

# **CODEBOOK**

## **SHARE – WAVE 2**

**Last update: October 2013**

<b>WARNINGS FOR USERS OF THE CODEBOOK.....</b>	<b>3</b>
Raw data .....	3
National specificities.....	3
<b>USERS GUIDE.....</b>	<b>4</b>
Objectives .....	4
Contextual elements wave 2 .....	4
Data file contents.....	5
How is the information presented? .....	6
<b>METHODOLOGICAL REMARKS .....</b>	<b>10</b>
Financial data .....	10
Respondent.....	10
Routing.....	11
Generated variables .....	12
Variable code .....	12
Loop .....	13
Showcards .....	13

# WARNINGS FOR USERS OF THE CODEBOOK

## Raw data

The figures presented here are intended to provide SHARE users with first descriptive results in the form of raw statistical data with no weighting. This unprocessed data is therefore not representative of the different populations concerned and should not be used to draw conclusions regarding these populations.

The only treatment applied to the raw data is the SHARETOM programme developed by the European coordination team at Mannheim. This automatic programme (sharetom.ado) can be downloaded with the data and its use is highly recommended. It treats missing codes and converts values associated with missing values (-1 for 'Don't know' for example) into codes that do not therefore enter into statistical calculations (.a for 'Don't know').

Furthermore, the quantile values for the quantitative variables were rounded to the closest whole number, with the exception of the variables ws011 and ws013 for which the statistics display two decimal places.

A first data cleaning has already been done at European level but some errors may remain. In addition to data cleaning, some variables have been corrected for this *codebook* (pf003 for example). A comment has been added beneath the tables concerned.

## National specificities

SHARE provides *ex-ante* and *ex-post* harmonized data (financial data converted into Euros for example). Due to the diversity of national systems however, certain questions, such as those regarding allowances or public benefits received, types of work contract or education, provide non-comparable results without additional processing.

The results presented in this document emanate from raw data that make no distinction between national specificities since the statistics are presented for all countries combined. For certain questions, the responses are thus presented under the generic appellation 'Country specific category'. It is thus indispensable to refer to the SHARE web site (heading *Item Correspondence*) for the data analysis.

Furthermore, the labels used in this document are taken from the English version of the questions and answers as they are presented in the generic questionnaire. National SHARE teams then translated and adapted them to their country's specificities.

# USERS GUIDE

For detailed information regarding the methodology and data file use, please refer to the user guide entitled *Release Guide 2.5.0 Waves 1 & 2* available from the SHARE web site: [www.share-project.org](http://www.share-project.org) (thumbnail *Documentation*). Certain elements of this guide are taken up here to emphasise the originality and objectives of this codebook.

## Objectives

The main objective is to present the whole set of questions asked during the wave 2 of the survey in each module together with response distributions (data version 2.5.0).

This codebook groups together statistical information (frequency distributions, averages, quartiles...) for SHARE wave 2 (2006-2007) and thus offers users an overall view of the survey data. It notably allows users to validate sub-sample numbers before carrying out more detailed analyses.

The statistical information presented here concerns all the European countries that participated in SHARE wave 2. It is important to note that this information is **not representative** of the population aged 50 and over as a whole in these countries. It is **raw data with no weighting**.

The complete datasets can be downloaded free of charge on request from the SHARE web site [www.share-project.org](http://www.share-project.org).

## Contextual elements wave 2

### Sample wave 2

SHARE interviews a sample of individuals<sup>1</sup> aged 50 and over and their partner, if applicable (whatever their age). During the course of the second wave, almost 35,000 individuals were interviewed in 14 European countries (Germany, Austria, Belgium, Denmark, Spain, France, Greece, Ireland, Italy, the Netherlands, Poland,

---

<sup>1</sup> In certain countries samples are composed of individuals whereas in others, such as France, they are composed of households as selection is based on housing, and thus households in which at least one member is aged 50 or over (Boersch-Supan et al (2005). The Survey of Health, Aging, and Retirement in Europe, Methodology, MEA, University of Mannheim (Germany), [www.share-project.org](http://www.share-project.org)).

the Czech Republic, Sweden and Switzerland). Of these, over 18,000 respondents had already participated in the first wave.

Wave 2 equally monitored wave 1 respondents that had subsequently been institutionalised.

