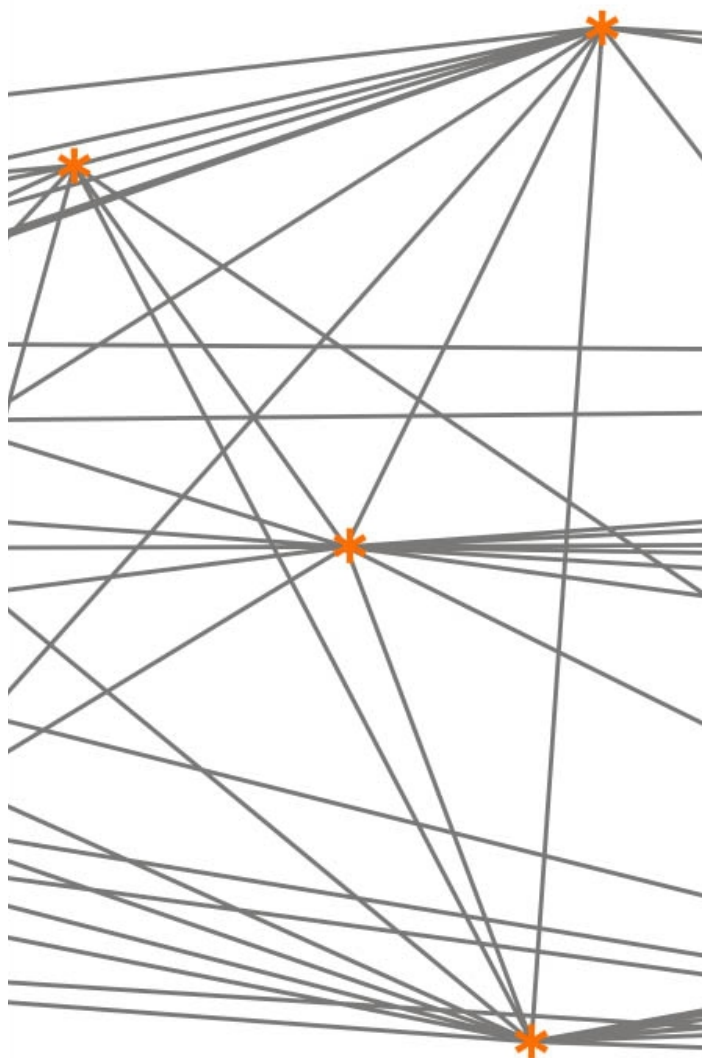




Release Guide 1.1.1

Wave 4



March 28, 2013

Table of Contents

1	Additional Sources of Information	3
1.1	SHARE Website	3
1.2	Questionnaires	3
2	SHARE data releases	3
3	What is New in SHARE Wave 4	4
3.1	What's updated in SHARE Wave 4 release 1.1.1	4
4	Countries in SHARE	4
5	Eligibility rules in SHARE wave 4	5
5.1	Refreshment Samples/New Countries	5
5.2	Longitudinal Sample	6
6	Composition of the Data Set and Types of Respondents in SHARE Wave 4 ...	6
6.1	Types of Data in Wave 4	7
6.2	Types of Respondents in Wave 4	7
7	Merging the Data	9
8	Self-Completion Questionnaires ("Drop-Off")	9
9	Citizenship and Country of Birth Coding	9
10	CH module: Looping over children	10
11	CF module: Ten Words List Learning in Wave 4	10
12	SP and FT Module: List of Relations	11
13	EX: Definition of life expectancy target age (ex009_)	14
14	Generated Variables	15
14.1	Social Networks	15
14.2	Imputations in SHARE Wave 4	16
14.3	Weights	27
14.3.1	Sampling design weights	27
14.3.2	Calibrated weights	28
14.4	NUTS	31
14.5	Biomarker	32

1 Additional Sources of Information

1.1 SHARE Website

This guide primarily gives an overview of changes and new measures in wave 4, compared to SHARE waves 1 and 2. Information on the naming of variables, the missing codes, and much other helpful material on the SHARE database can be found in the “SHARE Release Guide 2.5.0”:

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

In addition we will publish a volume focussing on the innovations and methodology of SHARE wave 4 soon:

Malter, Frederic, and Axel Börsch-Supan (2013). *SHARE Wave 4: Innovations & Methodology*. Munich: MEA, Max-Planck-Institute for Social Law and Social Policy.

More general information on SHARE data can be found on the website:

<http://www.share-project.org/data-access-documentation.html>

1.2 Questionnaires

Generic and country-specific questionnaires are downloadable from the SHARE-website:

<http://www.share-project.org/data-access-documentation/questionnaires/questionnaire-wave-4.html>

2 SHARE data releases

Wave 1 & Wave 2	
Release 1: April 28 th , 2005	Release 1.0.0: November 28 th , 2008
Release 2.0.0: June 19 th , 2007	Release 1.0.1: December 4 th , 2008
Release 2.0.1: July 5 th , 2007	
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	
SHARELIFE Wave 3	Wave 4
Release 1.0.0: November 24 th , 2010	Release 1.0.0: November 30 th , 2012 Release 1.1.1: March 28 th , 2013

3 What is New in SHARE Wave 4

- Four new countries: Estonia, Hungary, Portugal, and Slovenia
- New integrated social networks module with respective generated variables
- Biomarker variables (for Germany only)

3.1 What's updated in SHARE Wave 4 release 1.1.1

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

4 Countries in SHARE

In addition to almost all countries that participated in previous waves of SHARE four new countries joined in wave 4: Estonia, Hungary, Portugal, and Slovenia. The next table shows the list of countries, country identifiers, participation in waves, and when the data collection was conducted. The definition of the wave results from the questionnaire version used.

