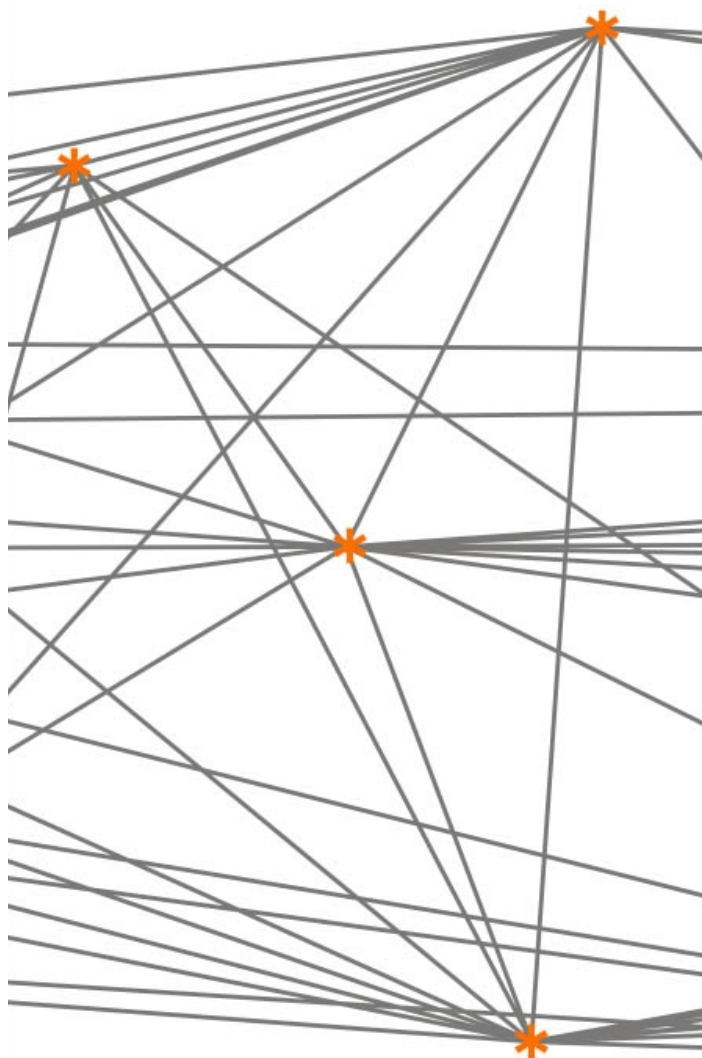




Release Guide 1.0.0

Wave 4



November, 30, 2012

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 Release 1.0.0	3
4	Countries	4
5	Eligibility rules in SHARE wave 4	5
5.1	Refreshment Samples/New Countries:	5
5.2	Longitudinal Sample:	5
6	Composition of the Data Set and Types of Respondents in SHARE Wave 4 ...	6
6.1	Types of Data in Wave 4	6
6.2	Types of Respondents in Wave 4	7
7	Merging the Data	7
8	Self-Completion Questionnaires ("Drop-Off")	8
9	Citizenship and Country of Birth Coding	8
10	CF module: Ten Words List Learning in Wave 4	9
11	SP and FT Module: List of Relations	10
12	EX: Definition of life expectancy target age (ex009_)	12
13	Generated Variables	13
13.1	Social Networks	13
13.2	Imputations in SHARE Wave 4	14
13.2.1	Key features of the SHARE wave 4 imputation model	15
13.2.2	List of variables included in the imputation database	18
13.2.3	Description of the Aggregation Step	21
13.2.4	Value of Flag Variables in the Imputations Module	22
13.3	NUTS	23
13.4	Biomarker	23

1 Additional Sources of Information

1.1 SHARE Website

More general information on SHARE data can be found on the website:
<http://www.share-project.org/data-access-documentation.html>

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.

1.2 Questionnaires

Generic and country-specific questionnaires are downloadable from the SHARE-website: <http://www.share-project.org/data-access-documentation.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

3 What is New in SHARE Wave 4 Release 1.0.0

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

4 Countries

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 overview 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.

