



## A Short Guide to SHARE Release 2.0

- 1 Cross-national data set and country specifics ..... 2
- 2 Data modules ..... 2
- 3 Respondents of the CAPI main questionnaire ..... 3
- 4 Merging the data ..... 6
- 5 Self completion questionnaire ..... 6
  - 5.1.Drop-offs ..... 6
  - 5.2.Vignettes ..... 6
- 6. Naming conventions ..... 7
- 7. Dummy variables ..... 7
- 8. Euro conversion ..... 8
- 9. Unfolding brackets ..... 9
- 10. CH module: Selection in child loop ..... 9
- 11. Top coding ..... 10
- 12. Specific variables ..... 10
- 13. Contact Info for general inquiries ..... 10

## 1 Cross-national data set and country specifics

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national database. One of its innovative features is the ex-ante harmonization of the data collection instruments in the generic SHARE questionnaire. The generic questionnaire served as a model for the translation of the country-specific questionnaires. However, some internationally highly diverse variables such as education require country-specific measurement.

**For a detailed overview on country specifics, see:**

<http://www.share-project.org/sharefb/>

## 2 Data modules

SHARE data collection is based on a computer-assisted personal interviewing technique (CAPI). The SHARE interview consists of various data collection instruments:

- The interview starts with a coverscreen interview on the household level, answered by one household member (filename **cv\_h**). Data on the individual level are available as well (**cv\_r**).
- The main questionnaire is based on 20 different CAPI modules (see questionnaire).
- In the *main* sample, the interview is finished with a self-completion paper & pencil questionnaires ("main drop off questionnaire") with additional questions in the areas of mental and physical health, health care, and social networks.
- *Extra* samples ("vignette samples") were taken in eight countries in order to collect a special self-completion questionnaire with anchoring vignette questions designed to improve cross-national comparability.

Households were either drawn

- in the main sample, completing a CAPI interview and a drop-off, or
- in the vignettes sample, answering the CAPI instruments and the vignettes.

To which sample a household belongs can be seen from the variable named *samptype* in the weights file. The sampling procedure in Sweden differed. See for details the file "Short information on generated variables: Weights" (SHARErel2\_GV\_weights.pdf).

Additionally, the SHARE data sets contains **various generated variables** that are documented separately in more detail.

**Table 1: Composition of the SHARE data set**

<b>Elements</b>	<b>Overview on documentation and comments</b>
Coverscreen interview cv_h	on the household level
Coverscreen interview cv_r	on the individual level for all household members, including non-eligible persons
CAPI Modules	See page 4 and main questionnaire, see for download <a href="http://www.share-project.org/index.php?page=Questionnaire&amp;menu=3&amp;sub=">http://www.share-project.org/index.php?page=Questionnaire&amp;menu=3&amp;sub=</a>
Drop-offs and vignettes	See for download <a href="http://www.share-project.org/index.php?page=Questionnaire&amp;menu=3&amp;sub=">http://www.share-project.org/index.php?page=Questionnaire&amp;menu=3&amp;sub=</a>
Generated variables: <ul style="list-style-type: none"> <li>• Health</li> <li>• Housing and region</li> <li>• Social support and household composition</li> <li>• Education, according to ISCED codes</li> <li>• Occupation and industries, according to ISCO and NACE codes</li> <li>• Weights</li> <li>• Imputed data</li> <li>• Gross sample</li> </ul>	See detailed documentation: <ul style="list-style-type: none"> <li>• health: SHARE1rel2-0_GV_health.pdf</li> <li>• housing: SHARE1rel2-0_GV_housing.pdf</li> <li>• social support and household composition: SHARE1rel2-0_GV_social_support.pdf</li> <li>• education: SHARE1rel2-0_GV_education.pdf</li> <li>• occupation and industries: SHARE1rel2-0_GV_occupation.pdf</li> <li>• weights: SHARE1rel2-0_GV_weights.pdf</li> <li>• imputed data: SHARE1rel2-0_imputations.pdf</li> <li>• available on request</li> </ul>

### **3 Respondents of the CAPI main questionnaire**

#### ***Types of respondents***

The SHARE CAPI main questionnaire is designed in such a way that not every eligible household member has to answer every CAPI module (see table 2, p. 4).

- **All respondents** answered the following modules: **DN, PH, BR, CF, MH, HC, EP, GS, WS, SP** *partly*, **AC**, and **EX** (see table 2, p. 4).
- **Selected household members** served as **family, financial or household respondents**. They answered questions about children and social support, financial issues or household features, on behalf of the couple or the household, respectively (see table 2, page 4).

## SHARE Release 2.0, June 2007

- The answers to finance, housing and family questions in modules FT and AS, HO, HH, CO, CH and the first part of SP are only coded for the financial, housing or family respondents, respectively.
- However, for the **generated variables**, the information is stored for all respondents, regardless of their status as regular or financial, housing or family respondent.
- **Proxy interviews** were allowed for most of the modules (see table 2, p.4).