Countries and Language Versions in SHARE Waves 1-4

ID	Country (Short)	Country and Language	Wave 1	Wave 2	Wave 3 SHARELIFE	Wave 4
11	AT	Austria	2004	2006/07	2008/09	2011
12	DE	Germany	2004	2006/07	2008/09	2011/12
13	SE	Sweden	2004	2006/07	2008/09	2011
14	NL	Netherlands	2004	2007	2008/09	2011
15	ES	Spain	2004	2006/07	2008/09	2011
16	IT	Italy	2004	2006/07	2008/09	2011
17	FR	France	2004/05	2006/07	2009	2011
18	DK	Denmark	2004	2006/07	2008/09	2011
19	GR	Greece	2004/05	2007	2008/09	-
20	Cg	Switzerland (German)	2004	2006/07	2008/09	2011
21	Cf	Switzerland (French)	2004	2006/07	2008/09	2011
22	Ci	Switzerland (Italian)	2004	2006/07	2008/09	2011
23	Bf	Belgium (French)	2004/05	2006/07	2008/09	2011
24	Bn	Belgium (Flemish)	2004/05	2006/07	2008/09	2011
25	Ih	Israel (Hebrew)	2005/06	-	-	-
26	Ia	Israel (Arabic)	2005/06	-	-	-
27	Ir	Israel (Russian)	2005/06	-	-	-
28	CZ	Czech Republic	-	2006/07	2008/09	2011
29	PL	Poland	-	2006/07	2008/09	2011/12
30	IE	Ireland	-	2007	-	-
32	HU	Hungary	-	-	-	2011
33	PT	Portugal	-	-	-	2011
34	SI	Slovenia	-	-	-	2011
35	EE	Estonia	-	-	-	2010/11

5 Eligibility rules in SHARE wave 4

5.1 Refreshment Samples/New Countries

The target population for the baseline samples consists of *all persons born 1960 or earlier* having their regular domicile in the respective country, together with their current partners/spouses, independent of age. As in wave 2 only one age-eligible member *plus* his/her partner/spouse has been interviewed within a household.

5.2 Longitudinal Sample

The target population for the longitudinal survey consists of all original 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). If respondents deceased since their last interview, the interviewers try to find a close relative or other proxy informant to conduct an end-of-life interview. 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 4

Please see "SHARE Release Guide 2.5.0" Chapter 6 for general information on this topic.

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

Using wave 4 data, users should be aware that some steady information (e.g. education or height) is only available in the first interview of a respondent, the so called baseline interview; other information is only asked again in follow-up interviews and updated if it has changed (e.g. marital status). Generated variables which use this kind of information show a high percentage of missing values because of this. E.g. the body mass index (bmi) is generated using height and weight. As height contains a missing value for the follow-up respondents of wave 4, their bmi is also missing.

This can be solved by using the information from the baseline interview in wave 1 or wave 2. 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 it after the merging.

6.1 Types of Data in Wave 4

Composition of the SHARE Wave 4 Data Set

CAPI Data	
Coverscreen interview cv_r	Data on the individual level for all household members, including non-eligible persons. It also includes information on deceased household members from previous waves.
Individual CAPI modules	See also chapter 6.2
End of life interviews	
Paper and pencil questionnaires	
Drop-off	Generic and country specific questions
Generated variables	
ISCED codes for education	
Physical and mental health, cognitive function	Includes generated variables on ten word list learning
Biomarker	Additional biomarker data available for Germany only
Social networks	
Housing and region	
Imputations	
Weights	

6.2 Types of Respondents in Wave 4

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 partners/other household members.

Note that in wave 4 financial and household respondents are factual the same respondents and can thus be used interchangeably.

Types of respondents and variable names

Type of respondent	Variable name in the cv_r module	Name of Filter in Questionnaire
Financial respondent	dumfinr	xx_NumFinR
Household respondent	dumhhr	xx_NumHHR
Family respondent	dumfamr	xx_NumFamR

Who Answers What in the CAPI Questionnaire?

CAPI Module	Name	All respondents	Financial/HH respondent	Family Respondent	non-proxy
CV_R	Coverscreen				
DN	Demographics	x			
SN	Social Networks (new in wave 4)	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			
GS	Grip Strength	x			x
PF	Peak Flow	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				
XT	End-of-Life Interview	proxy respondents			

7 Merging the Data

No changes with respect to previous waves 😊.

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

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

The paper and pencil questionnaire (drop-off) includes generic questions on health and health care. These generic variables have variable names starting with “q”. In wave 4 the drop-off questionnaires also include some country specific questions. Country specific variables are named cc_q*.

