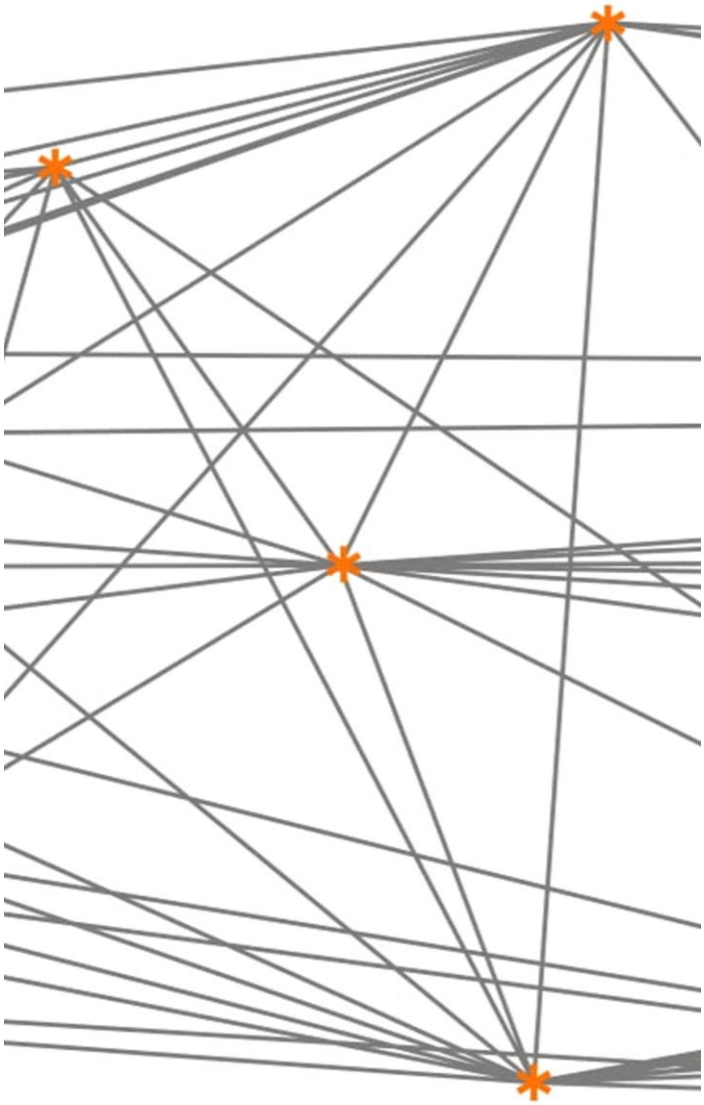
SHARE

Survey of Health, Ageing and Retirement in Europe

Release Guide 9.0.0



March 28, 2024

Table of Contents

1 Sources of Information 5
2 SHARE Data Releases 6
3 New in Release 9.0.0 6
4 Additional Files 6
4.1 SHARE Corona Survey 6
4.2 *easy*SHARE 7
4.3 Interviewer Survey 7
5 Countries, Languages and Fieldwork times in SHARE 8
6 Eligibility Rules 12
7 Data Collection 12
8 SHARELIFE 12
9 Types of Respondents 13
10 Merging the Data 16
11 Missing Codes 16
12 Variable Naming Conventions 18
12.1 General Conventions 18
12.2 Double Loops 18
12.3 Dummy Variables 18
12.4 Euro Conversion 19
12.5 Unfolding Brackets 20
12.6 Naming of Modules and Variables in Wave 3 20
13 Types of Questionnaires and Modules 21
13.1 Regular Questionnaire Modules: Wave Overview and Short Description 21
13.2 SHARELIFE Questionnaire Modules: Wave Overview and Short Description 25
13.3 Special Questionnaire Modules 26
13.3.1 Coverscreen 26
13.3.2 End-of-Life Interviews 28
13.3.3 Technical Variables 28
13.3.4 Paper and Pencil Drop-Off 29
13.3.5 Vignettes 29
14 Special Aspects of the Questionnaire Modules 30
14.1 DN Module: Citizenship/Country of Birth Coding and Education in Portugal 30
14.2 CH Module: Information on Children in SHARE 30

14.3 PH Module: *phrandom* 32
14.4 EX Module: Definition of Life Expectancy Target Age 32
14.5 CF Module: Ten Words List Learning 32
14.6 CO Module: Afford Expense Amount in CO206 33
14.7 SN Module (Waves 4, 6, 8, and 9) 34
14.8 SP and FT Module: List of Relations Differences due to SN 35
14.9 EP Module: EP005 & Re-Interview in Israel (Wave 1) 37
14.10 Social Exclusion Items (Wave 5) 37
15 Generated Variables 38
15.1 gv_allwaves_cv_r 40
15.2 gv_weights and gv_longitudinal_weights 40
15.3 gv_imputations 49
15.4 gv_isced 56
15.5 gv_health 56
15.6 gv_housing 58
15.7 gv_networks 58
15.8 gv_exrates 60
15.9 gv_job_episodes_panel 61
15.10 gv_grossnet 61
15.11 gv_isco 61
15.12 gv_ssw 62
15.13 gv_deprivation 63
15.14 gv_children 64
15.15 gv_linkage 66
15.16 gv_dbs 66
15.17 gv_big5 67
15.18 gv_accelerometer 67
APPENDIX 77

Index of Tables

Table 1: Links to Further Documentation Material 5
Table 2: Countries, Language Versions & Fieldwork Times 8
Table 3: Structure of Wave 7 Questionnaire 13
Table 4: Types of Respondents and Variable Names 14
Table 5: Who Answers What in the Regular Modules? 15
Table 6: Adjustment of ac004_ Variable Names in Wave 2 19

Table 7: Regular Questionnaire Modules of Waves 1, 2, 4, 5, 6, 7, 8, and 9 22
Table 8: Retrospective SHARELIFE Questionnaire Modules of Waves 3 and 7 25
Table 9: Variables in the Technical Variables Module 29
Table 10: Changes in Information on Children 31
Table 11: Country specific AROP values by year 33
Table 12: Relationship Type Coding 36
Table 13: Items on Social Exclusion 38
Table 14: Generated Variable Modules of Waves 1 to 9 39
Table 15: Generated Variables of gv_health 56
Table 16: Generated Housing Variables 58
Table 17: Generated Variables in Wave 4 gv_networks 59
Table 18: Generated Variables in Wave 6, 8 and 9 gv_networks 59
Table 19: Variables Used for ISCO and NACE Coding 62
Table 20: Items Used for Material Deprivation Index (depmat) 63
Table 21: Items Used for the Social Deprivation Index (depsoc) 63
Table 22: Generated Variables in gv_children 65
Table 23: Generated Variables in gv_big5 67
Table 24: Overview of ActiPASS variables in “day” and “hour” data 70
Table 25: Overview of ActiPASS variables in “sleep” 72
Table 26: Overview of measures, corresponding variables, and availability in datasets 74
Table 27: Overview of variables in epoch data 74

Index of Appendix Tables

Table 1: Digital Object Identifier of the SHARE Datasets 77
Table 2: SHARE Data Releases 78
Table 3: Overview of Additional Codes for Country of Birth and Citizenship 81
Table 4: Drop-Off Correspondence in Waves 1 & 2 83
Table 5: Calibrated cross-sectional weights of Wave t=1,…, 9 86
Table 6: Calibrated cross-sectional weights of pooled SHARELIFE sample 86
Table 7: Calibrated longitudinal weights of wave combination t-…-s 87
Table 8: Composition of aggregated imputation variables in wave 1 88
Table 9: Country-specific composition of imputed pension variables in wave 1 89
Table 10: Composition of aggregated imputation variables in wave 2 91
Table 11: Composition of aggregated imputation variables in wave 4 93
Table 13: Composition of aggregated imputation variables in waves 6, 7, 8 and 9 97
Table 14: List of variables in gv_imputations 99
Table 15: Description of flag variables associated with imputed variables 104
Table 16: Meaning of the categories of *htype* 104

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. This guide also contains information previously stored in a separate release guide for wave 3 so that the necessary information for all SHARE waves is comprised in one document.

Additional information to enable and facilitate the use of SHARE data is available on the SHARE website: 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 (https://share-eric.eu/data/data-documentation/share-data-releases).

The following overview contains links to further important documentation material.

Table 1: Links to Further Documentation Material

	W1	W2	W3	W4	W5	W6	W7	W8	SCS1	SCS2	W9
Release Guides	SHARE Release Guide 9.0.0 (PDF) SHARE Corona Survey Release Guide										
Questionnaires	X	X	X	X	X	X	X	X	X	X	X
Cross Wave Comparison	X										X
Methodology Volumes	X	Chapter 8 of W2 FRB	X	X	X	X	X	X	X	X	X
Data & Documentation Tool	Web interface for browsing and searching the SHARE meta data										
Scales Manual	Scales and Multi-Item Indicators										
Data Resource Profile	Börsch-Supan A. et al. (2013): Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE), Int J of Epidemiology										

In case you have questions that are neither answered by the content of these documents nor by the FAQ section, you can contact the SHARE Central-Team: 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: https://share-eric.eu/data/become-a-user

Updates of releases are indicated by the release number as follows:

- 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 **9**.0.0 with wave 9 being the latest regular SHARE wave.
- Major changes will be announced to users via e-mail and are indicated by the second digit, e.g. release 7.**1**.0.
- Minor changes will be indicated by the third digit, e.g. release 7.1.**1**.

Please regularly check the website for information on release updates.

The new Release 9.0.0 of waves 1 to 9 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 3). Moreover, the sample has been improved by conducting thorough consistency checks based on information from all waves. An overview on previous SHARE release updates is provided in the appendix.

3 New in Release 9.0.0

- Release 9.0.0 contains the first scientific release of SHARE wave 9 data
- The wave 8 sample is significantly larger than in Release 8.0.0 because it contains refresher households
- Update of the wave 8 accelerometer data
- Linking children across waves is now easier due to the new child ID variable "*childid_**"
- New variables make the linkage of children between CH and SN module easier
- New Interviewer ID "*intidwX*" for waves 1 to 8 allows linking interviewers across waves
- New variable *re012isco_** in wave 3 allows researchers to retrace a respondent's job history with the 4-digit ISCO classification (like in SHARELIFE of wave 7)

4 Additional Files

4.1 SHARE Corona Survey

The outbreak of the COVID-19 pandemic hit SHARE in the middle of its 8th Wave of data collection. The fieldwork had to be suspended in all participating countries in March 2020. At that time, about 70 percent of all expected longitudinal interviews across countries had been conducted. As a reaction, SHARE developed a specific questionnaire covering the same topics as the regular SHARE questionnaire - but considerably shortened and targeted to the living situation of people aged 50 and above during the pandemic. The SHARE Corona questionnaire

covers the topics health and health behaviour, mental health, COVID-19 related symptoms and healthcare, changes in work and the economic situation as well as changes in the social networks. The Wave 8 Corona questionnaires is available here, the Wave 9 Corona questionnaire here.

Based on methodological considerations in connection with the health protection of respondents and interviewers, the use of telephone-administered interviews (CATI) was the preferred alternative to face-to-face-interviewing (CAPI). For a more detailed description on the switch from CAPI to CATI, please see: Scherpenzeel, A., Axt, K., Bergmann, M., Douhou, S., Oepen, A., Sand, G., Schuller, K., Stuck, S., Wagner, M., & Börsch-Supan, A. (2020). Collecting survey data among the 50+ population during the COVID-19 outbreak: The Survey of Health, Ageing and Retirement in Europe (SHARE). *Survey Research Methods*, 14(2), 217-221.

The first round of the SHARE Corona Survey was conducted between June and September 2020. Data for the second round was collected one year later between June and August 2021 in order to observe developments in the course of the pandemic.

4.2 *easy*SHARE

*easy*SHARE is a simplified HRS-adapted dataset for student training, and for researchers who have little experience in quantitative analyses of complex survey data. *easy*SHARE stores information on all respondents and of all waves in one single dataset. For the subset of variables covered, the complexity was considerably reduced. *easy*SHARE 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 *easy*SHARE from the regular SHARE data. For further information see https://share-eric.eu/data/data-set-details/easyshare .

4.3 Interviewer Survey

The SHARE wave 5, wave 6 and wave 7 *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. In wave 7, more than 1100 interviewers from 22 countries and in wave 9 more than 930 interviewers from 23 countries participated. 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.

Please see https://share-eric.eu/data/data-set-details/share-interviewer-survey for further information.

5 Countries, Languages and Fieldwork times in SHARE

Table 2: Countries, Language Versions & Fieldwork Times in SHARE Waves 1-9 including SHARE Corona Survey 1 & 2

Country ID	Language ID	Country & language	W1	W2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	SCS1	SCS2	Wave 9
11	11	Austria	2004	2006/07	2008/09	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
12	12	Germany	2004	2006/07	2008/09	2011/12	2013	2015	2017	2019/2020	2020	2021	2021/2022
13	13	Sweden	2004	2006/07	2008/09	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
14	14	Netherlands	2004	2007	2008/09	2011	2013	2015[1]	2017	2019/2020	2020	2021	2021/2022
15	15	Spain (Castilian)	2004	2006/07	2008/09	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
15	65	Spain (Catalan)	2004	2006/07	2008/09	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
15	38	Spain/Girona (Castilian or Catalan)	-	-	-	-	2013	2015	2017	2019/2020	2020	2021	2021/2022
15	39	Spain/Girona (Catalan)	-	-	-	-	2013	2015	2017	2019/2020	2020	2021	2021/2022
15	40	Spain/Girona (Castilian)	-	-	-	-	2013	2015	2017	2019/2020	2020	2021	2021/2022
16	16	Italy	2004	2006/07	2008/09	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
17	17	France	2004/05	2006/07	2009	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
18	18	Denmark	2004	2006/07	2008/09	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
19	19	Greece	2004/05	2007	2008/09	-	-	2015	2017	2019/2020	2020	2021	2021/2022

[1] In SHARE Wave 6 and 7, the Netherlands did not participate in the regular SHARE wave but conducted a mixed mode experiment (see https://share-eric.eu/data/accompanying-datasets/dutch-mixed-mode-experiment for further information)

Country ID	Language ID	Country & language	W1	W2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	SCS1	SCS2	Wave 9
20	20	Switzerland (German)	2004	2006/07	2008/09	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
20	21	Switzerland (French)	2004	2006/07	2008/09	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
20	22	Switzerland (Italian)	2004	2006/07	2008/09	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
23	23	Belgium (French)	2004/05	2006/07	2008/09	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
23	24	Belgium (Flemish)	2004/05	2006/07	2008/09	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
25	25	Israel (Hebrew)	2005/06	2009/10	-	-	2013	2015	2017	2019/2020	2020	2021	2021/2022
25	26	Israel (Arabic)	2005/06	2009/10	-	-	2013	2015	2017	2019/2020	2020	2021	2021/2022
25	27	Israel (Russian)	2005/06	2009/10	-	-	2013	2015	2017	2019/2020	2020	2021	2021/2022
28	28	Czech Republic	-	2006/07	2008/09	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
29	29	Poland	-	2006/07	2008/09	2011/12	-	2015	2017	2019/2020	2020	2021	2021/2022
30	30	Ireland	-	2007	2009/10/11	-	-	-	-				
31	41	Luxembourg (French)	-	-	-	-	2013	2015	2017	2019/2020	2020	2021	2021/2022
31	42	Luxembourg (German)	-	-	-	-	2013	2015	2017	2019/2020	2020	2021	2021/2022

Country ID	Language ID	Country & language	W1	W2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	SCS1	SCS2	Wave 9
31	43	Luxembourg (Portuguese)					2013			2019/2020	2020	2021	2021/2022
32	32	Hungary	-	-	-	2011	-	-	2017	2019/2020	2020	2021	2021/2022
33	33	Portugal	-	-	-	2011	-	2015	2017/ 2018	2020[2]	2020	2021	2021/2022
34	34	Slovenia	-	-	-	2011	2013	2015	2017	2019/2020	2020	2021	2021/2022
35	35	Estonia (Estonian or Russian)	-	-	-	2010/11	2013 (XT only)	2015	2017	2019/2020	2020	2021	2021/2022
35	36	Estonia (Estonian)	-	-	-	-	2013	2015	2017	2019/2020	2020	2021	2021/2022
35	37	Estonia (Russian)	-	-	-	-	2013	2015	2017	2019/2020	2020	2021	2021/2022
47	47	Croatia	-	-	-	-	-	2015	2017	2019/2020	2020	2021	2021/2022
48	48	Lithuania	-	-	-	-	-	-	2017	2019/2020	2020	2021	2021/2022
51	51	Bulgaria	-	-	-	-	-	-	2017	2019/2020	2020	2021	2021/2022
53	53	Cyprus	-	-	-	-	-	-	2017	2019/2020	2020	2021	2021/2022
55	55	Finland (Finnish)	-	-	-	-	-	-	2017	2019/2020	2020	2021	2021/2022
55	56	Finland (Swedish)	-	-	-	-	-	-	2017	2019/2020	2020	2021	2021/2022
57	57	Latvia (Latvian)	-	-	-	-	-	-	2017	2019/2020	2020	2021	2021/2022

[2] In SHARE Wave 8, Portugal participated in the SHARE Corona Survey only.

Country ID	Language ID	Country & language	W1	W2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	SCS1	SCS2	Wave 9
57	58	Latvia (Russian)	-	-	-	-	-	-	2017	2019/2020	2020	2021	2021/2022
59	59	Malta (Maltese)	-	-	-	-	-	-	2017	2019/2020	2020	2021	2021/2022
59	60	Malta (English)	-	-	-	-	-	-	2017	2019/2020	2020	2021	2021/2022
61	61	Romania	-	-	-	-	-	-	2017	2019/2020	2020	2021	2021/2022
63	63	Slovakia	-	-	-	-	-	-	2017	2019/2020	2020	2021	2021/2022

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 from baseline or refreshment samples 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, 1964 or earlier in wave 6, 1966 or earlier in wave 7, 1969 or earlier in wave 8, and 1971 or earlier in wave 9. Please note that wave 3 did not contain any refreshment sample. 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

With exception of the SHARE Corona Survey that was collected via telephone (CATI), SHARE data collection is based on computer-assisted personal interviewing (CAPI). The interviewers conduct face-to-face interviews using a laptop 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 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 SHARELIFE

SHARELIFE data was collected in wave 3 and wave 7 focusing on respondents' life histories. Most of the information collected in the regular SHARE waves is about the current life circumstances. As a result, we know little about what happened earlier in the respondents' lives. SHARELIFE gathers more detailed information on important areas of our respondents'

lives, ranging from partners and children over housing and employment history to detailed questions on health and health care. SHARELIFE data complements the SHARE panel data by providing life history information to enhance our understanding of how early life experiences and events throughout life influence the circumstances of older people. With this variety SHARELIFE constitutes a unique cross-national, interdisciplinary database for research in the fields of sociology, economics, gerontology, and demography.