Overview: Countries and Language Versions in SHARE Waves 1-4

ID	Country (Short)	Country	Wave 1	Wave 2	Wave 3 SHARELIFE	Wave 4
11	AT	Austria	2004	2006/07	2008	2011
12	DE	Germany	2004	2006/07	2008	2011/12
13	SE	Sweden	2004	2006/07	2008	2011
14	NL	Netherlands	2004	2007	2008	2011
15	ES	Spain	2004	2006/07	2008	2011
16	IT	Italy	2004	2006/07	2008	2011
17	FR	France	2004/05	2006/07	2008	2011
18	DK	Denmark	2004	2006/07	2008	2011
19	GR	Greece	2004/05	2007	2008	-
20	Cg	Switzerland (German)	2004	2006/07	2008	2011
21	Cf	Switzerland (French)	2004	2006/07	2008	2011
22	Ci	Switzerland (Italian)	2004	2006/07	2008	2011
23	Bf	Belgium (French)	2004/05	2006/07	2008	2011
24	Bn	Belgium (Flemish)	2004/05	2006/07	2008	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	2011
29	PL	Poland	-	2006/07	2008	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. But 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 and below for changes in wave 4.

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

6.1 Types of Data in Wave 4

Overview: 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 and deceased respondents from previous waves
Individual CAPI modules	See also the overview in 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	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 only selected household members served as special respondents and answered some modules or questions also on behalf of other household members/partners. In wave 4 there is no selected household respondent anymore, but just a financial respondent.

Overview: Who Answers What in the CAPI Questionnaire?

CAPI Module	Name	All respondents	Financial 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 😊.

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 from:

<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.

Overview of 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 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.

11 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 networks member 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 network member. Network members can be found in the respective questions and loops in the SN module, see chapter 13.1 on the Social Networks module for more information.

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 coding	Wave 4 coding	Wave 1 & 2 variable names	Wave 4 variable names
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

12 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.

13 Generated Variables

13.1 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.

Overview: 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.

Overview: 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)

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

13.2 Imputations in SHARE Wave 4

Author: Giuseppe De Luca

Date: November 29, 2012

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 modules of the SHARE interview are affected to some extent by missing data. Imputations allow users to handle the potential bias and the loss of precision generated by missing data problems. Nevertheless, users should be aware of the fact that imputations are not generally the same as missing variable values (Little 1992; Meng 1994; Jones 1996; Horton and Kleinman 2007; Dardanoni et al 2011,

2012). First, if the missing data process is not missing-at-random or the imputation model is incorrectly specified with respect to the true data generating process, then imputations may lead to biased estimates. Second, valid inference requires taking into account the additional variability induced by the imputation process when assessing the precision of our estimators (Rubin 1987).

Like the imputation model used in Release 2.4 of waves 1 and 2 data (Christelis 2011), the imputation model used in Release 1 of wave 4 data relies on the fully conditional specification (FCS) method of van Buuren et al (2006). This is a multivariate iterative procedure which attempts to preserve the correlation structure of the imputed data. Since there are theoretical and computational reasons to limit the set of variables that are jointly imputed through the FCS method, we used this multivariate iterative procedure only for income, expenditure and wealth-related items that are affected by substantial amounts of missing data. Other important survey variables are instead imputed at an early stage by simple hot-deck and some of them are used as fully observed explanatory variables in FCS imputation model. For all imputed variables, we provide 5 (independent) imputations of their missing values to allow the assessment of the additional variability induced by the imputation process. A complete list of the variables imputed in Release 1 of wave 4 data is presented below (see chapter 13.2.2).

