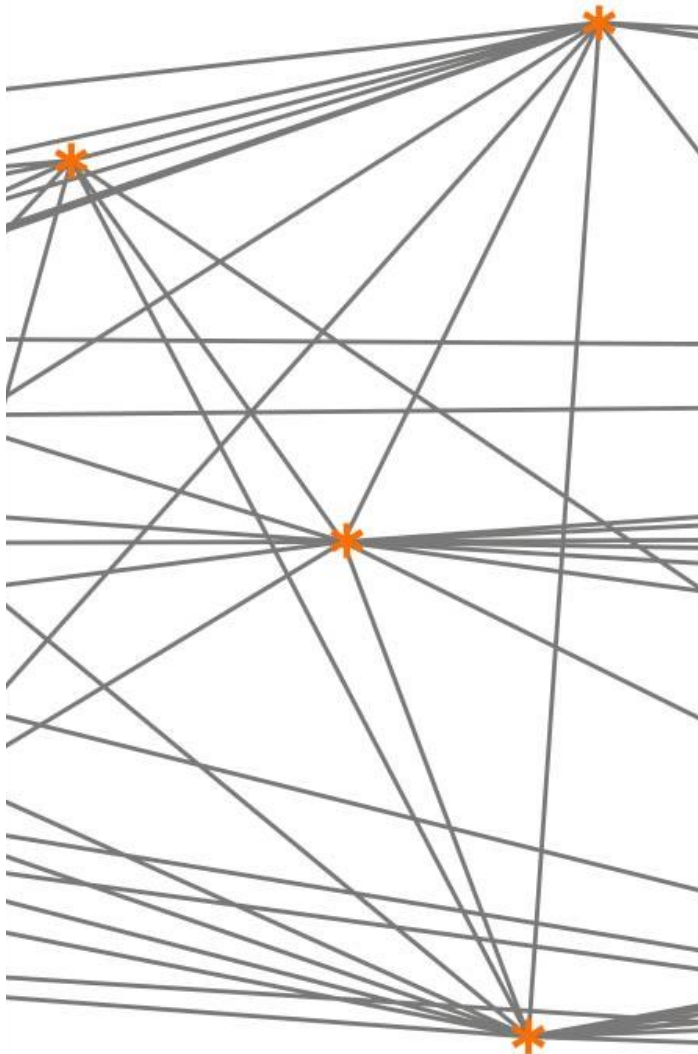




Release Guide 1.0.0

Wave 5



March 31, 2015

Table of Contents

1	Additional Sources of Information	3
1.1	SHARE Website	3
1.2	Methodology Volume	3
2	SHARE data releases	4
3	What is New in SHARE Wave 5	4
4	Countries	4
5	Eligibility rules in SHARE wave 5	6
5.1	Refreshment Samples/New Countries	6
5.2	Longitudinal Sample	6
6	Composition of the Data Set and Types of Respondents in SHARE Wave 5 ...	6
6.1	Types of Data & Questionnaires in Wave 5	7
6.2	Types of Respondents in Wave 5	7
7	New Modules & Items in Wave 5	9
7.1	Mini Childhood Module	9
7.2	IT Module	9
7.3	New Social Exclusion Items	9
8	Self-Completion Questionnaires ("Drop-Off")	10
9	Interviewer Survey	10
10	Citizenship and Country of Birth Coding	10
11	CH module: Looping over children	11
12	CF module: Ten Words List Learning in Wave 5	12
13	SP and FT Module: List of Relations	12
14	EX: Definition of life expectancy target age (ex009_)	14
15	Generated Variables	14
15.1	International Standard Classification of Education (ISCED)	14
15.2	Health	15
15.3	NUTS	15
15.4	Imputations in SHARE Wave 5	16
15.5	Weights	25
15.5.1	Sampling design weights	25
15.5.2	Calibrated weights	26

1 Additional Sources of Information

1.1 SHARE Website

This guide primarily gives an overview of changes and new measures in wave 5, compared to SHARE waves 1, 2 and 4. Information on the naming of variables, missing codes, merging and further general information on the SHARE database can be found in the "SHARE Release Guide 2.6.0" for waves 1 and 2:

http://www.share-project.org/fileadmin/pdf_documentation/SHARE_guide_release_2-6-0.pdf

Specific information on wave 4 is available in the "Release Guide 1.1.1 Wave 4"

http://www.share-project.org/fileadmin/pdf_documentation/SHARE_wave_4_release_guide_1.1.1.pdf

Generic and country-specific questionnaires used during fieldwork of wave 5 are downloadable from the SHARE-website <http://www.share-project.org/data-access-documentation/questionnaires/questionnaire-wave-5.html>.