Collecting accurate retrospective information is a challenge. Respondents may not remember the occurrence or the exact timing of past events perfectly. Therefore, SHARELIFE follows a *Life History Calendar* (LHC) approach, which has been designed to help respondents in remembering past events more accurately. Using this method, the life events of interest are displayed on a “calendar”, enabling interviewers and respondents to cross-reference certain life-events with others (e.g. "I moved from *A* to *B* the year after my first child was born"). Using the life history calendar technique has been shown to improve the accuracy of the retrospective information given by respondents.

In wave 3 all SHARE respondents answered the SHARELIFE questionnaire whereas in wave 7 SHARELIFE modules were asked to all respondents who did not participate in wave 3 (82% of respondents). In addition, in wave 7 regular panel modules were asked to all respondents who already participated in the SHARELIFE interview of wave 3 (18% of respondents). Please note that this complicated structure of the questionnaire leads to high amounts of missing values in specific variables due to routing.

Table 3: Structure of Wave 7 Questionnaire

Respondent type	Interview sections		No. of interviews
Respondent participated in Wave 3	**Regular Panel** DN CH PH BR CF MH HC EP GS SP FT CO AS HO HH AC EX IV		13.959 (18%)
Respondent did not participate in Wave 3	**SHARELIFE** RC RP RA CC RE WQ DQ FS HS RH GL	**Regular Panel** (condensed)	62.561 (82%)

9 Types of Respondents

In some of the regular modules 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 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 (defined by living with partner, no matter if married or not): The couple's first person interviewed is the family respondent. Note that the naming of *mergeid* 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 could 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 4: Types of Respondents and Variable Names

Type of respondent	Variable name in cv_r and technical variables module	Name of filter in w1 & w2 questionnaire	Name of filter in w4, w5, w6, w7, w8, w9 questionnaire
Family respondent	*fam_resp*	MN006_ (FAMILY RESPONDENT)	MN006_NumFamR
Financial respondent	*fin_resp*	MN007_ (FINANCIAL RESPONDENT)	MN007_NumFinR
Household respondent	*hou_resp*	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. 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 13.3.2). 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). Information on the relationship between proxy and respondent is stored in variable *iv020_* in the IV module.

Table 5: Who Answers What in the Regular Modules?

Module	Name	All respondents[1]	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
SR	Saving Regrets	x				x
TE	Time Expenditure	x				
IV	Interviewer Observations	Interviewer only				
XT	End-of-Life Interview	proxy only				
dropoff	Drop-off	x				x
vignettes	Vignettes	x				x

[1] "All respondents" only refers to types of respondents here. Due to other filters/routing not all modules are answered by all respondents. Please also note that the SHARELIFE modules are not listed here because they are answered by all SHARELIFE respondents.

10 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 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. The variable *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 15.1 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.

11 Missing Codes

Apart from so called system missing values due to filters/routing or an abandoned interview, SHARE uses the following 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"

For financial variables:
-9999991: "Don't know"
-9999992: "Refusal"

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 – gv_networks variables in Wave 6 and Wave 8

Wave 6	Wave 8	Wave 9
-20: “Missing information”	-20: “Missing information”	-20: “Missing information”
-21: “W4 only”	-21: “w6 or w4 only”	-21: “w8 only”
-22: “W6 only”	-22: “w8 only”	-22: “w9 only”
-23: “W4 SN size of 0”	-23: “w6 SN size of 0”	-23: “w8 SN size of 0”
-24: “W6 SN size of 0”	-24: “w8 SN size of 0”	-24: “w9 SN size of 0”
-25: “No SN in w6 & w4”	-25: “No SN in w8 & w6”	-25: “No SN in w9 & w8”

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://releases.sharedataportal.eu/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 *easy*SHARE. Please note that the new missing codes for gv_networks are not yet integrated in *sharetom5.*

12 Variable Naming Conventions

12.1 General Conventions

The naming of variables is harmonised across waves. Variable names in the individual interview 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” for 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)

12.2 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.

12.3 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:

  mmXXXd**no** “None of these”

  mmXXXd**ot** “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 6: Adjustment of *ac004_* Variable Names in Wave 2

wave 2: ac004_ questionnaire categories	variable names
1. To meet other people	*ac004d1_#*
2. To contribute something useful	*ac004d2_#*
3. Because I am needed	*ac004d**4**_#*
4. To earn money	*ac004d**5**_#*
5. To use my skills or to keep fit	*ac004d**7**_#*
96. None of these	*ac004dno_#*

12.4 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 15.8 for information about ppp-adjustment (purchasing power parity).

12.5 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.

12.6 Naming of Modules and Variables in Wave 3

Wave 3 was the first wave focusing on respondents' life histories providing data on sequences of life events in flat file format ("wide"). That is "year of marriage" is looped over all reported marriages and the information is stored as a set of variables in one observation. For analyses requiring a spell data format ("long"), the data matrix has to be reshaped accordingly.

Please be aware that some module names of wave 3 coincide with module abbreviations used in the regular SHARE waves, but the content differs. For this reason wave 3 variables are generally prefixed by "sl_". The variables stored in a module contain the module name abbreviation in their name, e.g. variables starting with "sl_ac" are stored in the ac module. There are, however, a few exceptions to this rule, where single variables are stored in a

module having a different name: sl_hh017_ as part of the sl_fs module and sl_ph003 as part of the sl_hs module.

Note that a few variables generically contain letters in their variable names. For example, the variables sl_xt040c_ sl_xt040d_ do not refer to a coded or dummy variable based on a variable xt040, respectively, but xt040c and xt040d are the generic variable names (as appearing in the questionnaire). It may also occur that such variables appear as dummy set in the data, such as sl_hc005ad1 to sl_hc005ad7, where the "a" stems from the generic variable name and the "d" indicates that the variable is coded as dummy.

13 Types of Questionnaires and Modules

The questionnaires for all waves and languages can be downloaded from the SHARE website: http://www.share-project.org/data-documentation/questionnaires.html

The SHARE interview consists of various thematic blocks or modules and starts with the coverscreen module (cv_r). The 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. For regular SHARE interviews, 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 versions baseline vs longitudinal (see also chapter 13.3.3). Please be aware that some questions are only asked in the baseline interview version (e.g. education, country of birth) and thus the information needs to be taken from earlier waves in which the baseline interviews took place. Also, 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 …"*.

In addition to the regular baseline and longitudinal SHARE questionnaire versions, retrospective SHARELIFE questionnaires were used in wave 3 and wave 7. In wave 3, all respondents answered the SHARELIFE questionnaire, whereas in wave 7 only those respondents who had not yet done a SHARELIFE interview in wave 3 received the SHARELIFE questionnaire. In wave 7, the variable *mn103_* indicates if a regular or a SHARELIFE interview was conducted.

13.1 Regular Questionnaire Modules: Wave Overview and Short Description

To pick up contemporary matters and due to time constraints not every module was part of every regular wave. The overview below lists the different questionnaire modules for the regular waves of SHARE and shows in which wave the corresponding module was included.

Table 7: Regular Questionnaire Modules of Waves 1, 2, 4, 5, 6, 7, 8, and 9

	Content of regular Questionnaire Modules	W1	W2	W4	W5	W6	W7	W8	W9
CV_R	Coverscreen on individual level	x	x	x	x	x	x	x	x
DN	Demographics and Networks	x	x	x	x	x	x	x	x
SN	Social Networks			x		x		x	x
CH	Children	x	x	x	x	x	x	x	x
PH	Physical Health	x	x	x	x	x	x	x	x
BR	Behavioral Risks	x	x	x	x	x	x	x	x
CF	Cognitive Function	x	x	x	x	x	x	x	x
MH	Mental Health	x	x	x	x	x	x	x	x
HC	Health Care	x	x	x	x	x	x	x	x
EP	Employment and Pensions	x	x	x	x	x	x	x	x
IT	Computer Use				x	x	x	x	x
MC	Mini Childhood				x				
GS	Grip Strength	x	x	x	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	x	x	x
FT	Financial Transfers	x	x	x	x	x	x	x	x
HO	Housing	x	x	x	x	x	x	x	x
HH	Household Income	x	x	x	x	x	x	x	x
CO	Consumption	x	x	x	x	x	x	x	x
AS	Assets	x	x	x	x	x	x	x	x
AC	Activities	x	x	x	x	x	x	x	x
EX	Expectations	x	x	x	x	x	x	x	x
SR	Saving Regrets							x	
TE	Time Expenditure							x	x
IV	Interviewer Observations	x	x	x	x	x	x	x	x
	Special Questionnaire Modules								
XT	End-of-Life Interview		x	x	x	x	x	x	x
DO	Drop-off	x	x	x	x	x	x	x	x
VI	Vignettes	x	x						
TC	Technical Variables	x	x	x	x	x	x	x	x
AX	Accelerometry							x	

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 9 provides information on whether and why social networks of respondents changed in comparison to waves 4, 6, and 8.

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, ES, 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. A list of amounts asked in FT002, FT009 and FT015 is available here.

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 as well as information on how people felt about various aspects of their life. In waves 7, 8 and 9, this module also contains the so-called "Big Five" items (see gv_big5).

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.

Saving Regrets (SR): Contains in wave 8 introduced questions about saving behaviour when respondents were younger and retrospective regrets about spending/saving behaviour.

Time Expenditure (TE): The module contains information on how respondents spend time.

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

13.2 SHARELIFE Questionnaire Modules: Wave Overview and Short Description

Table 8 lists the questionnaire modules of the retrospective SHARELIFE interviews. Please note that some modules have different module abbreviations in wave 3 and wave 7 to avoid confusion with regular modules. Furthermore, the modules IV, GS and XT were part of both wave 3 and wave 7 but they are not included in Table 8 as they do not contain retrospective information.

The cross-wave item correspondence list for waves 3 and 7 is available at http://www.share-project.org/data-documentation/questionnaires/cross-wave-comparison.html

Table 8: Retrospective SHARELIFE Questionnaire Modules of Waves 3 and 7

W3 module	W7 module	Content of Questionnaire Modules
CV_R	CV_R	Coverscreen on individual level
ST	DN	Demographics
AC	RA	Retrospective Accommodation
CS	CC	Childhood Section / Childhood Circumstances
DQ	DQ	Disability
FS	FS	Financial Section
GL	GL	General Life and Persecution
HC	RH	Retrospective Health Care
HS	HS	Health Section
RC	RC	Retrospective Children History
RE	RE	Retrospective Employment
RP	RP	Retrospective Partner History
WQ	WQ	Work Quality

Retrospective Accommodation (AC/RA): This module includes questions on former residences (country, region), moves, types of accommodation, ownership, etc.

Childhood Section / Childhood Circumstances (CS/CC): In this module, we ask questions about the respondent's childhood conditions such as accommodation features, number of books, school performance, the relationship with parents, neighbourhood and physical abuse in childhood.

Disability (DQ): Respondents are asked questions on, e.g. disability leaves, work reduction and disability pensions.

Financial Section (FS): This module contains a series of questions on respondents' past financial investments such as investments in stocks, funds, insurance uptake and retirement savings.

General Life and Persecution (GL): This module collects information about the general life of respondents such as periods of happiness, stress, financial hardship and hunger. Furthermore, questions on whether respondents or their parents had to face discrimination, persecution and oppression in their past are included in this module.

Health Care Section (HC/RH): This module contains questions about health care the respondents might have received covering topics such as vaccinations, doctor visits, preventive check-ups and health behaviours.

Health History (HS): Respondents are asked about events that relate to their health such as hospital stays, illnesses and diseases during their life.

Retrospective Children History (RC): Each SHARELIFE respondent is asked about all her/his children, including deceased ones. The module includes retrospective questions about births, children characteristics and maternity leave.

Retrospective Employment History (RE): This module covers many different aspects of people's past employment status, job characteristics, income, etc.

Retrospective Partner History (RP): This module collects information about each respondent's partner history and covers, e.g. living arrangements, marriages and divorces in the past.

Work Quality (WQ): This module collects information on work quality such as effort, demand, control and job circumstances in the respondent's past.

13.3 Special Questionnaire Modules

13.3.1 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 since release 6-0-0 is an indication of coverscreen proxy respondents for waves 5, 6, 7 and 8 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 *mergeid**p**`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 the same 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*
- "102" → wave 4
- "122" → wave 5
- "142" → wave 6
- "162" → wave 7
- "192" → wave 8
- "212" → wave 9

**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 nine waves of SHARE, they range from 1 to 9 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 9) have been renamed in line with the imputation datasets (since release 5.0.0.). 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* <= *int_month* and *age_int* = *int_year* – *yrbirth* – 1 if *mobirth* > *int_month*. *age_int* is only provided for those who did an interview.

Since Release 1.0.0 of wave 8, the coverscreen module contains the following additions:

- Relationship Matrix: new variables (*rel_**) indicate the relation between each household member
- Variable *cvdate* indicates whether the coverscreen was completed in the first field phase (= the personal interview conducted prior to the outbreak of the pandemic) or the second field phase (= SHARE Corona Survey done via telephone interviews after the outbreak of the pandemic).
- Update variables: When a household participated in both field phases, the coverscreen was completed in the first field phase only. However, some important information could be updated in the second field phase. Variables including such information are indicated by "**_update_ca*" indicating known changes in the second field phase. The variables affected are *deceased*, *hhsize*, *partnerinhh*, *pidcomp8*, *coupleid8* and *relrpers*.
- Variables *nursinghome* and *hhmove*" serve as a complement to *mn024_* and *mn104_* for those households that did not participate in the main interview (no first but only second field phase).

13.3.2 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 or 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.

The open answers to question XT012_ on cause of a respondent's death are coded. 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*.

13.3.3 Technical Variables

This module contains technical variables for waves 1, 2, 4, 5, 6, 7, 8 and 9 (named MN# in the questionnaires) that play an important role in the questionnaire routing. The module includes variables indicating the different respondent types, too: *fam_resp*, *fin_resp* and *hou_resp* (see chapter 9) that are additionally stored in the wave-specific cv_r modules.

Table 9: Variables in the Technical Variables Module

Variable	Label
fam_resp	Family respondent
fin_resp	Financial respondent
hou_resp	Household respondent
*mn005_*	Single or couple interview
*mn016_*	Mother in household
*mn017_*	Father in household
*mn018_*	Mother-in-law in household
*mn019_*	Father-in-law in household
*mn024_*	Nursing home interview
*mn026_*	First respondent from couple or single
mn028	Eligible for dried blood spots collection (bs &gv_dbs)
mn029	Eligible for linkage
mn030	Eligible for social networks module (sn)
*mn031_*	Eligible for mini childhood module
*mn032_*	Eligible for social exclusion items
*mn038_*	Eligible for Accelerometry
*mn040_*	Need to ask consent question (ex123)
*mn041_*	Need to ask retirement info
*mn101_*	Questionnaire version (longitudinal vs. baseline)
*mn103_*	SHARELIFE life history interview (only w7)
*mn104_*	Household moved

13.3.4 Paper and Pencil Drop-Off

In waves 1, 2, 4, 5, 6, 7, 8 and 9 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.

13.3.5 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, ES SE) and wave 2 (eleven countries: BE, CZ, DK, DE, FR, GR, IT, NL, PL, ES 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).

14 Special Aspects of the Questionnaire Modules

14.1 DN Module: Citizenship/Country of Birth Coding and Education in Portugal

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 country of birth of the respondent´s mother (dn504c) and father (dn505c) that were included in wave 5. 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 3 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.

Please note that wrong categories on respondents’ education as asked in DN010_ but also the educational level of children as asked in CH017_ entered the Portuguese questionnaire. However, the problem could be solved with the generated ISCED variables as provided in the gv_isced module. Therefore, for any analyses using educational information in Portugal we recommend using gv_isced instead of the original variables.

14.2 CH Module: Information on Children in SHARE

The CH module is answered by the family respondent on behalf of the couple (see chapter 9). 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”.

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.

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.

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.

Table 10: Changes in Information on Children

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

Since Release 9.0.0, the child ID variables *childid_ ** are available in the CH module of waves 1 to 9, the RC module of the SHARELIFE interviews in waves 3 and 7, and in gv_children of waves 4 to 9. The *childid_ ** variables can be used to precisely link and identify children across waves. The variable has the following format: "CC-123456-WWFF", where CC indicates the country and WWFF is a unique four-digit combination for each child put together of the first wave a child was mentioned (WW) and a counter of children within that first wave (FF). In addition, a new "*child_dead_ **" indicator was introduced in the CH module and the gv_children modules of waves 4 to 9 to accompany the *childid_ ** variables indicating that a child is deceased. In that case, a valid entry on *childid _ ** may exist, to identify the dead child, but all follow-up variables of the CH section are empty by design.

Furthermore, the variable *child_sn_loop_** introduced in release 9.0.0 in the CH and the gv_children module of waves 6, 8, and 9 facilitates the linkage of children between CH & SN module. In the SN and gv_networks modules, release 9.0.0 introduced the variables *sn_child_loop_ ** and *sn_childid_ ** to facilitate the linkage of CH and SN module.

14.3 PH Module: *phrandom*

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)

14.4 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.

14.5 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 from wave 4 onwards 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, 6, 7, 8 and 9 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 trail) 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, 6, 7, 8 and 9. This is due to the fact that these variables have been generated in waves 4, 5, 6, 7, 8 and 9 and were not regular CAPI items.

14.6 CO Module: Afford Expense Amount in CO206

Question CO206 asks: *"Could your household afford to pay an unexpected expense of [AffordExpenseAmount] without borrowing any money?"* The 'Afford Expense Amount', which is filled in the brackets, represents the at-risk-of-poverty threshold defined as 60 percent of the median equivalised disposable income. The only exception in this respect is Israel. In waves 5, 6 and 7, the at-risk-of-poverty threshold (AROP) was computed following the official Israeli calculation of 50 percent of the median equivalised disposable income. For reasons of harmonisation, this was changed from wave 8 onwards. Table 11 contains the country-specific values in each wave (the Israeli values in waves 5 and 6 are not available).

Table 11: Country specific AROP values by year

COUNTRY	Year of AROP calculation:					
		2010	2013	2015	2016	2019
	Currency	W5	w6	w7	w8	w9
Austria	EUR	1000	1100	1200	1200	1300
Belgium	EUR	1000	1000	1100	1100	1200
Bulgaria	BGN	n.a.	n.a.	300	300	400
Croatia	HRK	n.a,	2000	2000	2200	2700
Cyprus	EUR	n.a.	n.a.	700	700	800
Czech Republic	CZK	9000	9700	9900	10600	13000
Denmark	DKK	7000	10000	10600	10700	11000
Estonia	EUR	300	300	400	400	500
Finland	EUR	n.a.	n.a.	1200	1200	1200
France	EUR	1000	1000	1100	1100	1100
Germany	EUR	900	1000	1000	1100	1200
Greece	EUR	600	500	400	400	400
Hungary	HUF	61000	66000	70000	74300	90000
Israel	NIS	???	???	3200	4200	4900
Italy	EUR	800	800	800	800	900
Latvia	EUR	n.a.	n.a.	300	300	400
Lithuania	EUR	n.a.	n.a.	300	300	400
Luxembourg	EUR	1600	1700	1700	1700	1800
Malta	EUR	n.a.	n.a.	600	700	800
Netherlands	EUR	1000	1000	1100	1200	1200

Poland	PLN	900	1100	1100	1300	1500
Portugal	EUR	400	400	400	400	500
Slovakia	EUR	n.a.	n.a.	300	400	400
Slovenia	EUR	600	600	600	600	700
Spain	EUR	700	700	700	700	800
Romania	RON	n.a.	n.a.	500	600	900
Sweden	SEK	9000	11200	11800	12000	13000
Switzerland	CHF	2000	2400	2300	2400	2400

14.7 SN Module (Waves 4, 6, 8, and 9)

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 waves 6, 8 and 9, 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 waves 6, 8 and 9 SN modules trace changes that occurred within the network. Following the naming of the current network members, respondents were asked about the the social network members named in previous waves SN modules and asked to link them to their current members, if they mentioned them again. If a person was not mentioned again, respondents were asked to select reasons. The linkage information for social network members between wave 4, wave 6, wave 8, and wave 9 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 SN module.