13.2.1 Key features of the SHARE wave 4 imputation model

Below, we summarize the key features of the imputation model adopted in wave 4 and the most striking differences with respect to the imputation models adopted in the previous waves:

1. The new imputation model exploits the information collected in the variable HH017_ (the one shot monthly total household income question) to provide two alternative measures of total household income, namely versions A and B. Version A (named *thinc*) is obtained by a suitable aggregation of the individual and household income components collected in SHARE. Version B (named *thinc2*) is obtained from HH017_. These two alternative measures of total household income do not need to coincide for a variety of reasons (e.g. measurement errors). We allow users to decide which of the two measures of total household income is more accurate for their research questions.
2. In Release 1 of wave 4 data we treat as missing values both records corresponding to “Don’t know” or “Refusal” answers and outlier records that may have an undue effect on survey statistics. For each continuous variable, outliers are identified by symmetrically trimming 2 percent of the observed distribution within each country. This gives us lower and upper bounds for the observations to be imputed.
3. As in the previous waves, the imputation model used in the FCS method takes into account the partial information on missing amount items collected

through the sequence of unfolding bracket (UB) questions. Point UB information is treated as imputed values, while interval UB information is embedded in the bounds by restricting the range of feasible values placed on missing values.

4. To limit the effects of multi-collinearity problems that may arise in the estimation step of our FCS imputation method and the computing time required by this multivariate iterative procedure, the long list of income, wealth and expenditure items collected in SHARE has been aggregated into a considerably smaller subset of homogeneous variables (i.e. from 69 to 23 variables). A description of this aggregation step is given in chapter 13.2.3 below. Although aggregation is reasonable strategy to reduce the number of variables that are jointly imputed into in the FCS model, it is not exempt from criticisms. On the one hand, it is expected to increase robustness of the FCS model and to reduce undue delays in the release of imputations. On the other hand, we can only provide imputation on the aggregated variables, but not on their components. Notice that an aggregate variable is considered to be missing if any of its components is missing. This is a conservative definition that may lead to substantial amount of missing data on the aggregate variables, but the availability of partial information (i.e. cases where some components of the aggregate are observed and some are not) is always preserved by restricting the corresponding bounds for missing values.
5. Unlike the imputation models used in the previous waves, non-responding partners (NRP) are no longer included into the imputation sample. The main reason to include NRP in the imputation samples of the previous waves was to avoid a downward bias in the imputations of total household income (version A). In the new imputation model, this issue is taken in account by directly imputing the sum of individual income components of NRP (see stage 4 of point 6).
6. The new FCS imputation process is carried out separately by country and sample type. The sample types considered are people living as singles (sample 1), couples with both partners interviewed (sample 2) and all couples - with and without non-responding partners - (sample 3).¹ We no longer specify separate subsamples for panel and refreshment samples to avoid small imputation samples. The FCS imputation process consists of four stages:
 - 6.1. Stage 1 (singles): we first impute missing data in all income, expenditure and wealth items through the FCS method and then compute generated variables such as total household income (version A), total household expenditure, household real asset, household gross

¹ These sample types can be identified through the indicators `sample1`, `sample2`, and `sample3` included in the imputation database. Households with NRPs and the responding partner of a NRP can be instead identified by the variables `htype` and `p_nrp`, respectively.

and net financial wealth and household net worth through the sum of their components. As explanatory variables for the imputation of missing data in the various income, expenditure and wealth items we use gender, age, years of education, self-perceived health, number of children, number of chronic diseases, numeracy, employment status and willingness to answer. The imputation of each variable into the system is carried out through a parametric two-part model. First, we model the probability of owning a certain item using a probit model. Then, we model the amount of the item conditional on ownership using a linear regression model. Total household income (version B) is imputed at the end of this stage using as predictors the above set of explanatory variables, plus total household income (version A), total household expenditure and household net worth.