9 Citizenship and Country of Birth Coding

Citizenship (*dn008*) and country of birth (*dn005*) are coded according to ISO 3166-1 (numeric-3). The list is available on:

<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 additional countries/regions and for respondents with multiple citizenships.

Additional codes for country of birth and citizenship:

- 1010 - Congo (both)
- 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

- 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
- 1100 - Chechnya

10 CH module: Looping over children

The CH module which is only part of the CAPI interview for family respondents contains information about the children. 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 indicating such a change. The first variable indicates if a change occurred (yes/no). If there was a change reported, the second set of variables contains 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 respective child is specified.

Children changes

	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– ch515d21	ch516_1– ch516_20
parenthood	ch517	ch518d1– ch518d20	ch519_1– ch519_20 ch520_1– ch520_20
location	ch524_	ch525d1– ch525d20	ch526_1– ch526_20

11 CF module: Ten Words List Learning in Wave 4

As in previous waves the “ten words list learning” test has been conducted with a first trial and a delayed recall. Though, in 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 4 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* and *cf016tot* to be in line with the respective variables in wave 2. Both variables contain the totals over all four sets of “ten words list learning” in wave 4 for the first trail (*cf008tot*) and the delayed recall (*cf016tot*). Note that both variables can be found in the cognitive function (*cf*) module in wave 2, but are part of the generated health module (*gv_health*) in wave 4. This is due to the fact that these variables have been generated in wave 4 and were not regular CAPI items.

12 SP and FT Module: List of Relations

In SHARE wave 4 the new social network (*sn*) module was linked to the social support (*sp*) and financial transfer (*ft*) modules, so respondents could indicate for example which social network member provided help. As in wave 2 variables in the social support (*sp*) and the financial transfer (*ft*) module refer to a list of relations. 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 previous waves. To maintain comparability between waves the respective categories have been recoded to match the wave 2 coding. Categories referring to social network persons got new codes assigned. In addition four new response options were implemented in the wave 4 list of relations, i.e. categories 34-37 whereas others are not included anymore.

These changes in the wave 4 list of relations affect the following questions (see also wave 4 questionnaire for question wording):

- sp003_ Who gave you help
- sp009_ To whom did you give help
- sp015_ Parents from grandchildren
- sp019_ To whom given 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

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. a 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

The following table 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 values	Wave 4 values	Wave 1 & 2 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

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

Question ex009_ asks baseline respondents on “What are the chances that you will live to be age <fill> or more?” (for longitudinal respondents the question is not asked). 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). Year and month of interview are defined by filling in the coverscreen/sms part of the interview.

Based on the computed age, the <fill> for ex009_ is defined as shown here:

Respondents' age	Fill in ex009_
age < 65	75
age > 65 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 dataset as *ex009age*, while the substantive answers to the question is stored as *ex009_*. You will find that in the vast majority of cases the fill exactly matches the above procedure.

However, there are two possible scenarios causing the fill to deviate from the rule. While the CAPI software is programmed to adjust age if respondents change e.g. their year of birth in the DN section of the interview (originally year of birth is preloaded from the last interview or collected from the coverscreen respondent), the CAPI will not re-ask ex009_ in case the interviewer goes back to the DN section and corrects year of birth after ex009_ was answered. The second scenario is similar. If the country teams or central database management correct demographic information based on interviewer remarks, register information or plausibility checks after the interview is completed, the fill used in ex009_ might also deviate from the above rule.

14 Generated Variables

14.1 Social Networks

The social network module (SN) was implemented in the 4th wave of SHARE as an innovative means by which 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 new SN model employs a name generator that first identifies the respondent's self-reported meaningful relationships and then obtains the characteristics of the persons named. The information obtained in the SN module is a detailed description of study participants' personal social networks, that is, the persons who they consider to be their confidants.

The module begins with an initial probe; "Over the last 12 months, who are the people with whom you most often discussed important things?" Survey participants were permitted to list up to six names, and one additional name of a person important to them for any reason (i.e. maximum total of seven). The module records the role relationship of each social network member, and obtains information regarding each named person's gender, residential proximity to the participant, frequency of contact and level of emotional closeness of the relationship as perceived by the study participant.

Another innovation in SHARE wave 4 is the linking of information gathered in the SN module with two subsequent survey modules: social support (SP) and financial transfers (FT). The social network member names and role relationships, as provided by the survey respondent in the SN module, appear on the CAPI screen as specific persons with whom personal support or financial resources were exchanged, in addition to the standard role categories of others involved in exchange. This linkage distinguishes between the exchange of money and support within the personal social network of survey respondents and with people not named as confidants.

Central Variables in the New Social Network CAPI Module

Variables	Description
<i>sn005_X</i>	Relationship to network person X
<i>sn005a_X</i>	Gender of network person X
<i>sn006_X</i>	Geographical distance to network person X
<i>sn007_X</i>	Contact frequency with network person X
<i>sn009_X</i>	Emotional closeness to network person X
<i>sn012_ (sn017_)</i>	Satisfaction with (empty) network

The generated variables module "gv_networks" stores a total of 96 generated variables. These generated variables combine information from the CAPI SN module and information from the CH, SP, FT and DN module.

