, DE, GR, IL, IT, SI, SP, SE). All panel household members in the participating countries were eligible for DBS collection. Please note: in France eligibility was given only in a subsample (1/4) of panel households. The BS module provides information on the process of blood spots collection. The blood samples are currently being analysed in medical laboratories. Since these analyses are not yet completed, biomarker values are not yet publicly available. Nevertheless some generated variables on DBS are already provided in the gv\_dbs module. Please subscribe for the SHARE newsletter and/or check our homepage to be informed as soon as DBS data will be released.

**Peak Flow (PF):** Performance measurement of the respondent's expiratory lung force.

**Social Support (SP):** Information about any help the respondents might receive from or give to family or social network members or any other persons.

**Financial Transfers (FT):** Any financial transfers and payments given or received from others, also covering inheritances.

**Housing (HO):** Information about the respondent's current housing situation, including the size and quality of the accommodation. Owners are asked about the value of their property and, depending on the individuals' tenure, questions are asked about mortgages and rent payments. Further questions are related to the amenities and neighbourhood conditions.

**Household Income (HH):** Contains summary measures of various types of household income.

**Consumption (CO):** Brief module with information on household expenditures, e.g. on food.

**Assets (AS):** Questions about the amount of financial and non-financial assets held in various forms including income from these assets.

**Activities (AC):** Contains questions about a range of activities in the as well as information on how people felt about various aspects of their life.

**Expectations (EX):** Explores respondent's expectations, the level of certainty they feel about the future, and how they make financial decisions within their household. It further contains questions about trust in other people as well as about political and religious orientation.

**Interviewer Observations (IV):** Contains questions on the circumstances of the interview answered by the interviewers after the completion of an interview.

## 12.2 Special Questionnaire Modules

### 12.2.1 Coverscreen

The coverscreen is the first module of each interview. It collects basic demographic information about every person currently living in the household. The coverscreen questionnaire is completed by only one member of the household. Household members who are not eligible for a SHARE interview or do not participate are part of the coverscreen data, too, which explains the higher number of observations compared to the other CAPI interview modules.

New in release 6-0-0 is an indication of coverscreen proxy respondents for waves 5 and 6 in the *cv\_r* datasets. This occurs when somebody from outside the household completed the coverscreen on behalf of the household. These coverscreen respondents can be identified by the variable *cvresp* = 2 and by *mergeid* suffices such as “AT-123456-a5” or “SI-987654-b6”.

There have been considerable changes in the wave-specific coverscreen module since release 5.0.0. Some variables were improved or replaced by new variables in order to facilitate easier handling of the data.

Previously, a household member did not have a valid *mergeid* if he or she was not eligible or did not do an interview in the respective wave (e.g. “no int w.2” if no interview could be conducted in wave 2). Since release 5.0.0, each observation in the coverscreen has a valid *mergeid* which facilitates the handling of the data, for instance when merging coverscreen data with other waves and allows for a better monitoring of the household composition. In addition we included the variable *interview* that indicates whether or not a household member has done an interview (categories: 0 = “No interview”, 1 = “Main interview”, 2 = “End-of-life interview”).

Furthermore, it is easier now to identify and follow couples across waves. The variable *mergeidp`w`* (where `w` stands for the respective wave) indicates the *mergeid* of a respondent’s partner. In addition, since release 5.0.0, each couple has a *coupleid* indicated by the variable *coupleid`w`*. The *coupleid* is generated using *mergeid* of both partners and is therefore unique to each couple as well as fix across waves if the relationship stays the same.

The variable *waveid* indicates when a respondent entered SHARE and *waveid\_hh* when the household entered the study. Note that when talking about “waves”, we consider the questionnaire version used. All household members present in wave 1 have a wave 1 *waveid*. In case a new person moves in a wave 1 household in wave 2, this new household member gets a wave 2 *waveid*. Thus *waveid* and *waveid\_hh* can differ from each other. *waveid* has the following values corresponding to the following wave/questionnaire version:

- “42” and “52” (latter due to later fieldwork in Israel) → wave 1
- “62” and “92” (latter due to later fieldwork in Israel) → wave 2
- “82” → wave 3 (SHARELIFE)\*

- “102” → wave 4
- “122” → wave 5
- “142” → wave 6

\**waveid\_hh* is never “82” as there was no refreshment sample in wave 3.

In order to make it easier for users to see when a respondent or household first appeared in SHARE, the variables *firstwave* and *firstwave\_hh* have been introduced. In accordance with the six waves of SHARE, they range from 1 to 6 and are designed as a comprehensive supplement of the variables *waveid* and *waveid\_hh*.

The variables indicating respondent types (family respondent, financial respondent and household respondent; see also chapter 13) have been renamed in line with the imputation datasets. The new names are *fam\_resp*, *fin\_resp* and *hou\_resp* as opposed to the previous naming *dumfamr*, *dumfinr* and *dumhhr*.

The coverscreen contains various “age” variables. There are two concepts to be distinguished: *age* in a certain year and *age* at the date of interview. For example, the variable *age2004* is generated as follows:  $age2004 = 2004 - yrbirth$ . The same goes for the partner’s *age* (*agep2004*) and *age* variables in later waves are also generated along the same line. By contrast, *age\_int* indicates the age of the respondent at the time of the interview. The variable was generated following the rationale  $age\_int = int\_year - yrbirth$  if  $mobirth \leq int\_month$  and  $age\_int = int\_year - yrbirth - 1$  if  $mobirth > int\_month$ . *age\_int* is only provided for those who did an interview.

### 12.2.2 Paper and Pencil Drop-Off

In waves 1, 2, 4, 5 and 6 some interviews are supplemented with the self-completion of a paper & pencil questionnaire. Not all SHARE countries include a so-called drop off in each wave and the content of these questionnaires differs between waves (see table 0 of the appendix for wave 1 and 2 drop-off correspondence) and is partly or completely country-specific. These questionnaires include additional questions on e.g. mental and physical health, health care and social networks.

All drop-off questionnaires are translated into English and labelled accordingly. The generic variables have variable names starting with “q”, country-specific variables contain the country code as prefix, e.g. “at\_” for Austria.

### 12.2.3 Vignettes

Vignettes were filled in by a sub-sample of the respondents instead of the paper and pencil drop-off questionnaire. Vignettes were only part of wave 1 (eight countries: BE, DE, FR, GR, IT, NL, SP, SE) and wave 2 (eleven countries: BE, CZ, DK, DE, FR, GR, IT, NL, PL, SP, SE). The special self-completion questionnaire with anchoring vignette questions is supposed to assess cross-national comparability. Two types of vignettes indicated by the variable “type” were randomly assigned to the respondents. They differ with regard to question order,

gender and, in wave 2 only, age of the persons described in the statements. The labels of the variables show which questions correspond to the each type (A & B in wave 1 and B & C in wave 2).

#### 12.2.4 End-of-Life Interviews

In case a respondent deceased, interviewers conduct an end-of-life interview with a proxy-respondent. The interview is a special form of proxy interview conducted either in person or by phone. Proxy-respondents can be a family member, a household member, a neighbour or any other person of the closer social network of the deceased respondent. The end-of-life interview mainly contains information on respondent's last year of life and the circumstance of death like time and cause of death. The variables are stored in the XT module from wave 2 onwards.

One major innovation since release 5.0.0 is that the open answers to question XT012\_ on cause of a respondent's death are coded. XT012\_ is asked when proxy respondents answered "other" to question XT011\_. Whenever possible the answers to XT012\_ were back-coded to one of the options of XT011\_. If back-coding was not possible, a new variable *xt012c* (suffix "c" for coded) was generated containing new categories for the answers given to XT012\_. For getting a complete picture about the cause of respondents' death users simply have to combine the information stored in the variables *xt011\_* and *xt012c*.

#### 12.2.5 Technical Variables

This module is provided since release 5.0.0. It contains technical variables for waves 1, 2, 4, 5 and 6 (named MN# in the questionnaires) that play an important role in the questionnaire routing. In previous releases these were stored in the DN module. The module includes variables indicating the different respondent types, too: *fam\_resp*, *fin\_resp* and *hou\_resp* (see chapter 8) that are additionally stored in the wave-specific *cv\_r* modules.

**Table 7: Variables in the Technical Variables Module**

Variable	Label
<i>fam_resp</i>	Family respondent
<i>fin_resp</i>	Financial respondent
<i>hou_resp</i>	Household respondent
<i>mn005_</i>	Single or couple interview
<i>mn024_</i>	Nursing home interview
<i>mn026_</i>	First respondent from couple or single
<i>mn028</i>	Eligible for dried blood spots collection (bs &gv_dbs)
<i>mn029</i>	Eligible for linkage
<i>mn030</i>	Eligible for social networks module (sn)
<i>mn031_</i>	Eligible for mini childhood module
<i>mn032_</i>	Eligible for social exclusion items
<i>mn101_</i>	Questionnaire version (longitudinal vs. baseline)



## 13 Special Aspects of Questionnaire Modules

### 13.1 DN Module: Citizenship/Country of Birth Coding

Citizenship (dn008c) and country of birth (dn005c) are coded according to ISO 3166-1 (numeric-3; <http://unstats.un.org/unsd/methods/m49/m49alpha.htm>). The same applies for the since wave 5 newly included country of birth of the respondent's mother (dn504c) and father (dn505c). The introduction of the latter variables enables the identification of second-generation migrants.

The United Nations Statistics Division also provides codes for countries that no longer exist (see the above link). We added few additional codes for additional countries/regions and for respondents with multiple citizenships (see table 0 of the appendix).

An often-occurring case is a respondent born in the USSR, in a place that would now be part of Russia. In that case, we code by the mentioned birth country, not by the actual country at the time of birth. In this case: if the person answered "USSR", she or he will be coded as being born in the USSR. If she or he answered "Russia", her coding will read "Russia", although it was USSR when the respondent was born. If it is important for you to know the country name at the time of birth, you will have to deduce the code using the person's year of birth.

### 13.2 CH Module: Information on Children in SHARE

The CH module is answered by the family respondent (see chapter 8). All variables with the same suffix belong to the same child, so that e.g. for "child 2" the year of birth is stored in variable ch006\_2, the marital status in ch012\_2, etc. Due to the programming of the questionnaire it is possible that there are gaps in the numeration of the children, e.g. a respondent has two children: "child 1" and "child 3".

From wave 4 on longitudinal respondents were asked about changes concerning their children since the last interview. In the data there are three types of variables to capture such a change. The first variable indicates whether a change occurred (Yes/No). If there was a change reported, the consecutive variables contain the information for which child. This is stored in dummy-variables for each child (e.g. ch515d3 = 1 if the marital status of "child 3" changed). In the last step, the new status of the child is specified.

Please note that the suffix is not necessarily consistent for children across waves. We recommend using the children's gender and year of birth to identify children across waves.

**Table 8: Changes in Information on Children**

	Was there a change?	For which child?	New state
<b>Education</b>	ch508_	ch509d1– ch509d20	ch510_1– ch510_20
<b>Further education</b>	ch511_	ch512d1– ch512d20	ch513_1– ch513_20
<b>Marital status</b>	ch514_	ch515d1– ch515d20	ch516_1– ch516_20
<b>Parenthood</b>	ch517_	ch518d1– ch518d20	ch519_1– ch519_20 ch520_1– ch520_20
<b>Location</b>	ch524_	ch525d1– ch525d20	ch526_1– ch526_20

In waves 1 and 2 questions ch010 to ch020 only refer to a maximum of four children. When there are more than four children, the CAPI program selects the four children as follows:

1. Sort children in ascending order by
  - underage (defined as 0 for all children aged 18 and over and 1 for all others),
  - geographical proximity (ch007),
  - year of birth
2. Pick the first four children.
3. In case that all sorting variables of a group of two or more children are equal, children are selected randomly from this group.

The variables chselch1 up to chselch4 contain the corresponding loop number of the children who were selected by the program.

### **13.3 PH Module: *phrandom* and *ph006***

There are two types of answer categories for the question on self-perceived health in wave 1. Which type is asked at the beginning of section PH (questions ph002/003), and which (other) type for the end of this section (questions ph052/053), is randomized. The variable *phrandom* indicates which type is chosen:

- 1 for *ph002/ph052* (version 1)
- 2 for *ph003/ph053* (version 2)

### 13.4 EX Module: Definition of Life Expectancy Target Age

In ex009\_ baseline respondents are asked “What are the chances that you will live to be age <fill> or more?”. The <fill> used in this question is a function of the age of each respondent. Age in turn is computed by the system subtracting year and month of birth from year and month of interview (note that day is not used).

Based on the computed age, the <fill> for ex009\_ is defined as:

Respondents' age	Fill in ex009_
age < 65	75
age > 64 and age < 70	80
age > 69 and age < 75	85
age > 74 and age < 80	90
age > 79 and age < 85	95
age > 84 and age < 95	100
age > 94 and age < 100	105
age > 99 and age < 105	110
age > 104	120

The fill used in each interview is stored in the variable *ex009age*, while the substantive answer to the question is stored as *ex009\_*. In rare cases (e.g. if age was ex post corrected due to an interviewer remark) the fill might deviate from this rule.

### 13.5 CF Module: Ten Words List Learning

The “ten words list learning” test is conducted with a first trial and a delayed recall. In waves 1 and 2 the same list was used for all whereas in waves 4, 5 and 6 respondents were assigned randomly to one of four sets of “ten words list learning”. Thereby, a total of eight variables cover the “ten words list learning” items in the wave 4, 5 and 6 release data: the variables cf104tot, cf105tot, cf106tot and cf107tot refer to the four sets of the “ten words list learning” first trial, whereas the variables cf113tot, cf114tot, cf115tot and cf116tot refer to the four sets of the “ten words list learning” delayed recall. This means e.g. that the respective information for respondents who were assigned to the first set of “ten words list learning” (that is cf104\_Learn1 in the questionnaire), is stored in cf104tot and cf113tot, for respondents who were assigned to the second set it is cf105tot and cf114tot and so on.

In addition, the generated health module (gv\_health) provides the generated variables cf008tot (first trial) and cf016tot (delayed recall). Both variables contain the result for all respondents irrespective of which of the four sets of “ten words list learning” was used. Note that both “combined” variables are stored in the cognitive function (CF) module in waves 1 and 2, but are part of the generated health module (gv\_health) in waves 4, 5 and 6. This is due to the fact that these variables have been generated in waves 4, 5 and 6 and were not regular CAPI items.

### 13.6 SN Module (Waves 4 and 6)

The social network module (SN) was implemented in the 4th wave of SHARE as an innovative means to measure the personal social environment. This approach goes beyond the more common role-relational method of measuring social networks which is based mostly on socio-demographic proxies. The SN module employs a name generator that first identifies the respondent's self-reported meaningful relationships and then obtains the following characteristics of the persons named: gender, residential proximity to the respondent, frequency of contact, and level of emotional closeness. In wave 6, the module contains additional questions about the year of birth of each named person, their occupational status, as well as their partner status.

Moreover, the wave 6 SN module traces changes that occurred within the network since wave 4. Following the identification of the wave 6 roster (i.e. the list of persons named), respondents were shown a list of their wave 4 social network members and asked to link them to their wave 6 members. If a person was not mentioned again respondents were asked to select reasons. The linkage information for social network members between wave 4 and wave 6 appears in the `gv_networks` module.

Some follow-up questions on network members were skipped in the SN module in order to save interview time. This was done when the information was available from questions asked elsewhere in the interview, primarily for children, spouses, and parents. In such cases, the relevant information was aggregated into generated variables. It is therefore recommended to use the information on specific social network members from the `gv_networks` module instead of the data from the SN module. Although most of the information is available, some information has not yet been fully combined in release 6.0.0. Specifically, social network members who are children or spouses still lack some information, most of which is expected to be included in generated variables in a future release of SHARE data.

Furthermore, the SN information is linked to the subsequent questionnaire modules social support (SP) and financial transfers (FT). These modules collect information on the exchange of personal or financial support and the respondent can specifically assign social network members as providers or receivers of this kind of support.

For further information see:

Litwin H., Stoeckel K., Roll A., Shiovitz-Ezra S. & Kotte M. (2013): "Social Network Measurement in SHARE Wave Four". In: *Malter, F., Börsch-Supan, A. (Eds.) (2013). [SHARE Wave 4: Innovations & Methodology](#). Munich: MEA, Max Planck Institute for Social Law and Social Policy.*

Litwin, H., & Stoeckel, K. J. (2015). Social Network, Activity Participation, and Cognition: A Complex Relationship. *Research on Aging*, 38(1), 76–97.