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-EzraIn 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.

14.8 SP and FT Module: List of Relations Differences due to SN

In waves 1, 2, 4, 5, 6, 7, 8 and 9 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 waves 6, 8 and 9, 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 12).

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

Table 12: Relationship Type Coding

Value or variable label	Wave 1, 2, 5 values	Wave 4 values	Wave 6, 7, 8, 9 values	Wave 4 variables
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 Step-child/your current partner's child			10 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

14.9 EP Module: 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	**Israeli questionnaire**
1) Retired	1) Retired
2) Employed or self-employed	2) Employed or self-employed
3) Unemployed	3) Unemployed, looking for a job
	4) Unemployed, not looking for a job
4) Permanently sick or disabled	5) Temporarily sick or disabled
	6) Permanently sick or disabled
5) Homemaker	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_rel7-0-0_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".

14.10 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 13: 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.*

15 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 the EURO-D scale on depression and the CASP-12 index on subjective well-being. Further details on scales and multi-item indicators in SHARE are provided in the "Scales and Multi-Item Indicators Manual".

Table 14: Generated Variable Modules of Waves 1 to 9

Generated-Variable-Modules	Content	W1	W2	W3	W4	W5	W6	W7	W8	W9
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 and calibrated weights	X	X	X	X	X	X	X	X	X
gv_imputations	Multiple Imputations	X	X		X	X	X	X	X	X
gv_isced	International Standard Classification of Education (ISCED-97/since wave 5 additionally ISCED-11)	X	X		X	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	X	X	X
gv_housing	Housing and NUTS codes	X	X		X	X	X	X	X	X
gv_networks	Information on social networks				X		X		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	X	X	X	X	X
gv_linkage	Linkage to Statutory German Pension Insurance data	Cross-wave module								
gv_dbs	Dried Blood Spots						X			
gv_big5	Big Five personality traits							X	X	X
gv_accelerometer_day/_hour/_total/_sleep	Physical activity								X	

15.1 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.

15.2 gv_weights and gv_longitudinal_weights

Giuseppe De Luca & Paolo Li Donni – University of Palermo

This chapter describes the weighting strategies used for dealing with problems of unit nonresponse and sample attrition in the most recent SHARE studies: the ninth regular wave of SHARE (SHARE Wave 9) and the second wave of the SHARE Corona Survey. The documentation is organized as follows. Section a) provides a brief overview of the key features of these two studies that are relevant for the purposes of our weighting strategies. Section b) focuses on the calibrated survey weights that attempt to compensate for the potential selection effects generated by unit nonresponse and attrition. Section c) describes the structure of the SHARE weight datasets, while Section d) describes the supplementary material on calibrated weights that users can exploit to compute their own weights.

a) Composition of the sample in Wave 9

a1) SHARE Wave 9

The ninth regular wave of SHARE was fielded between October 2021 and September 2022 by means of a Computer-Assisted Personal Interview (CAPI) administered in the same 28 countries that had already participated in the eighth regular wave of the SHARE panel. Ignoring the End-of-life interviews, SHARE Wave 9 collected data from 69,154 individual interviews in 47,957 households. The sample size available in each country ranges from a minimum of 731 observations for Cyprus and a maximum of 4,802 observations for Poland.

As discussed in Chapter 2, the uncertainty generated by the ongoing COVID-19 pandemic in 2021 prevented the drawing of new refreshment samples. Hence, the gross sample of Wave 9 can be viewed as a follow-up of the sample originally drawn in Wave 8. Note that, in addition to the longitudinal samples from previous waves and the national refreshment samples from

batches that were already fielded in Wave 8, it also includes national refreshment samples from batches that were not fielded before the suspension of the Wave 8 fieldwork due to the COVID-19 outbreak in spring 2020.

In total, eighteen countries have drawn a refreshment sample in Wave 8: Austria, Belgium, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Israel, Latvia, Poland, Portugal, Slovenia, Spain, Sweden, and Switzerland. The target population of Wave 9 has been defined as the 50+ population in 2019 that survives up to 2021 (i.e., the beginning of the data collection process in Wave 9) because the national gross samples of all these countries cover cohorts of people born in 1969 or earlier who were already age-eligible at the time of the latest refreshment samples in Wave 8. As for other longitudinal studies, this definition of the target population accounts for the sizeable effects of mortality between Wave 8 and Wave 9. It excludes cohorts of people born in 1970 and 1971, aged 50 and 51 years in 2021, that are not covered by the national gross samples due to the lack of new refreshment samples in Wave 9. The representativeness of the cohorts of people born in 1968 and 1969 remains problematic for the ten countries (Bulgaria, Cyprus, Greece, Italy, Lithuania, Luxembourg, Malta, the Netherlands, Romania, and Slovakia) that have not drawn refreshment samples in Wave 8. Issues related to the coverage of these cohorts will be addressed in the refreshment samples of Wave 10.

In Section 8.3, we shall see that these survey design features have important implications on the calibrated weights of Wave 9. For example, unlike the other regular waves of SHARE, the calibrated cross-sectional weights of Wave 9 and the calibrated longitudinal weights of wave combination 8-9 aim to reproduce the same target population. These two different sets of calibrated weights differ only in relation to their subsamples of respondents and their sets of population margins.

a2) Second SHARE Corona Survey

The second wave of the SHARE Corona Survey was designed to study the long-term impact of the COVID-19 pandemic. It was fielded about one year later than the first wave, between June and August 2021, by means of a Computer-Assisted Telephone Interview (CATI) administered in the same 28 countries that had already participated in the first wave of this study.

The first wave of this study collected data from 57,560 individual interviews in 38,960 households, while its second wave involved 49,254 individual interviews in 33,109 households. The cross-country average household attrition rate is 15 percent, with the lowest retention rate of 68 percent in Sweden and the highest retention rate of 94 percent in Lithuania. About 2 percent of the interviews in the second wave refer to new entries such as new spouses/partners of age-eligible respondents and nonresponding spouses/partners from the first wave that were eligible for the second wave. The balanced sample of respondents who have participated in both waves of the study includes 48,357 individuals.

By design, the first wave was administered to the longitudinal sample of Wave 8, but not to the refreshment sample. The sample of the second SHARE Corona Survey is a follow-up of

those households that participated in the first SHARE Corona Survey, without refreshment samples in any of the participating countries. Unlike release 8.0.0, the target population of the first wave has been re-defined as the 50+ population in 2016 (i.e., the time of the latest baseline/refreshment samples drawn in Wave 7) that survives up to 2020 (i.e., the beginning of the data collection process in the first SHARE Corona Survey). Similarly, the target population of the second SHARE Corona Survey is defined as the 50+ population in 2016 that survives up to 2021.

b) Calibrated weights

In the ideal situation of complete response, the availability of design weights allows the users to account for the randomness of the sampling process by compensating for unequal selection probabilities of the various sampling units. Unfortunately, properties of inferential procedures based on the sampling design weights depend on the assumption of complete survey response, which is almost never satisfied in the practical implementation of surveys. SHARE is not an exception to this common situation. The baseline and refreshment samples of each wave suffer from problems of unit nonresponse (Groves & Peytcheva 2008). Moreover, the longitudinal part of the sample is subject to attrition problems (Lynn, 2009). Because of these nonsampling errors, we discourage the users from relying on sampling design weights for standard analyses of the SHARE data. These weights are included in the public release of the SHARE weights database only to favour the implementation and comparison of alternative statistical procedures for handling nonresponse and attrition errors.

The baseline strategy adopted by SHARE to handle problems of unit nonresponse and attrition relies on the calibration approach proposed by Deville and Särndal (1992). This approach allows align the sample and population distributions of some benchmark variables without the need of specifying an explicit model for the nonresponse mechanism. Under the assumption that the missing data mechanism is missing-at-random (Rubin 1987), calibrated weights may help reduce the potential selection bias generated by nonresponse errors. Thus, unless these sources of nonsampling errors are controlled for in other ways, this is the type of weight that we generally recommend using in standard analyses of the SHARE data. In the remainder of this section, we first discuss the key methodological advantages and limitations of the calibration procedure. Next, we describe the implementation of the calibration procedure for constructing the various types of calibrated cross-sectional and longitudinal weights available in the public release of SHARE Wave 9 data.

b1) The calibration procedure

Let $U = \{1, \dots, i, \dots, N\}$ be a finite population of N elements, from which a probability sample $s = \{1, \dots, i, \dots, n\} \subseteq U$ of size $n \leq N$ is drawn according to a probability-based sampling design. Unless otherwise specified, we shall assume that the inclusion probability $\pi_i = Pr\,(i \in s)$ is known and strictly positive for all population units. To describe the basic ideas and the key properties of the calibration approach, we first consider the ideal situation of a complete response where all units within sample s agree to participate in the survey. Then, we relax this

ideal setup to describe the key implications of nonresponse errors on the properties of this weighting method.

The sampling design weights $w_i = \pi_i^{-1}$ are typically used to account for the randomness of the sampling process and the variability of the inclusion probabilities across sample units due to stratification and clustering strategies. For example, one can estimate the population total $t_y = \sum_{i \in U} y_i$ of a variable of interest y by the Horvitz-Thompson estimator:

$$\hat{t}_y = \sum_{i \in s} w_i y_i. \qquad (1)$$

Under the ideal setup of complete response, this estimator is known to be design unbiased, that is $E_p(\hat{t}_y) = t_y$, where $E_p(\cdot)$ denotes the expectation with respect to the sampling design.

Let us assume now that the sampling frame or other external sources such as census data and administrative archives provide supplementary data on a q-vector of categorical auxiliary variables $x_i = (x_{i1}, \dots, x_{iq})^\top$ with known population totals $t_x = \sum_{i \in U} x_i$. We shall refer to the auxiliary variables x_i as calibration variables and to their population totals t_x as calibration margins. The basic idea of the calibration approach is to determine a set of *calibrated weights* w_i^* that are as close as possible to the design weights w_i and that satisfy the constraints

$$\sum_{i \in s} w_i^* x_i = t_x. \qquad (2)$$

Thus, given a distance function $G(w_i^*, w_i)$ and the availability of survey data on $(w_i, x_i^\top : i = 1, \dots, n)$ and population data on the calibration margins t_x, the aim of the procedure is to determine the calibrated weights w_i^* by minimizing the aggregate distance $\sum_{i \in s} G(w_i^*, w_i)$ with respect to w_i^* subject to the q equality constraints in (2). Under some regularity conditions on the distance function $G(w_i^*, w_i)$ (see Deville and Särndal 1992), the solution of this constrained optimization problem exists, is unique and can be written as

$$w_i^* = w_i F(\eta_i), \; i = 1, \dots, n, \qquad (3)$$

where $\eta_i = x_i^\top \lambda$ is a linear combination of the calibration variables x_i, $\lambda = (\lambda_1, \dots, \lambda_q)^\top$ is the q-vector of Lagrangian multipliers associated with the constraints (2), and $F(\cdot)$ is a calibration function, which is uniquely determined by the distance function $G(w_i^*, w_i)$.

A key feature of the calibration approach is that many traditional re-weighting methods such as post-stratification, raking, and generalized linear regression (GREG) correspond to special cases of the calibration estimator

$$\hat{t}_y^* = \sum_{i \in s} w_i^* y_i \qquad (4)$$

for particular choices of the calibration function $F(\cdot)$ (or, equivalently, of the distance function $G(\cdot,\cdot)$). Deville and Särndal (1992) present various functional forms for $G(w_i^*, w_i)$ and $F(\eta_i)$. The chi-square distance function $G(w_i^*, w_i) = (w_i^* - w_i)^2 / 2w_i$, which leads to the widely used GREG estimator, has the advantage of ensuring a closed form solution for the calibrated weights w_i^*. However, this distance function is unbounded and depending on the chosen set of calibration variables it may also lead to negative weights. Different specifications of the

calibration function may avoid these issues, but the underlying optimization problems may not admit a solution and the Lagrange multipliers must be computed numerically. In SHARE, we rely on the logit specification of the distance function

$$G(w_i^*, w_i) \propto \left(\frac{w_i^*}{w_i} - l\right) \ln\left(\frac{w_i^*/w_i - l}{1 - l}\right) + \left(u - \frac{w_i^*}{w_i}\right) \ln\left(\frac{u - w_i^*/w_i}{u - 1}\right),$$

which leads to a calibrated function of the form

$$F(\eta_i; u, l) = \frac{l(u-1) + u(1-l)\exp(a\eta_i)}{u - 1 + (1-l)\exp(a\eta_i)},$$

where $a = [(1-l)(u-1)]^{-1}(u-l)$. Unlike other distance functions, these functional forms restrict in advance the range of feasible values for the calibrated weights by suitable choices of the lower bound l and the upper bound u. Specifically, if a solution exists, then it must satisfy the restriction $w_i l \leq w_i^* \leq w_i u$.

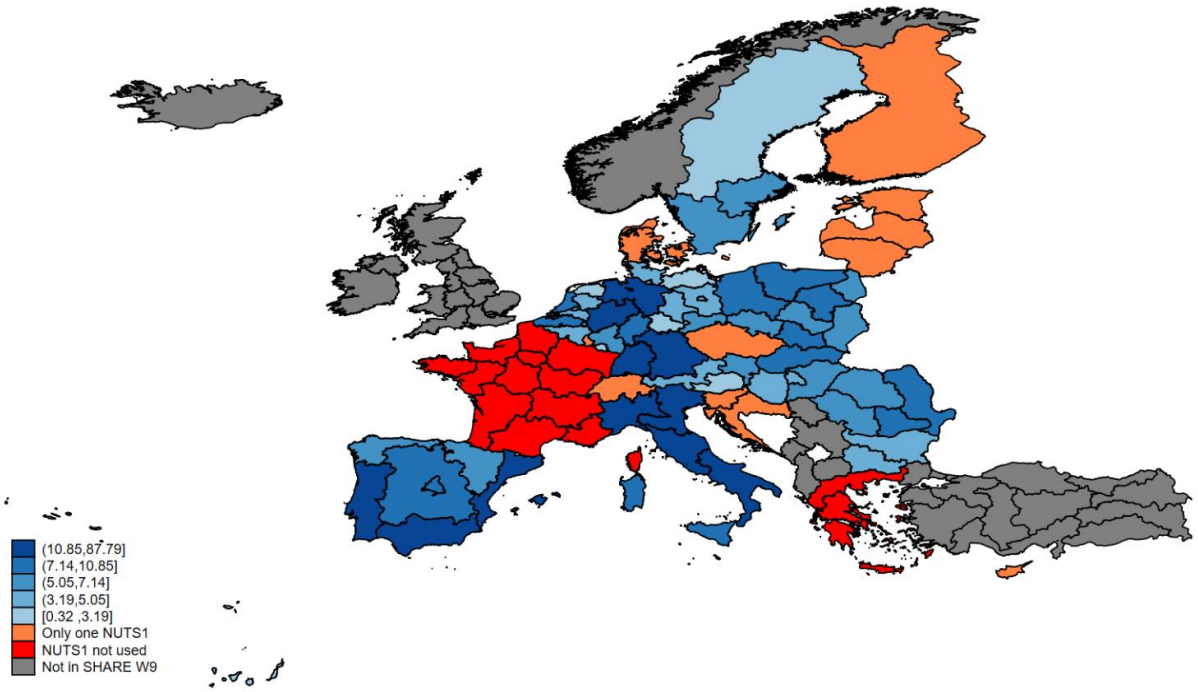
As discussed in Deville and Särndal (1992), effectiveness of the calibrated weights depends crucially on the correlation between the study variable y and the calibration variables x. In the extreme case when y can be expressed as a linear combination of x, it is clear that the calibrated estimator $\hat{t}_y^*$ gives an exact estimate of t_y for every realized sample s. Under suitable regularity conditions, the class of calibration estimators $\hat{t}_y^*$ satisfies other desirable asymptotic properties. For example, the estimators obtained by alternative specifications of the distance function are asymptotically equivalent to the GREG estimator based on a chi-squared distance function. Thus, in large samples, calibrated weights are robust to arbitrary choices of the calibration function $F(\cdot)$.

Unfortunately, this property does not necessarily extend to the more realistic cases where survey data are affected by nonresponse errors. Previous studies by Lundström and Särndal (1999) and Haziza and Lesage (2016) suggest that in these cases alternative specifications of the calibration function $F(\cdot)$ correspond in practice to imposing different parameterization of the relationship between response and calibration variables. Moreover, statistical properties of calibration estimators depend as usual on the validity of the missing at random assumption. Brick (2013), Molenberghs et al. (2015), Vermeulen and Vansteelandt (2015), and Haziza and Lesage (2016), among others, discuss a variety of robust weighting methods based on a propensity-score approach. One key issue in the implementation of these methods for SHARE is that selection probabilities and auxiliary variables are usually known for the subsample of respondents only.

b2) Calibrated cross-sectional weights of SHARE Wave 9

The calibrated cross-sectional weights of SHARE wave 9 were computed separately by country to match the size of the national 50+ populations in 2019 that survive up to 2021. In each country, we used a logit specification of the calibration function F(·) and a set of population margins for gender-age groups (i.e., males and females in the age classes ([50-59],[60-69], [70-79], [80+)). Mortality of the target population was taken into account by subtracting from each population margin the corresponding number of deaths between 2019 and 2021.

NUTS1 Population Margins for the Calibrated Cross-Sectional Weights of Wave 9



In 12 countries (Austria, Belgium, Bulgaria, Germany, Hungary, Italy, the Netherlands, Poland, Portugal, Romania, Spain, and Sweden), we included an additional set of population margins for the 2016 NUTS1 regional areas as illustrated in the figure above (Israel is excluded from the figure). This additional set of calibration margins was ineffective in all countries containing only one NUTS1 region. In Greece, NUTS1 calibration margins were excluded because of unsolved inconsistencies in the re-coding of NUTS1 codes over time. In Israel, where no NUTS nomenclature is available, we used instead an additional set of calibration margins for three population groups: Jewish Israeli, Arab Israeli, and immigrants from the former USSR. Population data about the calibration margins come from the Central Bureau of Statistics for Israel and the EUROSTAT regional database for all other countries.