Generated Variables from the Social Network Module

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

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

14.2 Imputations in SHARE Wave 4

Authors: Giuseppe De Luca, Elisabetta Trevisan, Chiara Dal Bianco, Martina Celidoni & Andrea Bonfatti

The new SHARE imputation procedure presents some common aspects with the previous imputation strategy implemented for Release 2.4 of waves 1 and 2 data (publicly available since March 2011, see [Christelis, 2011](#)), as well as some important innovations. 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 procedure only a subset of items are imputed using the fully conditional specification (FCS) method (van Buuren et al., 2006). Compared to wave 1 and 2, variables are no

longer imputed for non-responding partner; some items are imputed in aggregate terms, such as pension incomes and assets; variables, such as number of rooms in the main residence, dividends, self-rated reading skills, being depressed and risk preferences are no longer imputed; finally, additional variables used in the imputations process are provided. A set of tables provide a detailed description of the structure of the imputation data set. More precisely, the first table describes all the variables included in the data set and reports for each variable the correspondence with the variables included in the imputation datasets from previous waves and the corresponding variables in the questionnaire; the second table gives the details of the values taken by the flag variables; the third table describes in detail how the aggregated variables have been constructed and how they can be related to the variables from wave 1 and wave 2.

The decision to reduce the number of imputed items and to create some aggregate variables has been taken in order to reduce the multi-collinearity problems that may arise during the estimation stage of our imputation procedure, as the inclusion of a large set of covariates in the model could lead to a lower precision of the estimated coefficients. The strategy adopted in the new imputation process takes into account this trade-off and tries to summarize reasonably, where possible, the available information either by selecting specific variables (e.g. behavioural risks section: smoking and drinking behaviour variables) or aggregating various items, thus reducing the number of variables that are jointly imputed into the system. This reduction is easier to apply in wave 4 than in previous waves given the smaller set of variables collected, i.e. health care out of pocket expenditures and financial transfer amounts are no longer asked to respondents.

We decided to aggregate the long list of income, wealth and expenditure items collected in the SHARE questionnaire into a considerably smaller subset of key variables. Although this is a reasonable decision, it is not exempt from criticisms and problems. When aggregating items in fact, it could be that just a part of them is missing. In that case, even if the aggregated variable is considered as missing, the available information contained in the single items are preserved and used as individual lower bounds.

The imputation process is conducted separately by country and sample type. The samples considered are singles and 3rd respondents¹ (sample 1), couples with both partners interviewed (sample 2) and all couples - with and without non responding partners (NRPs) (sample 3). As mentioned before, unlike in previous imputation databases, non-responding partners (NRP) are not included in the imputation sample. Nevertheless, imputations of total household income (thinc)

¹ 3rd respondents are normally singles living with a couple, e.g. parents or relatives. These are typically individuals entered in the sample at the time of wave 1, when all members of the household over 50 years of age were interviewed.

takes into account the income components of NRP. Households with NRPs and the responding partner of a NRP can be identified by the variables `htype` and `p_nrp`, respectively. No separate subsamples for panel and refreshment are taken into account.

The variables are either imputed by simple hotdeck or jointly imputed in the system. The criteria determining which imputation method has to be used for each item is based on the prevalence of missing values. For each item the percentage of missing values (or the prevalence of missing values) for the total sample and at country level is computed. If this percentage is:

- lower than 5% for the entire sample and
- lower than 10% at the country level

then, the variable is imputed using simple hotdeck procedure. If the prevalence of missing values is higher than these thresholds the variable is included in the system and imputed using the fully conditional specification (FCS) method mentioned above. The definition of these thresholds implies that the list of variables to be jointly imputed is country and possibly subsample specific.

Monetary values, when missing, are imputed conditional on ownership² (which is imputed if missing) and according to individual upper and lower bounds (respectively UB and LB). 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.

Particular attention is devoted to the imputation of household income in the three subsamples. We distinguish between couples with and without NRPs to avoid underestimating total household income when only one of the two partners is interviewed. Since we want to use the additional information on monthly household income contained in `HH017_`, we will have two different definitions of total household income, version A (`thinc`) and B (`thinc2`). Version A is obtained by a suitable aggregation of all individual income components at the household level (this version is comparable with the ones provided in previous imputations data sets). Version B is obtained from the one shot question on total household income (`HH017_`). We let users decide which version of total household income is more suitable for their research.

The total household income imputation procedure has four-stages:

² Ownership is a dummy variable that identifies who has to respond to the specific question and depends on being financial/household/family respondent or on questionnaire routings. For income and asset questions the ownership is determined by specific questions. For instance, if respondent says 'yes' to question `EP204_` (Have you had any wages, salaries or other earnings from dependent employment in [previous year]?), then the related ownership for `EP205_` (After any taxes and contributions, what was your approximate annual income from employment in the year [previous year]?) will be 1.