### 13.7 SP and FT Module: List of Relations differences due to SN

In waves 1, 2, 4, 5 and 6 variables in the social support (SP) and the financial transfer (FT) module refer to a list of relations. This list differs in wave 4 because the social networks (SN) module was linked to the SP and FT module, so respondents could indicate which social network member e.g. provided help. In wave 4, information about social network persons has been forwarded and included in this list during the interview. Thus the list of relations in wave 4 consists of up to seven social network members plus the regular “list of relations”-categories. Therefore, the initial coding of these variables is different from other waves. To maintain comparability between waves the respective categories have been recoded. Categories referring to social network persons received new codes. In addition, four new response categories were implemented in the wave 4 list of relations, i.e. categories 34-37 whereas others are not included anymore.

These changes in the list of relations affect the following questions:

sp003_	Who gave you help
sp009_	To whom did you give help
sp019_	To whom given help in this household
sp021_	Who gave you help in this household
ft003_	To whom did you provide financial gift 250 or more
ft010_	From whom received financial gift 250 or more
ft017_	From whom inherited 5000 or more
ft027_	To whom given 5000 or more

In wave 4, they were renamed to accommodate these changes. Variables that refer to social network members get the additional suffix **sn**, e.g. sp019d1sn refers to the first mentioned social networks member. Dummy variables referring to the ‘standard’ categories of the list of relations get the additional suffix **sp** or **ft** after the loop counter.

Please note that the standard categories of the list of relations do not include all persons that have the respective relation to the respondent. If e.g. the partner that provided help is mentioned as the first social network member, he or she will not show up in the standard category ‘partner’ but only as first social network member.

In wave 6 both social network members and children were assessed via follow up questions subsequent to the list of relations. The wave 6 categories were partly recoded to match the categories of other waves. This is why the category numbers in the data might deviate from the numbers shown in the questionnaire (see table 9).

The following table illustrates the coding and variable naming across waves in the SP and FT module:

**Table 9: Relationship Type Coding**

Value or variable label	Wave 1, 2 & 5 values	Wave 4 values	Wave 6 values	Wave 4 variable suffix
Social network member1	-	101		d1sn
Social network member2	-	102		d2sn
Social network member3	-	103		d3sn
Social network member4	-	104		d4sn
Social network member5	-	105		d5sn
Social network member6	-	106		d6sn
Social network member7	-	107		d7sn
Spouse/partner	1	1	1	d1sp
Mother	2	2	2	d2sp
Father	3	3	3	d3sp
Mother-in-law	4	4	4	d4sp
Father-in-law	5	5	5	d5sp
Stepmother	6	6	6	d6sp
Stepfather	7	7	7	d7sp
Brother	8	8	8	d8sp
Sister	9	9	9	d9sp
Child			10	
Step-child/your current partner's child			11	
Child 1	10	-		-
Child 2	11	-		-
Child 3	12	-		-
Child 4	13	-		-
Child 5	14	-		-
Child 6	15	-		-
Child 7	16	-		-
Child 8	17	-		-
Child 9	18	-		-
Other child	19	19		d19sp
Son-in-law	20	20	20	d20sp
Daughter-in-law	21	21	21	d21sp
Grandchild	22	22	22	d22sp
Grandparent	23	23	23	d23sp
Aunt	24	24	24	d24sp
Uncle	25	25	25	d25sp
Niece	26	26	26	d26sp
Nephew	27	27	27	d27sp
Other relative	28	28	28	d28sp
Friend	29	29	29	d29sp
(Ex-)colleague	30	30	30	d30sp
Neighbour	31	31	31	d31sp
Ex-spouse/partner	32	32	32	d32sp
Other acquaintance	33	-		-
Step-child/your current partner's child	-	34		d34sp
Minister, priest, or other clergy	-	35	35	d35sp
Therapist or other professional helper	-	36	36	d36sp
Housekeeper/home health care provider	-	37	37	d37sp
NONE OF THESE	96	96	96	dno

## 13.8 EP Module

### 13.8.1 EP005 & Re-Interview in Israel (Wave 1)

Users working with the ep module will realise that there are two variables on the respondent's employment status in wave 1: *ep005\_* and *ep005raw*. *ep005raw* contains the original values. Most of the recodings implemented in *ep005\_* are due to the Israeli re-interviews. But also in other countries some cases were (re)coded mainly based on open answers that followed the "other specify" option. We recommend using the recoded variable *ep005\_* when doing any kind of analysis.

Since in wave 1 the question *EP005* in Israel differed from the generic version, there is an extra module to facilitate working with the data. While in all the other countries there are 5 response categories to question EP005, there are 7 in Israel: category 3 "Unemployed" was subdivided into "Unemployed, looking for a job" (Israeli category 3) and "Unemployed, not looking for a job" (Israeli category 4); category 4 "Permanently sick or disabled" (Israeli category 6) was complemented by the category "Temporarily sick or disabled" (Israeli category 5). This results in the following scheme:

#### Generic questionnaire

- 1) Retired
- 2) Employed or self-employed
- 3) Unemployed
- 4) Permanently sick or disabled
- 5) Homemaker

#### Israeli questionnaire

- 1) Retired
- 2) Employed or self-employed
- 3) Unemployed, looking for a job
- 4) Unemployed, not looking for a job
- 5) Temporarily sick or disabled
- 6) Permanently sick or disabled
- 7) Homemaker

Due to the generic programming of the CAPI the additional Israeli response categories caused some routing problems with follow-up questions in the EP module. The affected respondents were therefore re-interviewed by phone. In this phone interview the CAPI instrument was not used and only a subset of the questions was asked again. The created dataset *sharew1\_rel6-0-0\_ep\_ilextra* contains all respondents affected by routing problems (576 respondents of which 308 could be re-interviewed), irrespective of whether they participated in this second interview or not. Participation in the second interview is indicated by the variable *reint*. The variables *reint\_month* and *reint\_year* provide the date of the second interview. All Israeli variables in the regular EP-module which resulted from the wrong routing were recoded as "missing value".

### 13.8.2 Values in EP078\_3, EP078\_7, and EP078\_12 in Greece (Wave 6)

Although some pension types in the generic questionnaire version (*ep671d3*, *ep671d7* and *ep671d12*) are considered as not being relevant for Greece, there are values in the corresponding variables *ep078#3*, *ep078#7* and *ep078#12*. Please note that we did not include them in the imputation of pension variables in the *gv\_imputations* module as it remains unclear to which type of pension these values refer to. For very few respondents this was also the case in other countries (e.g. *ep078\_2* and *ep078\_6* in Slovenia).

## 13.9 Social Exclusion Items (Wave 5)

A new set of 19 questions was included in the wave 5 questionnaire to provide additional informative measures of respondents' economic situation and to allow the development of multidimensional measures of social exclusion.

The new items cover aspects of affordability of specific expenses and neighbourhood quality. They do not constitute a separate questionnaire module but are included in the consumption, behavioural risks and household income module:

**Table 10: Items on Social Exclusion**

Module	Variables
Consumption (CO)	co201_, co202_, co206_, co207_, co208_, co209_, co211_, co213_, co220_
Behavioural risks (BR)	br033_, br034_
Household income (HH)	hh022_, hh023_, hh024_, hh025_, hh026_, hh027_, hh028_, hh029_, hh030_

Most of the questions were answered by the household respondent on behalf of all household members. Please note that due to an unnoticed element in the instrument not all households received the set of questions. Altogether 94% of the households are covered. The variable *mn032\_* in the technical variables module contains information on who received the social exclusion items.

Further information on the social exclusion items is available in: Myck M., M. Oczkowska & D. Duda (2015): "Innovations for better understanding deprivation and social exclusion". In: *Malter, F. and A. Börsch-Supan (Eds.) (2015). [SHARE Wave 5: Innovations & Methodology](#). Munich: MEA, Max Planck Institute for Social Law and Social Policy.*

## 14 Generated Variables

To assure an easy and fast entry into cross-national data and high convenience while working with the data certain variables are readily provided for the SHARE users. In addition to internationally standardized codings (e.g. ISCED), there are several further generated variables (gv) that enhance working with the SHARE data including several indices and scales like EURO-D and the CASP-12 index.

In comparison to previous releases, several innovations were implemented for the generated variable modules since release 5.0.0. New are *gv\_allwaves\_cv\_r*, the *gv\_exrates* and *gv\_deprivation* module. Weights and imputations are updated for release 6.0.0. New generated modules for wave 6 are *gv\_children* and *gv\_dbs*.



**Table 11: Generated Variable Modules of Waves 1 to 6**

Generated-Variable-Modules	Content	W1	W2	W3	W4	W5	W6
gv_allwaves_cv_r	Coverscreen information across waves	Cross-wave module					
gv_longitudinal_weights	Longitudinal weights	Cross-wave module					
gv_weights	Cross-sectional sampling design weights and calibrated weights	X	X	X	X	X	X
gv_imputations	Multiple Imputations	X	X		X	X	X
gv_iscd	International Standard Classification of Education (ISCED-97/since wave 5 additionally ISCED-11)	X	X		X	X	X
gv_health	Physical and mental health variables and indices like BMI, EURO-D depression scale, etc.	X	X		X	X	X
gv_housing	Housing and NUTS codes	X	X		X	X	X
gv_networks	Information on social networks				X		X
gv_exrates	Exchange rates for all waves, incl. nominal and ppp-adjusted exchange rates	Cross-wave module					
gv_job_episodes_panel	Labour market status of each SHARELIFE respondent throughout her/his life	Cross-wave module					
gv_grossnet	Net income measures derived from reported gross incomes	X					
gv_isco	Classification of occupations via ISCO and of industries via NACE codes	X					
gv_ssw	Social security wealth				X		
gv_deprivation	Indices for material and social deprivation					X	
gv_children	Combined children information						X
gv_linkage	Linkage to Statutory Pension Insurance data (Germany only)	Cross-wave module					
gv_dbs	Dried Blood Spots						X

## 14.1 gv\_allwaves\_cv\_r

In addition to the wave-specific coverscreens, SHARE provides the *gv\_allwaves\_cv\_r* module. It is a dataset with merged and enriched information from all waves that allows monitoring household composition, changes of status (Is a respondent part of a couple? Is he or she dead or alive? , etc.) and the type of interviews conducted.

The variables *interview\_hh\_w`w'* (*`w'* stands for the wave) indicate whether or not a household participated in a specific wave. *interview\_ever* detects household members who have done at least one interview as opposed to those who have never done an interview because they are ineligible for a SHARE interview or refused to participate.

The variable *deadoralive\_w`w'* has been introduced in order to replace the former module *gv\_dol*. This generated status variable uses enriched information in addition to the variable *deceased* from the wave-specific coverscreens. The variables *deceased\_year*, *deceased\_month* and *deceased\_age* complement this section.

## 14.2 Weights

*Author: Giuseppe De Luca*

SHARE provides different types of weights. Which weights to use really depends on the concrete research question, i.e. the cross-sectional or longitudinal nature of the study, the waves under investigation, the unit of analysis (household or individual), and the reason for weighting sample observations. Thus, although it is not possible to give any general advice, we provide a large set of weights in order to satisfy most of the data users' needs.

### 14.2.1 Sampling design weights

The survey design of SHARE is based upon of the basic principles of probability-sampling with maximal population coverage (see, for example, De Luca and Rossetti 2008, Lynn et al 2013, and De Luca et al 2015). The availability of a probability sample ensures that every unit in the target population has a chance greater than zero of being selected into the sample. Further, it should be possible to compute the selection probability of each individual to enable valid inference on the target population by using (weighted) sample statistics. The extent to which full population coverage is strictly possible depends on the quality of national sampling frames available in each wave, but in all cases close to full coverage was achieved.

Notice that the sampling frame and the associated sampling design are not restricted to be the same in all countries. Moreover, countries are allowed to change sampling frame and sampling design across waves in order to choose the best available option at each wave. The national sampling designs adopted across waves of the SHARE panel vary from simple random sampling to rather complicated multi-stage designs. As a consequence, we are far from the ideal setting of having a constant selection probability for all units of the SHARE

sample. The design of the SHARE survey only ensures that selection probability is constant for all respondents belonging to the same households.

Sampling design weights, defined as the inverse of the probability of being included into the national cross-sectional sample of any given wave, allow the users to compensate for unequal selection probabilities of the sample units. As usual, the validity of these weights relies crucially on the assumption of full survey response. In practice, however, survey data are always affected by problems of unit nonresponse and attrition that may jeopardize the theoretical properties of inference based on the sampling design weights (Lessler and Kalsbeek 1992). Since SHARE is not an exception to this common situation, we strongly discourage users to rely on sampling design weights for standard analyses of the SHARE data. These weights are included in release 6.0.0 only for the implementation of specific statistical methods which account for nonresponse errors in other ways, or for other specific purposes.

The sampling design weights provided in release 6.0.0 present some worth noticing differences with respect to those provided in the previous releases of the SHARE weights.

**1. Gross sample database:** In several countries, we have carefully revised the gross sample database to ensure that it is consistent with the information on sample size, stratification and clustering available from the national sampling design documentation.

Although the gross sample database is not included in the scientific release, this database is considered a key ingredient for the analysis of sampling and nonsampling errors. Specifically, it allows to identify the various subsamples (i.e. sets of households drawn with the same sampling design) drawn in any specific country and wave of the SHARE panel. In addition, it contains the sampling design information (such as strata and primary sampling units) required to compute the sampling design weights and the auxiliary sampling frame information (such as age, gender, and NUTS/LAU regional codes of the selected household member) required to cope with unit nonresponse errors.

The sampling design weights included since release 5.0.0 may differ from those provided in the previous release of the SHARE weights because of the revisions applied to the gross sample database.

**2. Unknown cross-selection probabilities & pooling method:** Since the second wave of the study, the cross-sectional samples of many countries include two types of sample units: those sampled in some previous wave and those sampled in the current wave. We refer to these two sets of sample units as longitudinal and refreshment samples, respectively. In SHARE, the aim of drawing the refreshment sample can be twofold. First, it allows refreshing the longitudinal sample with the younger age-cohorts of the target population who were not age-eligible in the previous waves. Secondly, it allows compensating for the loss of observations in the longitudinal sample due to attrition across waves of the SHARE panel.

For an overview of the wave-specific composition of the national samples, we refer the reader to De Luca et al (2015).

For the specific purposes of the sampling design weights, we recognized that the longitudinal and refreshment samples are usually drawn from (partly) overlapping target populations. For example, the longitudinal sample of wave 2 is drawn from the (wave 1) target population of people born in 1954 or earlier, while the refreshment sample is drawn from the (wave 2) target population of people born in 1956 or earlier. As a consequence, the units belonging to the longitudinal sample may have a non-zero probability of being selected in the refreshment sample and the units belonging to the refreshment sample may have a non-zero probability of being selected in the longitudinal sample. From a practical point of view, the main problem is that these ‘cross-selection probabilities’ can be computed exactly only in few countries where sampling is carried out on the basis of simple designs.

Many alternative approaches are available to handle the issue of unknown ‘cross-selection probabilities’ (see, for example, Watson 2014). Since release 5.0.0, we implemented the so-called ‘pooling method’ in which the unknown cross-selection probabilities are estimated on the basis of the available sampling frame information such as strata, primary sampling unit (psu), age, gender and regional nuts codes. Unlike the previous releases of the weights, the sampling design weights of waves 2, 3, 4, 5 and 6 now reflect the estimated contribution of these ‘cross-selection probabilities’.

**3. Distinction between main & vignette samples:** In waves 1 and 2, we do not distinguish anymore between sampling design weights for the main sample alone, the vignette sample alone, and the two samples combined. Since such distinction was relevant only for the first two waves of the SHARE panel, we decided to simplify the scientific release of the weights by focusing on the more frequently used sampling design weights for the entire cross-sectional sample of each wave.

**4. Sampling design weights of Ireland:** As mentioned above, we do not currently know the sampling design adopted for drawing the Ireland sample of wave 2. To allow the computation of calibrated weights, for this country we computed fictitious sampling design weights based on the simplifying assumption of simple random sampling.

#### **14.2.2 Calibrated weights**

The strategy used by SHARE to cope with the potential selection effects associated with unit nonresponse and panel attrition is the calibration procedure proposed by Deville and Särndal (1992). This statistical re-weighting procedure produces calibrated weights which are as close as possible, according to a given distance measure, to the original sampling design weights while also respecting a set of known population totals (the calibration margins). Under the assumption the missing data mechanism is missing-at-random, calibrated weights may help reduce the potential selection bias generated by unit nonresponse and panel attrition. Unless nonresponse and attrition errors are controlled by alternative approaches,

this is the type of weights that we generally advise to use in standard analyses of the SHARE data.