Since 2013, the SHARE datasets are registered with a Digital Object Identifier (DOI) in order to enable a permanent identification and citability of the SHARE data. The DOI as well as more general information on data access and further documentation material can be found on the website: <http://www.share-project.org/data-access-documentation.html>

1.2 Methodology Volume

This volume documents the most important questionnaire innovations, methodological advancements and new procedures introduced during the fifth wave of SHARE. Users publishing with SHARE should draw special attention to chapter 9 that contains a detailed instruction on citation rules: http://share-dev.mpg.de/fileadmin/pdf_documentation/Method_vol5_31March2015.pdf

2 SHARE data releases

Wave 1 & Wave 2	
Release 1: April 28 th , 2005	
Release 2.0.0: June 19 th , 2007	
Release 2.0.1: July 5 th , 2007	
	Release 1.0.0: November 28 th , 2008
	Release 1.0.1: December 4 th , 2008
Release 2.2.0: August 19 th , 2009	
Release 2.3.0: November 13 th , 2009	
Release 2.3.1: July 28 th , 2010	
Release 2.4.0: March 17 th , 2011	
Release 2.5.0: May 11 th , 2011	
Release 2.6.0: November 29 th , 2013	
Wave 3 (SHARELIFE)	
Release 1.0.0: November 24 th , 2010	
Wave 4	
Release 1.0.0: November 30 th , 2012	
Release 1.1.1: March 28 th , 2013	
Wave 5	
Release 1.0.0: March 31 st , 2015	

3 What is New in SHARE Wave 5

- New country: Luxembourg
- Mini-childhood module
- IT module
- Social exclusion items
- Interviewer questionnaire
- Parents' country of birth

4 Countries

Most of the countries that participated in wave 4 were also part of wave 5. Exceptions are Poland, Portugal, and Hungary. Israel is back again in wave 5 and also one new country joined SHARE: Luxembourg.

Overview: Countries and Language Versions in SHARE Waves 1-5

Country ID	Language ID	Country & language	Wave 1	Wave 2	Wave 3 SHARELIFE	Wave 4	Wave 5
11	11	Austria	2004	2006/07	2008/09	2011	2013
12	12	Germany	2004	2006/07	2008/09	2011/12	2013
13	13	Sweden	2004	2006/07	2008/09	2011	2013
14	14	Netherlands	2004	2007	2008/09	2011	2013
15	15	Spain (Castilian)	2004	2006/07	2008/09	2011	2013
15	39	Spain/Girona (Catalan)	-	-	-	-	2013
15	40	Spain/Girona (Castilian)	-	-	-	-	2013
16	16	Italy	2004	2006/07	2008/09	2011	2013
17	17	France	2004/05	2006/07	2009	2011	2013
18	18	Denmark	2004	2006/07	2008/09	2011	2013
19	19	Greece	2004/05	2007	2008/09	-	-
20	20	Switzerland (German)	2004	2006/07	2008/09	2011	2013
20	21	Switzerland (French)	2004	2006/07	2008/09	2011	2013
20	22	Switzerland (Italian)	2004	2006/07	2008/09	2011	2013
23	23	Belgium (French)	2004/05	2006/07	2008/09	2011	2013
23	24	Belgium (Flemish)	2004/05	2006/07	2008/09	2011	2013
25	25	Israel (Hebrew)	2005/06	2009/10	-	-	2013
25	26	Israel (Arabic)	2005/06	2009/10	-	-	2013
25	27	Israel (Russian)	2005/06	2009/10	-	-	2013
28	28	Czech Republic	-	2006/07	2008/09	2011	2013
29	29	Poland	-	2006/07	2008/09	2011/12	-
30	30	Ireland	-	2007	-	-	-
31	41	Luxembourg (French)	-	-	-	-	2013
31	42	Luxembourg (German)	-	-	-	-	2013
32	32	Hungary	-	-	-	2011	-
33	33	Portugal	-	-	-	2011	-
34	34	Slovenia	-	-	-	2011	2013
35	35	Estonia (Estonian or Russian)	-	-	-	2010/11	2013 (XT only)
35	36	Estonia (Estonian)	-	-	-	-	2013
35	37	Estonia (Russian)	-	-	-	-	2013

5 Eligibility rules in SHARE wave 5

5.1 Refreshment Samples/New Countries

The target population for the baseline samples consists of *all persons born 1962 or earlier* having their regular domicile in the respective country. Current partners or spouses of target persons are also eligible for a SHARE interview independent of age.

5.2 Longitudinal Sample

The target population for the longitudinal sample consists of all sample members who were interviewed in any previous wave of SHARE and their current partners or spouses (independent of age and independent of their participation in previous waves). Age eligible respondents who already participated are traced and re-interviewed if they move within the country and end-of-life interviews are conducted if they decease. Respondents who moved within the country or moved into a nursing home, hospital or other old-age institution have been traced and re-interviewed, whereas respondents who have been incarcerated or moved abroad were not followed.

6 Composition of the Data Set and Types of Respondents in SHARE Wave 5

Please see waves 1 & 2 "Release Guide 2.6.0" Chapter 6 for general information on this topic.

Using wave 5 data, users should be aware that some steady information (e.g. education) is only available in the first interview of a respondent, the so-called baseline interview; other information is only asked in longitudinal interviews and updated if it has changed (e.g. marital status). Thus such variables have a high percentage of missing information in longitudinal interviews. This can be solved by using the information from the baseline interview in wave 1, 2 or 4. To merge the waves, you either have to take care of the order in which you merge them and use e.g. 'merge..., update' in Stata or rename the variables first (e.g. dn014_w1) and combine them after merging.

The variable mn101_ in the DN module contains the information if a baseline or a longitudinal interview was conducted.