- Stage 1 (singles and 3rd respondent): We impute all income components and thinc using a strategy similar to the one adopted in previous waves.
- Stage 2 (couples with both partners interviewed): We use partner's characteristics to impute incomes of both couple members. Here, we impute all income components of the two partners interviewed and thinc using a strategy similar to the one used in the previous waves.
- Stage 3 (all couples - with and without NRPs): Unlike stage 2, we use a smaller set of partner's characteristics. For couples with both partners interviewed, we impute thinc2 only. For couples with NRPs, we impute incomes of the responding partner and thinc2. In both group, thinc2 is imputed using unfolding brackets information from HH017.
- Stage 4 (all couples - with and without NRPs): For couples with NRPs, we impute thinc using thinc2 as a predictor and the sum of incomes of the responding partner as lower bound.

The status of each imputed variable in the 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 and should be aware of the following:

Unlike in previous imputations datasets 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.

Unlike in previous imputations datasets all monetary variables are expressed in annual terms, independently from the original periodicity indicated in the questionnaire. See the first and the third table for precise correspondence between variables.

References:

Christelis D. (2011). Imputation of missing data in waves 1 and 2 of SHARE. SHARE WP Series, N. 01-2011.

Meng X. (1994). Multiple-Imputation inferences with uncongenial source of input. Statistical Science, 9(4): 538-573.

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: Correspondence between imputation variables in wave 4, imputations variables in wave 2 and questionnaire items

Variable name	Description	Variable name Wave 2	Questionnaire
mergeid	Person ID	mergeid	
implicat	Implicat number	implicat	
hhid4	Household ID wave 4		
cvid	Wave specific person identifier	cvid	
cvidp	Wave specific person identifier of spouse/partner		
country	Country identifier	country	
language	Language of questionnaire	language	
htype	Household type	(only in wave 4)	
fam_resp	Family respondent	(only in wave 4)	
fin_resp	Financial respondent	(only in wave 4)	
hou_resp	Household respondent	(only in wave 4)	
excrate	Exchange rate	nomxWXYZ	
nursinghome	Living in nursing home	(only in wave 4)	
hysize	Household size	(only wave 4)	
single	Single	(only wave 4)	
couple	Couple	(only wave 4)	
partner	Partner in the couple	(only wave 4)	
p_nrp	Partner of non responding partner	nrpartn	
sample1	Imputation sample for single	(only wave 4)	
sample2	Imputation sample for couples with two respondents	(only wave 4)	
sample3	Imputation sample for all couples (both with two respondents and with non responding partner)	(only wave 4)	
ydip	Annual earnings from employment	ydipv	ep205_
yind	Annual earnings from self-employment	yindv	ep207_
yopen1	Annual old age & early retirement pensions	(see table 3)	(see table 3)
yopen2	Annual disability, unemployment, survivor and war pensions	(see table 3)	(see table 3)
yreg	Other regular payments	(see table 3)	(see table 3)

Variable name	Description	Variable name Wave 2	Questionnaire
ylsum	Lump sum payments	(see table 3)	(see table 3)
rhre	Annual rent and home-related expenditures	(see table 3)	(see table 3)
home	Value of main residence	homev	ho024_
mort	Mortgage on main residence	mortv	ho015_
ores	Value of other real estate – Amount	oresv	ho027_
ysrent	Annual income from rent or sublet	(see table 3)	(see table 3)
yaohm	Annual income from other household members	(see table 3)	(see table 3)
fahc	Annual food at home consumption	fahcv*12	co002_
fohc	Annual food outside home consumption	fohcv*12	co003_
hprc	Annual home produced consumption	hprcv*12	co011_
bacc	Bank accounts	hbaccv	as003_
bsmf	Bond, stock and mutual funds	(see table 3)	(see table 3)
slti	Savings for long-term investments	(see table 3)	(see table 3)
vbus	Value of own business	hownbv	as042_
sbus	Share of own business	sbusv	as044_
car	Value of cars	hcarv	as051_
liab	Financial liabilities	hliabv	as055_
yibacc	Interest income from bank accounts: computed multiplying a fixed interest rate (1%) by the total amount in bank account	hybaccv	
yibsmf	Interest income from bond, stock and mutual funds: computed multiplying a fixed interest rate (5%) by the total amount in bond, stock and mutual funds amount	hybondv, hystocv, hymutfv	
thinc	Total household net income - version A	hgtincv	
thinc2	Total household net income - version B	(only in wave 4)	hh017_
thexp	Total household expenditure (sum of rhre, fahc, fohc and hprc)	(only in wave 4)	
hrass	Household real assets (home*perho/100+vbus*sbus/100+car+ores-mor)	hrav	
hgfass	Household gross financial assets (sum of back, bsmf and slti)	hgfinv	
hnfass	Household net financial assets (hgfass-liab)	hnfinv	