As usual, calibrated cross-sectional 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. At the individual level, we assigned an individual-specific weight to each 50+ respondent that depends on the household design weight and the respondent's set of calibration variables (namely, gender, age class, and NUTS1 code). At the household level, we assigned instead a common calibrated weight to all interviewed household members which depends on the household design weight and the set of calibration variables for all 50+ respondents in that household.

By construction, calibrated cross-sectional weights are missing for respondents younger than 50 years (i.e., age-ineligible partners of an age-eligible respondent), for those with missing information on the calibration variables, and for those with missing sampling design weights (i.e., respondents from households for which we do not have sampling frame information). However, the number of these cases is negligible.

b3) Calibrated longitudinal weights

In addition to calibrated cross-sectional weights, the SHARE release 9.0.0 also includes calibrated longitudinal weights for the purposes of panel data analyses. Although these weights are based on the same calibration procedure, they differ from the cross-sectional weights in two important respects. First, calibrated longitudinal weights are usually computed for the balanced subsample of respondents who have participated in at least two waves of the study. Second, since mortality is a source of attrition that affects both the sample and the population, calibrated longitudinal weights aim to reproduce the target population at the beginning of a reference period that survives up to the end of the period considered (see, e.g., Lynn, 2009). As discussed above, SHARE Wave 9 is somehow an exception. Due to the lack of new refreshment samples in Wave 9, its target population coincides with that reproduced by the calibrated longitudinal weights of the wave combination 8-9. However, these two sets of calibrated weights differ in relation to their subsamples of respondents and their sets of population margins.

To simplify the structure of the public release of the data, we still provide calibrated longitudinal weights only for selected wave combinations of the SHARE panel. Those available in the SHARE release 9.0.0 are the 8 possible couples of any two adjacent waves (i.e., the wave combinations 1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, and 8-9) and the fully balanced panel (i.e., the wave combination 1-2-3-4-5-6-7-8-9). The weights of the generic wave combination t-…-s were always computed separately by country to represent the national 50+ populations of Wave t that survive up to the interview year of Wave s. For example, the wave combination 1-2 allows representing the 50+ national populations in 2004 that survive up to 2006, while the fully balanced panel allows representing the national 50+ populations in 2004 that survive up to 2021.

For the calibrated longitudinal weights of two adjacent waves, we used a logit specification of the calibration function $F(\cdot)$ and a set of calibration margins for the size of the target population across eight gender-age groups (i.e., males and females with age at the time of the starting wave in the four classes [50-59],[60-69], [70-79] and [80+)). Compared to the cross-sectional weights of previous waves, we did not control for the 2016 NUTS1 calibration margins due to the smaller number of observations usually available in the national longitudinal subsamples. Moreover, we always accounted for the mortality of the target population by subtracting from each calibration margin the corresponding number of deaths between Waves t and s.

For the calibrated longitudinal weights of the fully balanced panel, we further restricted the set of calibration margins to 6 gender-age groups (i.e., males and females with age in 2004 in the three classes [50-59],[60-69], and [70+)).

As with the calibrated cross-sectional weights, calibrated longitudinal weights are available both at the individual level and at the household level. For the individual weights, the balanced sample consists of respondents interviewed in each wave of the selected wave combination. For the household weights, the balanced sample consists of households with at least one

eligible member interviewed in each wave of the selected wave combination. These definitions imply that the balanced sample of households is larger than the balanced sample of individuals. For example, couples with one partner participating in Wave 8 and the other partner participating in Wave 9 belong to the balanced sample of households for the wave combination 8-9, even though none of the two partners belongs to the corresponding balanced panel of individuals.

b4) Calibrated cross-sectional and longitudinal weights of the SHARE Corona Survey

The SHARE release 9.0.0 includes two sets of calibrated cross-sectional weights for the first two waves of the SHARE Corona Survey and a set of calibrated longitudinal weights for the balanced panel of respondents who participated in both waves of the study.

A description of the calibrated cross-sectional weights for the first SHARE Corona Survey can be found in De Luca et al. (2021). As for release 8.0.0, the new release 9.0.0 includes separate sets of calibrated weights for the CAPI, CATI, and CAPI&CATI subsamples. The target population of the last two subsamples has been however redefined as the 50+ population in 2016 that survives up to 2020. As usual, the calibrated cross-sectional weights of each subsample were computed separately by country using a logit specification of the calibration function, a first set of population margins for the gender-age groups (i.e., males and females in the age classes [50-59],[60-69], [70-79], [80+)), and a second set of population margins for the 2016 NUTS1 regional areas. The weights of each subsample were also defined 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.

For the calibrated cross-sectional weights of the second SHARE Corona Survey, we maintained the distinction between individual-level and household-level weights, but not the distinction between the CAPI, CATI, and CAPI&CATI subsamples. These weights were computed for the cross-sectional sample of 49,254 respondents and 33,109 households who participated in the CATI of the second wave, irrespective of whether they also participated in the CATI of the first wave. The population margins are like those of the calibrated cross-section weights of the first wave, but they now refer to the national 50+ populations in 2016 that survive up to 2021.

Calibrated longitudinal weights were computed for the balanced panel of 48,357 respondents and 33,109 households who participated in the first and second SHARE Corona Survey. Compared to the two cross-sectional samples, this sample excludes the 9,203 respondents who participated only in the first and the 897 respondents who participated only in the second SHARE Corona Survey. The target population coincides with that of the second wave, but the calibrated longitudinal weights were constructed by controlling for the population margins of the gender-age groups only.

c) Structure of the SHARE weight datasets

The release of waves 1 to 9 contains 18 datasets of calibrated weights. The first nine datasets contain the cross-sectional weights of any single wave from 1 to 9 (named as sharewt_rel9-0-0_gv_weights for the generic Wave t=1,…,9) and are provided together with the other

modules of each corresponding wave. The remaining datasets with the calibrated longitudinal weights of the selected wave combinations 1-2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8, 8-9, 1-2-3-4-5-6-7-8-9 (named as sharewX_rel9-0-0_gv_longitudinal_weights_w*tws* for the generic wave combination t-…-s) are delivered as separate download files. Tables 5, 6 and 7 of the appendix provide a description of the variables included in these 18 datasets. In addition to the individual, household and country identifiers, each dataset includes:

- one variable for the relevant sampling design weights;
- two variables for the underlying type of calibrated weights at the household and the individual levels;
- one indicator which identifies the various subsamples drawn in any specific country and wave of the SHARE panel;
- four indicators for the information about stratification and clustering in each subsample.

Longitudinal datasets also contain a binary indicator that allows identifying the balanced sample of individuals within the balanced sample of households in the chosen wave combination. In this case, sampling design information (e.g. design weights, strata and clusters) refers to the initial wave of the chosen wave combination.

d) Supplementary material and user guide on calibrated weights

Since the SHARE panel now consists of nine waves, one can compute many different types of calibrated longitudinal weights depending on the selected combination of waves and the selected unit of analysis (either individuals or households). In addition, one can compute many different types of calibrated cross-sectional weights for specific subsamples of the data collected in each regular wave of the panel or other related studies, such as the SHARELIFE interviews of waves 3 and 7 or the two waves of the SHARE Corona Survey. These considerations make it clear why the strategy of providing all possible calibrated cross-sectional and longitudinal weights is not feasible, especially in the future when additional waves will be available. For cross-sectional studies based on specific subsamples and longitudinal studies based on other wave combinations, users are required to control for the potential selection effects of unit nonresponse and attrition by computing their own calibrated weights or by implementing some alternative correction methods.

To support users in the nontrivial methodological task, we provide a set of Stata do-files and ado-files that illustrate step-by-step how to compute calibrated cross-sectional and longitudinal weights. Our supplementary material on calibrated weights also includes a dataset with updated information on population size and number of deaths by year, gender, age, and NUTS1 code. Registered users can download this supplementary material on calibrated weights from the SHARE data dissemination website, under the link “Generate Calibrated Weights Using Stata (2020)”. A discussion of these step-by-step operations can also be found in the accompanying user guide “Computing Calibrated Weights”.

Please find the list of references of this chapter at the end of the appendix. For further questions on weights in SHARE, please contact giuseppe.deluca@unipa.it or paolo.lidonni@unipa.it.

15.3 gv_imputations

Author: Giuseppe De Luca

This chapter focuses on the construction of multiple imputations for the missing values due to item nonresponse errors. Specifically, we describe the imputation model for Wave 9 in Section a, and the imputation model for the second SHARE Corona Survey in Section b.

a) Imputations of missing values in SHARE Wave 9

Imputations of missing values due to item non-response errors in the regular face-to-face interview of Wave 9 were constructed using the same procedure adopted in the previous regular waves of SHARE (see, for example, De Luca et al., 2015). Of course, we adapted the imputation model to the specific features of the regular Wave 9 interview in terms of branching, skip patterns, proxy interviews, country-specific deviations from the generic version of the questionnaire, and availability of partial information from the sequence of unfolding bracket questions. However, we also attempted to preserve as much as possible the comparability of the imputations across different waves of the SHARE panel. The imputation procedure is essentially based on either the hot-deck method or the fully conditional specification (FCS) method depending on the prevalence of missing values for the variables collected in the regular interview of Wave 9.

a1) Hot-deck imputations. In SHARE, we always used the hot-deck method for variables affected by negligible fractions of missing values (usually, much less than 5 per cent of the respondents eligible to answer a specific item on the CAPI questionnaire). This method consists of replacing the missing values in one or more variables for a non-respondent (called the recipient) with the observed values in the same variables obtained from a respondent (called the donor) who is "similar" to the recipient according to some metric (see, for example, Andridge and Little, 2010).

In Wave 9, we computed hot-deck imputations in an early stage, separately by country, and according to a convenient order that accounts for branching and skip patterns in the various modules of the CAPI questionnaire. Donors were selected randomly within imputation classes based on observed auxiliary variables. We imputed first basic socio-demographic characteristics such as age and year of education, which contained very small fractions of missing values. These characteristics were then used as auxiliary variables to impute other variables. Our baseline set of auxiliary variables consisted of country, gender, five age classes ([– 49], [50 – 59], [60 – 69], [70 – 79], [80+]), five groups for years of education ([– 5], [6–10], [11–15], [16–20], [21+]), and two groups for self-reported good/bad health. For some variables, we exploited a larger set of auxiliary variables. For example, we also used the number of children to impute the number of grandchildren and an indicator for being

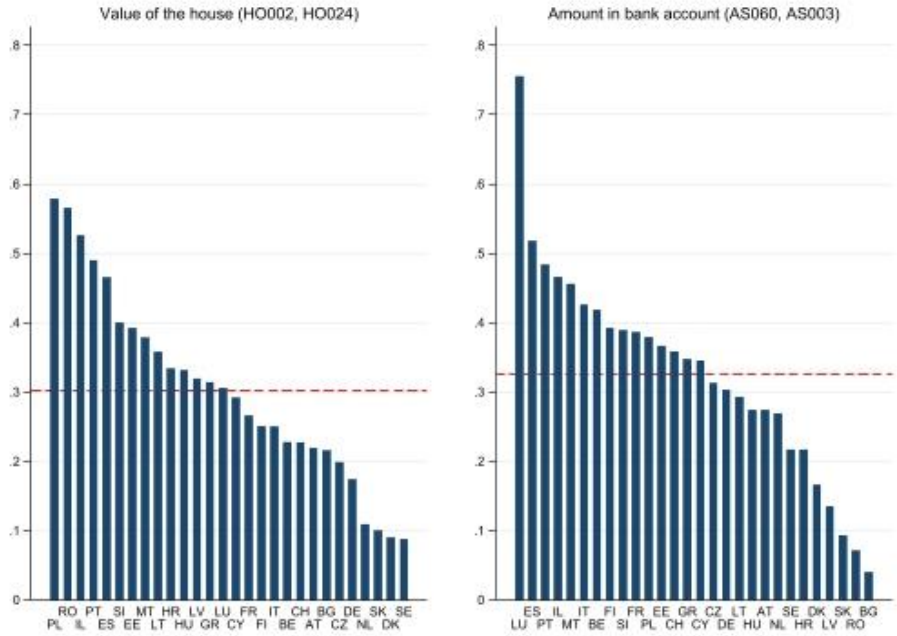
hospitalized overnight during the last year to impute other health-related variables. Variables that are known to be logically related, such as respondent's weight, height, and body mass index, were imputed jointly.

a2) FCS imputations. In the second stage of the imputation procedure, we dealt with the more worrisome issue of item non-response in monetary variables, such as income from various sources, real and financial assets, and consumption expenditures, which were collected by retrospective and open-ended questions that are sensitive and difficult to answer precisely.

Figure 2 shows the item non-response rates of two monetary variables: "Value of the house" (HO002, HO024), and "Amount in bank accounts" (AS060, AS003). For the first variable, the percentage of missing values among the eligible respondents ranges from a minimum of 8 per cent in Denmark and Sweden to a maximum of 57 per cent in Poland (30 per cent on average). For the second variable, the item non-response rate ranges from a minimum of 4 per cent in Bulgaria to a maximum of 76 per cent in Luxemburg (32 per cent on average). Similar patterns of item nonresponse were also observed in the previous waves (see, for example, De Luca et al. 2021). Thus, item nonsampling errors show some degree of persistency both over time and over country.

Since Wave 1, we handled these sizeable fractions of missing values on monetary variables by the FCS method of van Buuren et al. (1999). This method exploits a Gibbs sampling algorithm that imputes a set of variables jointly and iteratively through a sequence of regression models. Assume we want to impute arbitrary patterns of missing values on a set of J variables. At each step of the iterative process, we impute the missing values on the jth variable ($j=1,\ldots,J$) by drawing from the predictive distribution of a regression model that includes as predictors the most updated imputations of the other $J-1$ variables (as well as other fully observed predictors). The process is applied sequentially to the whole set of J variables and is repeated in a cyclical manner by overwriting at each iteration the imputed values computed in the previous iteration. Despite a lack of rigorous theoretical justification (see, for example, Arnold et al., 1999, 2001; and van Buuren, 2007), the FCS method is one of the most popular multivariate imputation procedures due to its flexibility in handling complicated data structures and its ability to preserve the correlations of the imputed variables (Raghunathan et al., 2001; van Buuren et al., 2006). Comparisons of the FCS method with other multivariate imputation techniques can be found in Lee and Carlin (2010).

Item nonresponse rates for "Value of the house" and "Amount in bank accounts" by country



In Wave 9, we computed FCS imputations separately by country and household type. The household types considered were singles and third respondents (sample 1), couples with both partners interviewed (sample 2), and all couples with and without a non-responding partner (sample 3). The distinction between the first two samples was primarily motivated by the fact of using socio-demographic characteristics of the partner of the designed respondent as additional predictors to impute the missing monetary amounts within couples. The overlapping partition of the last two samples was instead motivated by the need to impute properly total household income in the couples with a non-responding partner.

The set of monetary variables imputed jointly in the Gibbs sampling algorithm was country and sample-specific as we required a minimum number of donor observations for estimating the regression model associated with each variable. Variables that did not satisfy this requirement were imputed first (either by hot-deck or by regression imputations) and then used as fully observed predictors for computing the FCS imputations of missing values in the other monetary variables.

The imputation of each monetary variable was typically based on a two-part model that involved a probit model for ownership and a linear regression model for the amount conditional on ownership. Depending on eligibility and ownership, we converted (if needed) non-zero values of monetary variables in annual euro amounts to avoid modelling differences in the time reference periods of the various variables and the national currencies of non-euro countries.

In an early stage of the imputation process, we also symmetrically trimmed 2 per cent of the complete cases from the country-specific distribution of annual euro amounts to exclude (and then impute) outliers that may have a large influence on survey statistics. Moreover, we

applied logarithm or inverse hyperbolic sine transformations to reduce skewness in the right tails of the conditional distribution of each monetary variable.

The set of fully observed predictors was also sample-specific. For singles and third respondents (sample 1), our set of predictors consist of 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 (as perceived by the interviewer in the IV module of the CAPI instrument). For couples with both partners interviewed (sample 2), we added a similar set of predictors for the partner of the designed respondent. For couples with a non-responding partner (those remaining in sample 3 after excluding the couples in sample 2), we restricted the additional set of predictors referring to the non-responding partner to age and years of education only.

Imputations of the monetary amounts were always constrained to fall within individual-level bounds that incorporated the partial information available on the missing observations (e.g., country-specific thresholds used to trim outliers in the tails of the observed distribution of each monetary variable, bounds obtained from the sequence of unfolding bracket questions asked by design to non-respondents of open-ended monetary variables and lower bounds based on the observed components of aggregated monetary variables).

As usual, the imputation of total household income received special attention because the CAPI questionnaire provides two alternative measures of this variable. The first measure (*thinc*) can be obtained by a suitable aggregation at the household level of all individual income components, while the second (*thinc2*) can be obtained via the one-shot question on monthly household income (HH017). As argued by De Luca et al. (2015), it is not easy to find strong arguments to prefer one measure over the other. Moreover, the availability of two alternative measures may greatly improve the imputation process because each measure could contribute relevant information on the missing values of the other measure. To avoid understating the first measure of total household income in couples with a non-responding partner, we adopted the following three-stage algorithm:

Stage 1. For singles and third respondents (sample 1), we first imputed all monetary variables by the FCS method as discussed above. At the end of each iteration of the Gibbs sampling algorithm, we also computed total household income (*thinc*), household net worth (*hnetw*), and total household expenditure (*thexp*) by suitable aggregations of the imputed income, wealth, and expenditure items. Next, we imputed the second measure of total household income (*thinc2*) using as predictors *thinc*, *hnetw*, *thexp*, and the set of socio-demographic characteristics of the household respondent. The imputed values of *thinc2* were constrained to fall in the bounds derived from the sequence of unfolding bracket questions for the variable HH017.

Stage 2. For couples with both partners interviewed (sample 2), the imputation strategy is similar to the one adopted in stage 1 for the sample of singles and third respondents (sample 1). The only difference is that, at each iteration of the Gibbs sampling algorithm, we employed a larger set of predictors that also included the socio-demographic characteristics and the

most updated imputations of the monetary variables of the partner of the designed respondent.

Stage 3. Imputations of all monetary variables for the subsample of couples with both partners interviewed were obtained in stage 2. In stage 3, these couples were included in the imputation sample only as donor observations to impute the missing values in monetary variables for the remaining subsample of couples with a non-responding partner. As before, we imputed first all monetary variables for the responding partners using the FCS method. Unlike stage 2, the predictors referring to the non-responding partner now consisted, however, of age and years of education only. At the end of each iteration of the Gibbs sampling algorithm, we also imputed the *thinc2* using *hnetw*, *thexp*, and socio-demographic characteristics of the responding partner as predictors and the bounds obtained from the sequence of unfolding bracket questions for the variable HH017. Finally, we imputed *thinc* using *thinc2*, *hnetw*, *thexp*, and the set of socio-demographic characteristics of the responding partner as predictors, couples with two partners interviewed as donors, and the sum of imputed individual income sources of the responding partner as a lower bound.