### ***Selection of the financial, household and family respondent***

The financial, household and family respondent was identified during the interview as follows:

- **Family respondents** answer the questions of the **CH** module and the first part of the **SP** module (SP001 to SP017). They are indicated by the dummy variable *dumfamr*. They are selected by the chronological order of interviews per **couple** (married or not): The couple's first person interviewed is the family respondent. Note that the naming of the *cvid/respid* variable does not indicate the chronology of interviews within one household.
- The **financial respondent** is identified by the questions CM002 and CM003 at the start of the DN module. The financial respondent answers the modules FT and AS and is indicated by the dummy variable *dumfinr*. In case of a one-person household, the respondent is **always** the financial respondent. In multi-person households, the number of financial respondents may vary: respondents living without a partner in multi-person households are always financial respondents. Eligible couples, i.e. spouses and partners, may decide in CM002 to answer questions about their finances separately (this can be retrieved from *finsep*). Otherwise, one partner can answer on behalf of the **couple**. In this case, he or she is identified as the financial respondent for the couple in CM003, indicated by the dummy variable *dumfinr*.
- Similarly, only **household respondents** (picked in CV038) answer on behalf of the whole household questions about household features (HO, HH, CO). The household respondent is indicated by the dummy *dumhhr*.

**Table 2: Who answers what in the CAPI questionnaire?**

CAPI* Module	Name	All Respondents	Financial Respondent	Household Respondent	Family Respondent	non - proxy section
CV	Coverscreen					
DN	Demographics	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
WS	Walking Speed	x				x
CH	Children				x	
SP	Social Support	x			x	
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					

\*chronological order of the CAPI questionnaire

#### 4 Merging the data

The data for the coverscreen interview are available on the household level in the file **cv\_h**.

- Its key identifier is *sampid2*, the scrambled identification number of the household.
- The coverscreen data is also available on an individual level in the file **cv\_r** with the key identifiers *sampid2* AND *cvid*, the coverscreen identifier for an individual (within the household).

Datasets based on the main part of the instrument have two case identifiers other than *sampid2*:

- **cvid**: indicates the corresponding identifier from the coverscreen interview. The coverscreen interview comprises numbers for non-eligible or non-responding household members. They have no *respid* because they do not appear in the data set resulting from the main part of the instrument.
- **respid** (respondent identifier): indicates the case identifier for the main instrument (only eligible responding persons).

**Thus, when merging the data, one should use *sampid2* AND *cvid* as keys.**

#### 5 Self completion questionnaire

##### 5.1. Drop-offs

The status variable 'do\_status' indicates how the data from the drop-offs matches the data of the coverscreen and CAPI.

In Israel, the drop-off includes additional questions that are not asked in other countries. These variables have the prefix 'il'. The numeration of question in the Israeli paper version of the drop off differs from the generic version. In the data file the variables are harmonized according to the generic version. See for deviations of the Israeli drop-off the file "drop-off deviations Israel.pdf":

[http://www.share-project.org/index.php?menue=1&page=Questionnaire&link=Q-Israel\(Arabic\)](http://www.share-project.org/index.php?menue=1&page=Questionnaire&link=Q-Israel(Arabic))

##### 5.2. Vignettes

SHARE 2004 includes vignettes as self-administered questionnaires in Sweden, Belgium, Spain, France, Germany, Greece, Italy, and the Netherlands. Two types were randomly assigned to the respondents: type A and type B. They differ with regard to the question order and gender of the persons described in the statements. The variable 'type' contains the information on the vignette type.

## SHARE Release 2.0, June 2007

**Note that in the data set the answers are recoded to the order of the vignettes type A.** Which questions from type B correspond to type A can be seen from the variable label. The status variable 'vi\_status' indicates how the data from the vignettes matches the data of the coverscreen and CAPI.

### 6. Naming conventions

In general, the variable names in the main instrument data use the following format:

MMXXX\_YY

MM    module identifier, e.g. DN  
 XXX    question number, e.g. 001  
 \_    separation character, for **dummies** (see 6.) replaced by '**d**'  
 YY    optional digit(s) for category or loop indication; in case two digits are used the first digit refers to the 'outer loop' and the second digit refers to the 'inner loop'. When the loop number goes beyond 9 and there is only one character available, the numbering continues with a (=10), b (=11) etc.

The format slightly deviates in case of variables with Euro amounts (see 8) or variables for the unfolding brackets (see 9.)

### 7. Dummy variables

Answers to all questions that allow for multiple responses have dummy variables as final data. E.g. question BR005 ("What do or did you smoke") has three answer categories:

1. Cigarettes
2. Pipe
3. Cigars or cigarillos

The data set contains three dummies: BR005d1, BR005d2, and BR005d3 corresponding to the three categories. A value '1' means that the respondent chose the particular category as an answer, and in case of a value '0' the respondent did not choose the particular category as answer.