- 6.2. Stage 2 (couples with two responding partners): Same procedure as in Stage 1. The main difference is that the FCS model exploits information about the partner as predictors of the missing values in the various income, expenditure and wealth items.
- 6.3. Stage 3 (all couples – with and without NRPs): For the responding partner of a couple with NRP, we first impute missing data in all income, expenditure and wealth items using a FCS method similar to that used in Stage 2. In this case, the FCS model includes only a smaller subset of predictors related to the NRP (namely age and years of education). Individuals belonging to couples with two responding partners are included into the imputation sample, but only as donors (their missing data have been already imputed in Stage 2). After imputing the various expenditure and wealth items, we can compute expenditure and wealth related household aggregates as in Stage 2 since all their components are only recorded for the responding partner by design. The same is not true for total household income (version A) because we still need to account for the individual income components of the NRP (that are not included in the imputation sample). Therefore, we first impute total household income (version B) using explanatory variables related to the respondent and the NRP, total household expenditure and household net worth as predictors. Notice that total household income (version B) is easier to impute than total household income (version A) because we can exploit the UB information provided in the corresponding one-shot question (HH017_). Finally, we impute total household income (version A) using the explanatory variables related to the respondent and the NRP, total household expenditure, household net worth, and total household income (version B) as predictors. In this case, we also use the sum of the income components imputed so far as a lower bound for the missing values on total household income (version A).

7. The status of each variable in the imputation database is summarized by a flag variable named as varname_f. A list of all possible values of the flag variables is presented below.

References

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

Christelis, D. (2011), Imputation of missing data in waves 1 and 2 of SHARE. *SHARE Working Paper Series*: 1 – 50.

Dardanoni, V., G. De Luca, S. Modica and F. Peracchi (2012), A generalized missing-indicator approach to regression with imputed covariates, *Stata Journal* 12: 1-30.

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

Horton, N., and K. Kleinman (2007), Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician* 61: 79 – 90.

Jones, M. (1996), Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association* 91: 222 – 230.

Little, R.J.A. (1992), Regression with missing X's: A review. *Journal of the American Statistical Association* 87: 1227–1237.

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

Rubin D.B. (1987), *Multiple imputation for nonresponse in surveys*, John Wiley & Sons, New York.

13.2.2 List of variables included in the imputation database

Note: The variables listed below are imputed either by simple hot-deck or jointly in the fully conditional specification system. A variable is imputed using simple hot-deck imputation if the prevalence of missing values for an item is

- lower than 5% for the entire sample (all wave 4 observations) **and**
- lower than 10% for the country specific samples.

This implies that the list of variables to be jointly imputed in the system is country and possibly subsample specific.

Variable name	Description
pidcom	Person ID
implicat	Implicat number
hhidcom4	Household ID wave 4
country	Country identifier
language	Language of questionnaire
htype	Household type
fam_resp	Family respondent
fin_resp	Financial respondent

hou_resp	Household respondent
excrate	Exchange rate
nursinghome	Living in nursing home
hhsz	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 2R
sample3	Imputation sample for all couples
ydip	Earnings from employment
yind	Earnings from self-employment
yopen1	Old age & early retirement pensions
yopen2	Disability, unemployment, survivor and war pensions
yreg	Other regular payments
ylsum	Lump sum payments
rhre	Rent and home-related expenditures
home	Value of main residence
mort	Mortgage on main residence
ores	Value of other real estate – Amount
ysrent	Income from rent or sublet
yaohm	Income from other household members
fahc	Food at home consumption
fohc	Food outside home consumption
hprc	Home produced consumption
bacc	Bank accounts
bsmf	Bond, stock and mutual funds
slti	Savings for long-term investments
vbus	Value of own business
sbus	Share of own business
car	Value of cars
liab	Financial liabilities
yibacc	Interest income from bank accounts
yibsmf	Interest income from bond, stock and mutual funds
thinc	Total household income - version A
thinc2	Total household income - version B
thexp	Total household expenditure
hrass	Household real assets
hgfass	Household gross financial assets
hnfass	Household net financial assets
hnetw	Household net worth
gender	Gender
age	Age in 2010
age_p	Age of partner in 2010