Variable name	Description	Variable name Wave 2	Questionnaire
hnetw	Household net worth (sum of hrass and hnfass)	hnetwv	
gender	Gender	(only wave 4)	dn042_
age	Age in 2010	(only wave 4)	dn003_
age_p	Age of partner in 2010	(only wave 4)	dn003_
yeduc	Year of education	(only wave 4)	dn041_
yeduc_p	Year of education of partner	(only wave 4)	ex102_
sphus	Self-perceived health - US scale	srhealtha	ph003_
mstat	Marital status	(only wave 4)	dn014_
child	Number of children	nchild	ch001_
gchild	Number of grandchildren	n_gchild	ch201_
gali	Limitation with activities	gali	ph005_
chronic	Number of chronic diseases	(only wave 4)	ph006_
symptoms	Number of symptoms	(only wave 4)	ph010_
mobility	Mobility limitations	(only wave 4)	ph048_
adl	Limitations with activities of daily living	adlno	ph049_1
iadl	Limitations with instrumental activities of daily living	iadlno	ph049_2
esmoked	Ever smoked daily	(only wave 4)	br001_
drinking	More than 2 glasses of alcohol almost everyday	(only wave 4)	br019_
phactiv	Physical inactivity	(only wave 4)	br015_
meals	Number of meals every day	(only wave 4)	br025_
orienti	Score of orientation in time test	(only wave 4)	cf003_ - cf006_
memory	Score of memory test (only refresher)	(only wave 4)	cf103_
wllft	Score of words list learning test - trial 1	(only wave 4)	cf104_* - cf107_*
wllst	Score of words list learning test - trial 2	(only wave 4)	cf113_* - cf116_*
fluency	Score of verbal fluency test	(only wave 4)	cf010_
numeracy1	Score of first numeracy test (only refresher)	numeracy	cf012_ - cf015_
numeracy2	Score of second numeracy test	(only wave 4)	cf108_ - cf112_

Variable name	Description	Variable name Wave 2	Questionnaire
eurod	EURO depression scale	(only wave 4)	mh002_ - mh017_
doctor	Seen/Talked to medical doctor	(only wave 4)	hc002_
hospital	In hospital last 12 months	(only wave 4)	hc012_
thospital	Times being patient in hospital	(only wave 4)	hc013_
nhospital	Total nights stayed in hospital	(only wave 4)	hc014_
sn_num	Number of people within social network	(only wave 4)	sn013_
sn_sat	Satisfaction with social network	(only wave 4)	sn012_
cjs	Current job situation	(only wave 4)	ep005_
pwork	Did any paid work	(only wave 4)	ep002_
empstat	Employee or self-employed	(only wave 4)	ep009_
lookjob	Looking for job	(only wave 4)	ep337_
rhfo	Received help from others (how many)	(only wave 4)	sp002_, sp005_, sp007_
ghfo	Given help to others (how many)	(only wave 4)	sp008_, sp011_, sp013_
ghih	Given help in the household (how many)	(only wave 4)	sp018_
rhih	Received help in the household (how many)	(only wave 4)	sp020_
gfg	Number of given financial gifts 250 or more	(only wave 4)	ft002_, ft007_*
rfg	Number of received financial gifts 250 or more	(only wave 4)	ft009_, ft014_*
otr	Owner, tenant or rent free	(only wave 4)	ho002_
perho	Percentage of house owned	(only wave 4)	ho070_
fdistress	Household able to make ends meet	fdistress	co007_
lifesat	Life satisfaction	(only wave 4)	ac012_
lifehap	Life happiness	(only wave 4)	ac022_
naly	Number of activities last year	(only wave 4)	ac035_i
saly	Satisfied with no activities	(only wave 4)	ac038_
willans	Willingness to answer	(only wave 4)	iv004_
clarif	Respondent asked for clarifications	(only wave 4)	iv007_
undersq	Respondent understood questions	(only wave 4)	iv008_

Variable name	Description	Variable name Wave 2	Questionnaire
hnrsc	Help needed to read showcards	(only wave 4)	iv018_
nomxyear	Nominal exchange rate	nomxyear	
pppxyear	PPP adjusted exchange rates	pppxyear	
currency	Currency in which amounts are denominated	currency	
varname_f	Flag variables (see description in table 2)	varname_f	

Table 2: 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)

Table 3: Detailed description of aggregate variables

Aggregate variable	Definition	Varname Wave 2	Questionnaire
Old age and early retirement pensions (ypen1)	Public old age pension	annpen1v	ep078_1
	Public old age supplementary pension	annpen12v	ep078_2
	Private old age pension	annpen8v	ep078_11
	Private old age pension from a second job	annpen15v	ep078_12
	Private old age pension from a third job	annpen16v	ep078_13
	Public early retirement pension	annpen2v	ep078_3
	Private early retirement pension	annpen9v	ep078_14
Disability, unemployment, survivor and war pensions (ypen2)	Public disability insurance pension or sickness benefits	annpen3v	ep078_4
	Secondary public disability insurance pension or sickness benefits	annpen13v	ep078_5
	Private disability insurance pension	annpen10v	ep078_15
	Public long-term insurance payments	annpultv	ep078_10
	Unemployment benefit or insurance	annpen4v	ep078_6
	Public survivor pension from partner	annpen5v	ep078_7
	Secondary public survivor pension from partner	annpen14v	ep078_8
	Private survivor pension from partner's job	annpen11v	ep078_16
War pension	annpen7v	ep078_9	
Other regular payments (yreg)	Regular life insurance payments	annreg1v	ep094_1
	Regular payments from private annuity or personal pension	annreg2v	ep094_2
	Regular payments from alimony	annreg4v	ep094_3
	Regular payments from charities	annreg5v	ep094_4
	Regular payments from long-term care insurance	annprltv	ep094_5
Lump sum payments (ylsum)	It includes lump sum payments related to the previously defined variables.	(only wave 4)	ep082_1-16 ep209_1-5
Rent and home-related expenditures (rhre)	Amount rent paid	rentcv*12	ho005_
	Other home-related expenditures	ocscv*12	ho008_