To account for the additional variability generated by the imputation process, we provide five imputations of the missing values by independent replications of the hot-deck and FCS methods. Notice that neglecting this additional source of uncertainty by selecting only one of the five available replicates in the generated dataset of imputations may result in misleadingly precise estimates. After an initial set of burn-in iterations, convergence of the Gibbs sampling algorithm for FCS imputations was assessed by the Gelman-Rubin criterion (see, e.g., Gelman and Rubin 1992, and Gelman et al. 2004) applied to the mean, the median, and the 90th percentile of the five imputed distributions of each monetary variable.

b) Imputations of missing values in the second SHARE Corona Survey

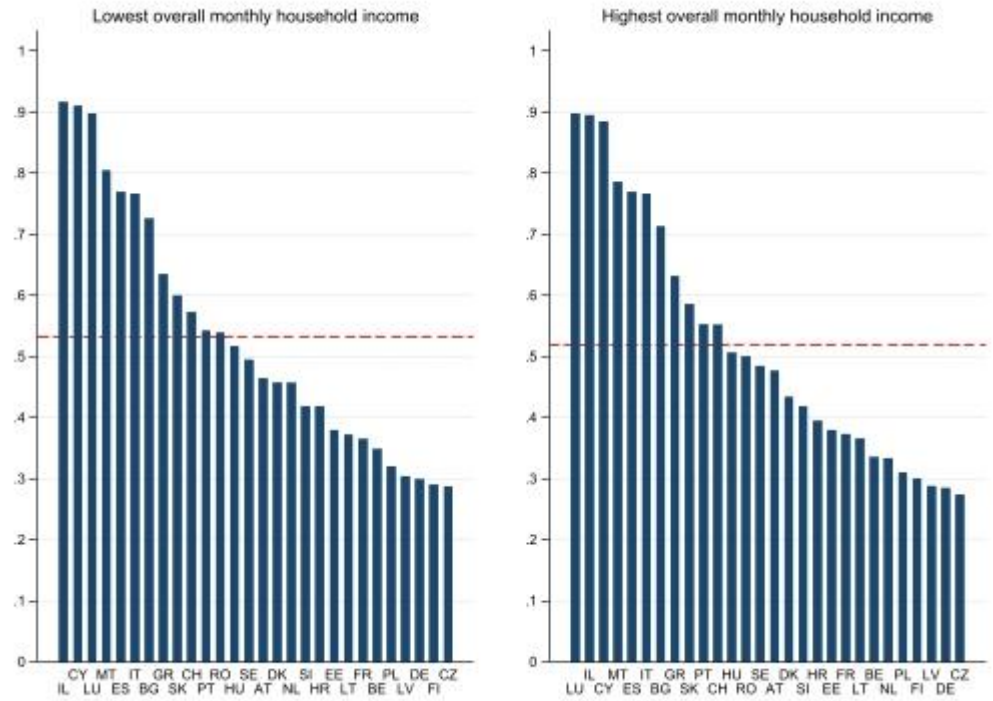
Since item nonresponse rates in the CATI data of the second SHARE Corona Survey were generally much less than 5 per cent, most variables were imputed by the hot-deck method. We used the FCS method only for 15 variables collected in Section E (Economic situation) and Section W (Work) of the questionnaire administered in the second wave. As for the first SHARE Corona Survey, the variables collected in these two sections suffer from somewhat larger amounts of item nonresponse errors. Moreover, Section E contains missing data by design due to the presence of a routing error in its filter variable CAE001 (see De Luca et al. 2021). Regarding possible issues of data comparability across the two waves of this survey, we note that seemingly similar questions may present relevant differences in terms of question wording, answer categories, time-reference period, branching, and skip patterns. To mark these differences within the generated dataset of imputations, we assigned slightly different variable names to items whose comparability is more doubtful.

b1) Hot-deck imputations. We first computed hot-deck imputations separately by country and according to a convenient order of the variables that accounts for branching and skip patterns in the CATI questionnaire of the second wave. The imputation classes for this method were generally based on the following set of auxiliary variables: country, gender, five age

classes ([– 49], [50– 59], [60– 69], [70 – 79], [80+]), a binary indicator for respondents living with a spouse/partner, five groups for years of education ([– 5], [6–10], [11–15], [16–20], [21+]), a binary indicator for good self-perceived health, and a binary indicator for changes in the self-perceived health status during the last three months. The first four auxiliary variables are fully observed, while the last three auxiliary variables contain very small fractions of missing values that were imputed first using only the first four variables. For some variables, we employed a larger set of auxiliary variables. For example, we used one additional binary indicator for keeping distance from others in public when imputing several variables included in Section H (Health and health behavior), Section C (Corona-related infection), and Section Q (Quality of healthcare) of the CATI questionnaire of the second SHARE Corona Survey. Furthermore, we jointly imputed missing values on the variables that are logically related. For example, we jointly imputed variables related to illness or health conditions since the last interview (CAH004) in Section H, those related to the COVID-19 symptoms (CAC102, CAC103) in Section C, and those related to forwent medical treatment since the outbreak (CAQ105 and CAQ106) in Section Q. In total, we imputed sequentially about 200 variables. As for the hot-deck imputations of the CAPI data collected in the regular SHARE waves, the imputation databases of the first and the second SHARE Corona Surveys contain five multiple imputations of the missing values and a flag variable for each imputed variable which allows the users to identify the imputed observations.

b2) FCS imputations. After hot-deck imputations, we constructed FCS imputations for fifteen variables: four of them related to changes in hours of work (namely CAW121, CAW122, CAW124, and CAW125), and the other eleven related to changes in the financial situation of the household (namely CAE100, CAE105, CAE107, CACO107, CAE111, CAE112, CAE103, and CAE104). As shown in Figure 3, the two most worrisome variables are the lowest (CAE107) and the highest (CAE105) overall amounts of monthly household income after taxes and contributions. In particular, the first respondent of each household was first asked whether monthly household income had been the same every month since the last interview (CAE100). Respondents who provided a negative answer to this question were then asked to report the lowest and the highest overall amounts of monthly household income. The unweighted cross-country average of the item nonresponse rates for these two variables are 53 and 51 per cent, respectively. In Luxemburg, Israel, Cyprus, and Malta, where the item nonresponse rates are around 90 per cent, we adopted a country-pooling strategy to increase the extremely low number of donors.

Item nonresponse rates for lowest and highest overall monthly household income by country



Except for these more problematic cases, FCS imputations were constructed separately by country. At each iteration of the Gibb sampling algorithm, we used a linear regression model for the continuous variables (CAE105 and CAE107), a simple hot-deck method for the lowest and the highest hours of work (CAW122 and CAW125), a logit model for five binary variables (CAW121, CAW124, CAE100, CAE111, and CAE112), a multinomial logit model for the categorical variable CACO107, and a multivariate hot-deck method for the six binary indicators related to financial support received since the outbreak (CAE103 and CAE104). For the variables CAE105 and CAE107, we symmetrically trimmed 2 per cent of the complete cases from the country-specific distribution of each variable to exclude (and then impute) outliers that may have a large influence on survey statistics. In addition to the variables imputed jointly within the Gibb sampling, our baseline set of observed predictors consists of age, years of education, and binary indicators for female respondents, living with a spouse/partner, and good self-perceived health. For all variables of Section E, we also used a binary indicator for being retired. For the variables imputed by either simple or multivariate hot-deck methods, all continuous predictors within the Gibb sampling were discretized to form the imputation classes. In some cases, we imposed a set of country and item-specific exclusion restrictions to avoid possible problems of collinearity, imprecise estimates, and convergence problems in the context of non-linear models. As for the other types of imputations provided by SHARE, we always provide five multiple imputations of the missing values. After an initial set of burn-in iterations, convergence of the Gibbs sampling algorithm was assessed by the Gelman – Rubin criterion applied to the mean, median, and 90th percentile of the distribution of each continuous variable and the mean of the distribution of each discrete variable.

Please find the list of references of this chapter at the end of the appendix. For further questions on imputations in SHARE, please contact giuseppe.deluca@unipa.it.

15.4 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, gv_isced from wave 5 onwards also contains 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.

The mapping of country-specific educational degrees to the corresponding ISCED code is available here for: wave 1, wave 2, wave 4, wave 5, wave 6, wave 7 and wave 8.

15.5 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 15: Generated Variables of gv_health

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

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

15.6 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 16: Generated Housing Variables

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

* *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). Because of privacy legislations, only NUTS1 level is provided. NUTS1 is **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.

15.7 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).

When children are mentioned as part of the respondents' social network, release 9.0 0 introduced two variables in gv_networks and the SN module of waves 6, 8, and 9 to facilitate the linkage to the CH module: *sn_child_loop_X* and *sn_childid_X*.

Table 17: Generated Variables in Wave 4 gv_networks

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

[1] e.g. *famnet1* = number of family members; *famnet2* = dummy based on *famnet1*; *famnet3* = percentage of family members in social network**.**

Table 18: Generated Variables in Wave 6, 8 and 9 gv_networks

Variables	Description
sn_size_w#	SN size wave 4/6/8 (0-7)
panel_change_size	in wave 6: w6 sn size - w4 sn size (if both done) in wave 8: w8 sn size – w6 sn size (if both done)
panel_status	In which waves was the SN modules done
panel_lost / new / continued	Count of lost/new/continued SN members
panel_changePattern_size *panel_changePattern_childnet* *panel_changePattern_friendnet*	Change patterns in SN between w4, w6 and w8 (only available in wave 8)
sn_scale	Scale of social connectedness, a summary scale of the social network data (higher = higher connectedness)
social_integration	Index of social integration, a summary index of the social network data (higher = higher social integration)
spousenet2	Is there a spouse in the social network?
famnet / childnet / siblingnet / parentnet / friendnet / formalnet / othernet	Count of different role relationship categories in a social network
womennet / mennet	Count of women/men in a social network
prx	Information on geographical distance to network members
contact	Information on contact frequencies with network members

Variables	Description
close	Information on emotional closeness to network members
year_mean	Average year of birth of network members
fin* / *gift* / *care	Information on given or received financial / personal help to / from network members
sn_satisfaction	Satisfaction with personal network (1-10)
partner_var	Relationship status, used to identify for whom the *spousenet2* variable does not apply because no partner was reported by the respondent.
W#_sn_mentioned_again_X	Was w4 SN member X mentioned again in w6?
W#_sn_mentioned_before_X	Was w6 SN member X mentioned before in w4?
w6_sn_w4_position_X *w8_sn_w6_position_X* *w8_sn_w4_position_X* *w9_sn_w6_position_X* *w9_sn_w8_position_X*	What was the position of w6 SN member X in the w4 SN? What was the position of w8 SN member X in the w6 SN? What was the position of w8 SN member X in the w4 SN? What was the position of w9 SN member X in the w6 SN? What was the position of w9 SN member X in the w8 SN?
sn_person_X	Was there an SN person x in the network?
rel_X	Relationship with each social network member
gender_X	Gender of each social network member
prx_X	Geographic proximity of each social network member
contact_X	Frequency of contact with each social network member
close_X	Emotional closeness of each social network member
year_X	Year of birth of each social network member
occ_X	Occupation of each social network member
occ_det_X	Detailed occupation of each social network member
partner_X	Partner status of each social network member
partner_det_X	Detailed partner status of each social network member
fin* / *gift* / *care _sn_X	Given or received financial / personal help to / from each social network member
sn_child_loop_X	Child link: sn person X mentioned as child (loop)
sn_childid_X	Child link: sn person X child identifier (fix across modules and waves)

15.8 gv_exrates

This module contains currencies (also pre-Euro) and exchange rates for non-Euro countries for SHARE waves 1 to 8. Additionally, the module stores nominal exchange rates as well as exchange rates that adjust for purchasing power parity (ppp) for the years 2003 to 2020. 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.

Differently from previous releases, two series of PPP exchange rates are provided, in current and constant terms. Current PPPs (pppc) are meant to be used for cross-sectional country comparisons of financial variables, with Gemany taken as a reference country. Constant PPPs (pppk) allow to compare financial amounts both between countries and over time: while preserving price relatives between countries in each year, like current PPPs, they also take into account annual inflation as measured by changes in the harmonized index of consumer prices in the reference country (Germany 2015=1). Therefore, they are conceptually broadly equivalent to PPP exchange rate series provided in previous releases, although figures cannot be directly compared because of major methodological differences in the construction.

PPP-adjustment to transform financial amounts into real ones can be done for both current and constant PPP series via:

financial_variable nomx#/ ppps#*

where *s* is either *c* or *k* for current or constant PPP rates, respectively.

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.

15.9 gv_job_episodes_panel

The Job Episodes Panel (JEP) is a generated dataset that contains the labor 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 here.

15.10 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).

15.11 gv_isco

ISCO codes for respondents, their former partner's and their parents' occupation are available in the "gv_isco" module only 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 EU.

Table 19: Variables Used for ISCO and NACE Coding

	ISCO	NACE	
		employed	self-employed
Respondent's first job	*ep016_1*	*ep018_1*	*ep023_1*
Respondent's second job	*ep016_2*	*ep018_2*	*ep023_2*
Respondent's last job	*ep052_*	*ep054_*	*ep060_*
Former partner's job	*dn025*		
Mother's job	*dn029_1*		
Father's job	*dn029_2*		

For SHARE waves 2, 4, and 5, the first digit of both the ISCO and the NACE classification (based on NACE Rev.1.1 of 2002) are available for specific variables in the ep module: the respondent's current job (*ep016* and *ep018*) and last job (*ep052* and *ep054*).

From wave 6 onwards, an in-field coding tool was used to capture the ISCO-08 classification. In the regular panel waves, ISCO codes are available for the respondent's current (*ep616isco*) and last main job (*ep152isco*) as well as for the jobs of the respondent's mother/father at the time when the respondent was aged 10 (*dn029isco_1*/*dn029isco_2*). In wave 7, ISCO codes are available for the respondent's current job (*ep616isco*) and the respondent's retrospective employment history (*re012isco_**).

15.12 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 15.9) 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: Belloni, M., Carrino, L., Orso, C.E., Buia, R.E., Cavapozzi, D., Pasini G., and Brugiavini A. (2016): "*Internationally comparable measures of individual Social Security Wealth in SHARE Wave 4*". SHARE Working Paper Series: 24-2016

15.13 gv_deprivation

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

Table 20: Items Used for Material Deprivation Index (*depmat*)

Item	Variable	Description
Meat	br033_	Eat meat, fish or chicken more often [*than three times per week*] because: *you cannot afford to eat it more often*
Fruit	br034_	Eat fruits or vegetables more often [*than three times per week*] because: *you cannot afford to eat it more often*
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 [*AffordExpenseAmount*] without borrowing any money?
In the last twelve months, to help you keep your living costs down, have you...		
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?

Table 21: Items Used for the Social Deprivation Index (depsoc)

Item	Variable(s)	Description
Room	ho032_, hhsize	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.

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.

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.

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, 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.*

15.14 gv_children

In wave 6, 7 and 8, 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 CAPI modules CH, SN (only waves 6 and 8), SP and FT. Please be aware that the gv_children variables are an aggregate of information from within waves 6, 7 and 8 but not 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. A generated module that aggregates child information from all waves of SHARE is planned to be published in one of the upcoming data releases. Also, note that wave 7 SHARELIFE information on children (RC module) was not included in gv_children. Please be aware that information collected on children in SHARELIFE differs from regular CH information. In SHARELIFE, respondents are asked to mention all children, including deceased children, whereas in the regular CH module the family respondent is asked about children who are still alive on behalf of both partners.

SHARE respondents can mention up to 20 children in the regular CH module, 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_isced module that summarises the educational attainments of children as well as respondents and their partners (see chapter 15.4). Table 22 contains an overview of the gv_children variables, their content, as well as the completeness of the information they provide.

To facilitate the linkage of children between CH & SN module, a new variable was added to gv_children and the CH module of waves w6, 8, and9: *child_sn_loop_**

Table 22: Generated Variables in gv_children

Variable name	**Information complete, values transferred to cohabitating partner**
*ch_gender_**	Gender of child *
*ch_yrbirth_**	Year of birth of child *
*ch_relation_**	Relation to child * (e.g. natural child, child of partner, adopted child, etc.)
*ch_occupation_**	Occupational status of child *
*ch_fin_gave_**	Financial help given to child *
*ch_fin_received_**	Financial help received from child *
*ch_gift_received_**	Financial gift received from child *
*ch_gift_gave_**	Financial gift given to child *
*child_sn_loop_**	Child X mentioned as social network member (loop)
Variable name	**Values transferred to cohabitating partner, information gathered in previous waves can be missing**
*ch_proximity_**	Residence of child *
*ch_move_out_year_**	Year when child * left the parental household
*ch_marital_status_**	Marital status of child *
*ch_partner_status_**	Partner status of child * (if not married/reg. partnership)
*ch_number_of_children_**	Number of children of child *
*ch_yrbirth_youngest_child_**	Year of birth of youngest child of child *
*ch_school_education_**	Highest school leaving certificate of child *
*ch_further_education_*_**	Further educational degrees of child * (multiple answers possible)
Variable name	**Information from respondent only (values not transferred to cohabitating partner)**
*ch_contact_**	Frequency of contact to child *
*ch_closeness_**	Emotional closeness to child * (only asked for children mentioned in social network module)

*ch_babysit_**	Did respondent look after child of child *?
*ch_outhh_receive_care_**	Personal/practical help received from child * who lives outside the respondent's household
*ch_outhh_gave_care_**	Personal/practical help given to child * who lives outside the respondent's household
*ch_hh_gave_care_**	Personal/practical help given to child * who lives within the respondent's household
*ch_hh_receive_care_**	Personal/practical help received from child * who lives within the respondent's household

15.15 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 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 (https://www.eservice-drv.de/FdzPortalWeb/antragAction.do). Further information on access conditions as well as user guide and codebook for SHARE-RV is available at https://share-eric.eu/data/accompanying-datasets/record-linkage-project.

15.16 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.

15.17 gv_big5

In waves 7, 8, and 9, the 10-item Big-Five inventory (BFI-10) was introduced for the first time, an established personality inventory measuring the "Big Five" personality dimensions with two items each. Introduced by Rammstedt and John (2007) the BFI-10 is an ultra-short measure of personality suitable especially for multi-theme surveys in which assessment time and questionnaire space are limited. For further information on the "Big Five" measurement, please see the corresponding chapter in the Wave 6 and 7 First Results Book.

Table 23: Generated Variables in gv_big5

Variable name	Trait	Variable	Question Text
bfi10_open	Openness	ac705_	I see myself as someone who has few artistic interests
		ac710_	I see myself as someone who has an active imagination
bfi10_consc	Conscientiousness	ac703_	I see myself as someone who tends to be lazy
		ac708_	I see myself as someone who does a thorough job
bfi10_extra	Extraversion	ac701_	I see myself as someone who is reserved
		ac706_	I see myself as someone who is outgoing, sociable
bfi10_agree	Agreeableness	ac702_	I see myself as someone who is generally trusting
		ac707_	I see myself as someone who tends to find fault with others
bfi10_neuro	Neuroticism	ac704_	I see myself as someone who is relaxed, handles stress well
		ac709_	I see myself as someone who gets nervous easily

15.18 gv_accelerometer

SHARE wave 8 used accelerometers in a subset of ten countries to collect physical activity data. Respondents' consent was collected in the CAPI (ax module). Participating respondents received the accelerometer after the CAPI interview via postal delivery. The used *Axivity AX3* accelerometer is a small and lightweight device. It was attached to the upper thigh by means of medical-style adhesive tape. Respondents were asked to wear the device for eight consecutive days (meaning 8x24 hours, day and night). Accelerometers were set to a sampling frequency of 50 Hz with a range of ±8*g*.[3]

For a detailed description of the SHARE accelerometer study see SHARE Wave 8 Methodology. The gv_accelerometer modules contain variables derived from the data collected in the SHARE accelerometer study processed with two different open-source software, ActiPASS and GGIR (see sections below). Raw sensor data is available upon request.