6.1 Types of Data & Questionnaires in Wave 5

Overview: Composition of the SHARE Wave 5 Data Set

CAPI data	
Coverscreen interview (cv_r)	Data for all household members on the individual level, including non-eligible persons and deceased respondents.
Individual CAPI modules	See the overview in chapter 6.2
End of life interviews (xt)	Interview with a proxy if the respondent deceased.
Generated variables	
ISCED codes for education	ISCED97 and ISCED2011 (see chapter 15.1)
Physical and mental health	See chapter 15.2
Housing and region	See chapter 15.3
Imputations	See chapter 15.4
Weights	See chapter 15.5
Paper and pencil questionnaires	
Drop-off	Only country specific questions in wave 5.
Online questionnaires	
Interviewer survey	Basic demographics of some interviewers as well as questions about interviewers' attitudes, expectations and experiences towards surveys (see chapter 9).
Linking SHARE with German administrative data	
Record Linkage (SHARE-RV)	Facilitates the direct linkage of the SHARE survey data with administrative records of the German Pension Fund (DRV). For further information see http://www.share-project.org/data-access-documentation/record-linkage-share-rv.html

6.2 Types of Respondents in Wave 5

As in previous waves, some selected household members serve as special respondents. They answer specific modules or questions concerning e.g. household income also on behalf of other household members/partners. The variables that indicate family respondent (dumfamr), financial respondent (dumfinr) and household respondent (dumhhr) are stored in the cv_r module.

Overview: Types of respondents

Type of respondent	Variable name	Name of Filter in Questionnaire
Family respondent	dumfamr	MN006_NumFamR
Financial respondent	dumfinr	MN007_NumFinR
Household respondent	dumhhr	MN008_NumHHR

Overview: Who Answers What in the CAPI Questionnaire?

CAPI Module	Name	All respondents ¹	Household Resp.	Financial Resp.	Family Resp.	non-proxy
CV_R	Coverscreen	CV-respondent				
DN	Demographics	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
CS	Chair Stand	x				x
SP	Social Support	x (partly)			x (partly)	
FT	Financial Transfers			x		
HO	Housing		x			
HH	Household Income		x			
CO	Consumption		x			
AS	Assets			x		
AC	Activities	x				x
EX	Expectations	x				x
IV	Interviewer Observations	interviewer				
XT	End-of-Life Interview	proxy respondents				

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

7 New Modules & Items in Wave 5

7.1 Mini Childhood Module

The mini childhood module (MC) is only asked if the respondent did not participate in the SHARELIFE interview of wave 3. The variable mn031_ in the DN module shows who was asked the mc module. It contains questions on living conditions in the respondents' childhood: accommodation, number of books in the residence, performance in Maths and language, financial situation and childhood health conditions.

Please note that the questions cover various reference periods. Some questions refer to the time when the respondent was ten years old, others to the period between birth and the age of 15.

See chapter 2.2 of the SHARE wave 5 Innovations and Methodology volume for further information on the mini childhood module.

7.2 IT Module

The IT module comprises four short questions on computer utilisation at work and computer skills.

7.3 New Social Exclusion Items

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:

Overview: 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 other household members/partners. Please note that due to an unnoticed element in the instrument not all households were asked the set of questions.

Altogether 94% of the households are covered. The variable `mn032_` in the DN module contains information on who was asked the social exclusion items.

Further information on the social exclusion items is available in chapter 2.3 of the SHARE wave 5 Innovations and Methodology volume.

8 Self-Completion Questionnaires (“Drop-Off”)

In wave 5 the drop-off questionnaires include country specific questions. Only three countries conducted a drop-off questionnaire: Austria, the Czech Republic, and Israel. There are two versions of the Czech drop-off. The first version (`cz_a_q*`) was asked to respondents who are retired whereas the second version (`cz_b_q*`) was asked to non-retired respondents.

9 Interviewer Survey

The SHARE wave 5 interviewer questionnaire 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.

Six of the wave 5 SHARE countries participated in the SHARE interviewer survey: Austria, Belgium, Germany, Spain, Sweden, and Slovenia. 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, an interviewer ID variable (`intid`) is provided.

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

10 Citizenship and Country of Birth Coding

Citizenship (`dn008`) and country of birth (`dn005`) are coded according to ISO 3166-1 (numeric-3). Same applies for the newly included country of birth of the respondents’ mother (`dn504`) and father (`dn505`). The introduction of the latter variables allows identifying second-generation migrants now. Further information on this topic is provided in chapter 2.5 of the Innovations and Methodology volume. The list of ISO codes is available at:

<http://unstats.un.org/unsd/methods/m49/m49alpha.htm>

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

Overview of additional codes for country of birth and citizenship:

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

11 CH module: Looping over children

The CH module is answered by singles and family respondents on behalf of the couple. All variables with the same suffix belong to the same child, so that e.g. for "child 2" the year of birth is 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".

Longitudinal respondents were asked about changes concerning their children since the last interview. In the data there are three types of variables for indicating such a change. The first variable indicates if 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.

Overview: children change

	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

12 CF module: Ten Words List Learning in Wave 5

As in previous waves, the “ten words list learning” test has been conducted with a first trial and a delayed recall. Similar to wave 4, 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 5 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.