### Themes treated

The collected data explores the following themes:

- Health status (physical health, behavioural risks, cognitive functions, mental health, health care system, grip strength, peak flow, walking speed...);
- Labour force participation (employment and retirement);
- Material living conditions (wealth, income, consumption, financial assets, financial transfers, housing);
- Demographics (household composition, children);
- Social environment (family and social support, social activities, expectations).

### **Data file contents**

During the data collection process, questions are organised into modules respecting the following order:

CV	<i>Coverscreen (household composition)</i>
DN	<i>Demographics</i>
PH	<i>Physical Health</i>
BR	<i>Behavioural Risks</i>
CF	<i>Cognitive Function</i>
MH	<i>Mental Health</i>
HC	<i>Health Care</i>
EP	<i>Employment and Pensions</i>
GS	<i>Grip Strength</i>
PF	<i>Peak Flow</i>
CS	<i>Chair Stand</i>
WS	<i>Walking Speed</i>
CH	<i>Children</i>
SP	<i>Social Support</i>
FT	<i>Financial Transfers</i>
HO	<i>Housing</i>
HH	<i>Household Income</i>
CO	<i>Consumption</i>
AS	<i>Assets</i>
AC	<i>Activities</i>
EX	<i>Expectations</i>
IV	<i>Interviewer Observations</i>

XT	<i>End-of-Life Interview</i>
Drop-off	<i>Drop-off</i>
Vignettes	<i>Vignettes</i>

A data file<sup>2</sup> (in STATA or SPSS format) corresponding to each of the modules includes:

- variables common to all modules and identifiers (mergeid, hhid, hhid1, hhid2, country, waveid, split, language, cvid) on the one hand
- and on the other, variables generated from each of the questions.

Variables specific to each of the modules are identified via a code containing the question number. This code is made up of a two character abbreviation of the module name in English (1<sup>st</sup> column of the preceding table) and a 3 figure number. For example, the variable 'FT002\_' corresponds to question '002' of the module 'FT – Financial Transfers'.

In total, almost 2 000 variables are available, all data files combined.

Variables are classified into three major categories (see examples on the following pages):

- (1) Qualitative variables: the values taken by the variables are not quantifiable;
- (2) Quantitative variables: the values taken by the variables are quantifiable;
- (3) Dummy variables: for the multiple choice questions (MCQ), several variables each corresponding to one of the possible responses were created; each variable takes the value 1 if the response has been selected, otherwise 0.

For each of these categories, a different statistical treatment was effectuated.

## How is the information presented?

The document is organised by module respecting the chronological order of a wave 2 interview (see the list of modules in the preceding above). The first chapter presents the variables available in each of the detailed data files and the following chapters, each corresponding to a module, contain the totality of questions asked.

For each question, the information is presented in two phases.

First, a table allows the user to visualize the question asked:

---

<sup>2</sup> The section entitled *Coverscreen* resulted in the creation of 2 files: CV-H at household level and CV-R at individual level.

<b>Variable</b>	<b>Variable number or list of variables for MCQ</b>
Label	<i>Variable label in STATA (in English)</i>
Question	<i>Question exactly as posed to the respondent</i>
Instruction	<i>Instruction given to the interviewer if applicable</i>
Filter	Filter applied to the question during the one-to-one interview. In certain cases, this can explain the large number of 'item non-responses'.

Then, following the type of variable, a table presents the responses provided by the respondents in the different European countries together with statistical data.

### Qualitative variable

These variables generally correspond to questions requesting respondents' to choose one single response out of several possible responses.

For example, for the question 'ph003 – Would you say your health is...'

<b>Responses</b>	<b>Frequency</b>	<b>Percentage<sup>(a)</sup></b>	<b>Cumulated percentage</b>
1. Excellent	3 170	9.2	9.2
2. Very good	6 516	19.0	28.3
3. Good	12 642	36.9	65.1
4. Fair	8 365	24.4	89.5
5. Poor	3 590	10.5	100.0
Don't know	10	.	.
Refusal	14	.	.
Missing value <sup>(b)</sup>	108	.	.
	<b>34 415<sup>(c)</sup></b>	<b>100,0</b>	

(a) The percentage is calculated in relation to completed observations.

(b) The 'Missing value' field contains missing values and filtered individuals.

(c) The last line corresponds to the total number of observations.

### (1) Quantitative variable

Certain responses are given as quantities, for example in questions regarding income, age, age at which an event occurred, the respondent's height, weight...

'ph009\_10 – About how old were you when you were first told by a doctor that you had cancer?'