As in the previous releases of the SHARE weights, release 6.0.0 includes calibrated cross-sectional weights to be used in the context of cross-sectional analyses and calibrated longitudinal weights to be used in the context of longitudinal analyses. Both types of calibrated weights are computed at the individual level for inference to the target population of individuals and at the household level for inference to the target population of households. On the other hand, release 6.0.0 does not include any more separate calibrated weights for the main sample alone, the vignette sample alone and the two samples combined. Further, because of changes in the underlying imputation procedure (see the documentation about imputations), it does not include any more calibrated cross-sectional weights for non-responding partners.

Notice that the calibrated weights provided in the scientific release only compensate for unit nonresponse and panel attrition in the main CAPI interview by ignoring additional problems of unit nonresponse sample attrition that are present in the drop-off questionnaires. Also notice that, as a general rule, calibrated cross-sectional weights can be missing for respondents who are younger than 50 years (i.e. age-ineligible partners of an age-eligible respondent), those with missing sampling design weights, and those with missing information on the set of calibration variables (i.e. age, gender and NUTS1 code).

#### **a) Calibrated cross-sectional weights**

For each wave, we provide 'calibrated cross-sectional individual weights' (*cciw\_w#*) and 'calibrated cross-sectional household weights' (*cchw\_w#*). The former assign a calibrated weight to each 50+ respondent that depends on the underlying sampling design weight and the individual-specific set of calibration variables, while the latter assign a common calibrated weight to all respondents in the same household that depends on the underlying sampling design weight and the set of calibration variables for all 50+ household members.

Both types of calibrated cross-sectional weights are computed separately by country to reproduce the size of the national target populations in each wave of the study. In each country and wave, the set of calibration margins reflects the size of the target population across 8 gender-age groups (i.e. males and females in the age groups [50-59], [60-69], [70-79], [80+]) and across NUTS1 regional areas. For countries involved in wave-specific oversampling of age-cohorts that are not represented in the longitudinal sample, there is also a finer partition of [50-59] age group. For example, in wave 2, we use separate calibration margins for male and female respondents with age in the intervals [50-51] and [52-59]. In wave 4, we use instead separate calibration margins for male and female respondents with age in the intervals [50-53] and [54-59].

Calibration margins about the size of target population of each wave are taken from the EUROSTAT regional database.

## **b) Calibrated longitudinal weights**

Calibrated longitudinal weights differ from calibrated cross-sectional weights in two important respects. First, they are only defined for the subsample of units who participated in two or more waves of the panel. Second, since mortality is a source of attrition which affects both the sample and the population, the calibrated longitudinal weights account for mortality of the original target population across waves. The target population for longitudinal analyses is implicitly defined as the target population at the beginning of a time reference period that survives up to the end of the period considered.

Since the SHARE panel now consists of six waves, one can compute a lot of different types of calibrated longitudinal weights depending on the selected wave combination (i.e., 1-2,...,5-6, 1-2-3,...,4-5-6, 1-2-3-4,..., 2-3-4-5-6, 1-2-3-4-5-6) and the basic unit of analysis (either individuals or households). Furthermore, the number of possible calibrated longitudinal weights will increase rapidly as the panel goes ahead and additional waves will be available. These considerations clarify why the strategy of providing calibrated longitudinal weights for all possible wave combinations of the panel is not feasible, especially in the long run. Since wave 4, we provide calibrated longitudinal weights only for selected wave combinations. The wave combinations considered in release 6.0.0 are: 1-2, 2-3, 3-4, 4-5, 5-6 and 1-2-3-4-5-6. For longitudinal analyses exploiting other wave combinations, users are required to control for attrition either by computing their own calibrated weights or by implementing some alternative correction method. To support users in this nontrivial methodological task, we provide a Stata ado-file called ``cweight.ado'` which implements the calibration procedure of Deville and Särndal (1992), a Stata do-file called ``weighting.do'` which illustrates step-by-step how to compute calibrated longitudinal weights at the individual and the household level, and a Stata data file ``population_margins.dta'` which contains information about the size and the number of dead of the national 50+ populations by age, gender and NUTS1 regional area. The necessary datasets and Stata dofiles can be downloaded by registered SHARE users from the SHARE data dissemination website. For further information see the "[Generating Calibrated Weights](#)" user guide.

The calibrated longitudinal weights associated with the wave combination  $t-s$  are computed separately by country to represent the national populations of wave  $t$  that survives up to the interview year of wave  $s$ . For example, in several countries, the wave combination 1-2 allows to represent the national population of people aged 50+ in 2004 that survives up to 2006/2007. The wave combination 1-2-3-4-5-6 (the fully balanced panel) allows instead to represent the national population of people aged 50+ in 2004 that survives up to 2013. For the longitudinal calibrated weights, the set of calibration margins reflects the size of the target population by 8 gender-age groups (i.e. males and females with age in the beginning of a time reference period in the intervals [50-59], [60-69], [70-79], [80+]). Compared to calibrated cross-sectional weights, we do not control for NUTS1 regional codes and finer partitioning of the [50-59] age interval due to the relatively lower sample size. Mortality is

accounted for by subtracting from each population margin the corresponding number of deaths occurred in time reference period considered.

Similarly to calibrated cross-sectional weights, our set of longitudinal calibrated weights is available both at the individual and the household level. Notice that for the weights defined at the household level, we only require that there is at least one eligible household member interviewed in each wave of the chosen wave combination. This implies that balanced sample of households is larger than the corresponding balanced sample of individuals. For example, households consisting of one partner participating in wave 1 and the other partner participating in wave 2 belong to the balanced sample of households for the wave combination 1-2, even if neither partner belongs to the corresponding balanced sample of individuals.

### 14.2.3 Structure of the SHARE weights in release 6.0.0

Release 6.0.0 includes five datasets for the cross-sectional weights of waves 1 to 6 (sharew#\_rel6-0-0\_gv\_weights). These datasets come together with the other modules of each wave. Moreover, six datasets with longitudinal weights are included in a separate download file. The first five refer to the wave combinations 1-2, 2-3, 3-4, 4-5 and 5-6. These datasets are named e.g. sharewX\_rel6-0-0\_gv\_longitudinal\_weights\_w1w2 for wave combination 1-2. The others are named accordingly. Finally, there is one dataset named sharewX\_rel6-0-0\_gv\_longitudinal\_weights\_w1w-w6 for the longitudinal weights of the fully balanced panel (i.e. wave combination 1-2-3-4-5-6). Tables 18 and 19 provide, respectively, a description of the variables contained in the five cross-sectional and the five longitudinal datasets.

**Table 12: Sampling design and calibrated cross-sectional weights**

Variable	Description	Unit of analysis
dw_w#	Design weight – wave #	Household & individual
cchw_w#	Calibrated cross-sectional household weight – wave #	Household
cciw_w#	Calibrated cross-sectional individual weight – wave #	Individual
subsample	Subsamples within country	Household & individual
stratum1	First stratum	Household & individual
stratum2	Second stratum	Household & individual
psu	Primary sampling unit	Household & individual
ssu	Secondary sampling unit	Household & individual

In addition to the individual, household and country identifiers, the cross-sectional weights include:

- one variable for the sampling design weights (variable name *dw\_w#*) of the cross-sectional sample of wave # (both household and individual levels),
- two variables for the corresponding calibrated cross-sectional household weights (variable name *cchw\_w#*) and the calibrated cross-sectional individual weights (variable name *cciw\_w#*),
- one indicator (variable name *subsample*) which identifies the various subsamples drawn in any specific country and wave of the SHARE panel,
- four indicators (variables names *stratum1*, *stratum2*, *psu*, and *ssu*) for the information about stratification and clustering in each subsample,

**Table 13: Sampling design and calibrated longitudinal weights**

Variable	Description	Unit of analysis
<i>dw_w#</i>	Design weight - wave #	Household & individual
<i>panel_resp</i>	Respondent participation in the selected panel	Individual
<i>clhw_a</i>	Calibrated long. household weight - panel: 1-2-3-4-5-6	Household
<i>cliw_a</i>	Calibrated long. individual weight - panel: 1-2-3-4-5-6	Individual
<i>clhw_b</i>	Calibrated long. household weight - panel: 1-2	Household
<i>cliw_b</i>	Calibrated long. individual weight - panel: 1-2	Individual
<i>clhw_c</i>	Calibrated long. household weight - panel: 2-3	Household
<i>cliw_c</i>	Calibrated long. individual weight - panel: 2-3	Individual
<i>clhw_d</i>	Calibrated long. household weight - panel: 3-4	Household
<i>cliw_d</i>	Calibrated long. individual weight - panel: 3-4	Individual
<i>clhw_e</i>	Calibrated long. household weight - panel: 4-5	Household
<i>cliw_e</i>	Calibrated long. individual weight - panel: 4-5	Individual
<i>clhw_f</i>	Calibrated long. household weight - panel: 5-6	Household
<i>cliw_f</i>	Calibrated long. individual weight - panel: 5-6	Individual
<i>subsample</i>	Subsamples within country	Household & individual
<i>stratum1</i>	First stratum	Household & individual
<i>stratum2</i>	Second stratum	Household & individual
<i>psu</i>	Primary sampling unit	Household & individual
<i>ssu</i>	Secondary sampling unit	Household & individual

Similarly, the longitudinal weights include:

- a binary indicator (variable name *panel\_resp*) which is equal to 1 for the balanced panel of individuals (i.e. the respondents participated to all waves of the selected wave combination) and equal to zero otherwise,
- one variable for the sampling design weights (variable name *dw\_w#*) of the starting wave #,



- two variables for the calibrated longitudinal household weights and the calibrated longitudinal individual weights ,
- one indicator (variable name *subsample*) which identifies the various subsamples drawn in any specific country and wave of the SHARE panel,
- four indicators (variables names *stratum1*, *stratum2*, *psu*, and *ssu*) for the information about stratification and clustering in each subsample,

Notice that the longitudinal databases contain one observation for each respondent belonging to the balanced sample of households in the selected wave combination. Within each dataset, the balanced sample of individuals can be identified through the binary indicator *panel\_resp*.

## References

De Luca G. and C. Rossetti (2008), *Sampling design and weighting strategies in the second wave of SHARE*, in Börsch-Supan A., Brügiavini A., Jürges H., Mackenbach J., Siegrist J. and Weber G. (ed.), *Health, Aging and Retirement in Europe (2004-2007) – Starting the Longitudinal Dimension*, Mannheim.

De Luca G., Rossetti C. and Malter F. (2015), *Sample design and weighting strategies in SHARE Wave 5*, in Malter F. and Börsch-Supan A. (ed.), *SHARE Wave 5: Innovations & Metodology*, Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Deville J.C. and Särndal C.E. (1992), *Calibration Estimators in Survey Sampling*, *Journal of the American Statistical Association* 87: 376-382.

Lessler J. and Kalsbeek W. (1992), *Nonsampling Error in Survey*, New York: John Wiley & Sons.

Lynn P., De Luca G., Ganninger M. and Häder S. (2013), *Sampling Design in SHARE Wave Four*, in Malter F. and Börsch-Supan A. (ed.), *SHARE Wave 4: Innovations & Metodology*, Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Rubin D.B. (1976), *Inference and missing data*, *Biometrika*, 63, 581-592.

Solon G., Haider S.J., Wooldrige J. (2013), "What are we weighting for?", *NBER Working Paper Series N. 18859*, <http://www.nber.org/papers/w18859>.

Watson N. (2014), *Evaluation of Weighting Methods to Integrate a Top-Up Sample with an Ongoing Longitudinal Sample*, *Survey Research Methods* 8: 195-208.

## 14.3 Imputations

*Author: Giuseppe De Luca*

This section describes the multiple imputations available in release 6.0.0 of the SHARE data. Please note that SHARE does not provide imputations for the third wave (SHARELIFE) which collects life history data.

### 14.3.1 Practical decisions about imputations

Handling item nonresponse in a cross-national, multi-disciplinary and longitudinal survey like SHARE is a challenging task that involves many different decisions that have to be balanced against each other. For this reason, we find it important to inform data users of the key steps that were necessary to implement the imputation model and the rationale driving the construction of harmonised imputations for all regular waves of SHARE.

#### a) Dimensionality of the imputation model

Due to the large number of variables collected in each wave, the first issue to be addressed is how to select a feasible subset of core variables that accommodates a wide variety of analyses that data users might want to perform on the imputed data. Preliminary choices regarding the dimensionality of the imputation model are important in the context of multivariate imputation procedures that attempt to preserve the correlation structure of the imputed variables. In these methods, multiple variables are imputed simultaneously on the basis of some Markov Chain Monte Carlo (MCMC) technique. The main problem is that as the number of variables to be imputed jointly increases, these iterative techniques often require significant effort in programming and fine tuning. A compromise between generality and complexity of the imputation model is therefore needed.

To deal with this issue, the new imputation procedure exploits a simple hot-deck method for various types of variables affected by negligible fractions of missing data and a more sophisticated multivariate imputation method for monetary variables affected by more relevant fractions of missing data.

#### b) Preliminary transformations and standardizations of the raw data

After selecting a set of core variables to be imputed in each wave, we construct a set of binary eligibility indicators to identify the respondents who are eligible to answer each question. This nontrivial step requires taking into account possible inconsistencies in the raw data, country-specific deviations from the generic versions of the CAPI questionnaires, branching, skip patterns and proxy interviews.

For open-ended questions on monetary variables, which are usually preceded by one or more ownership questions, we also construct a set of binary ownership indicators to identify the subset of eligible respondents with a non-zero monetary amount. Conditional on eligibility and ownership, non-zero values of monetary variables are converted (if needed) in annual Euro amounts to avoid differences in the time reference period of each question and

the national currencies of non-Euro countries. Moreover, we symmetrically trim 2 percent of complete cases from the country-specific distribution of annual Euro amounts to exclude outliers that may have a disproportional influence on the imputation process. Thus, in addition to non-substantial answers ('Don't know' and 'Refusal'), we also impute outliers in the tails of the distribution of each monetary variable.

Notice that these preliminary transformations and standardizations of the raw data are carried out on the basis of common procedure for all waves of the panel by taking into account possible changes in the structure of the underlying CAPI questionnaires.

### **c) Logical constraints**

To avoid unreasonable combinations of the imputed data, we require that complete cases and imputed values satisfy a set of (logical) constraints on the ownership conditions regarding items that are logically related. For example, the ownership indicators of some financial assets (bonds, stocks and mutual funds) are set to zero (no ownership) if we know that the household/couple does not own a bank account. Other examples include the interests from financial assets that may have positive ownership indicator only if the household/couple owns a bank account and (possibly) bonds, stocks and mutual funds, or mortgage on the main residence that may have a positive ownership indicator only if the household owns the main residence.

### **d) Sequence of unfolding-bracket questions**

In case of initial nonresponse to open-ended questions on monetary variables, respondents are usually asked a sequence of unfolding-bracket (UB) questions to recover partial information on the missing monetary amounts. Specifically, the respondent is asked whether the amount is larger than, smaller than, or about equal to three predefined thresholds defined at the country level. The threshold of the first UB question is assigned randomly and the sequence of UB questions either stops or continues with the next threshold depending on the answer provided in the previous questions. The information collected in this way allows obtaining either an approximate point estimate (i.e. about equal to one of the three country-specific thresholds) or an interval estimate for the original missing amounts. The sequence of UB questions is considered as uninformative only if the respondent does not give a substantial answer to the first question of the sequence.

Our procedure to exploit the individual-specific information collected by the sequence of UB questions is as follows. For approximate point estimates, the missing amounts are imputed using the thresholds selected by the respondents throughout the sequence of UB questions. For interval estimates, the available UB information is instead combined with the additional information from logical constraints and percentiles of the country distribution to shrink as much as possible the individual-specific bounds placed on the missing values.

### **e) Aggregation**

To reduce the number of monetary variables that are imputed jointly, the extensive list of income, wealth and consumption expenditure items collected in each wave is condensed

into a smaller subset of ‘aggregated variables’. This simplification has both theoretical and practical implications. From a theoretical viewpoint, aggregation corresponds to imposing linear restrictions on the imputation model and this may undermine validity of the analyses performed on the basis of imputed data (see, for example, Rubin 1996). From a practical viewpoint, the scientific release contains imputations only for the chosen subset of aggregated variables, but not for their specific components.