Aggregate variable	Definition	Varname Wave 2	Questionnaire
Income from rent or sublet (ysrent)	Income from sublet	yrentv	ho074_
	Income from rent of real estate		ho030_
Income from other household members (yaohm)	Other household members' net income	yohmv	hh002_
	Other household members' net income from other sources	yohbv	hh011_
Bond, stock and mutual funds (bsmf)	Government/corporate bonds	hbondv	as007_
	Stocks	hstocv	as011_
	Mutual funds	hmutfv	as017_
Savings for long-term investments (slti)	Individual retirement accounts from respondent	irav	as021_
	Individual retirement accounts from partner	irav	as024_
	Contractual saving	Hcontv	
	Whole life insurance holdings	Hlinsv	as030_

14.3 Weights

Authors: Giuseppe De Luca & Claudio Rossetti

14.3.1 Sampling design weights

Sampling design weights are defined as the inverse of the probability of being included in the sample of any specific wave. These weights 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.

Sampling design weights are computed separately by country to account for the peculiar features of each national sampling design. A detailed documentation of the national sampling designs used in the fourth wave can be found in Lynn et al (2013).

For most countries, the sample consists of two parts: a baseline sample drawn in the first two waves and a refreshment sample drawn in the fourth wave. Among the countries which participated in the fourth wave, those with no refreshment sample are Germany, Poland and Sweden. Four new countries (Estonia, Hungary, Portugal and Slovenia) entered SHARE for the first time in the fourth wave. These countries had to construct their baseline samples that will ultimately form their "first wave" panel cases.

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. The main difference with respect to the previous waves is that SHARE does not collect any more vignette data in wave 4. Therefore, there is no reason to distinguish between sampling design weights for the main sample alone, the vignette sample alone and the two samples combined. The public release of the wave 4 data include just one sampling design weight.

14.3.2 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 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 4 SHARE data and the most important differences with respect to the calibrated weights of the previous waves.

- As in the previous waves, the wave 4 data include calibrated cross-sectional weights to be used in the context of cross-sectional analyses and calibrated longitudinal weights to be used for longitudinal analyses. A description of these calibrated weights is given below in Sections 2.1 and 2.2 respectively.
- Since the basic units of analysis can be either individuals or households, calibrated cross-sectional and longitudinal 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.
- As for design weights, the wave 4 data do not include any more calibrated weights for alternative variants of its sample (i.e. main sample alone, vignette sample alone and the two samples combined).
- Because of changes in the imputation procedure of wave 4 (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 4 data, SHARE does not provide any more calibrated longitudinal weights for all possible wave combinations. Guidelines for longitudinal analyses based on wave combinations that are not available in the public release SHARE data are given in Section 2.2.

- A list of the calibrated weights available in the public release of the wave 4 data is presented in Table 1. The structure of variable names for design and calibrated weights are different from those used in the previous 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_w#`, calibrated household weights are named as `chw_w#`, and calibrated individual weights are named as `ciw_w#`.

14.3.2.1 Calibrated cross-sectional weights

Calibrated cross-sectional weights are defined for the sample of 50+ respondents (either individuals or households) in wave four by ignoring the distinction between longitudinal and refreshment samples. At the individual level (i.e. variable `chi_w4`), 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_w4`), 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 1960 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 `(-1930]`, `[1931-40]`, `[1941-50]`, `[1951-60]`) 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).

14.3.2.2 Calibrated longitudinal weights

Calibrated longitudinal weights differ from calibrated cross-sectional weights in two important respects.

First, these weights are only defined for the balanced subsample of eligible units who participated in two or more waves of the panel. Second, since mortality is a source of attrition which affects both the sample and the population, calibrated longitudinal weights take into account mortality of the original target population across waves. Hence, the target population for longitudinal analyses is the original population at the beginning of the time reference period that survives up to the end of period.

Since the SHARE panel now consists of four waves, one can currently compute 30 different types of calibrated longitudinal weights depending on the selected wave combination (i.e., 1-2, 1-3,...,3-4, 1-2-3,...,2-3-4, 1-2-3-4) and the basic unit of analysis (either individuals or households). Furthermore, the number of possible calibrated longitudinal weights will increase rapidly as the panel goes ahead and additional waves will be available. These considerations clarify why the strategy of providing calibrated longitudinal weights for all possible wave combinations of the panel is no more feasible, especially in the long run. Since wave 4, the new strategy adopted in SHARE is to provide calibrated longitudinal weights only for the fully balanced sample (i.e. the sample of 50+ respondents participating to all waves).³ For longitudinal analyses based on other wave combinations, users are required to control for sample attrition either by computing their own calibrated weights or by implementing some other method. To support users in this methodological task, we plan to provide a set of Stata routines which implement and illustrate the weighting calibration procedure of Deville and Särndal (1992). This user-support material will be available soon through the SHARE website.