[3] Some devices were unintentionally set to a deviating sampling frequency, indicated in the *frequency* variable.

Available data

There are several generated modules available that describe the accelerometer data on different levels. Dataset “gv_accelerometer_total” contains measures describing physical activity over the entire observation time. The “gv_accelerometer_day” dataset describes the daily physical activity of respondents. Sleep spells detected by ActiPASS are available in the “gv_accelerometer_sleep” dataset. Detailed information on the respondents’ physical activity patterns is available on an hourly level in the “gv_accelerometer_hour” dataset, on one-second level as activities and postures in the “gv_accelerometer_epoch_1s” datasets (see section “ActiPASS”) and on 5-second epoch level in the “gv_accelerometer_epoch_5s” datasets (see section “GGIR”).

Some basic information on the accelerometer measurement is included in the “total”, “day”, and “hour” datasets. The period in which the respondent wore the device is described with the variables *month* and *year* (month and year of the first day of measurement), *weekday*, as well as *reported_N_valid_hours_total, ActiPASS_N_valid_hours*, and *GGIR_N_valid_hours* which indicate duration of wear time reported by the respondent and detected by ActiPASS and GGIR algorithm respectively. Additional information includes the wearing position, i.e. left or right thigh (*position*), an indicator for the device used (*device_id*), and the sampling frequency (*frequency*).

The generated modules contain only days (midnight to midnight) with at least 16 hours of wear time reported by the respondent and detected by ActiPASS or GGIR. Metrics derived from ActiPASS and GGIR are only available for days with a minimum of 16 hours valid wear time according to the respective software. Therefore, the number of valid days of observation can differ between ActiPASS and GGIR.

ActiPASS

Since Release 9.0.0, SHARE provides measures on postures and activities derived from ActiPASS Version 1.61beta (Hettiarachchi and Johansson 2023). ActiPASS is an open source software that is based on the Acti4 algorithm for posture and activity recognition for thigh-worn accelerometer data (Skotte et al. 2014)

The detection of sitting, standing, walking, running, cycling, and walking stairs (Stemland et al. 2015) as well as the number of steps (Jorgen Ingebrigtsen et al. 2013) have been validated. As an advancement of Acti4, ActiPASS includes algorithms to identify lying and sleeping (Hettiarachchi et al. 2021; Johansson et al. 2023).

The names of variables that are derived by ActiPASS start with “ActiPASS_”.

If you use ActiPASS variables, please report your work and publications at:
www.github.com/Ergo-Tools/ActiPASS/

Please also cite all relevant methodological resources. An updated list of ActiPASS related publications is available at www.github.com/Ergo-Tools/ActiPASS/.
Relevant methodological references by the time at SHARE Release 9.0.0 are:

Software used to generate the Variables:

- Hettiarachchi, P.; Johansson, P. (2023). ActiPASS (v1.61-beta). Zenodo. https://doi.org/10.5281/zenodo.7701098

Original reference to the core activity detection algorithm Acti4:

- Skotte, J; Korshøj, M; Kristiansen, J; Hanisch, C; Holtermann, A. (2014): Detection of physical activity types using triaxial accelerometers. J Phys Act Health. Jan;11(1):76–84.

Validation of Acti4 in free living settings:

- Stemland, I; Ingebrigtsen, J; Christiansen, CS; Jensen, BR; Hanisch, C; Skotte, J et al. (2015): Validity of the Acti4 method for detection of physical activity types in free-living settings: comparison with video analysis. Ergonomics.58(6):953–65.

Validation of Acti4 together with ActivPAL and Axivity:

- Crowley, P; Skottem J; Stamatakism E; Hamer, M; Aadahl, M; Stevens, ML et al. (2019): Comparison of physical behavior estimates from three different thigh-worn accelerometers brands: a proof-of-concept for the Prospective Physical Activity, Sitting, and Sleep consortium (ProPASS). International Journal of Behavioral Nutrition and Physical Activity. 16;16(1):65.

Validation of step counts in Acti4:

- Ingebrigtsen, J; Stemland, I; Christiansen, C; Jorgen, S; Hanisch, C; Krustrup, P et al. (2013): Validation of a Commercial and Custom Made Accelerometer-Based Software for Step Count and Frequency during Walking and Running. J Ergonomics

Validation of lying down:

- Hettiarachchi, P; Aili, K; Holtermann, A; Stamatakis, E; Svartengren, M; Palm, P. (2021): Validity of a Non-Proprietary Algorithm for Identifying Lying Down Using Raw Data from Thigh-Worn Triaxial Accelerometers. Sensors 21(3):904.

Validation of thigh accelerometer sleep algorithm:

- Johansson, P. J.; Crowley, P.; Axelsson, J.; Franklin, K.; Garde, A. H.; Hettiarachchi, P.; Holtermann, A.; Keckl und, G.; Lindberg, E.; Ljunggren, M.; Stamatakis, E.; Theorell Haglöw, J.; Svartengren, M. (2022): Development and performance of a sleep estimation algorithm using a single accelerometer placed on the thigh: an evaluation against polysomnography. Journal of Sleep Research, e13725.

ActiPASS Measures

ActiPASS variables describe the time spent in activities and postures (e.g. sitting, standing, lying, sleeping, walking, running, cycling) and in four intensity categories[4] based on activities

[4] For the methodology of intensity classification see: https://github.com/Ergo-Tools/ActiPASS/wiki/ActiPASS-variables-and-other-quality-check-files.

that may correspond to sedentary, light, moderate and vigorous activity per day and hour. In addition, the number of steps and sleep intervals are derived. For an overview of all variables and definitions of "gv_accelerometer_day", see Table 24. The dataset "gv_accelerometer_total" includes averages across all days of the number of steps and the time spent in intensity categories.

For proper performance, the ActiPASS algorithm is dependent on the correct positioning of the device on the thigh and some walking to perform calibration. As described above, days with less than 16 hours of wear time are excluded from the data. Some other indicators of poor posture identification are included in the dataset to enable users to define their own inclusion criteria. Days with no sleep, little walking, as well as suspiciously long bedtime, stair walking, and "other" activities are flagged. The decision of whether to exclude these cases or not depends on the research question. Some variables, like sit, lie and upright, are quite robust, while others, such as stair or cycle, are more sensitive to issues related to positioning and calibration.

Table 24Overview of ActiPASS variables in "day" and "hour" data

Variable	Definition
Quality indicator	
ActiPASS_NoSleepInt_day	ActiPASS Check: No sleep interval
ActiPASS_NoWalk_day	ActiPASS Check: < 30 sec walking
ActiPASS_TooMuchBedtime_day	ActiPASS Check: > 16 h bedtime
ActiPASS_TooMuchOther_day	ActiPASS Check: > 30 min other
ActiPASS_TooMuchStair_day	ActiPASS Check: > 120 min stair
Posture, Activities, Intensities	
ActiPASS_Nonwear_day	Non-wear time detected by ActiPASS
ActiPASS_Bedtime_day	The time within this calendar day flagged as bedtime
ActiPASS_LieStill_day	The total time of possible sleep outside a bedtime
ActiPASS_Sleep_day	Total sleep time within this calendar day
ActiPASS_SleepInterval_day	The time within this calendar day which is flagged as sleep interval. A sleep interval is defined as the time from first sleep onset until last awakening, including possible awake intervals in-between.
ActiPASS_Awake_day	The time within this calendar day which is not flagged as sleep interval
ActiPASS_Walk_Slow_day	Walking with a cadence lower than 100
ActiPASS_Walk_Fast_day	Walking with a cadence equal or higher than 100
ActiPASS_Lie_day	Lying
ActiPASS_Sit_day	Sitting
ActiPASS_SitLie_day	Sitting or lying
ActiPASS_Stand_day	Standing
ActiPASS_Move_day	'Move' is defined as periods where person is in standing posture with certain movements above the

	threshold for standing posture (ex. intermittent steps), but no purposeful walking.
ActiPASS_StandMove_day	Standing or moving
ActiPASS_Walk_day	Walking
ActiPASS_Run_day	Running
ActiPASS_Stair_day	All stair walking like activities are counted. i.e. terrain walking etc.
ActiPASS_Cycle_day	Cycling
ActiPASS_Upright_day	Time of all upright activities/postures (i.e. Stand + Move + Walk + Run + Stair)
ActiPASS_Other_day	Time of activities in "Other" postures/activities with movements which cannot be classified into known activities/postures
ActiPASS_INT1_day	Sedentary: "Sit" and "Lie"*
ActiPASS_INT2_day	Low physical activity*: "Stand", "Move", "Walk_Slow" (walking with a cadence lower than 100/min), and "Other" with no periodic movements and "Other" with periodic movements with a cadence lower than 100/min*
ActiPASS_INT3_day	Moderate physical activity: "Walk" (with a cadence between 100-135/min) and "Other" with periodic movements with a cadence between 100-135/min*
ActiPASS_INT4_day	Vigorous physical activity: "Run", "Cycle", "Stair", "Walk" with a cadence higher than 135/min , "Other" with periodic movement with a cadence higher than or equal to 135/min*
ActiPASS_INT34_day	Moderate and vigorous physical activity: INT3 + INT4*
ActiPASS_Steps_day	Number of steps

*For the methodology of intensity classification see: https://github.com/Ergo-Tools/ActiPASS/wiki/ActiPASS-variables-and-other-quality-check-files

Dataset “gv_accelerometer_sleep” includes spells of sleep intervals detected by ActiPASS, where a sleep interval is defined by the time between sleep onset and the last awakening. An overview of variables is available in Table 25. Both, sleep at night and day are included. Also, sleep intervals which start at the day before the first or end after the last valid day of observation, are included, e.g. start of a sleep interval can be *measurementday* = 0.

The dataset provides two flag variables that show possible wrong detected sleep intervals. First, *ActiPASS_LongBedtime* indicates bedtimes longer than 16 hours. Second, *ActiPASS_LongSleep* indicates sleep intervals longer than 16 hours. When using these data, it is recommended to define inclusion criteria suitable for the research question at hand and therefore also consider the quality indicators in the “gv_accelerometer_day” dataset.

Please note that all variables that describe time (e.g. ActiPASS_Time_SleepStart) are denoted as decimal hour, i.e. time 15.25 = 3:15pm. Durations are denoted in decimal hour or minute, i.e. 0.5 hours = 30 minutes and 0.1 minute = 6 seconds.

Table 25: Overview of ActiPASS variables in "sleep"

Variable	Definition
ActiPASS_LongBedtime	Dummy indicating bedtime longer than 16 hours
ActiPASS_LongSleep	Dummy indicating sleep duration longer than 16 hours
measurementday_SleepStart	Day the sleep interval starts
ActiPASS_Time_SleepStart	Time the sleep interval starts
measurementday_SleepEnd	Day the sleep interval ends
ActiPASS_Time_SleepEnd	Time the sleep interval ends
ActiPASS_SleepInterval	Duration of sleep interval
ActiPASS_TotalSleep	Total sleep time during sleep interval

Detailed information on the activities detected by ActiPASS are available on one-second level in the "epoch_1s" files. Here, the data from each respondent is stored in a separate file. The file names include the ID of the respondent. These "gv_accelerometer_epoch_1s" datasets are available for download in the SHARE Research Data Center in a separate file. Datasets are available in Stata, SPSS, and csv format. These datasets are intended for use by experienced users only. Prior to utilization, thorough consultation of the ActiPASS documentation is strongly advised.

The "epoch_1s" datasets include the *mergeid* to identify respondents and variable *measurementday* and *timestamp* which specify the day and time (on second level). Activities and non-wear time is indicated in variable *activity* with following values: 0 = NonWear; 1 = Lie; 2 = Sit; 3 = Stand; 4 = Move; 5 = Walk; 6 = Run; 7 = Stair; 8 = Cycle; 9 = Other; 10 = Sleep; 11 = LieStill. For definition of these categories, see Table 24. Variable *steps* indicates the number of steps per second. Variables *bedtime* and *sleepinterval* are dummies that indicate if the second was classified as bedtime and sleep interval (variable = 1, otherwise 0), respectively. Please note that only the postures and activities during no-sleep time are used for the calculation of activities and postures in the "day" and "hour" datasets.

GGIR

SHARE provides datasets with various measures that are generated with R-package GGIR version 3.0-5 (Migueles et al. 2019), running with R version 4.3.2 (R Core Team 2023).[5] All measures calculated with GGIR are based on the Euclidian norm minus one with negative values set to zero (ENMO) (Sabia et al. 2014; van Hees et al. 2013).

Raw sensor data is aggregated to 5-second epochs. Auto-calibration is performed (van Hees et al. 2014). Non-wear time is detected by bouts of 15 minutes and imputation is performed for non-wear time by means of valid measurements of the same person at the same time on different days.

[5] Earlier versions of GGIR 1-11.0 and 2.4-0 and R 4.1.0 and 4.3.2 were used in release 1.0.0 of SHARE wave 8 and release 8.0.0, respectively.

The names of variables that are derived by GGIR start with "GGIR_".

If you use GGIR variables, please cite:

- Migueles JH, Rowlands AV, et al. GGIR: A Research Community–Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data. Journal for the Measurement of Physical Behaviour. 2(3) 2019. doi: 10.1123/jmpb.2018-0063.
- van Hees VT, Fang Z, et al. Auto-calibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. J Appl Physiol 2014.
- van Hees, Vincent T.; Gorzelniak, Lukas; Dean León, Emmanuel Carlos; Eder, Martin; Pias, Marcelo; Taherian, Salman et al. (2013): Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. In: PloS one 8 (4), S. e61691. DOI: 10.1371/journal.pone.0061691.
- Sabia, Séverine; van Hees, Vincent T.; Shipley, Martin J.; Trenell, Michael I.; Hagger-Johnson, Gareth; Elbaz, Alexis et al. (2014): Association between questionnaire- and accelerometer-assessed physical activity: the role of sociodemographic factors. In: American journal of epidemiology 179 (6), S. 781–790. DOI: 10.1093/aje/kwt330

GGIR: Measures

The "hour", "day", and "total" datasets contain the average vector magnitude per hour/day/total as Euclidian norm minus one with negative values set to zero (ENMO), describing the volume of activity (*GGIR_mean_ENMO),* see van Hees et al. (2013). The Euclidean norm is the vector magnitude, combining the acceleration of the three axes to one vector ($\sqrt{x^2 + y^2 + z^2}$). As accelerometers also record gravity, 1 gravity unit (1*g* = $9.81\frac{m}{s^2}$) is subtracted from the Euclidean norm. ENMO is calculated with formula $\sqrt{x^2 + y^2 + z^2} - 1$ (negative values set to zero). In the generated modules, ENMO is denoted in milligravity units (1 m*g* = $9.81\frac{\mathrm{m}}{\mathrm{s}^2}/1000$).

LXMX describes the X-hours with the least and most activity (based on ENMO) during a day. Starting time (*GGIR_LX_hr*, *GGIR_MX_hr*) and average ENMO (*GGIR_LX_mean_ENMO*, *GGIR_MX_mean_ENMO*) of the most and least active 16, 10, 8, 5, 2, 1, and 0.5 hours are reported. Most and least active X hours were calculated for start times in 10 minutes intervals, e.g. for most/least active one hour for 0:00 a.m. – 1:00 a.m., 0:10 a.m. – 1:10 a.m., 0:20 a.m. – 1:20 a.m., etc. Please note that time is denoted as decimal, e.g. 10.5 = 10:30 a.m.

Time (in minutes) spent per hour/day/total in intensity level between X and Y milligravity is aggregated in intervals of 25 m*g* and indicated by variables *GGIR_X_Y_ENMO*. For example, the module "gv_accelerometer_hour" contains a variable called *GGIR_300_325_ENMO_hour* reporting the minutes spent in ENMO Level [300-325) m*g* during a specific hour for each respondent and for each measurement day. Please note that time (GGIR_LX_hr, GGIR_MX_hr)

and duration (e.g. GGIR_X_Y_ENMO*, GGIR_N_valid_hours) is denoted as decimal (e.g. 0.5 hours = 30 minutes).

The intensity gradient (IG) describes the intensity profile based on the intensity intervals. The more negative the gradient, the longer the time spent inactive and the shorter the time spent active. The more negative the gradient, thus, the worse the intensity profile. See Rowlands et al. (2018).

Mean ENMO and intensity intervals are available in "hour", "day", and "total" datasets. Information on X hours with the least and most activity (LXMX) are only included in the day level data, the intensity gradient (IG) is available on day and total level, see Table 26.

Table 26: Overview of measures, corresponding variables, and availability in datasets

Measure	Variables[1]	Dataset		
		total	day	hour
Mean ENMO	GGIR_mean_ENMO*	X	X	X
LXMX	GGIR_LX_hr GGIR_LX_mean_ENMO GGIR_MX_hr GGIR_MX_mean_ENMO		X	
Intensity intervals	GGIR_X_Y_ENMO*	X	X	

[1] * is a placeholder for "total", "day", and "hour". X and Y are placeholder for duration in hours (LXMX) and interval thresholds in milligravity (m*g*) (intensity intervals).

Since release 8.0.0, additional information on the accelerometer measurements is available. SHARE provides a dataset for each participant of the accelerometer study, containing the average acceleration (ENMO) in 5 second intervals, so-called epochs. The epoch data is generated with GGIR. These "gv_accelerometer_epoch_5s" datasets are available for download in the SHARE Research Data Center in a separate file. Datasets are available in Stata, SPSS, and csv format.

The file names include the ID of the respondent. The datasets consist of the following variables:

Table 27: Overview of variables in epoch data

Name of variable	Content
mergeid	ID
measurementday	Consecutive days of wear time
timestamp	Time (beginning of 5 second interval)
GGIR_ENMO	Average ENMO in *g*
GGIR_imputation	Flag for imputed ENMO values (1=imputed;0=original value)

Please note that in the epoch data, the average acceleration (ENMO) is denoted in gravity (not milligravity). Information form “hour”, “day”, “total” datasets (e.g. day of the week) can be merged by *mergeid* and measurementday.

Oxford COVID-19 Government Response Tracker

Some of the accelerometer measurements were taken during the onset of the COVID-19 pandemic and thus influenced by different governmental regulations that may have affected respondents’ physical behavior. For this reason, two variables from the Oxford COVID-19 Government Response Tracker (OxCGRT) are added to the datasets to capture the country-specific situation at the time respondents wore the accelerometers. These include an indicator for “stay at home requirements” and an overall “stringency index”, describing the degree of containment and closure policies (Hale et al. 2021).