Notice that, in aggregating several items, it is often the case that only some of them are missing. In this respect, we assume that an aggregated variable is missing if any of its components is missing. However, the partial information that may be available on each single item is not lost. This information is used to define lower bounds for the missing aggregated values.

Detailed information on the aggregation process across waves and countries is summarised through a set of tables in the appendix. Specifically, tables 5 and 6 show, respectively, the list & composition of aggregated variables and the country-specific composition of the pension variables in wave 1. Tables 7, 8, 9 and 10 show instead the list & composition of aggregated variables in waves 2, 4, 5 and 6, respectively. Notice that, although some variables have common names in the wave-specific imputation databases, the underlying components are not always the same because of possible changes in the underlying CAPI questionnaires. In such respect, it is important to stress that our attempt to harmonise as much as possible the information available in each wave of the SHARE panel does not guarantee that imputations are fully comparable across waves. This is especially true for wave 1 where there are major differences in the country-specific deviations and the structure of the CAPI questionnaire.

### **14.3.2 The imputation procedure**

#### **a) Hot-deck versus FCS imputations**

In each wave, missing values can be imputed by either (univariate) hot-deck method or (multivariate) fully conditional specification (FCS) method (Raghunathan et al. 2001; van Buuren et al. 2006; van Buuren 2007). Hot-deck imputations are performed in an early stage, separately by country, and according to a convenient order of the variables which accounts for branching and skip patterns in the questionnaires. We first impute socio-demographic characteristics such as age and education that are affected by very small fractions of missing values. These variables are then used as predictors in the imputation of the other variables. The set of predictors used in the hot-deck imputation stage typically included gender, age group, years of education and self-reported health. For some variables, we use a larger set of predictors. For example, we use the number of children when imputing the number of grandchildren and an indicator for being a patient in a hospital overnight during the last year when imputing health-related variables.

FCS imputations are performed separately by country and household type to account for heterogeneity across these different groups. The household types considered are:

- Sample 1: singles and third respondents;
- Sample 2: couples with both partners interviewed;
- Sample 3: all couples – with and without NRPs.

Notice that sample 2 is embedded into sample 3. This overlapping partitioning of the sample allows estimating total household income in couples with NRPs. The basic idea is that, in an early stage, we first impute total household income of couples belonging to sample 2. In a later stage, we use the couples with both partners interviewed as valid observations (donor observations) to impute total household income of couples with NRPs in sample 3.

The list of variables that are imputed jointly by the FCS method is country and sample-specific. For using this imputation method, we require that each monetary variable has at least 100 donor observations in sample 1 and 150 donor observations in samples 2 and 3. Monetary variables that do not satisfy this requirement are imputed first and then used as predictors in the imputation of the other monetary variables.

#### **b) FCS imputations: the imputation of a single monetary variable**

In the FCS method, the imputation of the  $j$ -th monetary variable is carried out by a two-part model that involves a probit model for ownership and a regression model for the amount conditional on ownership.

In the regression model to impute the amount, we apply variable transformations to account for skewness in the right tails of the underlying distributions. In particular, we apply the logarithm transformation for the variables that are strictly positive and the inverse hyperbolic sine transformation for the variables that may assume negative values (namely, income from self-employment, bank account, and value of own business).

The set of predictors used to impute the  $j$ -th variable is sample-specific. Specifically, there are two subsets of predictors. One is the subset of endogenous predictors that includes all variables that are imputed jointly except the  $j$ -th variable which is currently imputed. The other is the subset of exogenous predictors, which is also sample-specific. In sample 1, the exogenous predictors include gender, age, years of education, self-perceived health, number of children, number of chronic diseases, score of the numeracy test, employment status and willingness to answer of the designated respondent. In sample 2, we also have a similar set of exogenous predictors for the partner of the designated respondent. Finally, in sample 3, the exogenous predictors referring to the NRPs are confined to age and years of education only.

In the prediction step, imputations of missing amounts are always constrained to fall within the individual-level bounds that incorporate the available partial information on missing values.

#### **c) FCS imputations: imputation of total household income**

In the FCS imputations, special attention is given to total household income – especially in waves 2, 4, 5 and 6 where SHARE provides two alternative measures of this variable. The first measure (thinc) is obtained by aggregating at the household level all individual income

components, while the second (thinc2) is obtained from the one-shot question on monthly household income (hh017). Since the choice between these two alternative measures is not obvious, we let the users decide which of the two measures is more suitable for their research questions.

For our purposes, the availability of these two alternative measures is extremely useful because each measure contributes relevant information on the missing values of the other measure. A detailed description of the three-stage procedure used to impute the two measures of total household income can be found in De Luca et al. (2015).

#### **d) FCS imputations: iterations and convergence criterion**

The FCS procedure is repeated several times until the iterative algorithm gives a set of stationary distributions. Convergence is assessed by the Gelman-Rubin criterion (Gelman and Rubin 1992; Gelman et al. 2004) applied to the mean, the median and the 90th percentile of the distributions of the imputed variables. Convergence is also assessed for the generated variables such as total household income (thinc), total household expenditure (thexp) and household net worth (hnetw). After an initial set of 7 burn-in iterations, this criterion suggests that convergence is usually achieved for most of the statistics considered before reaching the maximum number of 30 iterations.

#### **e) Multiple imputations**

To allow data users to take into account the additional variability induced by the imputation process, **the imputation database of each wave includes five imputations** (or implicates) of the missing values. These multiple imputations are constructed using five independent replications of imputation procedure discussed above. Notice that neglecting the uncertainty of the imputed data (by selecting for instance only one of the five available replications) may lead to misleadingly precise estimates.

#### **f) Full list of imputed variables**

The full list of variables included in the imputation dataset of each wave is presented in table 11 of the appendix. For each imputed variable we also provide a flag variable (named as `variablename_f`) which summarises the status of the imputation process as illustrated in table 12 of the appendix.

### **14.3.3 Additional comments on the usage of imputations**

We find it important to stress that imputations are not the same as missing variable values. Although the use of imputed data is a quite common empirical strategy for handling missing data problems, validity of the underlying assumptions should not be taken for granted.

Validity of the so-called fill-in approach (i.e. the simple approach of fill-in the missing values with imputations) is indeed based on two important conditions. The first is that the model used to create the imputations is correctly specified, including the assumptions on the posited missing-data mechanism. The second is that the imputation model is congenial in the sense of Meng (1994), i.e. the imputation model cannot be more restrictive than the

model used to analyze the filled-in data. Uncongeniality may occur, for instance, when the model of interest and the imputation model are based either on different parametric assumptions or on different sets of explanatory variables. If these two conditions hold, then the use of imputed data may help to account for possible nonresponse bias and loss of precision due to item nonresponse. However, the fill-in approach may also lead to biased and inconsistent estimates whenever the imputation model is either incorrectly specified or uncongenial (see, for example, Dardanoni et al. 2011, 2015).

Judgments about the validity of these assumptions in the context of concrete applications remain a researcher's duty. To our experience, comparing the results from different approaches (such as complete data analysis, simple and generalized missing indicator approaches, pattern-mixture modeling and sample selection models) may give important hints about robustness of the empirical findings.

Please note that for some imputed variables information that was collected only in the baseline interview was transferred to later waves. This also concerns potentially time-variant information (e.g. numeracy). We strongly recommend checking the questionnaire routing and the regular interview data before deciding which imputed variables to include in which kind of analysis.

## References

Dardanoni V., Modica S., and Peracchi F. (2011). *Regression with imputed covariates: a generalized missing-indicator approach*. *Journal of Econometrics* 162: 362-368.

Dardanoni V., De Luca G., Modica S., and Peracchi F. (2015). *Model averaging estimation of generalized linear models with imputed covariates*. *Journal of Econometrics* 184: 452-463.

De Luca G., Celidoni M., and Trevisan E. (2015). *Item nonresponse and imputation strategies in SHARE Wave 5*, in Malter F. and Börsch-Supan A. (ed.), *SHARE Wave 5: Innovations & Methodology*, Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Gelman A., and Rubin D.B. (1992). *Inference from iterative simulation using multiple sequences*. *Statistical Science* 7: 457-511.

Gelman A., Carlin J.B., Stern H.S., and Rubin D.B. (2004). *Bayesian Data Analysis, Second Edition*. Boca Raton, FL: Chapman and Hall.

Meng X.L., (1994). *Multiple-imputation inferences with uncongenial sources of input*. *Statistical Science* 9: 538-558.

Raghunathan T.E., Lepkowski J.M., van Hoewyk J., and Solenberger P. (2001). *A multivariate technique for multiply imputing missing values using a sequence of regression models*. *Survey Methodology* 27: 85-95.

Rubin D.B. (1996). *Multiple imputations after 18+ years*. *Journal of the American Statistical Association* 91: 473-489.

Van Buuren S. (2007). *Multiple imputation of discrete and continuous data by fully conditional specification. Statistical Methods in Medical Research* 16: 219-242.

Van Buuren S., Brands J.P.L., Groothuis-Oudshoorn C.G.M., and Rubin D.B. (2006). *Fully conditional specification in multivariate imputation. Journal of Statistical Computation and Simulation* 76: 1049-1064.

## 14.4 gv\_isced

Education is one of the most diverse international variables. SHARE is using the International Standard Classification of Education (ISCED) which allows for the standardised reporting of education statistics according to an internationally agreed set of definitions and concepts (for further information see <http://uis.unesco.org/en/topic/international-standard-classification-education-isced>). The *gv\_isced* module contains the 1997 International Standard Classification of Education (ISCED-97). It is not only provided for respondents' educational level but also for respondents' children and former spouses' as well as interviewers (the latter only in wave 1). In wave 1 and 2 the education of only up to four selected children was asked. In 2011, a revision to ISCED was adopted by the UNESCO Member States. From wave 5 onwards both ISCED versions (97 and 2011) are provided in *gv\_isced*. Furthermore, waves 5 and 6 also contain the educational level of the respondents' parents. Please note that we collect information on education only in the baseline interview of a respondent indicated by the variable *mn101\_* in the technical variables module. For the respondents' children information on educational degrees is updated up to a certain age.

## 14.5 gv\_health

The *gv\_health* module contains a broad range of physical and mental health measures and indices. Therefore, it uses information from six different CAPI modules: physical health (PH), behavioural risks (BR), cognitive function (CF), mental health (MH), grip strength (GS), and walking speed (WS).

**Table 14: Generated Variables of *gv\_health***

Variable	Description
<b>Activities (AC) / Drop-off (in wave 1)</b>	
<i>casp</i>	quality of life and well-being
<b>Cognitive Function (CF)</b>	
<i>numeracy(2)</i>	numeracy score; in waves 4, 5 and 6 also <i>numeracy2</i> available
<i>orienti</i>	orientation to date, month, year and day of week
<i>cf008tot</i>	ten words list learning first trial (in waves 1 & 2 stored in CF module)
<i>cf016tot</i>	ten words list learning delayed recall (in waves 1 & 2 stored in CF module)



<b>Mental Health (MH)</b>	
<i>euro1-euro12</i>	variables forming the EURO-D scale: depression, pessimism, suicidality, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment, tearfulness
<i>eurod</i>	depression scale EURO-D
<i>eurodcat</i>	EURO-D caseness
<i>loneliness</i>	loneliness scale (in waves 5 and 6 only)
<b>Physical Health (PH)</b>	
<i>gali</i>	limitations with activities (GALI)
<i>spheu</i>	self-perceived health European version (wave 1 only)
<i>spheu2</i>	spheu - less than good health (wave 1 only)
<i>sphus</i>	self-perceived health US version
<i>sphus2</i>	sphus - less than very good health
<i>chronicw#</i>	number of chronic diseases
<i>chronic2w#</i>	2+ chronic diseases
<i>symptomsw#</i>	number of symptoms (not available in waves 5 and 6)
<i>symptoms2w#</i>	2+ symptoms (not available in waves 5 and 6)
<i>bmi</i>	body mass index (BMI)
<i>bmi2</i>	BMI categories
<i>mobility</i>	mobility, arm function and fine motor limitations
<i>mobilit2</i>	1+ mobility, arm function and fine motor limitations
<i>mobilit3</i>	3+ mobility, arm function and fine motor limitations
<i>adl</i>	number of limitations with activities of daily living (ADL)
<i>adl2</i>	1+ ADL limitations
<i>iadl</i>	number of limitations with instrumental activities of daily living (IADL)
<i>iadl2</i>	1+ IADL limitations
<b>Behavioural Risks (BR)</b>	
<i>cusmoke</i>	current smoking (not available in waves 5 and 6)
<i>drinkin2</i>	drinking more than 2 glasses of alcohol almost every day (wave 1 only)
<i>phactiv</i>	physical inactivity
<b>Walking Speed (WS)</b>	
<i>wspeed</i>	walking speed (waves 1 and 2 only)
<i>wspeed2</i>	walking speed: cut-off point (waves 1 and 2 only)
<b>Grip Strength (GS)</b>	
<i>maxgrip</i>	maximum of grip strength measures

## 14.6 gv\_housing

If the interview takes place in the house of the respondent, the interviewer does not ask for the type of accommodation in the HO module. Instead, the interviewer fills in this kind of information in the IV module him/herself.

For user convenience, we generated the variables *areabldgi*, *typebldgi* (*typebldgi6* in wave 6), *floorsbli* (not available in wave 6), and *nstepsi* that combine the data from the HO and the IV module. The information is stored in the *gv\_housing* module for all responding household members (regardless whether they are household respondents or not).

**Table 15: Generated Housing Variables**

Generated Variable	Description	Derived from	
<i>areabldgi</i>	Area of Building	<i>iv009_</i>	<i>ho037_</i>
<i>typebldgi(6)*</i>	Type of Building	<i>iv010_</i> (w6: <i>iv610_</i> )	<i>ho036_</i> (w6: <i>ho636_</i> )
<i>floorsbli*</i>	Number of Floors of Building	<i>iv011_</i>	<i>ho042_</i>
<i>nstepsi</i>	Number of Steps to Entrance	<i>iv012_</i>	<i>ho043_</i>

\* *typebldgi6*: In wave 6, answer option 8 in *iv610\_* and *ho636\_* differ from *iv010\_* and *ho036\_*

\* *floorsbli* is not available for wave 6

The Nomenclature of Territorial Units for Statistics (NUTS) is used to indicate in which territorial unit the household was located at time of sampling. In waves 1 and 2 NUTS version 2003 is used (*nuts#\_2003*). In waves 4 and 5 the suffix is referring to the NUTS classification of 2010 (*nuts#\_2010*), in wave 6 to the classification of 2015 (*nuts#\_2015*). Depending on the country-specific privacy legislations, not all NUTS levels are provided for every country. NUTS are **only available for the baseline samples** in the respective waves. Please be aware that the codes apply to the place where respondents lived at the time of sampling. For the first time in wave 5, regional codes analogous to NUTS are also available for Israel. They differentiate between seven regions (North, Haifa, Center, Tel-Aviv, Jerusalem, South, Judea and Samaria) that are similar in its population size and comparable to the level 1 NUTS regions in European countries.

## 14.7 gv\_networks

The generated “*gv\_networks*” module combines information from the SN, CH, DN, SP and FT modules and contains a new measure of social connectedness, a summary scale of the social network data that has been used previously in research (Litwin & Stoeckel, 2015).

The module stores variables that summarise information on the different attributes of respondent’s social networks. In wave 6, the variables additionally summarise panel information and provide information on each social network member. The derived summary variables are created only for respondents with data on all their social network members for that variable (e.g., mean proximity will have a value only for respondents with information on the proximity of all their social network members).

**Table 16: Generated Variables in Wave 4 *gv\_networks***

Variables w4	Description
<i>sizeofsocialnetwork</i>	Minimum: 0, Maximum: 7
<i>spousenet#</i>	Respondent's spouse in social network?
<i>famnet#</i> <sup>1</sup> , <i>childnet#</i> / <i>gchildnet#</i> / <i>siblingnet#</i> / <i>parentnet#</i> / <i>friendnet#</i> / <i>formalnet#</i> / <i>othernet#</i>	Amount / percentage of family members in a social network
<i>womennet#</i> / <i>mennet#</i>	Count of women/men in a social network
<i>*prx*</i>	Information on geographical distance to network members
<i>*contact*</i>	Information on contact frequencies with network members
<i>*close*</i>	Information on emotional closeness to network members
<i>*fin*</i> / <i>*gift*</i> / <i>*care*</i>	Information on given or received financial / personal help to / from network members
<i>sn_satisfaction</i>	Satisfaction with personal network (1-10)
<i>partner</i>	Relationship status, combined information from different waves. Is used to identify for whom the <i>spousenet#</i> variables do not apply because no partner was reported by the respondent.