13 SP and FT Module: List of Relations

Analogous to waves 1 and 2, in wave 5 variables in the social support (SP) and the financial transfer (FT) module refer to a list of relations. Note that in wave 4 the list of relations includes potential social network members from the SN module. The following overview illustrates the coding and variable naming across waves in the SP and FT module:

Relationship type coding and variable names across waves

Value or variable label	Wave 1, 2 & 5 values	Wave 4 values	Wave 1, 2 & 5 variable suffix	Wave 4 variable suffix
social network member1	-	101	-	d1sn
social network member2	-	102	-	d2sn
social network member3	-	103	-	d3sn
social network member4	-	104	-	d4sn
social network member5	-	105	-	d5sn
social network member6	-	106	-	d6sn
social network member7	-	107	-	d7sn
spouse/partner	1	1	d1	d1sp
mother	2	2	d2	d2sp
father	3	3	d3	d3sp
mother-in-law	4	4	d4	d4sp
father-in-law	5	5	d5	d5sp
stepmother	6	6	d6	d6sp
stepfather	7	7	d7	d7sp
brother	8	8	d8	d8sp
sister	9	9	d9	d9sp
child 1	10	-	d10	-
child 2	11	-	d11	-
child 3	12	-	d12	-
child 4	13	-	d13	-
child 5	14	-	d14	-
child 6	15	-	d15	-
child 7	16	-	d16	-
child 8	17	-	d17	-
child 9	18	-	d18	-
other child	19	19	d19	d19sp
son-in-law	20	20	d20	d20sp
daughter-in-law	21	21	d21	d21sp
grandchild	22	22	d22	d22sp
grandparent	23	23	d23	d23sp
aunt	24	24	d24	d24sp
uncle	25	25	d25	d25sp
niece	26	26	d26	d26sp
nephew	27	27	d27	d27sp
other relative	28	28	d28	d28sp
friend	29	29	d29	d29sp
ex-colleague	30	30	d30	d30sp
neighbour	31	31	d31	d31sp
ex-spouse/partner	32	32	d32	d32sp
other acquaintance	33	-	d33	-
step-child/your current partner's child	-	34	-	d34sp
minister, priest, or other clergy	-	35	-	d35sp
therapist or other professional helper	-	36	-	d36sp
housekeeper/home health care provider	-	37	-	d37sp
NONE OF THESE	96	96	dno	dno

14 EX: Definition of life expectancy target age (ex009_)

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 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 actually 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 deviates from respondents' age.

15 Generated Variables

15.1 International Standard Classification of Education (ISCED)

Education is one of the most diverse international variables. SHARE is using the International Standard Classification of Education (ISCED) which allows for the standardized reporting of education statistics according to an internationally agreed set of definitions and concepts. For further information see <http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx>.

So far, SHARE used the ISCED-97 version. In wave 5 we additionally provide ISCED 2011. This revised version takes into account significant changes in education systems worldwide since the last ISCED revision in 1997. Please note that the official ISCED 2011 mappings done by the UNESCO Institute for Statistics in close cooperation with member states are not yet publicly available. Thus our ISCED 2011 codes should be considered as preliminary. As in previous waves, we collect information on education only in the baseline interview of a respondent indicated by the variable *mn101_* in the DN module.

15.2 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).

In addition to the indices that were part of previous SHARE waves like the body mass index (BMI), self-perceived health (`sphus`), limitations with (instrumental) activities of daily living (ADL and IADL), the numeracy score or the EURO-D depression scale, a scale for loneliness is included in wave 5. The measure was developed by Hughes and colleagues in 2004 to assess loneliness in large scale surveys (Hughes et al. 2004). The variable "loneliness" in the `gv_health` module contains a sum score based on reverse-coding of three items on loneliness from the mental health module: `mh034_`, `mh035_`, and `mh036_`.

References:

Hughes M., Waite L., Hawkey L., Cacioppo J. (2004). A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies. *Res Aging*. 2004;26(6):655-672

15.3 NUTS

As in previous waves the Nomenclature of Territorial Units for Statistics (NUTS) is used to indicate in which territorial unit the household is located. In wave 5 the variables are named analogous to wave 4 `nuts1_2010`, `nuts2_2010` and `nuts3_2010`. The 2010 suffix is referring to the NUTS classification of 2010 used since wave 4. Be aware that the NUTS codes of previous waves might be different, due to a different NUTS classification being in use at the time of sampling.

For the first time in wave 5, regional codes analogous to NUTS are also available for Israel. In cooperation with SHARE Central, the Israeli country team subdivided the country into seven regions (North, Haifa, Center, Tel-Aviv, Jerusalem, South, Judea and Samaria) that are similar in its population size and thus comparable to the level 1 NUTS regions in the European countries.

NUTS codes are only provided for households that joined SHARE in wave 5 (refreshment sample). Please note that due to country-specific privacy legislations not all NUTS levels are provided for every country.

15.4 Imputations in SHARE Wave 5

Authors: Giuseppe De Luca, Andrea Bonfanti, Martina Celidoni, and Elisabetta Trevisan

The imputation procedure used in wave 4 and wave 5 presents important innovations with respect to the procedure adopted in wave 1 and wave 2. Two differences are particularly striking. First, some items are now imputed in aggregate terms to simplify the computational burden of the imputation model. For similar reasons, separate imputations for longitudinal and refreshment subsamples are no longer considered and lagged variables from previous waves are not used as predetermined predictors any more. The second important difference is that we handle the problem of non-responding partners (NRPs) differently, namely the fact that only one of the two partners may have agreed to be interviewed. Unlike the strategy adopted in the first two waves, NRPs are now viewed as a problem of unit nonresponse due to the limited information available to cope with this type of nonresponse error. Our imputation procedure provides only an indirect estimate of the income from NRPs to avoid understating total household income when only one of the two partners was interviewed. As discussed below, the strategy used to recover this information exploits the distinction between couples with and without NRPs and additional information obtained from a one shot question on monthly household income (hh017). Harmonized imputations for all waves of SHARE are planned to be delivered in the near future.