The names of variables derived by the Oxford COVID-19 Government Response Tracker start with “OxCGRT_”.

If you use variables from the Oxford COVID-19 Government Response Tracker, please cite:

- Thomas Hale , Noam Angrist , Rafael Goldszmidt, Beatriz Kira , Anna Petherick , Toby Phillips, Samuel Webster, Emily Cameron-Blake , Laura Hallas, Saptarshi Majumdar, and Helen Tatlow. (2021). “A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker).” Nature Human Behaviour.

References

Hale, Thomas, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster et al. 2021. “A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker).” *Nature human behaviour* 5 (4): 529–38. doi:10.1038/s41562-021-01079-8.

Hettiarachchi, Pasan, Katarina Aili, Andreas Holtermann, Emmanuel Stamatakis, Magnus Svartengren, and Peter Palm. 2021. “Validity of a Non-Proprietary Algorithm for Identifying Lying down Using Raw Data from Thigh-Worn Triaxial Accelerometers.” *Sensors (Basel, Switzerland)* 21 (3). doi:10.3390/s21030904.

Hettiarachchi, Pasan, and Peter Johansson. 2023. *ActiPASS (Version 1.61):* Zenodo. doi:10.5281/zenodo.7701098.

Ingebrigtsen, Jorgen, Ingunn Stemland, Caroline Christiansen, Skotte Jorgen, Christiana Hanisch, Peter Krustrup, and Andreas Holtermann. 2013. “Validation of a Commercial and Custom Made Accelerometer-Based Software for Step Count and Frequency During Walking and Running.” *J Ergonomics* 03 (02). doi:10.4172/2165-7556.1000119.

Johansson, Peter, Patrick Crowley, John Axelsson, Karl Franklin, Anne H. Garde, Pasan Hettiarachchi, Andreas Holtermann et al. 2023. “Development and Performance of a Sleep

Estimation Algorithm Using a Single Accelerometer Placed on the Thigh: An Evaluation Against Polysomnography." *Journal of sleep research* 32 (2): e13725. doi:10.1111/jsr.13725.

Migueles, Jairo H., Alex V. Rowlands, Florian Huber, Séverine Sabia, and Vincent T. van Hees. 2019. "GGIR: A Research Community–Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes from Multi-Day Raw Accelerometer Data." *Journal for the Measurement of Physical Behaviour* 2 (3): 188–96. doi:10.1123/jmpb.2018-0063.

R Core Team. 2023. "R: A Language and Environment for Statistical Computing." https://www.R-project.org/.

Sabia, Séverine, Vincent T. van Hees, Martin J. Shipley, Michael I. Trenell, Gareth Hagger-Johnson, Alexis Elbaz, Mika Kivimaki, and Archana Singh-Manoux. 2014. "Association Between Questionnaire- and Accelerometer-Assessed Physical Activity: The Role of Sociodemographic Factors." *American journal of epidemiology* 179 (6): 781–90. doi:10.1093/aje/kwt330.

Skotte, Jørgen, Mette Korshøj, Jesper Kristiansen, Christiana Hanisch, and Andreas Holtermann. 2014. "Detection of Physical Activity Types Using Triaxial Accelerometers." *Journal of physical activity & health* 11 (1): 76–84. doi:10.1123/jpah.2011-0347.

Stemland, Ingunn, Jørgen Ingebrigtsen, Caroline S. Christiansen, Bente R. Jensen, Christiana Hanisch, Jørgen Skotte, and Andreas Holtermann. 2015. "Validity of the Acti4 Method for Detection of Physical Activity Types in Free-Living Settings: Comparison with Video Analysis." *Ergonomics* 58 (6): 953–65. doi:10.1080/00140139.2014.998724.

van Hees, Vincent T., Zhou Fang, Joss Langford, Felix Assah, Anwar Mohammad, da Silva, Inacio C M, Michael I. Trenell, Tom White, Nicholas J. Wareham, and Søren Brage. 2014. "Autocalibration of Accelerometer Data for Free-Living Physical Activity Assessment Using Local Gravity and Temperature: An Evaluation on Four Continents." *Journal of applied physiology (Bethesda, Md. : 1985)* 117 (7): 738–44. doi:10.1152/japplphysiol.00421.2014.

van Hees, Vincent T., Lukas Gorzelniak, Emmanuel C. Dean León, Martin Eder, Marcelo Pias, Salman Taherian, Ulf Ekelund et al. 2013. "Separating Movement and Gravity Components in an Acceleration Signal and Implications for the Assessment of Human Daily Physical Activity." *PloS one* 8 (4): e61691. doi:10.1371/journal.pone.0061691.

APPENDIX

Table 1: Digital Object Identifier of the SHARE Datasets

Dataset	Release	Date	DOI
Wave 1	9.0.0	28th March 2024	10.6103/SHARE.w1.900
Wave 2	9.0.0	28th March 2024	10.6103/SHARE.w2. 900
Wave 3	9.0.0	28th March 2024	10.6103/SHARE.w3. 900
Wave 4	9.0.0	28th March 2024	10.6103/SHARE.w4. 900
Wave 5	9.0.0	28th March 2024	10.6103/SHARE.w5. 900
Wave 6	9.0.0	28th March 2024	10.6103/SHARE.w6. 900
Wave 7	9.0.0	28th March 2024	10.6103/SHARE.w7. 900
Wave 8	9.0.0	28th March 2024	10.6103/SHARE.w8. 900
Wave 8 COVID-19 Survey	9.0.0	28th March 2024	10.6103/SHARE.w8ca. 900
AdditionalWave 8 COVID-19 Interview date data	9.0.0	28th March 2024	10.6103/SHARE.w8caintd. 900
Wave 9 COVID-19 Survey	9.0.0	28th March 2024	10.6103/SHARE.w9ca. 900
Additional Wave 9 COVID-19 Interview date data	9.0.0	28th March 2024	10.6103/SHARE.w9caintd. 900
Wave 9	9.0.0	28th March 2024	10.6103/SHARE.w9. 900
All Waves Coverscreen	9.0.0	28th March 2024	10.6103/SHARE.wXcvr. 900
Longitudinal Weights	9.0.0	28th March 2024	10.6103/SHARE.wXweights. 900
***easy*SHARE**	8.0.0	10th Feb. 2022	10.6103/SHARE.easy.800
SHARE-RV	8.0.0	11th March 2022	10.6103/SHARE.SHARE-RV.800
Job Episodes Panel	9.0.0	28th March 2024	10.6103/SHARE.jep.900
Dutch Mixed Mode Experiment w6	1.0.0	12th Sept. 2017	10.6103/SHARE.w6NLmmExp.100
Dutch Mixed Mode Experiment w7	1.0.0	26th June 2020	10.6103/SHARE.w7NLmmExp.100

Table 2: SHARE Data Releases

Wave 1 & Wave 2*	
Release 1: April 28 th, 2005	
Release 2.0.0: June 19th, 2007	
Release 2.0.1: July 5th th, 2007	
	Release 1.0.0: November 28th, 2008
	Release 1.0.1: December 4th, 2008
Release 2.2.0: August 19th, 2009*	
Release 2.3.0: November 13th, 2009	
Release 2.3.1: July 28th, 2010	
Release 2.4.0: March 17th, 2011	
Release 2.5.0: May 11th, 2011	
Release 2.6.0: November 29th, 2013	
Wave 3 (SHARELIFE)	
Release 1.0.0: November 24th, 2010	
Wave 4	
Release 1.0.0: November 30th, 2012	
Release 1.1.1: March 28th, 2013	
Wave 5	
Release 1.0.0: March 31st, 2015	
Release 5.0.0 of Waves 1 to 5	
Release 5.0.0: May 10th 2016	
Release 6.0.0 of Waves 1 to 6	
Release 6.0.0: March 31st 2017	
Release 6.1.0: March 29th 2018	
Release 6.1.1: June 19th 2018	
Release 7.0.0 of Waves 1 to 7	
Release 7.0.0: April 3rd 2019	
Release 7.1.0: June 26th 2020	
Release 7.1.1: December 17th 2020	
Release 1.0.0 of Wave 8	
Release 1.0.0: June 23rd 2021	
Release 8.0.0 of Waves 1 to 8	
Release 8.0.0: February 10th 2022	
Release 9.0.0 of Waves 1 to 9	
Release 9.0.0: March 28th 2024	

* 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

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
- Parents´ country of birth
- Mini-childhood module
- IT module
- New social exclusion items

New in Release 5.0.0

- New modules: gv_allwaves_cv_r, gv_exrates and gv_ssw
- Improved coverscreen module
- Cause of death (xt012) now coded (xt012c)
- 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

New in SHARE Wave 6

- A new country joined SHARE: Croatia
- Longitudinal information on social networks: SN module and gv_networks
- New generated variables module for children: gv_children
- Collection of Dried Blood Spots: BS module) and gv_dbs
- Updated imputations
- Updated weights

New in release 6.1.1

- Update of gv_isced
- Minor changes in the AC, RC and RE modules of wave 3

New in release 6.1.0

- Children's country of residence (ch008c*) added to the CH module of waves 1 and 2
- New routing variables added to the technical variables module of waves 1, 2, and 4
- Minor changes in the PH and the EP module of wave 4
- Wave 6 drop-off data for Switzerland now available
- More cases in the Czech drop-off of wave 6
- Minor changes in the DN, CH, PH, SP and EP module of wave 6
- Minor changes in the following generated variable modules: gv_allwaves_cv_r, gv_imputations, gv_weights, gv_children, gv_linkage, and gv_housing

New in SHARE Wave 7

- In wave 7, Hungary was recovered and eight new countries joined SHARE with Bulgaria, Cyprus, Finland, Latvia, Lithuania, Malta, Romania and Slovakia entailing full coverage of all continental EU Member States.
- The wave 7 questionnaire contains SHARELIFE modules for all respondents who did not participate in wave 3 (82% of respondents), as well as regular panel modules for all respondents who already participated in the SHARELIFE interview of wave 3 (18% of respondents). Please note that this complicated structure of the questionnaire leads to high amounts of missing values in specific variables due to routing.
- Big-Five inventory (10-item) as part of the AC module and as generated variables module (gv_big5)

New in Release 7.1.0 including other waves

- Latest state of data cleaning
- Updated weights and updated imputations of all SHARE waves
- Complete Portuguese sample included
- Updated versions of *easy*SHARE and the Job Episodes Panel (JEP)
- Interviewer survey of Wave 7
- Updated information on the vital status of respondents in gv_allwaves_cv_r by using national register data or administrative records where available: Denmark, Estonia, Sweden, Belgium, the Netherlands, France, Austria and Germany
- Minor corrections in gv_isced for Croatian respondents

New in Release 7.1.1 of Wave 7

- Corrections in variables ch007, ch014 and ch019

New in SHARE Wave 8

- Information on social networks was collected for the third time in SHARE
- New modules on saving regrets (SR) and time expenditure (TE)
- New accelerometer data measuring respondents' physical activity in a subset of countries including the brand new accelerometry epoch data
- SHARE Corona Survey

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
1133	British-Swiss
1134	French-American
1135	Dutch-Belgian
1136	Finnish-Greek
1137	English-Irish
1138	French-Swiss
1139	Macedonian-Bulgarian
1140	British-Australian
1141	French-Australian
1142	Canadian-Italian
1143	Serbian-Russian
1300	Other European country
1400	Other Asian country
1500	Other African country
1600	Other Non-European country

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 (Life Orientation Test: pessimism/optimism)	3 a)	q3_a	1 a)	q3_a	
	3 b)	q3_b	1 b)	q3_b	
	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 (CES_D	4a)	q4_a	-		ac027(3)
	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: Responsibility for different tasks	10	q10	6	q10	
	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 General practitioner/ usual source of care questions and checks	12	q12	-		
	13 a)	q13_a	8 a)	q13_a	
	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 Talk about problems	-		9 a)	q39_a	
	-		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	32 a)	q32_a	-		ho056(1)
accommodation	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: Calibrated cross-sectional weights of Wave t=1,…, 9

Variable	Description	Unit of analysis
dw_wt	Design weight - wave t	Household & individual
cchw_wt	Calibrated cross-sectional household weight - wave t	Household
cciw_wt	Calibrated cross-sectional individual weight - wave t	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

Table 6: Calibrated cross-sectional weights of pooled SHARELIFE sample (dataset name sharelife_rel9-0-0_gv_weights)

Variable	Description	Unit of analysis
dw_sl	Design weight - waves 3 or 7	Household & individual
cchw_sl	Calibrated cross-sectional household weight - wave #	Household
cciw_sl	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

Table 7: Calibrated longitudinal weights of wave combination t-…-s

$t-\cdots-s$	Variable	Description	Unit of analysis
$1-2$	dw_w1	Design weight - wave 1	Household & individual
	clhw_wc_b	Calibrated longitudinal household weight	Household
	cliw_wc_b	Calibrated longitudinal individual weight	Individual
$2-3$	dw_w2	Design weight - wave 2	Household & individual
	clhw_wc_c	Calibrated longitudinal household weight	Household
	cliw_wc_c	Calibrated longitudinal individual weight	Individual
$3-4$	dw_w3	Design weight - wave 3	Household & individual
	clhw_wc_d	Calibrated longitudinal household weight	Household
	cliw_wc_d	Calibrated longitudinal individual weight	Individual
$4-5$	dw_w4	Design weight - wave 4	Household & individual
	clhw_wc_e	Calibrated longitudinal household weight	Household
	cliw_wc_e	Calibrated longitudinal individual weight	Individual
$5-6$	dw_w5	Design weight - wave 5	Household & individual
	clhw_wc_f	Calibrated longitudinal household weight	Household
	cliw_wc_f	Calibrated longitudinal individual weight	Individual
$6-7$	dw_w6	Design weight - wave 6	Household & individual
	clhw_wc_g	Calibrated longitudinal household weight	Household
	cliw_wc_g	Calibrated longitudinal individual weight	Individual
$7-8$	dw_w7	Design weight - wave 7	Household & individual
	clhw_wc_g	Calibrated longitudinal household weight	Household
	cliw_wc_g	Calibrated longitudinal individual weight	Individual
$8-9$	dw_w8	Design weight - wave 8	Household & individual
	clhw_wc_g	Calibrated longitudinal household weight	Household
	cliw_wc_g	Calibrated longitudinal individual weight	Individual
$1-\cdots-9$	dw_w1	Design weight - wave 1	Household & individual
	clhw_wc_a	Calibrated longitudinal household weight	Household
	cliw_wc_a	Calibrated longitudinal individual weight	Individual
ALL	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
	panel_resp	Respondent participation in the selected panel	Individual

Table 8: 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 9: Country-specific composition of imputed pension variables in wave 1

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

Table 9: Country-specific composition of imputed pension variables in wave 1 (continued)

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

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 10: Composition of aggregated imputation variables in wave 2

Aggregates	Components	Description
ypen1[(a)]	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[(b)]	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[(c)]	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[(d)]	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[(e)]	LSP - Long-term care insurance from private insurance company

Table 10: Composition of aggregated imputation variables in wave 2 (continued)

Aggregates	Components	Description
yreg2	EP094_3	RP - Alimony
	EP094_4	RP - Charities
ylsr2	EP209_3	LSP - Alimony
	EP209_4[(f)]	LSP - 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[(g)]	Amount rent paid
	HO008[(g)]	Other home-related expenditures: charges and services
yaohm	HH002[(g)]	Total income received by other household members
	HH011[(g)]	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: 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) 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 11: Composition of aggregated imputation variables in wave 4

Aggregates	Components	Description
ypen1[(a)]	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[(b)]	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[(c)]	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[(d)]	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 11: Composition of aggregated imputation variables in wave 4 (continued)

Aggregates	Components	Description
yreg2	EP094_3	RP – Alimony
	EP094_4	RP – Charities
ylsr2	EP209_3	LSP – Alimony
	EP209_4	LSP – Charities
rhre	HO005[(e)]	Amount rent paid
	HO008[(e)]	Other home-related expenditures: charges and services
	HO065[(f)]	Out of pocket payment for nursing home accommodation
ysrent	HO030[(e)]	Income from rent of secondary/holiday homes and other real estate
	HO074[(e)]	Income from sublet of accommodation
yaohm	HH002[(e)]	Total income received by other household members
	HH011[(e)]	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[(g)]	Contractual savings
	AS030	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 12: Composition of aggregated imputation variables in wave 5

Aggregates	Components	Description
ypen1[(a)]	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[(b)]	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[(c)]	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[(d)]	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 12: Composition of aggregated imputation variables in wave 5 (continued)

Aggregates	Components	Description
rhre	HO005[(e)]	Amount rent paid
	HO008[(e)]	Other home-related expenditures: charges and services
	HO065[(f)]	Out of pocket payment for nursing home accommodation
	HO008[(f)]	Other home-related expenditures: charges and services in nursing home
ores	HO027[(e)]	Value of secondary/holiday homes and other real estate
	HO076[(f)]	Value of secondary/holiday homes and other real estate (respondents in nursing home)
ysrent	HO030[(e)]	Income from rent of secondary/holiday homes and other real estate
	HO074[(e)]	Income from sublet of accommodation
	HO078[(f)]	Income from rent of secondary/holiday homes and other real estate (respondents in nursing home)
yaohm	HH002[(e)]	Total income received by other household members
	HH011[(e)]	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[(g)]	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[(e)]	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, ylsp1 also includes: EP082_11, EP082_12, EP082_13, and EP082_14. In LU, ylsp1 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, ylsp2 includes only: EP082_15 and EP082_16. In LU, ylsp2 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 13: Composition of aggregated imputation variables in waves 6, 7,8 and 9

Aggregates	Components	Description
ypen1[(a)]	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[(b)]	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	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 13: Composition of aggregated imputation variables in waves 6, 7,8 and 9 (continued)

Aggregates	Components	Description
rhre	HO005[(c)]	Amount rent paid
	HO008[(c)]	Other home-related expenditures: charges and services
	HO065[(d)]	Out of pocket payment for nursing home accommodation
ores	HO027[(c)]	Value of secondary/holiday homes and other real estate
	HO076[(d)]	Value of secondary/holiday homes and other real estate (respondents in nursing home)
ysrent	HO030[(c)]	Income from rent of secondary/holiday homes and other real estate
	HO074[(c)]	Income from sublet of accommodation
	HO078[(d)]	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[(e)]	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[(e)]	Out of pocket payment for nursing home accommodation
	HC129	Out of pocket payment for personal care, domestic tasks, meals on wheels, other help
aapt	HC142	Out of pocket payment for aids and appliances
	HC143	Out of pocket payment for physical therapy

Notes: RP denotes 'regular payments'. LSP denotes 'Lump-sum payments'.
(a) In SE, ypen1 includes only: EP078_1, and EP078_9. EP078_2 is included in In PL and LU, ypen1 includes only: EP078_1, EP078_3, EP078_9 and EP078_11.
(b) In SE, ylsp1 includes only: EP082_1, and EP082_9. In PL and LU, ylsp1 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 EE.