<sup>1</sup> e.g. *famnet1* = number of family members; *famnet2* = dummy based on *famnet1*; *famnet3* = percentage of family members in social network.

**Table 17: Generated Variables in Wave 6 gv\_networks**

Variables	Description
<i>sn_size_w6</i>	SN size wave 6 (0-7)
<i>sn_size_w4</i>	SN size wave 4 (0-7)
<i>panel_change_size</i>	W6 sn size - w4 sn size (if both done)
<i>panel_status</i>	In which waves was the SN modules done
<i>panel_lost / new / continued</i>	Count of lost/new/continued SN members
<i>sn_scale</i>	Scale of social connectedness, a summary scale of the social network data (higher = higher connectedness)
<i>spousenet2</i>	Is there a spouse in the social network?
<i>famnet / childnet / siblingnet / parentnet / friendnet / formalnet / othernet</i>	Count of different role relationship categories in a social network
<i>womennet / mennet</i>	Count of women/men in a social network
<i>*prx*</i>	Information on geographical distance to network members
<i>*contact*</i>	Information on contact frequencies with network members
<i>*close*</i>	Information on emotional closeness to network members
<i>year_mean</i>	Average year of birth of network members
<i>*fin* / *gift* / *care*</i>	Information on given or received financial / personal help to / from network members
<i>sn_satisfaction</i>	Satisfaction with personal network (1-10)
<i>partner_var</i>	Relationship status, used to identify for whom the <i>spousenet2</i> variable does not apply because no partner was reported by the respondent.
<i>w4_sn_mentioned_again_X</i>	Was w4 SN member X mentioned again in w6?
<i>w6_sn_mentioned_before_X</i>	Was w6 SN member X mentioned before in w4?
<i>w6_sn_w4_position_X</i>	What was the position of w6 SN member X in the w4 SN?
<i>sn_person_X</i>	Was there an SN person x in the network?
<i>rel_X</i>	Relationship with each social network member
<i>gender_X</i>	Gender of each social network member
<i>prx_X</i>	Geographic proximity of each social network member
<i>contact_X</i>	Frequency of contact with each social network member
<i>close_X</i>	Emotional closeness of each social network member
<i>year_X</i>	Year of birth of each social network member
<i>occ_X</i>	Occupation of each social network member
<i>occ_det_X</i>	Detailed occupation of each social network member
<i>partner_X</i>	Partner status of each social network member
<i>partner_det_X</i>	Detailed partner status of each social network member
<i>*fin* / *gift* / *care*_sn_X</i>	Given or received financial / personal help to / from each social network member

## 14.8 gv\_exrates

This module contains currencies (also pre-Euro) and exchange rates for non-Euro countries for SHARE waves 1 to 6. Additionally, the module stores nominal exchange rates as well as exchange rates that adjust for purchasing power parity (ppp) for the years 2003 to 2015. The ppp-adjusted exchange rates allow researchers to estimate what the exchange rate between two currencies would have to be in order for the exchange to be at par with the purchasing power of different countries' currencies. PPP-adjustment to transform financial amounts into real ones – denominated in prices obtained in Germany in 2005 – can be done via:

$$\text{financial\_variable} * \text{nomx\#} / \text{pppx\#}.$$

Please note that the year # to be used depends on the time frame referred to by the financial variable you use. As in previous releases, the respective exchange rate variables are also stored in the wave-specific financial modules as well as in gv\_imputations.

## 14.9 gv\_job\_episodes\_panel

The Job Episodes Panel (JEP) is a generated dataset that contains the labour market status of each SHARELIFE respondent throughout her/his life. A detailed description of the methodology and assumptions underlying the construction of the dataset is available in the SHARE working paper 11-2013: "[Working life histories from SHARELIFE: a retrospective panel](#)", by Agar Brugiavini, Danilo Cavapozzi, Giacomo Pasini, and Elisabetta Trevisan.

## 14.10 gv\_grossnet

In wave 1 most income variables have been collected *before* taxes and social insurance contributions whereas in the following waves most income variables have been gathered *after* taxes and social contributions, to capture the notion of take-home pay. To make the different income measures comparable across waves and to facilitate longitudinal analyses, the module *gv\_grossnet* contains net income measures that are derived from reported gross incomes for SHARE wave 1. The instrument chosen to carry out this task is EUROMOD, the EU tax-benefit micro-simulation model.

A detailed description of the dataset and the method used is available in the [SHARE Working Paper 25-2016](#) by Bertoni et al. (2016).

## 14.11 gv\_isco

ISCO codes for respondents, their former partner's and their parents' occupation are only available for SHARE wave 1. SHARE uses the current (1988) International Standard Classification of Occupations (ISCO-88) by the International Labour Organization (ILO) (<http://www.ilo.org/public/english/bureau/stat/isco/isco88/>). Corresponding industries are classified according to the NACE Codes (Version 4 Rev. 1 1993), created by the European Union.

Please note that since release 5.0.0 the codes for “No answer” (“0009” for isco# and “79” for nace#) have been recoded to system missing values leading to different case numbers in the respective variables.

**Table 18: Variables Used for ISCO and NACE Coding**

	ISCO	NACE	
		employed	self-employed
<b>Respondent’s first job</b>	<i>ep016_1</i>	<i>ep018_1</i>	<i>ep023_1</i>
<b>Respondent’s second job</b>	<i>ep016_2</i>	<i>ep018_2</i>	<i>ep023_2</i>
<b>Respondent’s last job</b>	<i>ep052_</i>	<i>ep054_</i>	<i>ep060_</i>
<b>Former partner’s job</b>	<i>dn025</i>		
<b>Mother’s job</b>	<i>dn029_1</i>		
<b>Father’s job</b>	<i>dn029_2</i>		

## 14.12 gv\_ssw

Wave 4 includes a generated module containing two measures of individual accrued social security wealth. The module provides a set of internationally comparable measures of pension wealth computed for a large number of countries. The measures are lower bounds of first-pillar social security wealth (SSW) for individuals who declared in wave 4 to be either a retiree or a worker. The SSW of workers could be computed only for those countries and individuals participating in both wave 3 (SHARELIFE) and wave 4.

A key issue when constructing individual SSW for workers in wave 4 based on retrospective data from the Job Episode Panel (JEP; see chapter 14.9) is that the JEP provides information on net of taxes earnings, while pension rules are often computed on the basis of gross of taxes bases. Therefore, considering the difficulty of grossing up wages especially when they refer to periods far in the past, two versions of the SSW are provided. The two variables are *SSW\_nw* and *SSW\_gw* respectively. The former is based on net wages earned by individuals during their working career. The latter is based on their approximately grossed-up wages, and additionally takes into account minimum pension benefits whenever the individual is entitled to that benefit. Note that since no information from the JEP was required to compute the SSW for retirees, the two variables *SSW\_nw* and *SSW\_gw* are equal for this group.

Further information on the SSW measures is available in: Michele Belloni, Ludovico Carrino, Cristina Elisa Orso, Raluca Elena Buia, Danilo Cavapozzi, Giacomo Pasini and Agar Brugiavini (2016): [\*“Internationally comparable measures of individual Social Security Wealth in SHARE Wave 4”\*](#). SHARE Working Paper Series: 24-2016

### 14.13 gv\_deprivation

This module is available in wave 5 and contains three variables on material and social deprivation: *depmat*, *depsoc* and *depsev*.

*depmat* is an aggregate measure of material conditions using a set of 11 items that refer to two broad domains: the failure in the affordability of basic needs and financial difficulties.

**Table 19: Items Used for Material Deprivation Index (*depmat*)**

Item	Variable	Description
Meat	br033_	Eat meat, fish or chicken more often [ <i>than three times per week</i> ] because: <i>you cannot afford to eat it more often</i>
Fruit	br034_	Eat fruits or vegetables more often [ <i>than three times per week</i> ] because: <i>you cannot afford to eat it more often</i>
Groceries	co201_	Can your household afford to regularly buy necessary groceries and household supplies?
Holiday	co202_	Could your household afford to go for a week long holiday away from home at least once a year?
Expense	co206_	Could your household afford to pay an unexpected expense of [ <i>AffordExpenseAmount</i> ] without borrowing any money?
<i>In the last twelve months, to help you keep your living costs down, have you...</i>		
Clothing	co207_	... continued wearing clothing that was worn out because you could not afford replacement?
Shoes	co208_	... continued wearing shoes that were worn out because you could not afford replacement?
Heating	co209_	... put up with feeling cold to save heating costs?
Glasses	co213_	... gone without or not replaced glasses you needed because you could not afford new ones?
Dentist	co211_	... postponed visits to the dentist?
Doctor	hc114_	Was there a time in the past 12 months when you needed to see a doctor but could not because of cost?

*depsoc* is an index for measuring social deprivation. For this purpose 15 items were used. Some of them were newly included in the SHARE wave 5 questionnaire

**Table 20: Items Used for the Social Deprivation Index (*depsoc*)**

Item	Variable(s)	Description
Room	ho032_, hhsz	Less than one room per person in HH.
Literacy	cf001_, cf002_	Poor reading or writing skills.
IT skills	it003_	Poor computer skills or never used a computer.
Feeling part	hh022_	Not feeling part of the local area.
Vandalism	hh023_	Vandalism in the local area.
Clean area	hh024_	Local area not clean.
Help in area	hh025_	No helpful people in local area.
Bank access	hh027_	Difficult access to bank.
Shop access	hh028_	Difficult access to grocery shop.
Pharmacy access	hh030_	Difficult access to pharmacy.
Doctor	hc115_	Waiting too long to see a doctor.
Course	ac035_	Not attending any course in the past 12 months.
Organisation	ac035_	Not taking part in any organisation in the past 12 months.
Trust	ex026_	People cannot be trusted.
Isolation	ac016_	Feeling left out of things.

As most of the questions on material and social deprivation were asked at household level, information was imputed for the partner in the case of couple households. *depmat* and *depsoc* which can take values between 0 and 1 are therefore also available for partners. The variables are missing in case of missing information (including “Don’t know” and “Refusal”) in at least one of the included items. Note that both indices are weighted using hedonic weights.

*depsev* is a single two-dimensional indicator that identifies those with high levels of deprivation in each dimension. The threshold is the 75th percentile of the total distribution of each deprivation index. Individuals with deprivation measures placing them above the threshold in both dimensions are classified as being “severely deprived”.

For further information on this topic, please see: Börsch-Supan, A., T. Kneip, H. Litwin, M. Myck, G. Weber (Eds.) (2015). [\*Ageing in Europe - Supporting Policies for an Inclusive Society\*](#). Berlin: De Gruyter.

#### 14.14 gv\_children

In wave 6, information on the respondents’ children is collected in various parts of the SHARE questionnaire. The variables in the *gv\_children* module were generated in an attempt to make this information more easily accessible to SHARE users. The module combines information from the wave 6 CAPI modules CH, SN, SP and FT. Please be aware that the *gv\_children* variables are an aggregate of information from within wave 6 but not of information from previous waves. This means that not all information in the module is complete: some information, such as residence, marital and educational status or number of children of the respondents’ children is only collected in baseline interviews or if the



respondent mentions new children or some changes in this information in comparison to a previous interview. Thus, information that is not updated in wave 6 cannot be found in wave 6. A generated module that aggregates child information from all waves of SHARE is planned to be published in one of the upcoming data releases.

Up to 20 children can be mentioned by SHARE respondents, hence each type of information (gender, year of birth, etc.) consists of 20 variables (e.g. *ch\_gender\_1* – *ch\_gender\_20*). In addition, the CH module of the SHARE questionnaire is only completed by the family respondent. Where objective information was collected (e.g. gender or marital status of a child), the values of the child information were transferred from the family respondent to the cohabitating partner. The same approach is applied for information from the FT module: this module is only done by the financial respondent, and the information of some variables was transferred to the cohabitating partner. Rather subjective information (e.g. emotional closeness or frequency of personal contact to children) was not transferred and is thus missing for cohabitating partners. Furthermore, the variables in the dataset as well as their labels were assigned with “speaking” names in order to provide the user with additional information on how the variables were computed (e.g. “Contact to child 1, based on *ch014\_1* & *sn007\_1*”)

As for child education (*school\_education\_\** and *further\_education\_\**), the information can be cumbersome to use; the categories of educational degrees are country-specific, and the loop-in-loop structure of further education, where multiple degrees can be selected, highly inflates the number of variables. Therefore, we recommend the usage of the *gv\_iscd* module that summarises the educational attainments of children as well as respondents and their partners (see chapter 14.4).

Table 25 contains an overview of the *gv\_children* variables, their content, as well as the completeness of the information they provide.

**Table 21: Generated Variables in Wave 6 *gv\_children***

Variable name	Information complete within wave 6, values transferred to cohabitating partner
<i>ch_gender_*</i>	Gender of child *
<i>ch_yrbirth_*</i>	Year of birth of child *
<i>ch_relation_*</i>	Relation to child * (e.g. natural child, child of partner, adopted child, etc.)
<i>ch_occupation_*</i>	Occupational status of child *
<i>ch_fin_gave_*</i>	Financial help given to child *
<i>ch_fin_received_*</i>	Financial help received from child *
<i>ch_gift_received_*</i>	Financial gift received from child *
<i>ch_gift_gave_*</i>	Financial gift given to child *

Variable name	Wave 6 values transferred to cohabitating partner, information gathered in waves 1-5 can be missing
<i>ch_proximity_*</i>	Residence of child *
<i>ch_move_out_year_*</i>	Year when child * left the parental household
<i>ch_marital_status_*</i>	Marital status of child *
<i>ch_partner_status_*</i>	Partner status of child * (if not married/reg. partnership)
<i>ch_number_of_children_*</i>	Number of children of child *
<i>ch_yrbirth_youngest_child_*</i>	Year of birth of youngest child of child *
<i>ch_school_education_*</i>	Highest school leaving certificate of child *
<i>ch_further_education_ *_*</i>	Further educational degrees of child * (multiple answers possible)
Variable name	Wave 6 information, from respondent only (values not transferred to cohabitating partner)
<i>ch_contact_*</i>	Frequency of contact to child *
<i>ch_closeness_*</i>	Emotional closeness to child * (only asked for children mentioned in social network module)
<i>ch_babysit_*</i>	Did respondent look after child of child *?
<i>ch_outhh_receive_care_*</i>	Personal/practical help received from child * who lives outside the respondent's household
<i>ch_outhh_gave_care_*</i>	Personal/practical help given to child * who lives outside the respondent's household
<i>ch_hh_gave_care_*</i>	Personal/practical help given to child * who lives within the respondent's household
<i>ch_hh_receive_care_*</i>	Personal/practical help received from child * who lives within the respondent's household

### 14.15 gv\_linkage

Upon respondents' written consent, administrative data of the German Pension Insurance (DRV) can be linked to the survey data of the German sample of SHARE. In wave 3, all German respondents were asked for consent. In wave 5 and 6, all respondents who have not consented before or were new in the German SHARE sample were asked for consent. The generated data file "sharewX\_rel6-0-0\_gv\_linkage" summarises useful information about the availability of the data, e.g. who consented and whether and which administrative data are available. It can be used in the early stages of analysis to find out for which respondents the administrative data are available. To get access to the longitudinal administrative data, which provides detailed information on respondents' employment histories, the researcher has to apply directly at the data center of the DRV. Further information on access conditions as well as user guide and codebook for SHARE-RV is available at <http://www.share-project.org/special-data-sets/record-linkage-project.html>.

#### 14.16 gv\_dbs

The dried blood spots (DBS) samples SHARE collected in 12 countries are currently being analysed for biomarkers related to diseases and conditions that are typical for older people and/or influenced by lifestyle, for example cardiovascular diseases, cognitive decline, diabetes, or markers of stress and inflammation. More information can be found on the [SHARE website](#).