<b>Statistics</b>	<b>Value</b>
Mean <sup>(a)</sup>	58
Standard deviation <sup>(a)</sup>	13
Minimum	1
1 <sup>st</sup> quartile	50
Median	58
3 <sup>rd</sup> quartile	67
Maximum	85
Number of observations <sup>(b)</sup>	689
Don't know	1
Refusal	2
Missing value	33 723

(a) The average and the standard deviation are not weighted.

(b) The number of observations corresponds to the number of completed observations.

## (2) Indicator variable (MCQ)

For the multiple choice questions, the respondent has the possibility of selecting several responses. In order to stock this information, an indicator variable corresponding to each of the possible responses was created.

In this case, we present the response frequency for each of the possible responses and the percentage is calculated according to the number of respondents.

/!\ In the case of a multiple choice question, the sum of percentages is generally superior to 100%.

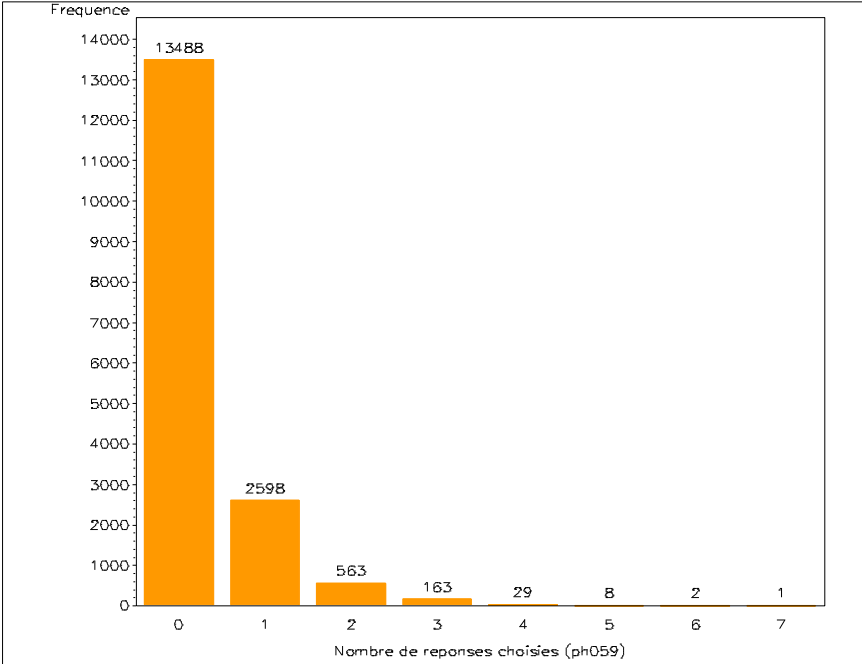
'ph059 – Please look at card 13. Do you use any of the items listed on this card?'  
(Only the first 3 possible responses are presented below):

<b>Responses</b>	<b>Frequency</b>	<b>Percentage</b>
1. A cane or walking stick	2 489	14.8
2. A zimmer frame or walker	821	4.9
3. A manual wheelchair	378	2.2
...	79	0.5
Number of observations	16 852	.
Don't know	4	.
Refusal	5	.
Missing value	17 554	.



This information is completed by a graph presenting the number of possible responses chosen by each of the respondents.

Question	ph059 (continued)
Histogram	of the number of ticked answers by person (among the respondents)



Reading	guide
Ex. 1:	13488 people have ticked the answer 96. None of these
Ex. 2:	2598 people have chosen only one answer for this question

# METHODOLOGICAL REMARKS

## Financial data

All financial data (modules EP and AS for example) were collected in the local currencies. They were then converted into Euros (exchange rates available in the tables containing the detailed data).

## Respondent

In certain question headings or certain interviewer instructions, the respondent may be referred to as 'The respondent' or the letter 'R'.

In the main questionnaire, different question modules are asked according to the respondent's category within the household. The table below recapitulates the question modules according to population.

There are three main types of respondent:

- The **financial respondent** is determined prior to the beginning of module DN. This respondent answers modules FT and AS and is identified by the variable dumfin<sup>3</sup>;
- The **household respondent** is determined prior to the individual interview. This respondent answers questions concerning the household in general (modules HO, HH, CO), and is identified by the variable dumhhr<sup>8</sup> ;
- The **family respondent** (first person in a couple to start the main interview) answers module CH and the beginning of module SP (SP001 to SP017) and is identified by the variable dumfamr<sup>8</sup>.