Table 14: List of variables in gv_imputations

Variable	Description	Questionnaire	Waves[6]
implicat	Implicat number		1 2 4 5 6 7 (R-S) 8 9
htype	Household type		1 2 4 5 6 7 (R-S) 8 9
fam_resp	Family respondent	MN006_	1 2 4 5 6 7 (R-S) 8 9
fin_resp	Financial respondent	MN007_	1 2 4 5 6 7 (R-S) 8 9
hou_resp	Household respondent	MN008_	1 2 4 5 6 7 (R-S) 8 9
exrate	Exchange rate		1 2 4 5 6 7 (R-S) 8 9
nursinghome	Living in nursing home	MN024_	1 2 4 5 6 7 (R-S) 8 9
perho	Percentage of house owned	HO070_	4 5 6 7 (R) 8 9
single	Single		1 2 4 5 6 7 (R-S) 8 9
couple	Couple		1 2 4 5 6 7 (R-S) 8 9
partner	Partner in the couple		1 2 4 5 6 7 (R-S) 8 9
p_nrp	Partner of non-responding partner		1 2 4 5 6 7 (R-S) 8 9
sample1	Imputation sample for single		1 2 4 5 6 8 9
sample2	Imputation sample for couples with two partners interviewed		1 2 4 5 6 8 9
sample3	Imputation sample for all couples		1 2 4 5 6 8 9
Sample_SH1	First imputation sample for SHARELIFE interviews: single		7 (S)
Sample_SH2	Second imputation sample for SHARELIFE interviews: all couples		7 (S)
Sample_RE1	First imputation sample for regular interviews: single and C2R		7 (R)
Sample_RE2	Second imputation sample for regular interviews: all couples		7 (R)
inpat#	Out-of-pocket payment for inpatient care	HC095, HC695 in w6+7	1 2 5 6 7 (R)
outpa#	Out-of-pocket payment for outpatient care	HC083, HC683 in w6+7	1 2 5 6 7 (R)
drugs#	Out-of-pocket payment for drugs	HC089, HC689 in w6+7	1 2 5 6 7 (R)
nurs#	Out-of-pocket payment for nursing home / home care	HC097	1 2 5 6 7 (R)
aapt#	Out-of-pocket payment for aids, appliances, physical therapy	HC140	6 7 (R)
hinsu	Paid for voluntary health insurance	HC061	1
ydip	Earnings from employment	EP205	1 2 4 5 6 7 (R) 8 9
yind	Earnings from self-employment	EP207	1 2 4 5 6 7 (R) 8 9
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 7 (R) 8 9
ypen2[(a)]	Annual private occupational pensions	EP078_11-16 (EP678 in w6)	1 2 4 5 6 7 (R) 8 9
ypen3	Annual disability pension and benefits	EP078_5-6	1 6 7 (R) 8 9
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 7 (R) 8 9
ypen5	Annual payment from social assistance	EP078_10(12-13 in w6)	1 2 4 5 6 7 (R) 8 9
ypen6	Sickness benefits and pensions	EP078_3_6_10 (4-7 in w6)	1 6 7 (R) 8 9
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 7 (R) 8 9

[6] R-S refers to the interview type in wave 7: R = Regular, S = SHARELIFE

Variable	Description	Questionnaire	Waves
ylsp2[(b)]	Lump sum payments for private occupational pension	EP082_11-16; EP682_11-16 in w6+7	2 4 5 6 7 (R) 8 9
ylsp3	Lump sum payments for disability pension and benefits	EP082_5-6	6 7 (R) 8 9
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 7 (R) 8 9
ylsp5	Lump sum payments for social assistance	EP082_10 (12-13 in w6)	2 4 5 6 7 (R) 8 9
ylsp6	Lump sum payments for sickness benefits	EP082_4-7	6 7 (R) 8 9
yreg1	Other regular payments from private pensions	EP094_1-2-5	1 2 4 5 6 7 (R) 8 9
yreg2	Other regular payment from private transfer	EP094_3-4	1 2 4 5 6 7 (R) 8 9
ylsr1	Lump sum from private payments	EP209_1-2-5	2 4 5 6 7 (R) 8 9
ylsr2	Lump sum from private transfers	EP209_3-4	2 4 5 6 7 (R) 8 9
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 7 (R) 8 9
home	Value of main residence	HO024	1 2 4 5 6 7 (R) 8 9
mort	Mortgage on main residence	HO015	1 2 4 5 6 7 (R) 8 9
ores	Value of other real estate – Amount	HO027	1 2 4 5 6 7 (R) 8 9
ysrent	Annual income from rent or sublet	HO074, HO030	1 2 4 5 6 7 (R) 8 9
yaohm[(c)]	Annual income from other household members	HH002, HH011	1 2 4 5 6 7 (R) 8 9
fahc	Annual food at home consumption	CO002	1 2 4 5 6 7 (R-S) 8 9
fohc	Annual food outside home consumption	CO003	1 2 4 5 7 (R-S) 8 9
telc	Amount spent on telephones	CO004	1 2
hprf	Annual home produced consumption	CO011	2 4 5 6 7 (R-S) 8 9
bacc	Bank accounts	AS003	1 2 4 5 6 7 (R) 8 9
bsmf	Bond, stock and mutual funds	AS007, AS011, AS017	1 2 4 5 6 7 (R) 8 9
ybabsmf	Interest / dividend from bank account, bond, stock, and mutual funds		1 2 4 5 6 7 (R) 8 9
slti	Savings for long-term investments	AS021, AS023, AS27, AS030	1 2 4 5 6 7 (R) 8 9
vbus	Value of own business	AS042	1 2 4 5 6 7 (R) 8 9
sbus	Share of own business	AS044	1 2 4 5 6 7 (R) 8 9
car	Value of cars	AS051	1 2 4 5 6 7 (R) 8 9
liab	Financial liabilities	AS055	1 2 4 5 6 7 (R) 8 9
thinc	Total household net income - version A (sum of ydip, ypen1, ypen2, ypen3, ypen4, ypen5, ypen6, yreg1, yreg2, yind, ysrent, yaohm, ybabsmf)		1 2 4 5 6 7 (R) 8 9
thinc2	Total household net income - version B	HH017	2 4 5 6 7 (R-S) 8 9
thexp	Total household expenditure (sum of rhre, fahc, fohc, hprf, inpat, outpa, drugs, nurs)		1 2 4 5 6 7 (R) 8 9
yincnrp	Income from non-responding partner		1 2 4 5 6 7 (R) 8 9
hrass	Household real assets (home*perho/100+vbus*sbus/100+car+ ores – mort)		1 2 4 5 6 7 (R) 8 9
hgfass	Household gross financial assets (sum of bacc, bsmf and slti)		1 2 4 5 6 7 (R) 8 9

Variable	Description	Questionnaire	Waves
hnfass	Household net financial assets (hgfass – liab)		1 2 4 5 6 7 (R) 8 9
hnetw	Household net worth (hnfass + hrass)		1 2 4 5 6 7 (R) 8 9
hmem	How much needed to make ends meet	CO020	6 7 (R)
gender	Gender	DN042	1 2 4 5 6 7 (R-S) 8 9
age	Age of respondent (based on interview year)	DN003	1 2 4 5 6 7 (R-S) 8 9
age_p	Age of partner (based on interview year)	DN003	1 2 4 5 6 7 (R-S) 8 9
yedu	Years of education	DN041 (in w1 based on ISCED)	1 2 4 5 6 7 (R-S) 8 9
yedu_p	Year of education of partner	EX102	1 2 4 5 6 7 (R-S) 8 9
isced	ISCED 97 coding		1 2 4 5 6 7 (R-S) 8 9
sphus	Self-perceived health – US scale	PH003	1 2 4 5 6 7 (R-S) 8 9
mstat	Marital status	DN014	1 2 4 5 6 7 (R-S) 8 9
nchild	Number of children	CH001	1 2 4 5 6 7 (R-S) 8 9
ngrchild	Number of grandchildren	CH201	1 2 4 5 6 7 (R-S) 8 9
gali	Limitation with activities	PH005	1 2 4 5 6 7 (R-S) 8 9
chronic	Number of chronic diseases	PH006	1 2 4 5 6 7 (R-S) 8 9
symptoms	Number of symptoms	Ph010	1 2 4
eyesightr	Eyesight reading	PH044	1 2 4 5 6 7 (R) 8 9
hearing	Hearing	PH046	1 2 4 5 6 7 (R) 8 9
bmi	Body mass index	PH012, PH013	1 2 4 5 6 7 (R-S) 8 9
weight	Weight	PH012	1 2 4 5 6 7 (R-S) 8 9
height	Height	PH013	1 2 4 5 6 7 (R-S) 8 9
mobility	Mobility limitations	PH048	1 2 4 5 6 7 (R-S) 8 9
adl	Limitations with activities of daily living	PH049_1	1 2 4 5 6 7 (R-S) 8 9
iadl	Limitations with instrumental activities of daily living	PH049_2	1 2 4 5 6 7 (R-S) 8 9
esmoked	Ever smoked daily	BR001	1 2 4 5 6 7 (R) 8 9
drinking	More than 2 glasses of alcohol almost everyday	BR019	1 2 4 5
phinact	Physical inactivity	BR015	1 2 4 5 6 7 (R) 8 9
reading	Self-rated reading skills	CF001	1 2 4 5 6 7 (R) 8 9
writing	Self-rated writing skills	CF002	1 2 4 5 6 7 (R) 8 9
orienti	Score of orientation in time test	CF003 - CF006	1 2 4 5 6 7 (R) 8 9
wllft	Score of words list learning test – trial 1	CF104_*-CF107_*	1 2 4 5 6 7 (R-S) 8 9
wllst	Score of words list learning test – trial 2	CF113_*-CF116_*	1 2 4 5 6 7 (R-S) 8 9
fluency	Score of verbal fluency test	CF010	1 2 4 5 6 7 (R) 8 9
numeracy	Score of first numeracy test	CF012-CF015	1 2 4 5 6 7 (R) 8 9
numeracy2	Score of second numeracy test	CF108-CF112	4 5 6 7 (R) 8 9
memory	Score of memory test	CF103	4 5 6 7 (R) 8 9
maxgrip	Maximum of grip strength measures	GS006-GS009	1 2 4 5 6 7 (R-S) 8 9
eurod	EURO depression scale	MH002-MH017	1 2 4 5 6 7 (R) 8 9
doctor	Seen/Talked to medical doctor	HC002	1 2 4 5 6 7 (R-S) 8 9
hospital	In hospital last 12 months	HC012	1 2 4 5 6 7 (R-S) 8 9
thospital	Times being patient in hospital	HC013	1 2 4 5 6 7 (R-S) 8 9
nhospital	Total nights stayed in hospital	HC014	1 2 4 5 6 7 (R-S) 8 9
dairyp	How often consume dairy products	BR026	5 6 7 (R) 8 9

Variable	Description	Questionnaire	Waves
legeggs	How often consume legumes, beans or eggs	BR027	5 6 7 (R) 8 9
meat	How often consume meat, fish or poultry	BR028	5 6 7 (R) 8 9
fruit	How often consume fruits or vegetables	BR029	5 6 7 (R) 8 9
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
vacc15	Received vaccinations in first 15 years	MC015	5 6
cjs	Current job situation	EP005	1 2 4 5 6 7 (R) 8 9
pwork	Did any paid work	EP002	1 2 4 5 6 7 (R) 8 9
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 7 (R) 8 9
lookjob	Looking for job	EP337	5 6 7 (R) 8 9
rhfo	Received help from others (how many)	SP002, SP005, SP007	1 2 4 5 6 7 (R) 8 9
ghto	Given help to others (how many)	SP008, SP011, SP013	1 2 4 5 6 7 (R) 8 9
ghih	Given help in the household (how many)	SP018	1 2 4 5 6 7 (R) 8 9
rhih	Received help in the household (how many)	SP020	1 2 4 5 6 7 (R) 8 9
otrf	Owner, tenant or rent free	HO002	1 2 4 5 6 7 (R) 8 9
gfg	Number of given financial gifts 250 or more	FT002, FT007_*	4 5 6 7 (R) 8 9
rfg	Number of received financial gifts 250 or more	FT009, FT014_*	4 5 6 7 (R) 8 9
rggp	Number of received gifts, goods, properties 5000+	FT015, FT020_*	4 5 6 7 (R) 8 9
gggp	Number of given gifts, goods, properties 5000+	FT025, FT031_*	4 5 6 7 (R) 8 9
fdistress	Household able to make ends meet	CO007	1 2 4 5 6 7 (R-S) 8 9
nalm	Number of activities last month	AC002	1 2
naly	Number of activities last year	AC035_*	4 5 6 7 (R-S) 8 9
saly	Satisfied with no activities	AC038	4 5 6 7 (R-S) 8 9
lifesat	Life satisfaction	AC012	2 4 5 6 7 (R-S) 8 9
lifehap	Life happiness	AC022	2 4 5 6 7 (R-S) 8 9
lifex	Living in ten years	EX009	5 6 7 (R-S) 8 9
politics	Left or right in politics	EX028	5 6 7 (R)
bfi10_extra	Big Five – Extraversion	AC701, AC706	7 (R-S) 8 9
bfi10_agree	Big Five – Agreeableness	AC702, AC707	7 (R-S) 8 9
bfi10_consc	Big Five – Conscientiousness	AC703, AC708	7 (R-S) 8 9
bfi10_neuro	Big Five – Neuroticism	AC704, AC709	7 (R-S) 8 9
bfi10_open	Big Five – Openness	AC705, AC710	7 (R-S) 8 9
te_weekday	Time expenditure: what day yesterday	TE002	8 9
te_special	Time expenditure: normal day yesterday	TE003_	8 9

Variable	Description	Questionnaire	Waves
te_chores	Time expenditure: chores (mins)	TE006_ _	8 9
te_pcare	Time expenditure: personal care (mins)	TE012_	8 9
te_children	Time expenditure: children (mins)	TE015_	8 9
te_parents	Time expenditure: parents (mins)	TE018_	8 9
te_partner	Time expenditure: partner (mins)	TE021_	8 9
te_family	Time expenditure: other family members (mins)	TE024_	8 9
te_leisure	Time expenditure: leisure (mins)	TE027_	8 9
te_admin	Time expenditure: administration (mins)	TE033_	8 9
te_pwork	Time expenditure: paid work (mins)	TE036_	8 9
te_vwork	Time expenditure: voluntary work (mins)	TE039_	8 9
te_travel	Time expenditure: traveling from/to work (mins)	TE042_	8 9
te_nap	Time expenditure: napping (mins)	TE048_	8 9
te_sleep	Time expenditure: sleeping (mins)	TE051_	8 9
te_other	Time expenditure:	TE056	8 9
tppdi	Third person present during the interview	IV002	1 2 4 5 6 7 (R-S) 8 9
willans	Willingness to answer	IV004	1 2 4 5 6 7 (R-S) 8 9
clarif	Respondent asked for clarifications	IV007	1 2 4 5 6 7 (R-S) 8 9
undersq	Respondent understood questions	IV008	1 2 4 5 6 7 (R-S) 8 9
hnrsc	Help needed to reed showcards	IV018	1 2 4 5 6 7 (R-S) 8 9
currency	Currency in which amounts are denominated		1 2 4 5 6 7 (R-S) 8 9
nomx*year*	Nominal exchange rate		1 2 4 5 6 7 (R-S) 8 9
pppc*year*	Current PPP exchange rate (national currency/Euro)		1 2 4 5 6 7 (R-S) 8 9
pppk*year*	Constant PPP exchange rate (national currency/Euro), (Germany 2015=1)		1 2 4 5 6 7 (R-S) 8 9

*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), ylsp2 includes EP682 and EP082_2. In LU (wave 6), ylsp2 also includes EP682, EP082_2, and ep082_10. In PL (wave 6) ylsp2 includes only EP082_2.
(c) HH002 not asked in waves 6 and 7.

Table 15: 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 uniformative	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)

Table 16: Meaning of the categories of *htype*

Category	Meaning
1 S	single household
1 CNRP	household with a couple and one non-responding partner
1 C2R	household with a couple, both are respondents
1 CNRP + 1 S	household with a couple including one non-responding partner + one single person
Multi S	household that consists of several singles
1 C2R + 1 S	household with a couple, both are respondents + one single person
Multi C2R	household that consists of several responding couples
Multi CNRP	household with several couples with one non-responding partner

List of references cited in chapters 15.2 (gv_weights) and 15.3 (gv_imputations)

Andridge, R. and Little, R. (2010). A Review of Hot Deck Imputation for Survey Non-Response. International statistical review, 78. 40-64.

Arnold, B.C., Castillo, E. and Sarabia, J.M. (1999). Conditional Specification of Statistical Models. New York: Springer.

Arnold, B.C., Castillo, E. and Sarabia, J.M. (2001). Conditionally specified distributions: An introduction. Statistical Science, 16: 249-274.

Brick, J.M. (2013). Unit nonresponse and weighting adjustments: a critical review. Journal of Official Statistics, 29: 329–353.

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. (eds.), SHARE Wave 5: Innovations & Metodology, Munich: MEA, Max Planck Institute for Social Law and Social Policy.

De Luca, G. and Rossetti, C. (2018). Stata program to compute calibrated weights from scientific use file and additional database. SERISS report D2.10.

De Luca, G., & Rossetti, C. (2019). Weights and imputations. In: M. Bergmann, A. Scherpenzeel & A. Börsch-Supan (eds.), SHARE Wave 7 methodology: Panel innovations and life histories. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

De Luca, G., Li Donni, P., & Rashidi, M. (2021). Weights and imputations in SHARE Wave 8. In: Bergmann, M. & Börsch-Supan, A. (eds.), SHARE Wave 8 Methodology: Collecting cross-national survey data in times of COVID-19. 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.

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.

Haziza, D. and Lesage, E. (2016). A discussion of weighting procedures for unit nonresponse. Journal of Official Statistics 32: 129–145.

Lee, K.J. and Carlin, J.B. (2010). Multiple imputation for missing data: Fully conditional specification versus multivariate normal imputation. American Journal of Epidemiology, 171: 624-632.

Lundström, S. and Särndal, C.E. (1999). Calibration as a standard method for treatment of nonresponse. Journal of Official Statistics, 15: 305-327.

Lynn, P. (2009). Methods for longitudinal surveys. In P. Lynn (Ed.), Methodology of longitudinal surveys (pp. 1-19). Chichester: Wiley.

Molenberghs, G., Fitzmaurice, G., Kenward, M.G., Tsiatis, A. and Verbeke, G. (2015). Handbook of Missing Data Methodology. New York: CRC Press.

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. (1987). Multiple Imputation for Nonresponse in Surveys. New York: Wiley.

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., Boshuizen, H.C. and Knook, D.L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. Statistics in Medicine, 18: 681–694.

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.

Vermeulen, K. and Vansteelandt, S. (2015). Biased-reduced doubly robust estimation. Journal of the American Statistical Association, 110: 1024-1036.

Watson, N. (2014). Evaluation of weighting methods to integrate a top-up sample with an ongoing longitudinal sample. Survey Research Methods, 8: 195-208.