Calibrated longitudinal weights for the fully balanced sample are computed separately by country to match the size of the national populations of individuals born in 1954 or earlier that survive up to 2011. Due to the limited number of observations in the fully balanced sample, we used only one set of calibration margins for the size of the target population across eight gender-age groups (i.e. males and females with year of birth in the classes (-1924], [1925-34], [1935-44], [1945-54]). Mortality is accounted for by subtracting from each population margin the estimated number of deaths between 2004 and 2010. These weights are available at the individual (i.e. variable `ciw_w1234`) and the household level (i.e. variable `chw_w1234`). Notice that, for the weights at the household level, we only require that there is at least one eligible respondent in each wave. Thus, households with one partner participating in the first wave and the other partner participating in the other waves belong to the balanced sample of households, even if neither partner belongs to the balanced panel of individuals. The balanced panels of households and individuals can be identified through the binary indicators `balanced_hh` and `balanced_ind`, respectively. The dataset also includes flag variables for missing calibrated longitudinal weights. These weights are missing for respondents younger than 50 years, those with missing information on the set of calibration variables, those with missing sampling design weights, and those not belonging to the selected balanced sample.

³ Calibrated longitudinal weights for the wave combination 1-2 can always be found in the data file `sharew2_rel#_gv_weights`, those for the wave combinations 1-3, 2-3, and 1-2-3 in the data file `sharew3_rel#_gv_weights`.

Sampling design and calibrated weights in the fourth wave of SHARE

Variable	Description	Unit of analysis
dw_w4	Design weights wave 4	Household & individual
chw_w4	Calibrated cross-sectional weight wave 4	Household
ciw_w4	Calibrated cross-sectional weight wave 4	Individual
chw_w1234	Calibrated longitudinal weight for balanced panel	Household
ciw_w1234	Calibrated longitudinal weight for balanced panel	Individual

References

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

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

Lynn P., G. De Luca, M. Ganninger, S. Häder (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.

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

14.4 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 4 these variables are named *nuts1_2010*, *nuts2_2010* and *nuts3_2010*, whereas the 2010 suffix is referring to the NUTS classification of 2010. Depending on the country-specific privacy legislations, not all NUTS level are provided for every country.

NUTS 2010 are only available for the refreshment samples and new countries. For panel cases, please use the NUTS variables stored in the datasets of previous waves. Be aware that the NUTS codes of previous waves might be different, due to a different NUTS classification being at use at the time of sampling.

14.5 Biomarker

From the first wave on, SHARE combined self-reports on health with two physical performance measurements, namely grip strength and walking speed. Additionally, respondents reported their height and weight. In wave 2, SHARE added peak-flow and chair stand to the questionnaire programme. In wave 3 (SHARELIFE) grip strength was the only physical measurement included in the questionnaire. In wave 4, the physical measurements included grip strength, peak-flow, self-reported height and weight as well as a completely *new set of biomarkers that were measured in the German part of the study*.

The new measures include:

- *Measured Height*: (in addition to the self-reported height of the respondents): Allows validating the self-reported height and thus enhances the accuracy of the computed body mass index (BMI).
- *Measured Waist Circumference*: Allows the computation of the height-to-waist-ratio (HTWR), which is an indicator for the distribution of body fat.
- *Blood Pressure*: Allows identifying respondents with high blood pressure. BMI, HTWR, and blood pressure serve as indicators for the risk of developing cardiovascular diseases.
- *Dried Blood Spots (DBS)*: If written consent has been given by the respondent, the DBS are collected on a filter card by pricking a respondent's finger with a lancet. The DBS include information on total cholesterol, C-reactive protein (CRP, which is a marker for inflammations in the body), and HbA1c (which is a measure for blood sugar levels over the last 120 days). Both CRP and total cholesterol are associated with the development of cardio-vascular diseases, whereas HbA1c allows us to identify respondents who have diabetes.

As said above these additional biomarker measures were only collected in Germany. Hence, the generated dataset *gv_biomarker* is only available for the German subsample of SHARE wave 4.

The Central Variables in the New Biomarker Module

Variables	Description
<i>b1_*</i>	Written consent for the single measurements
<i>b2_*</i>	Information on height measurement
<i>b3_*</i>	Information on waist measurement
<i>b4_*</i>	Information on blood pressure measurements
<i>hba1c</i>	Blood sugar
<i>crpmgl</i>	C-reactive protein [mg/l]
<i>cholmgdl</i>	Cholesterol [mg/dl]

Data sources

The biomarker part of the interview was not programmed as CAPI. All questions from the biomarker section were included in a paper & pencil booklet that was filled out by the interviewers. All results from the measurements of height, waist circumference, and blood pressure were noted down in this booklet. After the interview, the interviewers sent the booklet to the survey agency.

Furthermore, an external laboratory processed the filter cards with the DBS and then sent us the results for HbA1c, cholesterol and CRP. These lab results were linked to the booklet data.

Finally, the dataset consisting of booklet data and lab results was linked to the CAPI data of the SHARE interview.

Responsible for database management and preparation of the releases:

Stephanie Stuck, Sabrina Zuber,
Morten Schuth, Markus Kotte,
Christian Hunkler, Thorsten Kneip,
Fabio Franzese, Stefan Gruber
and all SHARE country team operators.

Contact: info@share-project.org