SHARE contains a huge amount of detailed information about demographics, physical and mental health, cognitive abilities, social activities, expectations, employment status and incomes, housing and assets, and all these sections are affected by missing data. As in the previous waves, most of the variables collected in the fifth wave of SHARE are affected by small amounts of missing data (usually lower than 5%). Non-negligible amounts of missing data occurred instead for monetary variables about incomes, assets and consumption expenditures. To simplify complexity of the imputation model, only this subset of monetary variables has been imputed using the fully conditional specification (FCS) method (van Buuren et al. 2006) which attempts to preserve the correlation structure of the imputed variables. Moreover, we reduced the number of monetary variables that had to be imputed jointly by aggregating 55 items on income, wealth and consumption expenditure into 17 aggregated variables. Each aggregated variable is obtained by summing two or more original items as illustrated in Table 1.

Variables with a low prevalence of missing values has been imputed by univariate methods such as hot-deck and regression imputations. These imputations were performed in an early stage separately by country. We first imputed socio-demographic characteristics such as age and education that were affected by a small fraction of missing values so that these variables could then

be used as exogenously observed predictors in the imputation of the other variables. Our set of predictors for hot-deck imputations typically included gender, age group, years of education and self-reported health. For some variables additional predictors were used. For example, we also employed the number of children when imputing the number of grandchildren and an indicator for being a patient in a hospital overnight during the last year when imputing health-related variables. Variables that were known to be logically related, such as respondent's weight, height and body mass index, were imputed simultaneously by hot-deck.

FCS imputations were performed separately by country and household type to allow for heterogeneity across these different groups. The household types considered were singles and third respondents (sample 1), couples with both partners interviewed (sample 2), and all couples – with and without NRPs (sample 3). The set of variables imputed jointly by the FCS method was country- and sample-specific as we required that the sample used in the estimation step includes at least 100 donor observations in sample 1 and 150 donor observations in samples 2 and 3. Monetary variables that did not satisfy this requirement were imputed first and then used as observed predictors in the imputation of the other variables.

The imputation of each monetary variable was always carried out on the basis of a two-part model that involved a probit model for ownership and a regression model for the amount conditional on ownership. Strictly positive variables were transformed in logarithms, while variables that may also take negative values (e.g. income from self-employment, bank account and value of own business) were transformed by the inverse hyperbolic sine transformation.

The set of exogenous predictors was also sample-specific. For singles and third respondents, it included 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. For couples with both partners interviewed, we used a larger set of predictors that also included the mentioned variables for the partner of the designated respondent. For couples with NRPs, the predictors referring to the NRPs were confined to age and years of education only.

Imputed monetary values were always constrained to fall within the individual-level bounds that incorporated the partial information available on missing observations. These bounds incorporate the available information about unfolding brackets, country distributions, single item amounts for the aggregated variables or individual amounts for household variables. Imputed values are necessarily included between those two bounds.

As mentioned above, particular attention was devoted to the imputation of total household income because SHARE provides two alternative measures of this variable. The first measure ("thinc") could be obtained by a suitable aggregation

at the household level of all individual income components, while the second ("thinc2") could be obtained from the one-shot question on monthly household income (hh017). The choice between these two alternative measures is not obvious. There are reasons to believe that none of the two measures of total household income can be strictly preferred to the other and thus we let the users decide which of the two measures was more suitable for their research questions. For our imputation purposes, the availability of these two alternative measures greatly improved the imputation process because each measure could contribute relevant information on the missing values of the other measure. Our procedure to impute these two measures of total household income consisted of three stages.

Stage 1 (singles and 3rd respondent). We imputed all monetary variables by the FCS method discussed before. At the end of each iteration, 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. We finally imputed the second version of total household income (thinc2) using total household income (thinc), household net worth (hnetw), total household expenditure (thexp), and characteristics of the household respondent as predictors.

Stage 2 (couples with both partners interviewed). We used an imputation strategy similar to that adopted in stage 1, but with a larger set of predictors that also includes characteristics of the partner of the designed respondent.

Stage 3 (all couples – with and without NRPs). Imputed values of all variables for the subsample of couples with both partners interviewed were obtained from stage 2. In stage 3, these couples entered the imputation sample only as observations available for the imputation of missing values on the other subsample of couples with NRPs. Similarly to the previous stages, we first imputed all monetary variables for the responding partners by standard implementation of the FCS method. Predictors referring to the NRPs now consisted however of age and years of education only. At the end of each iteration, we also imputed total household income (thinc2) using household net worth (hnetw), total household expenditure (thexp) and characteristics of the responding partner as predictors. For all couples with NRPs, we finally imputed the total household income (thinc) using the second version of total household income (thinc2), household net worth (hnetw), total household expenditure (thexp) and characteristics of the responding partner as predictors, couples with two partners interviewed as observations available for the estimation step, and the imputed sum of incomes of the responding partner as lower bound.

For additional information on the imputation procedure we refer the reader to De Luca et al (2015). The list of variables included in the imputation dataset is provided in Table 2. The status of each imputed variable in this dataset is summarized by a flag variable (*variablename_f*) described in table 2. All

variables in the dataset already contain labels. Given the differences with previous imputations datasets, users should check carefully the tables included in the documentation in order to individualize the correct correspondences between variables.

Notice that like in wave 4 amounts are all expressed in Euro and not in local currency. This implies that when applying PPP exchange rates to monetary variables for non-Euro countries users should first convert them in local currency using "excrate" and then divide them by the PPP exchange rate. Moreover, all monetary variables are expressed in annual terms, independently from the original periodicity indicated in the questionnaire.