Even if the analyses of the dried blood spots are not yet completed some generated variables are already provided in this module. The most important one is *dbs\_values\_exp* (“Expected availability of laboratory results”). Results will only be available if (a) there is proof of written consent by the respondent, (b) the DBS sample is linkable to the CAPI interview via its barcode number, and (c) the DBS filter card contains enough blood material for at least one analysis. Given all these conditions are met, *dbs\_values\_exp*= 1. NOTE: not all collected DBS samples contain enough blood for the analyses of *all* biomarkers of interest. Even if *dbs\_values\_exp*=1, analytical values for only a subset of the above mentioned biomarkers may be available. Further variables in *gv\_dbs* are *spots\_nr* (“Number of blood spots collected”), and *spots\_co* (“Number of blood spots filling pre\_printed circle”). The latter indicates how many of the blood spots contain the amount of blood covering the size of the pre-printed circle (1 cm in diameter) on the blood collection card.

Responsible for database management and preparation of the releases:

Stephanie Stuck, Sabrina Zuber,  
Markus Kotte, Fabio Franzese  
Stefan Gruber, Tim Birkenbach

and all SHARE country team operators.

Contact: [info@share-project.org](mailto:info@share-project.org)

## 15 APPENDIX

**Table 1: Digital Object Identifier of the SHARE Datasets**

<b>Dataset</b>	<b>Release</b>	<b>Date</b>	<b>DOI</b>
<b>Wave 1</b>	6.0.0	31st March 2017	<a href="https://doi.org/10.6103/SHARE.w1.600">10.6103/SHARE.w1.600</a>
<b>Wave 2</b>	6.0.0	31st March 2017	<a href="https://doi.org/10.6103/SHARE.w2.600">10.6103/SHARE.w2.600</a>
<b>Wave 3 (SHARELIFE)</b>	6.0.0	31st March 2017	<a href="https://doi.org/10.6103/SHARE.w3.600">10.6103/SHARE.w3.600</a>
<b>Wave 4</b>	6.0.0	31st March 2017	<a href="https://doi.org/10.6103/SHARE.w4.600">10.6103/SHARE.w4.600</a>
<b>Wave 5</b>	6.0.0	31st March 2017	<a href="https://doi.org/10.6103/SHARE.w5.600">10.6103/SHARE.w5.600</a>
<b>Wave 6</b>	6.0.0	31st March 2017	<a href="https://doi.org/10.6103/SHARE.w6.600">10.6103/SHARE.w6.600</a>
<b>All Waves Coverscreen</b>	6.0.0	31st March 2017	<a href="https://doi.org/10.6103/SHARE.wXcvr.600">10.6103/SHARE.wXcvr.600</a>
<b>Longitudinal Weights</b>	6.0.0	31st March 2017	<a href="https://doi.org/10.6103/SHARE.wXweights.600">10.6103/SHARE.wXweights.600</a>
<b>Job Episodes Panel</b>	6.0.0	31st March 2017	<a href="https://doi.org/10.6103/SHARE.jep.600">10.6103/SHARE.jep.600</a>
<b><i>easy</i>SHARE</b>	5.0.0	27th July 2016	<a href="https://doi.org/10.6103/SHARE.easy.500">10.6103/SHARE.easy.500</a>
<b>SHARE-RV</b>	6.0.0	31st March 2017	<a href="https://doi.org/10.6103/SHARE.SHARE-RV.600">10.6103/SHARE.SHARE-RV.600</a>

**Table 2: SHARE Data Releases**

<b>Wave 1 &amp; Wave 2*</b>	
Release 1: April 28 <sup>th</sup> , 2005	
Release 2.0.0: June 19 <sup>th</sup> , 2007	
Release 2.0.1: July 5 <sup>th</sup> , 2007	
	Release 1.0.0: November 28 <sup>th</sup> , 2008
	Release 1.0.1: December 4 <sup>th</sup> , 2008
Release 2.2.0: August 19 <sup>th</sup> , 2009	
Release 2.3.0: November 13 <sup>th</sup> , 2009	
Release 2.3.1: July 28 <sup>th</sup> , 2010	
Release 2.4.0: March 17 <sup>th</sup> , 2011	
Release 2.5.0: May 11 <sup>th</sup> , 2011	
Release 2.6.0: November 29 <sup>th</sup> , 2013	
<b>Wave 3 (SHARELIFE)</b>	
Release 1.0.0: November 24 <sup>th</sup> , 2010	
<b>Wave 4</b>	
Release 1.0.0: November 30 <sup>th</sup> , 2012	
Release 1.1.1: March 28 <sup>th</sup> , 2013	
<b>Wave 5</b>	
Release 1.0.0: March 31 <sup>st</sup> , 2015	
<b>Release 5.0.0 of Waves 1 to 5</b>	
Release 5.0.0: May 10 <sup>th</sup> 2016	
<b>Release 6.0.0 of Waves 1 to 6</b>	
Release 6.0.0: March 31 <sup>st</sup> 2017	

\* combined releases of wave 1 and wave 2 from release 2.2.0 onwards

## **New in SHARE Wave 2**

### **New in release 2.6.0**

- Israel wave 2 data included
- Austrian interviews that could not be confirmed by the agency excluded

### **New in release 2.5.0**

- New imputations for Israel
- Newly computed cross-sectional and longitudinal weights
- Correction of children's ISCED mapping in wave 1
- New naming convention for double-looped variables in ep module w2
- Correction of variable finresp in cv\_r module in wave 1

### **New in release 2.4.0**

- Corrected imputations for non-Euro countries in waves 1 & 2
- Correction of erroneous values in as024e

### **New in release 2.3.1**

- New imputations for waves 1 & 2

### **New in release 2.3.0**

- Imputations for wave 2 and new imputations for wave 1
- Longitudinal weights and new structure of weights files
- Generated health variables added for wave 2 and revised for wave 1
- Some corrections in ph (w1 & w2), ep (w2), hc (w2) modules
- Correction of minor coding error in respondents' ISCED variable (w2)
- Greek vignettes (wave 2)
- Euro conversions for wave 1 ex module

### **New in release 2.2.0**

- Ireland included
- IDs and merging of data files
- Missing codes
- Naming of dummy variables and variables in drop-off and vignettes
- Coding of "other" citizenship and country of birth
- Israel: ISCO & NACE codes and additional modules
- Corrections of known problems in coding etc.

## **New in SHARE Wave 4**

- Four new countries: Estonia, Hungary, Portugal, and Slovenia
- New integrated social networks module (SN) with respective generated variables (see chapter **Fehler! Verweisquelle konnte nicht gefunden werden.**)

### **New in release 1.1.1**

- Corrected imputations
- Corrected weights
- NUTS variables included in the generated housing module
- Revised ISCED coding for Estonia
- Variable formats changed in the children module
- Additional drop-offs in Italy and Portugal

### **New in SHARE Wave 5**

- New country: Luxembourg
- Interviewer survey (see chapter 7)
- Parents' country of birth
- Mini-childhood module (see chapter **Fehler! Verweisquelle konnte nicht gefunden werden.**)
- IT module
- New social exclusion items (see chapter 13.9)

**Table 3: Overview of Additional Codes for Country of Birth and Citizenship**

Code	Citizenship / country of birth
1010	Congo
1011	Stateless
1012	Cypriote-American
1015	EU-Citizenship
1016	Argentinean-Italian
1017	Serbian-Bosnian
1018	Austrian-Italian-Czech
1019	American-Irish
1020	Galicia-Central Europe
1021	Italian-Croatian
1022	Italian-Slovenian
1023	Portuguese-Swiss
1024	Afghan-Turkish
1025	Turkish-Kurdish
1026	Italian-Austrian
1027	German-Italian
1028	British - Estonian
1029	Dutch-Czech
1110	Tunisian-French
1030	Former Territories of German Reich
1031	Former Eastern Territories of German Reich
1040	Kosovo
1050	Minor Asia
1060	Former Netherlands-East Indies
1070	Former Austria-Hungary
1080	Kurdistan (region)
1090	Borneo-Island
1095	Former Protectorate of Northern Rhodesia
1100	Chechnya
1101	German-Spanish
1103	Caucasus
1110	Tunisian-French
1130	French-German
1131	Italian-Uruguayan
1132	Mexican-Swiss



**Table 4: Drop-Off Correspondence in Waves 1 & 2**

Topic	Wave 1 (question)	Wave 1 (variable)	Wave 2 (question)	Wave 2 (variable)	Wave 2 CAPI (variable name)
	number of question in drop off	name of variable in data	number of question in drop off	name of variable in data	question correspondence: (1) = accurate (2) = almost equal (3) = related
Life satisfaction	1	q1	-		ac012 (3)
CASP-12*	2 a)	q2_a	-		ac014 (2)
	2 b)	q2_b	-		ac015 (2)
	2 c)	q2_c	-		ac016 (2)
	2 d)	q2_d	-		ac017 (2)
	2 e)	q2_e	-		ac018 (2)
	2 f)	q2_f	-		ac019 (2)
	2 g)	q2_g	-		ac020 (2)
	2 h)	q2_h	-		ac021 (2)
	2 i)	q2_i	-		ac022 (2)
	2 j)	q2_j	-		ac023 (2)
	2 k)	q2_k	-		ac024 (2)
	2 l)	q2_l	-		ac025 (2)
LOT-R	3 a)	q3_a	1 a)	q3_a	
(Life Orientation Test:	3 b)	q3_b	1 b)	q3_b	
pessimism/optimism)	3 c)	q3_c	1 c)	q3_c	
	3 d)	q3_d	1 d)	q3_d	
	3 e)	q3_e	1 e)	q3_e	
	3 f)	q3_f	1 f)	q3_f	
	3 g)	q3_g	1 g)	q3_g	
Depression/feelings	4a)	q4_a	-		ac027(3)
(CES_D	4b)	q4_a	-		ac028(3)
	4c)	q4_a	-		ac029(3)
	4d)	q4_a	-		ac030(3)
	4e)	q4_a	-		ac031(3)
	4f)	q4_a	-		
	g)	q4_a	-		ac032(3)
	4h)	q4_a	-		ac033(3)
	4i)	q4_a	-		
	4j)	q4_a	-		ac034(3)
	4k)	q4_a	-		
	4l)	q4_a	-		

Topic	Wave 1 (question)	Wave 1 (variable)	Wave 2 (question)	Wave 2 (variable)	Wave 2 CAPI (variable name)
	4m)	q4_a	-		
	4n)	q4_a	-		
Expectations of others	5 a)	q5_a	-		
	5 b)	q5_b	-		
	5 c)	q5_c	-		
	5 d)	q5_d	-		
Family duties	6 a)	q6_a	2 a)	q6_a	
	6 b)	q6_b	2 b)	q6_b	
	6 c)	q6_c	2 c)	q6_c	
	6 d)	q6_d	2 d)	q6_d	
Family/state responsibility	7 a)	q7_a	3 a)	q7_a	
	7 b)	q7_b	3 b)	q7_b	
	7 c)	q7_c	3 c)	q7_c	
Conflicts with others	8 a)	q8_a	4 a)	q8_a	
	8 b)	q8_b	4 b)	q8_b	
	8 c)	q8_c	4 c)	q8_c	
	8 d)	q8_d	4 d)	q8_d	
	8 e)	q8_e	4 e)	q8_e	
	8 f)	q8_f	4 f)	q8_f	
	9	q9	5	q9	
Ever lived with partner:	10	q10	6	q10	
Responsibility for different tasks	11 a)	q11_a	7 a)	q11_a	
	11 b)	q11_b	7 b)	q11_b	
	11 c)	q11_c	7 c)	q11_c	
	11 d)	q11_d	7 d)	q11_d	
Health	12	q12	-		
General practitioner/ usual source of care	13 a)	q13_a	8 a)	q13_a	
questions and checks	13 b)	q13_b	8 b)	q13_b	
	13 c)	q13_c	8 c)	q13_c	
	13 d)	q13_d	8 d)	q13_d	
	13 e)	q13_e			
	13 f)	q13_f	8 e)	q13_f	
Health	-		9 a)	q39_a	
Talk about problems	-		9 b)	q39_b	
	-		9 c)	q39_c	
	-		9 d)	q39_d	
	-		9 e)	q39_e	

Topic	Wave 1 (question)	Wave 1 (variable)	Wave 2 (question)	Wave 2 (variable)	Wave 2 CAPI (variable name)
Health Explanations/listening	-		10 a)	q40_a	
	-		10 b)	q40_b	
	-		10 c)	q40_c	
Health Prevention	14	q14	-		
	15	q15	-		
	16	q16	12	q16	
	17	q17	13	q17	
	18	q18	-		
	19	q19	-		
	20	q29	-		
	21	q21	-		
			11 a)	q41_a	
			11 b)	q41_b	
			11 c)	q41_c	
			11 d)	q41_d	
Health Joint pain	22	q22	-		
	23	q23	-		
	24	q24	-		
	25	q25	-		
	26	q26	-		
	27	q27	-		
	28 a)	q28_a	-		
	28 b)	q28_b	-		
	28 c)	q28_c	-		
	29 a)	q29_a	-		
	29 b)	q29_b	-		
	29 c)	q29_c	-		
Accommodation	30 a)	q30_a	-		ho050(1)
	30 b)	q30_b	-		ho051(1)
	30 c)	q30_c	-		ho052(1)
	30 d)	q30_d	-		ho053(1)
	30 e)	q30_e	-		ho054(1)
	30 f)	q30_f	-		ho055(1)
	31 a)	q31_a	-		
	31 b)	q31_b	-		
	31 c)	q31_c	-		
	31 d)	q31_d	-		

Topic	Wave 1 (question)	Wave 1 (variable)	Wave 2 (question)	Wave 2 (variable)	Wave 2 CAPI (variable name)
Area of accommodation	32 a)	q32_a	-		ho056(1)
	32 b)	q32_b	-		ho057(1)
	32 c)	q32_c	-		ho058(1)
	32 d)	q32_d	-		ho059(1)
Pet animals	33 a)	q33_a	14 a)	q33_a	
	33 b)	q33_b	14 b)	q33_b	
	33 c)	q33_c	14 c)	q33_c	
	33 d)	q33_d	14 d)	q33_d	
	33 e)	q33_e	14 e)	q33_e	
	33 f)	q33_f	-		
Religion	34	q34	-		
	35	q35	-		ex029 (1)
	36	q36	-		

\* For more information about CASP see: Hyde, M. (2003) A measure of quality of life in early old age: The theory, development and properties of a needs satisfaction model (CASP-19). *Aging and mental health*, 7 (3), 186-194

**Table 5: Composition of aggregated imputation variables in wave 1**

Aggregates	Components	Description
yreg1	EP094_1	Life insurance payment
	EP094_2	Private annuity/private personal pension
	EP094_3	Private health insurance payment
yreg2	EP094_4	Alimony
	EP094_5	Regular payments from charities
aftgiv	FT004_1	Financial gift given 250€ or more: first amount
	FT004_2	Financial gift given 250€ or more: second amount
	FT004_3	Financial gift given 250€ or more: third amount
aftrec	FT011_1	Financial gift received 250€ or more: first amount
	FT011_2	Financial gift received 250€ or more: second amount
	FT011_3	Financial gift received 250€ or more: third amount
aftinh	FT018_1	Inheritance/gift received: first amount
	FT018_2	Inheritance/gift received: second amount
	FT018_3	Inheritance/gift received: third amount
	FT018_4	Inheritance/gift received: fourth amount
	FT018_5	Inheritance/gift received: fifth amount
rhre	HO005	Amount rent paid
	HO008	Other home-related expenditures: charges and services
yaohm	HH002	Total income received by other household members
	HH011	Other monetary benefits received: housing allowances, child benefits, poverty relief, etc.
bsmf	AS007	Government/corporate bonds
	AS011	Stocks
	AS017	Mutual funds
ybabsmf	AS005	Interest from bank accounts
	AS009	Interest from government/corporate bonds
	AS015	Dividends from stocks
	AS058	Interests or dividends from mutual funds
slti	AS021	Individual retirement accounts from respondent
	AS024	Individual retirement accounts from spouse/partner
	AS027	Contractual savings
	AS030	Whole life insurance holdings

Notes: The country-specific composition of the aggregates ypen1, ypen2 and ypen3 is presented in Table 6.