yeduc	Year of education
yeduc_p	Year of education of partner
sphus	Self-perceived health - US scale
mstat	Marital status
child	Number of children
gchild	Number of grandchildren
gali	Limitation with activities
chronic	Number of chronic diseases
symptoms	Number of symptoms
bmi	Body mass index
weight	Weight
height	Height
mobility	Mobility limitations
adl	Limitations with activities of daily living
iadl	Limitations with instrumental activities of daily living
esmoked	Ever smoked daily
drinking	More than 2 glasses of alcohol almost everyday
phactiv	Physical inactivity
meals	Number of meals every day
orienti	Score of orientation in time test
memory	Score of memory test
wllft	Score of words list learning test - trial 1
wllst	Score of words list learning test - trial 2
fluency	Score of verbal fluency test
numeracy1	Score of first numeracy test
numeracy2	Score of second numeracy test
eurod	EURO depression scale
doctor	Seen/Talked to medical doctor
hospital	In hospital last 12 months
thospital	Times being patient in hospital
nhospital	Total nights stayed in hospital
sn_num	Number of people within social network
sn_sat	Satisfaction with social network
cjs	Current job situation
pwork	Did any paid work
empstat	Employee or self-employed
lookjob	Looking for job
rhfo	Received help from others (how many)
ghfo	Given help to others (how many)
ghih	Given help in the household (how many)
rhih	Received help in the household (how many)
gfg	Number of given financial gifts 250 or more
rfg	Number of received financial gifts 250 or more
otr	Owner, tenant or rent free
perho	Percentage of house owned
fdistress	Household able to make ends meet

lifesat	Life satisfaction
lifehap	Life happiness
naly	Number of activities last year
saly	Satisfied with no activities
willans	Willingness to answer
clarif	Respondent asked for clarifications
undersq	Respondent understood questions
hnrsc	Help needed to read show-cards

13.2.3 Description of the Aggregation Step

Aggregate Variables	Definition
Old age and early retirement pensions (yapen1)	Public old age pension (pen1), public old age supplementary pension (pen12), private old age pension (pen8), private old age pension from a second job (pen15), private old age pension from a third job (pen16), public early retirement pension (pen2), and private early retirement pension (pen9).
Disability, unemployment, survivor and war pensions (yapen2):	Public disability insurance pension or sickness benefits (pen3), secondary public disability insurance pension or sickness benefits (pen13), private disability insurance pension (pen10), public long-term insurance payments (pult), unemployment benefit or insurance (pen4), public survivor pension from partner (pen5), secondary public survivor pension from partner (pen14), private survivor pension from partner's job (pen11), war pension (pen7).
Other regular payments (yareg):	Regular life insurance payments (reg1), regular payments from private annuity or personal pension (reg2), regular payments from alimony (reg3), regular payments from charities (reg4), regular payments from long-term care insurance (reg5).
Lump sum payments (ylsum):	Lspen1, lspen2, lspen3, lspen4, lspen5, lspen7, lspen8, lspen9, lspen10, lspen11, lspen12, lspen13, lspen14, lspen15, lspen16, lspult, lsreg1, lsreg2, lsreg3, lsreg4, lsreg5.
Rent and home-related expenditures (rhre):	Amount rent paid (rentc) and other home-related expenditures (ocsc).
Income from rent or sublet (ysrent):	Income from sublet (ysub) and income from rent of real estate (yrent).
Income from other household members (yaohm):	Other household members' net income (yohm), and other household Members' net income from other sources (yohb).
Bond, stock and mutual funds (bsmf):	Government/corporate bonds (bond), stocks (stoc), mutual funds (mutf).
Savings for long-term investments (slti):	Individual retirement accounts from respondent (ira) individual retirement accounts from partner (ira2), contractual saving (cont), and whole life insurance holdings (lins).

13.2.4 Value of Flag Variables in the Imputations Module

Varname_f	Label	Description
-99	"Missing by design"	(because of skip patterns in the questionnaire)
1	"Not designed resp"	Missing is due to survey design
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 chapter 13.2.3
13	"Imp: NRP"	(only for thinc)
14	"Imp: missing value"	(only for explanatory variables imputed ex-ante by hot-deck)

13.3 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 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.

13.4 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.

Overview: 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>crp_{mg/l}</i>	C-reactive protein [mg/l]
<i>chol_{mg/dl}</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
and all SHARE country team operators.

Contact: info@share-project.org