References:

De Luca, G., Celidoni, M., and Trevisan, E. (2011). Item Nonresponse and Imputation Strategies in Wave 5. in Malter F. and A. Börsch-Supan (ed.), SHARE Wave 5: Innovations & Metodology, Munich: MEA, Max Planck Institute for Social Law and Social Policy.

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(2): 1049-1064.

Table 1: Aggregate variables in wave 5

Aggregate variables	Components	Variable Name
Regular payments from public old age, early retirement, survivor and war pensions	Public old age pension Public old age supplementary pension Public early retirement pension Main public survivor pension Secondary public survivor pension Public war pension	EP078_1 EP078_2 EP078_3 EP078_7 EP078_8 EP078_9
Regular payments from private occupational pensions	Occupational old age pension from last job Occupational old age pension from second job Occupational old age pension from third job Occupational early retirement pension Occupational disability or invalidity insurance Occupational survivor pension	EP078_11 EP078_12 EP078_13 EP078_14 EP078_15 EP078_16
Regular payments from disability pensions and benefits	Main public disability insurance pension Secondary public disability insurance pension	EP078_4 EP078_5
Regular payments of other private pensions	Regular life insurance payments Regular private annuity or personal pension payments Long-term care payments from private insurance	EP094_1 EP094_2 EP094_5
Regular payments from private transfers	Alimony Regular payment from charities	EP094_3 EP094_4

Aggregate variables	Components	Variable Name
Lump-sum payments from public old age, early retirement, survivor and war pensions	Lump-sum payments from public old age pension	EP082_1
	Lump-sum payments from public old age supplementary pension	EP082_2
	Lump-sum payments from public early retirement pension	EP082_3
	Lump-sum payments from main public survivor pension	EP082_7
	Lump-sum payments from secondary public survivor pension	EP082_8
	Lump-sum payments from public war pension	EP082_9
Lump-sum payments from private occupational pensions	Lump-sum payments from occupational old age pension from last job	EP082_11
	Lump-sum payments from occupational old age pension from second job	EP082_12
	Lump-sum payments from occupational old age pension from third job	EP082_13
	Lump-sum payments from occupational early retirement pension	EP082_14
	Lump-sum payments from occupational disability or invalidity insurance	EP082_15
	Lump-sum payments from occupational survivor pension	EP082_16
Lump-sum payments from disability pensions and benefits	Lump-sum payments from main public disability insurance pension	EP082_4
	Lump-sum payments from secondary public disability insurance pension	EP082_5
Lump-sum payments of other private pensions	Lump-sum payments from life insurance	EP209_1
	Lump-sum payments from private annuity or personal pension	EP209_2
	Lump-sum payments from long-term care private insurance	EP209_5
Lump-sum payments from private transfers	Lump-sum payments from alimony	EP209_3
	Lump-sum payments from charities	EP209_4
Rent and home-related expenditures	Amount rent paid	HO005
	Other home-related expenditures	HO008
Income from rent or sublet	Income from sublet	HO074
	Income from rent of real estate	HO030
Income from other household members	Other household members' net income	HH002
	Other household members' net income from other sources	HH011
Bond, stock and mutual funds	Government/corporate bonds	AS007
	Stocks	AS011
	Mutual funds	AS017
Savings in long term Investments	Individual retirement accounts from respondent	AS021
	Individual retirement accounts from partner	AS024
	Contractual savings	AS027
	Whole life insurance holdings	AS030
Paid out-of-pocket for outpatient care	Paid out-of-pocket for doctor visits	HC083
	Paid out-of-pocket for dental care	HC093
Paid out-of-pocket for nursing home and home-based care	Paid out-of-pocket for home-based care	HC129
	Paid out-of-pocket for nursing home	HC097

Table 2: List of variables included in the imputation dataset of wave 5

Variable name	Description	Questionnaire
mergeid	Person ID	
implicat	Implicat number	
hhidcom5	Household ID wave 5	
country	Country identifier	
language	Language of questionnaire	
htype	Household type	
fam_resp	Family respondent	MN006_
fin_resp	Financial respondent	MN007_
hou_resp	Household respondent	MN008_
excrate	Exchange rate	
nursinghome	Living in nursing home	MN024_
hsize	Household size	
single	Single	
couple	Couple	
idcouple	Couple ID	
partner	Partner in the couple	
p_nrp	Partner of non-responding partner	
sample1	Imputation sample for single	
sample2	Imputation sample for couples with two partners interviewed	
sample3	Imputation sample for all couples	
inpat	Out-of-pocket payment for inpatient care	HC095
outpat	Out-of-pocket payment for outpatient care	HC083
drugs	Out-of-pocket payment for drugs	HC089
nurs	Out-of-pocket payment for nursing home / home care	HC097
ydip	Earnings from employment	EP205
yind	Earnings from self-employment	EP207
ypen1	Annual old age, early retirement pensions, survivor and war pension	EP078_1-2-3-7-8-9
ypen2	Annual private occupational pensions	EP078_11-16
ypen3	Annual disability pension and benefits	EP078_4-5
ypen4	Annual unemployment benefits and insurance	EP078_6
ypen5	Annual payment from social assistance	EP078_10
ylsp1	Lump sum payments for old age, early retirement, survivor and war pension	EP082_1-2-3-7-8-9
ylsp2	Lump sum payments for private occupational pension	EP082_11-16
ylsp3	Lump sum payments for disability pension and benefits	EP082_4-5
ylsp4	Lump sum payments for unemployment benefits and insurance	EP082_6
ylsp5	Lump sum payments for social assistance	EP082_10