**Table 6: Country-specific composition of imputed pension variables in wave 1**

Country	Aggregates					
	ypen1	ypen2	ypen3	ypen4	ypen5	ypen6
AT	EP078_1	EP078_8	EP078_3	EP078_4	EP078_6	
	EP078_2	EP078_9				
	EP078_5	EP078_10				
	EP078_7	EP078_11				
BE	EP078_1	EP078_8	EP078_3 EP078_6	EP078_4		
	EP078_2	EP078_9				
	EP078_5	EP078_10				
	EP078_7	EP078_11				
CH	EP078_1	EP078_8	EP078_3 EP078_10	EP078_4	EP078_6	
	EP078_2	EP078_9				
	EP078_5	EP078_11				
DE	EP078_1	EP078_8	EP078_3	EP078_4		
	EP078_2	EP078_10				
	EP078_5	EP078_11				
	EP078_7					
DK	EP078_1	EP078_8	EP078_6	EP078_4		EP078_3
	EP078_2	EP078_9				
	EP078_5	EP078_10				
	EP078_7	EP078_11				
ES	EP078_1	EP078_8	EP078_3	EP078_4		EP078_6
	EP078_2	EP078_9				
	EP078_5	EP078_10				
	EP078_7	EP078_11				
FR	EP078_1	EP078_8	EP078_6 EP078_10	EP078_4		
	EP078_2	EP078_9				
	EP078_5	EP078_11				
	EP078_7					
GR	EP078_1	EP078_8	EP078_3	EP078_4		EP078_6
	EP078_2	EP078_10				
	EP078_5	EP078_11				
	EP078_7					
IL	EP078_1	EP078_8	EP078_3 EP078_6	EP078_4		
	EP078_2	EP078_9				
	EP078_5	EP078_10				
	EP078_7	EP078_11				

**Table 6: Country-specific composition of imputed pension variables in wave 1 (continued)**

Country	Aggregates					
	ypen1	ypen2	ypen3	ypen4	ypen5	ypen6
IT	EP078_1	EP078_8				
	EP078_2	EP078_9				
	EP078_5	EP078_10	EP078_3	EP078_4		EP078_6
	EP078_7	EP078_11				
NL	EP078_1					
	EP078_2					
	EP078_5	EP078_10				
	EP078_7	EP078_11	EP078_3	EP078_4	EP078_6	
	EP078_8					
	EP078_9					
SE		EP078_4				
	EP078_1	EP078_5				
	EP078_3	EP078_6	EP078_2	EP078_9		EP078_10
		EP078_7				
		EP078_8				

Notes: A description of the original variables EP078\_j (j=1,...,11) can be found in the SHARE questionnaire of wave 1. The set of 11 original variables does not necessarily apply to all countries. For each country, the table shows only the set of original variables with a valid positive amount in the wave 1 raw data.

**Table 7: Composition of aggregated imputation variables in wave 2**

Aggregates	Components	Description
ypen1 <sup>(a)</sup>	EP078_1	RP - Public old age pension
	EP078_2	RP - Public old age supplementary pension/public old age second pension
	EP078_3	RP - Public early retirement/pre-retirement pension
	EP078_7	RP - Main public survivor pension from spouse/partner
	EP078_8	RP - Secondary public survivor pension from spouse/partner
	EP078_9	RP - Public war pension
ylsp1 <sup>(b)</sup>	EP082_1	LSP - Public old age pension
	EP082_2	LSP - Public old age supplementary pension/public old age second pension
	EP082_3	LSP - Public early retirement/pre-retirement pension
	EP082_7	LSP - Main public survivor pension from spouse/partner
	EP082_8	LSP - Secondary public survivor pension from spouse/partner
	EP082_9	LSP - Public war pension
ypen2 <sup>(c)</sup>	EP078_11	RP - Occupational old age pension from last job
	EP078_12	RP - Occupational old age pension from a second job
	EP078_13	RP - Occupational old age pension from a third job
	EP078_14	RP - Occupational early retirement pension
	EP078_15	RP - Occupational disability or invalidity insurance
	EP078_16	RP - Occupational survivor pension from spouse/partner's job
ylsp2 <sup>(d)</sup>	EP082_11	LSP - Occupational old age pension from last job
	EP082_12	LSP - Occupational old age pension from a second job
	EP082_13	LSP - Occupational old age pension from a third job
	EP082_14	LSP - Occupational early retirement pension
	EP082_15	LSP - Occupational disability or invalidity insurance
	EP082_16	LSP - Occupational survivor pension from spouse/partner's job
ypen36	EP078_4	RP - Main public disability insurance pension/sickness benefits
	EP078_5	RP - Secondary public disability insurance pension/sickness benefits
ylsp36	EP082_4	LSP - Main public disability insurance pension/sickness benefits
	EP082_5	LSP - Secondary public disability insurance pension/sickness benefits
yreg1	EP094_1	RP - Life insurance
	EP094_2	RP - Private annuity/private personal pension
	EP094_5	RP - Long-term care insurance from private insurance company
ylsr1	EP209_1	LSP - Life insurance
	EP209_2	LSP - Private annuity/private personal pension
	EP209_5 <sup>(e)</sup>	LSP - Long-term care insurance from private insurance company



**Table 7: Composition of aggregated imputation variables in wave 2 (continued)**

Aggregates	Components	Description
yreg2	EP094_3 EP094_4	RP - Alimony RP - Charities
ylsr2	EP209_3 EP209_4 <sup>(f)</sup>	LSP - Alimony LSP - Charities
aftgiv	FT004_1 FT004_2 FT004_3	Financial gift given 250€ or more: first amount Financial gift given 250€ or more: second amount Financial gift given 250€ or more: third amount
aftrec	FT011_1 FT011_2 FT011_3	Financial gift received 250€ or more: first amount Financial gift received 250€ or more: second amount Financial gift received 250€ or more: third amount
aftinh	FT018_1 FT018_2 FT018_3 FT018_4 FT018_5	Inheritance/gift received: first amount Inheritance/gift received: second amount Inheritance/gift received: third amount Inheritance/gift received: fourth amount Inheritance/gift received: fifth amount
rhre	HO005 <sup>(g)</sup> HO008 <sup>(g)</sup>	Amount rent paid Other home-related expenditures: charges and services
yaohm	HH002 <sup>(g)</sup> HH011 <sup>(g)</sup>	Total income received by other household members Other monetary benefits received: housing allowances, child benefits, poverty relief , etc.
bsmf	AS007 AS011 AS017	Government/corporate bonds Stocks Mutual funds
ybabsmf	AS005 AS009 AS015 AS058	Interest from bank accounts Interest from government/corporate bonds Dividends from stocks Interests or dividends from mutual funds
slti	AS021 AS024 AS027 AS030	Individual retirement accounts from respondent Individual retirement accounts from spouse/partner Contractual savings Whole life insurance holdings

Notes: RP denotes 'regular payments'. LSP denotes 'Lump-sum payments'.

(a) In NL, ypen1 also includes: EP078\_11, EP078\_12, EP078\_13, and EP078\_14.

(b) In NL, ylspl also includes: EP082\_11, EP082\_12, EP082\_13, and EP082\_14.

(c) In NL, ypen2 includes only: EP078\_15 and EP078\_16.

(d) In NL, ylspl2 includes only: EP082\_15 and EP082\_16.

(e) The item EP209\_5 is excluded from ylsr1 due to lack of valid observations.

(f) The item EP209\_4 is excluded from ylsr2 due to lack of valid observations.

(g) Questions asked only to respondents living in private households.

**Table 8: Composition of aggregated imputation variables in wave 4**

Aggregates	Components	Description
ypen1 <sup>(a)</sup>	EP078_1	RP – Public old age pension
	EP078_2	RP – Public old age supplementary pension/public old age second pension
	EP078_3	RP – Public early retirement/pre-retirement pension
	EP078_7	RP – Main public survivor pension from spouse/partner
	EP078_8	RP – Secondary public survivor pension from spouse/partner
	EP078_9	RP – Public war pension
ylsp1 <sup>(b)</sup>	EP082_1	LSP – Public old age pension
	EP082_2	LSP – Public old age supplementary pension/public old age second pension
	EP082_3	LSP – Public early retirement/pre-retirement pension
	EP082_7	LSP – Main public survivor pension from spouse/partner
	EP082_8	LSP – Secondary public survivor pension from spouse/partner
	EP082_9	LSP – Public war pension
ypen2 <sup>(c)</sup>	EP078_11	RP – Occupational old age pension from last job
	EP078_12	RP – Occupational old age pension from a second job
	EP078_13	RP – Occupational old age pension from a third job
	EP078_14	RP – Occupational early retirement pension
	EP078_15	RP – Occupational disability or invalidity insurance
	EP078_16	RP – Occupational survivor pension from spouse/partner's job
ylsp2 <sup>(d)</sup>	EP082_11	LSP – Occupational old age pension from last job
	EP082_12	LSP – Occupational old age pension from a second job
	EP082_13	LSP – Occupational old age pension from a third job
	EP082_14	LSP – Occupational early retirement pension
	EP082_15	LSP – Occupational disability or invalidity insurance
	EP082_16	LSP – Occupational survivor pension from spouse/partner's job
ypen36	EP078_4	RP – Main public disability insurance pension/sickness benefits
	EP078_5	RP – Secondary public disability insurance pension/sickness benefits
ylsp36	EP082_4	LSP – Main public disability insurance pension/sickness benefits
	EP082_5	LSP – Secondary public disability insurance pension/sickness benefits
yreg1	EP094_1	RP – Life insurance
	EP094_2	RP – Private annuity/private personal pension
	EP094_5	RP – Long-term care insurance from private insurance company
ylsr1	EP209_1	LSP – Life insurance
	EP209_2	LSP – Private annuity/private personal pension
	EP209_5	LSP – Long-term care insurance from private insurance company

**Table 8: Composition of aggregated imputation variables in wave 4 (continued)**

Aggregates	Components	Description
yreg2	EP094_3 EP094_4	RP – Alimony RP – Charities
ylsr2	EP209_3 EP209_4	LSP – Alimony LSP – Charities
rhre	HO005 <sup>(e)</sup> HO008 <sup>(e)</sup> HO065 <sup>(f)</sup>	Amount rent paid Other home-related expenditures: charges and services Out of pocket payment for nursing home accommodation
ysrent	HO030 <sup>(e)</sup> HO074 <sup>(e)</sup>	Income from rent of secondary/holiday homes and other real estate Income from sublet of accommodation
yaohm	HH002 <sup>(e)</sup> HH011 <sup>(e)</sup>	Total income received by other household members Other monetary benefits received: housing allowances, child benefits, poverty relief , etc.
bsmf	AS007 AS011 AS017	Government/corporate bonds Stocks Mutual funds
slti	AS021 AS024 AS027 <sup>(g)</sup> AS030	Individual retirement accounts from respondent Individual retirement accounts from partner Contractual savings Whole life insurance holdings

Notes: RP denotes 'regular payments'. LSP denotes 'Lump-sum payments'.

(a) In NL, ypen1 also includes: EP078\_11, EP078\_12, EP078\_13, and EP078\_14.

(b) In NL, ylsp1 also includes: EP082\_11, EP082\_12, EP082\_13, and EP082\_14.

(c) In NL, ypen2 includes only: EP078\_15 and EP078\_16.

(d) In NL, ylsp2 includes only: EP082\_15 and EP082\_16.

(e) Questions asked only to respondents living in private households.

(f) Questions asked only to respondents living in nursing home.

(g) Question not asked in NL and EE.

**Table 9: Composition of aggregated imputation variables in wave 5**

Aggregates	Components	Description
ypen1 <sup>(a)</sup>	EP078_1	RP – Public old age pension
	EP078_2	RP – Public old age supplementary pension/public old age second pension
	EP078_3	RP – Public early retirement/pre-retirement pension
	EP078_7	RP – Main public survivor pension from spouse/partner
	EP078_8	RP – Secondary public survivor pension from spouse/partner
	EP078_9	RP – Public war pension
ylsp1 <sup>(b)</sup>	EP082_1	LSP – Public old age pension
	EP082_2	LSP – Public old age supplementary pension/public old age second pension
	EP082_3	LSP – Public early retirement/pre-retirement pension
	EP082_7	LSP – Main public survivor pension from spouse/partner
	EP082_8	LSP – Secondary public survivor pension from spouse/partner
	EP082_9	LSP – Public war pension
ypen2 <sup>(c)</sup>	EP078_11	RP – Occupational old age pension from last job
	EP078_12	RP – Occupational old age pension from a second job
	EP078_13	RP – Occupational old age pension from a third job
	EP078_14	RP – Occupational early retirement pension
	EP078_15	RP – Occupational disability or invalidity insurance
	EP078_16	RP – Occupational survivor pension from spouse/partner's job
ylsp2 <sup>(d)</sup>	EP082_11	LSP – Occupational old age pension from last job
	EP082_12	LSP – Occupational old age pension from a second job
	EP082_13	LSP – Occupational old age pension from a third job
	EP082_14	LSP – Occupational early retirement pension
	EP082_15	LSP – Occupational disability or invalidity insurance
	EP082_16	LSP – Occupational survivor pension from spouse/partner's job
ypen36	EP078_4	RP – Main public disability insurance pension/sickness benefits
	EP078_5	RP – Secondary public disability insurance pension/sickness benefits
ylsp36	EP082_4	LSP – Main public disability insurance pension/sickness benefits
	EP082_5	LSP – Secondary public disability insurance pension/sickness benefits
yreg1	EP094_1	RP – Life insurance
	EP094_2	RP – Private annuity/private personal pension
	EP094_5	RP – Long-term care insurance from private insurance company
ylsr1	EP209_1	LSP – Life insurance
	EP209_2	LSP – Private annuity/private personal pension
	EP209_5	LSP – Long-term care insurance from private insurance company
yreg2	EP094_3	RP – Alimony
	EP094_4	RP – Charities
ylsr2	EP209_3	LSP – Alimony
	EP209_4	LSP – Charities

**Table 9: Composition of aggregated imputation variables in wave 5 (continued)**

Aggregates	Components	Description
rhre	HO005 <sup>(e)</sup>	Amount rent paid
	HO008 <sup>(e)</sup>	Other home-related expenditures: charges and services
	HO065 <sup>(f)</sup>	Out of pocket payment for nursing home accommodation
	HO008 <sup>(f)</sup>	Other home-related expenditures: charges and services in nursing home
ores	HO027 <sup>(e)</sup>	Value of secondary/holiday homes and other real estate
	HO076 <sup>(f)</sup>	Value of secondary/holiday homes and other real estate (respondents in nursing home)
ysrent	HO030 <sup>(e)</sup>	Income from rent of secondary/holiday homes and other real estate
	HO074 <sup>(e)</sup>	Income from sublet of accommodation
	HO078 <sup>(f)</sup>	Income from rent of secondary/holiday homes and other real estate (respondents in nursing home)
yaohm	HH002 <sup>(e)</sup>	Total income received by other household members
	HH011 <sup>(e)</sup>	Other monetary benefits received: housing allowances, child benefits, poverty relief , etc.
bsmf	AS007	Government/corporate bonds
	AS011	Stocks
	AS017	Mutual funds
slti	AS021	Individual retirement accounts from respondent
	AS024	Individual retirement accounts from partner
	AS027 <sup>(g)</sup>	Contractual savings
	AS030	Whole life insurance holdings
outpa	HC083	Out of pocket payment for doctor visits
	HC093	Out of pocket payment for dentist care
nurs	HC097 <sup>(e)</sup>	Out of pocket payment for nursing home accommodation
	HC129	Out of pocket payment for personal care, domestic tasks, meals on wheels, other help

Notes: RP denotes 'regular payments'. LSP denotes 'Lump-sum payments'.

(a) In NL, ypen1 also includes: EP078\_11, EP078\_12, EP078\_13, and EP078\_14. In LU, ypen1 includes only: EP078\_1, EP078\_3, EP078\_7, and EP078\_9.

(b) In NL, yls1 also includes: EP082\_11, EP082\_12, EP082\_13, and EP082\_14. In LU, yls1 includes only: EP082\_1, EP082\_3, EP082\_7, and EP082\_9.

(c) In NL, ypen2 includes only: EP078\_15 and EP078\_16. In LU, ypen2 also includes: EP078\_2 and EP078\_8.

(d) In NL, yls2 includes only: EP082\_15 and EP082\_16. In LU, yls2 also includes: EP082\_2 and EP082\_8.

(e) Questions asked only to respondents living in private households.

(f) Questions asked only to respondents living in nursing home.

(g) Question not asked in NL and EE.