Depending on the module, this explains the fact that the number of respondents per question changes.

---

<sup>3</sup> These variables are available in the data file CV-R.

### Who answers which module?

N°	Module	Question	All	Financial	Household	Family	Non-proxy
0	CV	Coverscreen					
1	DN	Demographics	x				
2	PH	Physical health	x				
3	BR	Behavioural risks	x				
4	CF	Cognitive function	x				x
5	MH	Mental health	x				x (partial)
6	HC	Health care	x				
7	EP	Employment and pensions	x				
8	GS	Grip strength	x				x
9	PF	Peak flow	x				x
10	CS	Chair stand (< 75 years old)	x				x
11	WS	Walking speed (> 75 years old)	x				x
12	CH	Children	x				
13	SP	Social support <sup>(a)</sup>	x (partial)			x (partial)	
14	FT	Financial transfers		x			
15	HO	Housing			x		
16	HH	Household income			x		
17	CO	Consumption			x		
18	AS	Assets		x			
19	AC	Activities	x				x
20	EX	Expectations	x				x
21	IV	Interviewer observations <sup>(b)</sup>					
22	XT	'End of Life' interview					

(a) Certain questions regarding help received are asked to the family respondent only.

(b) To be completed by the interviewer after each interview.

## Routing

As in all surveys, certain questions are filtered. Following the respondent's answer to the previous question, depending on age or whether or not they live in a private household or a nursing home for example, the following question will either be asked or not which can lead to a high number of item non-responses. The fact of belonging to a longitudinal sample equally constitutes an important filter. A certain number of questions are not asked if respondents were previously interviewed in wave 1 of the survey.

**Information regarding filters is presented in the same way as it is presented in the generic questionnaire (English) used in SHARE wave 2.**

## Generated variables

Generated variables have been created for certain modules so as to be able to compare results (for example, in the case of 'Country specific categories' mentioned in the Warning section), carry out item non-response adjustments and impute missing values. A large number of generated variables are available (interview date, total income, assets, health, housing...). They will be described in a future version of the codebook.

Data files containing the generated variables or imputations are available for download. The list of available files is as follows:

- gv\_health.dta: generated health variables;
- gv\_housing.dta: generated housing variables;
- gv\_isced.dta: generated education variables;
- gv\_weights.dta: weighting taking respondents into account only;
- gv\_weights\_nrp.dta: weighting taking into account partners that did not respond;
- imputations.dta: imputed variables.

## Variable code

The code assigned to a variable generally corresponds to the question to which it is associated. The code is made up of a two character abbreviation of the module name in English (see the table listing the modules in the user's guide) and a three figure number.

There are, however, three exceptions:

- Loops: a same question can be asked several times with variations in certain fields. The variable code then contains the question number followed by *underscore* and occurrence number.

*Example*: ph068\_3 signifies the 3<sup>rd</sup> occurrence of the question ph068\_.

- Double-loop: questions ep111\_ to ep116\_ are placed within two simultaneous loops. The first loop corresponds to the number of one of the responses given to question ep110\_ (type of pension or public benefit received, ranging from 1 to 6), and the second, to the period of reference for the received benefit (ranging from 1 to 20).

*Example*: ep112\_2\_3 corresponds to the year in which the 2<sup>nd</sup> type of pension or public benefit was first received (see questionnaire for details) for a third period of entitlement.

- Multiple choice questions: each variable corresponds to one of the possible responses to the question and its code corresponds to the question number followed by 'd' and the response number.

*Example*: ph010d3 corresponds to the 3<sup>rd</sup> possible answer to the question ph010\_.

## Loop

Certain questions were asked several times. In this codebook, supplementary variables are presented only if the information they provide is different and non-redundant.

For example, in the module 'PH – Physical health', the 4 variables associated with question PH067 were presented as they corresponded to the appearance of a different physical problem (heart attack, stroke, cancer or hip fracture).

*A contrario*, in the module 'CH –Children', only the 1<sup>st</sup> variable (out of 14) associated with question 'CH005 - is {child's name} male or female?' was presented.

## Showcards

During the data collection process, a book containing possible answers to certain questions was made available to respondents by the interviewer.

This book is available in PDF format by clicking on 'Showcards' on the bottom left hand side of each page of the summary.