Variable name	Description	Questionnaire
ylsr1	Lump sum from private payments	EP209_1-2-5
ylsr2	Lump sum from private transfers	EP209_3-4
yreg1	Other regular payments from private pensions	EP094_1-2-5
yreg2	Other regular payment from private transfer	EP094_3-4
home	Value of main residence	HO024
mort	Mortgage on main residence	HO015
rhre	Annual rent and home-related expenditures	HO005, HO008
ores	Value of other real estate - Amount	HO027
ysrent	Annual income from rent or sublet	HO074, HO030
yaohm	Annual income from other household members	HO002, HO011
fahc	Annual food at home consumption	CO002
fohc	Annual food outside home consumption	CO003
hprc	Annual home produced consumption	CO011
bacc	Bank accounts	AS003
bsmf	Bond, stock and mutual funds	AS007, AS011, AS017
ybabsmf	Interest / dividend from bank account, bond, stock, and mutual funds	
slti	Savings for long-term investments	AS021, AS023, AS27, AS030
vbus	Value of own business	AS042
sbus	Share of own business	AS044
car	Value of cars	AS051
liab	Financial liabilities	AS055
thinc	Total household net income - version A	
thinc2	Total household net income - version B	HH017
thexp	Total household expenditure (sum of rhre, fahc, fohc and hprc)	
yincnrp	Income from non-responding partner	
hrass	Household real assets (home*perho/100+vbus*sbus/100+car+ ores - mort)	
hgfass	Household gross financial assets (sum of back, bsmf and slti)	
hnfass	Household net financial assets (hgfass - liab)	
hnetw	Household net worth	
gender	Gender	DN042
age	Age in 2013	DN003
age_p	Age of partner in 2013	DN003
yeduc	Year of education	DN041
yeduc_p	Year of education of partner	EX102
iscled	ISCED 97 coding	
sphus	Self-perceived health - US scale	PH003
mstat	Marital status	DN014

Variable name	Description	Questionnaire
nchild	Number of children	CH001
ngcchild	Number of grandchildren	CH201
gali	Limitation with activities	PH005
chronic	Number of chronic diseases	PH006
eyesightr	Eyesight reading	PH044
hearing	Hearing	PH046
bmi	Body mass index	PH012, PH013
weight	Weight	PH012
height	Height	PH013
mobility	Mobility limitations	PH048
adl	Limitations with activities of daily living	PH049_1
iadl	Limitations with instrumental activities of daily living	PH049_2
esmoked	Ever smoked daily	BR001
drinking	More than 2 glasses of alcohol almost everyday	BR019
phinactiv	Physical inactivity	BR015
dairyp	How often consume dairy products	BR026
legeggs	How often consume legumes, beans or eggs	BR027
meat	How often consume meat, fish or poultry	BR028
fruit	How often consume fruits or vegetables	BR029
reading	Self-rated reading skills	CF001
writing	Self-rated writing skills	CF002
orienti	Score of orientation in time test	CF003 - CF006
memory	Score of memory test	CF103
wllft	Score of words list learning test - trial 1	CF104_* - CF107_*
wllst	Score of words list learning test - trial 2	CF113_* - CF116_*
fluency	Score of verbal fluency test	CF010
numeracy1	Score of first numeracy test	CF012 - CF015
numeracy2	Score of second numeracy test	CF108 - CF112
maxgrip	Maximum of grip strength measures	GS006 - GS009
eurod	EURO depression scale	MH002 - MH017
doctor	Seen/Talked to medical doctor	HC002
hospital	In hospital last 12 months	HC012
thospital	Times being patient in hospital	HC013
nhospital	Total nights stayed in hospital	HC014
prre10	Private residence at the age of 10	MC002
room10	Rooms at home at the age of 10	MC003
people10	Number of people at home at the age of 10	MC004
book10	Number of books at home at the age of 10	MC005
math10	Math performance at the age of 10	MC006
lang10	Language performance at the age of 10	MC007

Variable name	Description	Questionnaire
finance15	Financial situation of the family in first 15 years	MC009
health15	Health status in first 15 years	MC010
disease15	Number of childhood diseases in first 15 years	MC012
illness15	Number of childhood illnesses in first 15 years	MC013
vacc15	Received vaccinations in first 15 years	MC015
cjs	Current job situation	EP005
pwork	Did any paid work	EP002
empstat	Employee or self-employed	EP009
lookjob	Looking for job	EP337
rhfo	Received help from others (how many)	SP002, SP005, SP007
ghfo	Given help to others (how many)	SP008, SP011, SP013
ghih	Given help in the household (how many)	SP018
rhih	Received help in the household (how many)	SP020
gfg	Number of given financial gifts 250 or more	FT002, FT007_*
rfg	Number of received financial gifts 250 or more	FT009, FT014_*
gggp	Number of given gifts, goods, properties 5000 or more	FT025, FT031_*
rggp	Number of received gifts, goods, properties 5000 or more	FT015, FT020_*
otr	Owner, tenant or rent free	HO002
perho	Percentage of house owned	HO070
fdistress	Household able to make ends meet	CO007
lifesat	Life satisfaction	AC012
lifehap	Life happiness	AC022
lifex	Living in ten years	EX009
poltics	Left or right in politics	EX028
naly	Number of activities last year	AC035_*
saly	Satisfied with no activities	AC038
tppdi	Third person present during the interview	IV002
clarify	Respondent asked for clarifications	IV007
undersq	Respondent understood questions	IV008
hnrsc	Help needed to read showcards	IV018
nomxyear	Nominal exchange rate	
pppxyear	PPP adjusted exchange rates	
currency	Currency in which amounts are denominated	

Table 3: Description of the flag variable

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

15.5 Weights

Authors: Giuseppe De Luca & Claudio Rossetti

15.5.1 Sampling design weights

Sampling design weights, defined as the inverse of the probability of being included in the sample of any specific wave, compensate for unequal selection probabilities of the sample units. They allow obtaining unbiased estimators of the population parameters only under the ideal situation of complete response.