- In case the respondent answers with a "none of these", "don't know" or "refusal", the naming of the dummy names has the following structure:

MMXXXdno	"none of these"
MMXXXdrf	"refusal"
MMXXXddk	"don't know"

## SHARE Release 2.0, June 2007

- In case the question requires **loop indication**, the second last digit indicates the number of the loop (#), the last letter stands for the answer categories "none of these", "refusal", "don't know":

MMXXXd#n "none of these"

MMXXXd#r "refusal"

MMXXXd#d "don't know"

### 8. Euro conversion

All answers about an amount of money are converted into Euro values. For non-Euro countries a frozen exchange rate is chosen. For Euro countries the Euro value is either the given value or the converted pre-Euro value because respondents in Euro countries were given the option to report in either Euro or the pre-Euro currency.

The format of the variable name is mentioned in 4 except for the separation identifier '\_' ; this is now replaced by 'e'. Possible digits that follow reflect loop numbers.

When the respondent gives a "don't know" (DK) or "refusal" (RF) as an answer to a question indicating a financial amount, the following values are included in the dataset:

-9999998: Refusal

-9999999: Don't know

The following exchange rates were used for the EURO conversion:

**Table 3: Frozen Euro exchange rates used in SHARE**

Country	Currency	Old currency	Exchange rate (x to the Euro)
<b>Austria</b>	Euro	Austrian Schilling	13.7603
<b>Germany</b>	Euro	German Mark	1.95583
<b>Sweden</b>	Swedish Krona	-	9.1803
<b>Netherlands</b>	Euro	Dutch Guilder	2.20371
<b>Spain</b>	Euro	Spanish Peseta	166.386
<b>Italy</b>	Euro	Italian Lire	1936.27
<b>France</b>	Euro	French Franc	6.55957
<b>Denmark</b>	Danish Krone	-	7.4388
<b>Greece</b>	Euro	Greek Drachma	340.75
<b>Switzerland</b>	Franc	-	1.5342
<b>Belgium</b>	Euro	Belgium Franc	40.3399



## SHARE Release 2.0, June 2007

### 9. Unfolding brackets

When a respondent does not know (DK) or refuses (RF) the answer to a question about an amount of money, usually an unfolding sequence of bracket questions starts. There are three entry points, and the starting point is chosen randomly. All details of the sequence are stored in the dataset. However, in the public release only a few (summary) variables are included. For all sequences we have the **country-specific** bracket values (in Euros) and the final category where the respondent ended. When a DK or RF is given during the unfolding bracket sequence, the value for the final category is set to either DK or RF.

The format of the summarizing unfolding bracket variable is as follows:

MMXXXubY

with

**MM** module identifier, e.g. HC  
**XXX** question number, e.g. 045  
**Y** optional digit for loop indication

The variable indicating where the respondent finally ends can take seven values:

1. Less than low entry point
2. About low entry point
3. Between low and mid entry point
4. About mid entry point
5. Between mid and high entry point
6. About high entry point
7. More than high entry point

The country-specific bracket values are indicated as:

MMXXXv1, MMXXXv2, and MMXXXv3

In the case of a loop, the digit for the loop indication **precedes** the 'v1', 'v2', and 'v3'.

### 10. CH module: Selection in child loop

Questions CH009 to CH020 in module CH about children are only asked for at most four children. When there are more than four children, the CAPI program selects the four children as follows:

## **SHARE Release 2.0, June 2007**

1. Sort children **in ascending order** by
  - minor (defined as 0 for all children aged 18 and over and 1 for all others),
  - geographical proximity (CH007),
  - birth year.
2. Pick the first four children.

When all sorting variables are equal, the CAPI program chooses a child randomly.

The variables chselch1 up to chselch4 contain the numbers of the children who were selected by the program. The numbers refer to the order in which the respondent listed the children.

### **11. Top coding**

In the public release top coding was done in the case of Sweden due to legal constraints, according to the Swedish Secrecy Act. The following variables had to be top coded (with the value displayed to the right):

ho024e: 1,000,000 Euros (or 9,000,000 SEK)  
ho027e: 2,750,000 Euros (or 25,000,000 SEK)

### **12. Specific variables**

*PH module: PHrandom*

There are two types of answer categories for the question about self-perceived health. Which type is asked at the beginning of section PH (questions PH002/003), and which (other) type for the end of this section (questions PH052/053), is randomized. The variable phrandom indicates which type is chosen:

1 for PH002/PH052 (version 1)  
2 for PH003/PH053 (version 2)

*EX module: EX009age, EX012val*

EX009age: age used in question EX009  
EX012val: value used in question EX012

### **13. Contact Info for general inquiries**

Mannheim Research Institute for the Economics of Aging  
Stephanie Stuck, M.A.  
[stuck@mea.uni-mannheim.de](mailto:stuck@mea.uni-mannheim.de)