**Table 10: Composition of aggregated imputation variables in wave 6**

Aggregates	Components	Description
ypen1 <sup>(a)</sup>	EP078_1	RP – Public old age pension
	EP078_2	RP – Public old age supplementary pension/public old age second pension
	EP078_3	RP – Public early retirement/pre-retirement pension
	EP078_9	RP – Main public survivor pension from spouse/partner
	EP078_10	RP – Secondary public survivor pension from spouse/partner
	EP078_11	RP – Public war pension
ylsp1 <sup>(b)</sup>	EP082_1	LSP – Public old age pension
	EP082_2	LSP – Public old age supplementary pension/public old age second pension
	EP082_3	LSP – Public early retirement/pre-retirement pension
	EP082_9	LSP – Main public survivor pension from spouse/partner
	EP082_10	LSP – Secondary public survivor pension from spouse/partner
	EP082_11	LSP – Public war pension
ypen3	EP078_5	RP – Main public disability insurance pension
	EP078_6	RP – Secondary public disability insurance pension
ylsp3 <sup>(c)</sup>	EP082_5	LSP – Main public disability insurance pension
	EP082_6	LSP – Secondary public disability insurance pension
ypen5	EP078_12	RP – Public long-term care insurance
	EP078_13	RP – Social assistance
ylsp5	EP082_12	LSP – Public long-term care insurance
	EP082_13	LSP – Social assistance
ypen6	EP078_4	RP – Main public sickness benefits
	EP078_7	RP – Secondary public sickness benefits
ylsp6	EP082_4	LSP – Main public sickness benefits
	EP082_7	LSP – Secondary public sickness benefits
yreg1	EP094_1	RP – Life insurance from a private insurance company
	EP094_2	RP – Private annuity/private personal pension
	EP094_5	RP – Long-term care insurance from private insurance company
ylsr1	EP209_1	LSP – Life insurance from a private insurance company
	EP209_2	LSP – Private annuity/private personal pension
	EP209_5	LSP – Long-term care insurance from private insurance company
yreg2	EP094_3	RP – Alimony
	EP094_4	RP – Charities
ylsr2	EP209_3	LSP – Alimony
	EP209_4	LSP – Charities

**Table 10: Composition of aggregated imputation variables in wave 6 (continued)**

Aggregates	Components	Description
rhre	HO005 <sup>(c)</sup>	Amount rent paid
	HO008 <sup>(c)</sup>	Other home-related expenditures: charges and services
	HO065 <sup>(d)</sup>	Out of pocket payment for nursing home accommodation
ores	HO027 <sup>(c)</sup>	Value of secondary/holiday homes and other real estate
	HO076 <sup>(d)</sup>	Value of secondary/holiday homes and other real estate (respondents in nursing home)
ysrent	HO030 <sup>(c)</sup>	Income from rent of secondary/holiday homes and other real estate
	HO074 <sup>(c)</sup>	Income from sublet of accommodation
	HO078 <sup>(d)</sup>	Income from rent of secondary/holiday homes and other real estate (respondents in nursing home)
bsmf	AS007	Government/corporate bonds
	AS011	Stocks
	AS017	Mutual funds
slti	AS021	Individual retirement accounts from respondent
	AS024	Individual retirement accounts from partner
	AS027 <sup>(e)</sup>	Contractual savings
	AS030	Whole life insurance holdings
outpa6	HC083	Out of pocket payment for doctor visits
	HC093	Out of pocket payment for dentist care
nurs	HC097 <sup>(e)</sup>	Out of pocket payment for nursing home accommodation
	HC129	Out of pocket payment for personal care, domestic tasks, meals on wheels, other help

Notes: RP denotes 'regular payments'. LSP denotes 'Lump-sum payments'.

(a) In SE, ypen1 includes only: EP078\_1, and EP078\_9. In PL and LU, ypen1 includes only: EP078\_1, EP078\_3, EP078\_9 and EP078\_11.

(b) In SE, ylspl includes only: EP082\_1, and EP082\_9. In PL and LU, ylspl includes only: EP082\_1, EP082\_3, EP082\_9, and EP082\_11.

(c) Questions asked only to respondents living in private households.

(d) Questions asked only to respondents living in nursing home.

(e) Question not asked in NL and EE.

**Table 11: List of variables in gv\_imputations**

Variable	Description	Questionnaire	Waves
implicat	Implicat number		1 2 4 5 6
htype	Household type		1 2 4 5 6
fam_resp	Family respondent	MN006_	1 2 4 5 6
fin_resp	Financial respondent	MN007_	1 2 4 5 6
hou_resp	Household respondent	MN008_	1 2 4 5 6
exrate	Exchange rate		1 2 4 5 6
nursinghome	Living in nursing home	MN024_	1 2 4 5 6
perho	Percentage of house owned		4 5 6
single	Single		1 2 4 5 6
couple	Couple		1 2 4 5 6
partner	Partner in the couple		1 2 4 5 6
p_nrp	Partner of non-responding partner		1 2 4 5 6
sample1	Imputation sample for single		1 2 4 5 6
sample2	Imputation sample for couples with two partners interviewed		1 2 4 5 6
sample3	Imputation sample for all couples		1 2 4 5 6
inpat	Out-of-pocket payment for inpatient care	HC095	1 2 5 6
outpa	Out-of-pocket payment for outpatient care	HC083	1 2 5 6
drugs	Out-of-pocket payment for drugs	HC089	1 2 5 6
nurs	Out-of-pocket payment for nursing home / home care	HC097	1 2 5 6
hinsu	Paid for voluntary health insurance	HC061	1
ydep	Earnings from employment	EP205	1 2 4 5 6
yind	Earnings from self-employment	EP207	1 2 4 5 6
ypen1	Annual old age, early retirement pensions, survivor and war pension	EP078_1-2-3-7-8-9 (1-2-3-9-10-11 in w6)	1 2 4 5 6
ypen2 <sup>(a)</sup>	Annual private occupational pensions	EP078_11-16 (EP678 in w6)	1 2 4 5 6
ypen3	Annual disability pension and benefits	EP078_5-6	1 6
ypen36	Annual disability/sickness pension and benefits	EP078_4-5	2 4 5
ypen4	Annual unemployment benefits and insurance	EP078_6 (8 in w6)	1 2 4 5 6
ypen5	Annual payment from social assistance	EP078_10 (12-13 in w6)	1 2 4 5 6
ypen6	Sickness benefits and pensions	EP078_3_6_10 (4-7 in w6)	1 6
ylsp1	Lump sum payments for old age, early retirement, survivor and war pension	EP082_1-2-3-7-8-9 (1-2-3-9-10-11 in w6)	2 4 5 6
ylsp2 <sup>(b)</sup>	Lump sum payments for private occupational pension	EP082_11-16 (EP682 in w6)	2 4 5 6
ylsp3	Lump sum payments for disability pension and benefits	EP082_5-6	6
ylsp36	Lump sum payments for disability/sickness pension and benefits	EP082_4-5	2 4 5
ylsp4	Lump sum payments for unemployment benefits and insurance	EP082_6 (8 in w6)	2 4 5 6
ylsp5	Lump sum payments for social assistance	EP082_10 (12-13 in w6)	2 4 5 6
ylsp6	Lump sum payments for sickness benefits	EP082_4-7	6



Variable	Description	Questionnaire	Waves
yreg1	Other regular payments from private pensions	EP094_1-2-5	1 2 4 5 6
yreg2	Other regular payment from private transfer	EP094_3-4	1 2 4 5 6
ylsr1	Lump sum from private payments	EP209_1-2-5	2 4 5 6
ylsr2	Lump sum from private transfers	EP209_3-4	2 4 5 6
aftgiv	Financial transfers given	FT004	1 2
aftrec	Financial transfers received	FT011	1 2
aftinh	Inheritance received	FT015	1 2
rhre	Annual rent and home-related expenditures	HO005, HO008	1 2 4 5 6
home	Value of main residence	HO024	1 2 4 5 6
mort	Mortgage on main residence	HO015	1 2 4 5 6
ores	Value of other real estate – Amount	HO027	1 2 4 5 6
ysrent	Annual income from rent or sublet	HO074, HO030	1 2 4 5 6
yaohm <sup>(c)</sup>	Annual income from other household members	HO002, HO011	1 2 4 5 6
fahc	Annual food at home consumption	CO002	1 2 4 5 6
fohc	Annual food outside home consumption	CO003	1 2 4 5 6
telc	Amount spent on telephones	CO004	1 2
hprf	Annual home produced consumption	CO011	2 4 5 6
bacc	Bank accounts	AS003	1 2 4 5 6
bsmf	Bond, stock and mutual funds	AS007, AS011, AS017	1 2 4 5 6
ybabsmf	Interest / dividend from bank account, bond, stock, and mutual funds		1 2 4 5 6
slti	Savings for long-term investments	AS021, AS023, AS27, AS030	1 2 4 5 6
vbus	Value of own business	AS042	1 2 4 5 6
sbus	Share of own business	AS044	1 2 4 5 6
car	Value of cars	AS051	1 2 4 5 6
liab	Financial liabilities	AS055	1 2 4 5 6
thinc	Total household net income - version A		1 2 4 5 6
thinc2	Total household net income - version B	HH017	2 4 5 6
thexp	Total household expenditure (sum of rhre, fahc, fo hc and hprc)		1 2 4 5 6
yincnrp	Income from non-responding partner		1 2 4 5 6
hrass	Household real assets (home*perho/100+vbus*sbus/100+car+ ores – mort)		1 2 4 5 6
hgfass	Household gross financial assets (sum of bacc, bsmf and slti)		1 2 4 5 6
Hnfass	Household net financial assets (hgfass – liab)		1 2 4 5 6
hnetw	Household net worth		1 2 4 5 6
gender	Gender	DN042	1 2 4 5 6
age	Age of respondent (based on interview year)	DN003	1 2 4 5 6
age_p	Age of partner (based on interview year)	DN003	1 2 4 5 6
yedu	Years of education	DN041 (in w1 based on ISCED)	1 2 4 5 6
yedu_p	Year of education of partner	EX102	1 2 4 5 6
iscd	ISCED 97 coding		1 2 4 5 6
sphus	Self-perceived health – US scale	PH003	1 2 4 5 6

Variable	Description	Questionnaire	Waves
mstat	Marital status	DN014	1 2 4 5 6
nchild	Number of children	CH001	1 2 4 5 6
ngrchild	Number of grandchildren	CH201	1 2 4 5 6
gali	Limitation with activities	PH005	1 2 4 5 6
chronic	Number of chronic diseases	PH006	1 2 4 5 6
symptoms	Number of symptoms	Ph010	1 2 4
eyesightr	Eyesight reading	PH044	1 2 4 5 6
hearing	Hearing	PH046	1 2 4 5 6
bmi	Body mass index	PH012, PH013	1 2 4 5 6
weight	Weight	PH012	1 2 4 5 6
height	Height	PH013	1 2 4 5 6
mobility	Mobility limitations	PH048	1 2 4 5 6
adl	Limitations with activities of daily living	PH049_1	1 2 4 5 6
iadl	Limitations with instrumental activities of daily living	PH049_2	1 2 4 5 6
esmoked	Ever smoked daily	BR001	1 2 4 5 6
drinking	More than 2 glasses of alcohol almost everyday	BR019	1 2 4 5 6
phinact	Physical inactivity	BR015	1 2 4 5 6
reading	Self-rated reading skills	CF001	1 2 4 5 6
writing	Self-rated writing skills	CF002	1 2 4 5 6
orienti	Score of orientation in time test	CF003 - CF006	1 2 4 5 6
wllft	Score of words list learning test – trial 1	CF104_*-CF107_*	1 2 4 5 6
wllst	Score of words list learning test – trial 2	CF113_*-CF116_*	1 2 4 5 6
fluency	Score of verbal fluency test	CF010	1 2 4 5 6
numeracy	Score of first numeracy test	CF012-CF015	1 2 4 5 6
numeracy2	Score of second numeracy test	CF108-CF112	4 5 6
memory	Score of memory test	CF103	4 5 6
maxgrip	Maximum of grip strength measures	GS006-GS009	1 2 4 5 6
eurod	EURO depression scale	MH002-MH017	1 2 4 5 6
doctor	Seen/Talked to medical doctor	HC002	1 2 4 5 6
hospital	In hospital last 12 months	HC012	1 2 4 5 6
thospital	Times being patient in hospital	HC013	1 2 4 5 6
nhospital	Total nights stayed in hospital	HC014	1 2 4 5 6
diaryp	How often consume dairy products	BR026	5 6
legeggs	How often consume legumes, beans or eggs	BR027	5 6
meat	How often consume meat, fish or poultry	BR028	5 6
fruit	How often consume fruits or vegetables	BR029	5 6
room10	Rooms at home at the age of 10	MC003	5 6
people10	Number of people at home at the age of 10	MC004	5 6
book10	Number of books at home at the age of 10	MC005	5 6
math10	Math performance at the age of 10	MC006	5 6
lang10	Language performance at the age of 10	MC007	5 6
health15	Health status in first 15 years	MC010	5 6
diseas15	Number of childhood diseases in first 15 years	MC012	5 6
illness15	Number of childhood illnesses in first 15 years	MC013	5 6

Variable	Description	Questionnaire	Waves
vacc15	Received vaccinations in first 15 years	MC015	5 6
cjs	Current job situation	EP005	1 2 4 5 6
pwork	Did any paid work	EP002	1 2 4 5 6
afwork	Away from work during last month	EP003	1
mtoj	More than one job	EP007	1
empstat1	Employee or self-employed first job	EP009_1	1
empstat2	Employee or self-employed second job	EP009_2	1
empstat	Employee or self-employed	EP009	2 4 5 6
lookjob	Looking for job	EP337	5 6
rhfo	Received help from others (how many)	SP002, SP005, SP007	1 2 4 5 6
ghfo	Given help to others (how many)	SP008, SP011, SP013	1 2 4 5 6
ghih	Given help in the household (how many)	SP018	1 2 4 5 6
rhih	Received help in the household (how many)	SP020	1 2 4 5 6
otrf	Owner, tenant or rent free	HO002	1 2 4 5 6
gfg	Number of given financial gifts 250 or more	FT002, FT007_*	4 5 6
rfg	Number of received financial gifts 250 or more	FT009, FT014_*	4 5 6
rggp	Number of received gifts, goods, properties 5000 or more	FT015, FT020_*	4 5 6
gggp	Number of given gifts, goods, properties 5000 or more	FT025, FT031_*	4 5 6
fdistress	Household able to make ends meet	CO007	1 2 4 5 6
nalm	Number of activities last month	AC002	1 2
naly	Number of activities last year	AC035_*	4 5 6
saly	Satisfied with no activities	AC038	4 5 6
lifesat	Life satisfaction	AC012	2 4 5 6
lifehap	Life happiness	AC022	2 4 5 6
lifex	Living in ten years	EX009	5 6
politics	Left or right in politics	EX028	5 6
tpmdi	Third person present during the interview	IV002	1 2 4 5 6
willans	Willingness to answer	IV004	1 2 4 5 6
clarif	Respondent asked for clarifications	IV007	1 2 4 5 6
undersq	Respondent understood questions	IV008	1 2 4 5 6
hnrsc	Help needed to read showcards	IV018	1 2 4 5 6
currency	Currency in which amounts are denominated		1 2 4 5 6
nomxyear	Nominal exchange rate		1 2 4 5 6
pppxyear	PPP adjusted exchange rates		1 2 4 5 6

\*Note:

(a) In SE (wave 6), ypen2 includes EP678 and EP078\_2. In LU (wave 6), ypen2 also includes EP678, EP078\_2, and ep078\_10. In PL (wave 6) ypen2 includes only EP078\_2.

(b) In SE (wave 6), yls2 includes EP682 and EP082\_2. In LU (wave 6), yls2 also includes EP682, EP082\_2, and ep082\_10. In PL (wave 6) yls2 includes only EP082\_2.

(c) HO002 not asked in wave 6.

**Table 12: Description of flag variables associated with imputed variables**

Varname_f	Label	Description
-99	Missing by design	Missing values depends from skip patterns in the questionnaire
1	Not designed resp	Missing values depends on the type of respondents designed to respond
2	No ownership	No declared ownership
3	Regular obs.	Regular observation
4	Imp: ub point	Imputation based on specific declared amounts in the unfolding brackets routing
5	Imp: ub range	Imputation is based on unfolding brackets range information
6	Imp: ub incomplete	Imputation is based on unfolding brackets partial information
7	Imp: ub uninformative	Unfolding brackets uninformative
8	Imp: ownership	Ownership has been imputed
9	Imp: amount	Imputed amount
10	Imp: outlier LB	Imputed value if lower than LB
11	Imp: outlier UB	Imputed value if lower than UB
12	Imp: aggregate	Imputation of aggregate variable
13	Imp: NRP	(only for thinc)
14	Imp: missing value	(only for explanatory variables imputed ex-ante by hot-deck)