Unfortunately, the SHARE data are affected by problems of unit non-response (i.e., eligible sample units fail to participate in the survey because of either noncontact or explicit refusal to cooperate) and sample attrition (i.e., responding units in a given wave of the panel drop out in a subsequent wave). Therefore, estimators constructed using sample design weights alone, and ignoring unit non-response and attrition, may be biased (Lessler and Kalsbeek 1992). Although sample design weights are included in the public release of the SHARE data, we strongly discourage users to rely on these weights unless they are used for the implementation of specific statistical methods which account for non-response errors in other ways, or for other specific purposes.

In this release of the data, sampling design weights are computed under the simplifying assumption of simple random sampling in each country. Proper sampling design weights are planned to be delivered in the near future. A documentation of the national sampling designs used in the fifth wave can be found in De Luca et al (2015). Similarly to the previous waves, the sampling design weights at the individual and the household levels coincide because the inclusion probability of any eligible household member is by design the same as the inclusion probability of the whole household. Thus, the public release of the wave 5 weighting dataset includes just one sampling design weight.

15.5.2 Calibrated weights

The strategy used by SHARE to cope with the potential selection bias generated by unit non-response and panel attrition relies on the ex-post calibration procedure of Deville and Särndal (1992). This statistical re-weighting procedure gives calibrated weights which are as close as possible, according to a given distance measure, to the original design weights while also respecting a set of known population totals (the calibration margins). Under certain assumptions about the missing data process, calibrated weights may help reduce the potential selectivity bias generated by unit non-response and panel attrition. The key assumption is that, after conditioning on a set of variables (the calibration variables), there is no relation between the response probability and the other key survey variables excluded from the conditioning set. Using the terminology introduced by Rubin (1987) this corresponds to assuming that the process generating missing observations is missing-at-random (MAR).

For methodological details about the calibration procedure used in SHARE, we refer the reader to De Luca et al (2015), Lynn et al (2013) and the SHARE release guide 2.5.0. In what follows, we only outline the main features of the calibrated weights included in the public release of the wave 5 SHARE data and the most important differences with respect to the calibrated weights of the previous waves.

- As in the previous waves, the wave 5 data include calibrated cross-sectional weights to be used in the context of cross-sectional analyses. Calibrated longitudinal weights for the balanced panel sample of the last two waves (i.e. the sample of 50+ respondents of wave 4 and wave 5) are planned to be provided to be delivered in the near future.
- Since the basic units of analysis can be either individuals or households, calibrated cross-sectional weights can be 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.
- Because of changes in the imputation procedure of wave 5 (see the documentation about imputations), SHARE does not provide any more calibrated cross-sectional weights for non-responding partners (i.e.

calibrated cross-sectional weights at the individual level where non-responding partners are treated as true respondents). Once the new imputation procedure will be extended to the previous waves, calibrated cross-sectional weights for non-responding partners of waves 1 and 2 will be deleted from the SHARE data archive.

- To simplify the structure of the wave 5 data, SHARE will not provide any more calibrated longitudinal weights for all possible wave combinations.
- The structure of variable names for design and calibrated weights are different from those used in the first two waves because we do not distinguish any more between alternative variants of the SHARE sample (i.e. main sample alone, vignette sample alone and the two samples combined). Design weights are named as `dw_w5`, calibrated household weights are named as `chw_w5`, and calibrated individual weights are named as `ciw_w5`.

Calibrated cross-sectional weights are defined for the sample of 50+ respondents (either individuals or households) in wave 5 by ignoring the distinction between longitudinal and refreshment samples. At the individual level (i.e. variable `ciw_w5`), each 50+ respondent receives a calibrated weight that depends on the household design weight and the respondent's set of calibration variables. At the household level (i.e. variable `chw_w5`), each interviewed household member receives a common calibrated weight that depends on the household design weight and the calibration variables of all 50+ respondents in the same household.

Calibrated weights are always computed separately by country to match the size of the national populations of individuals born in 1962 or earlier. Within each country, we used a set of calibration margins for the size of the target population across 8 gender-age groups (i.e. males and females with year of birth in the classes `[-1932]`, `[1933-42]`, `[1943-52]`, `[1953-62]`) and across NUTS1 regional areas.

For each type of calibrated weight, we also provide a flag variable which is equal to 1 whenever the corresponding calibrated weight is missing. This occurs for respondents younger than 50 years (i.e. age-ineligible partners of an age-eligible respondent), those with missing information on the set of calibration variables (i.e. year of birth, gender and NUTS1 code), and those with missing sampling design weights (i.e., respondents with missing sampling frame information).

References

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

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

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

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

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

Responsible for database management and preparation of the releases:

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

and all SHARE country team operators.

Contact: info@share-project.org