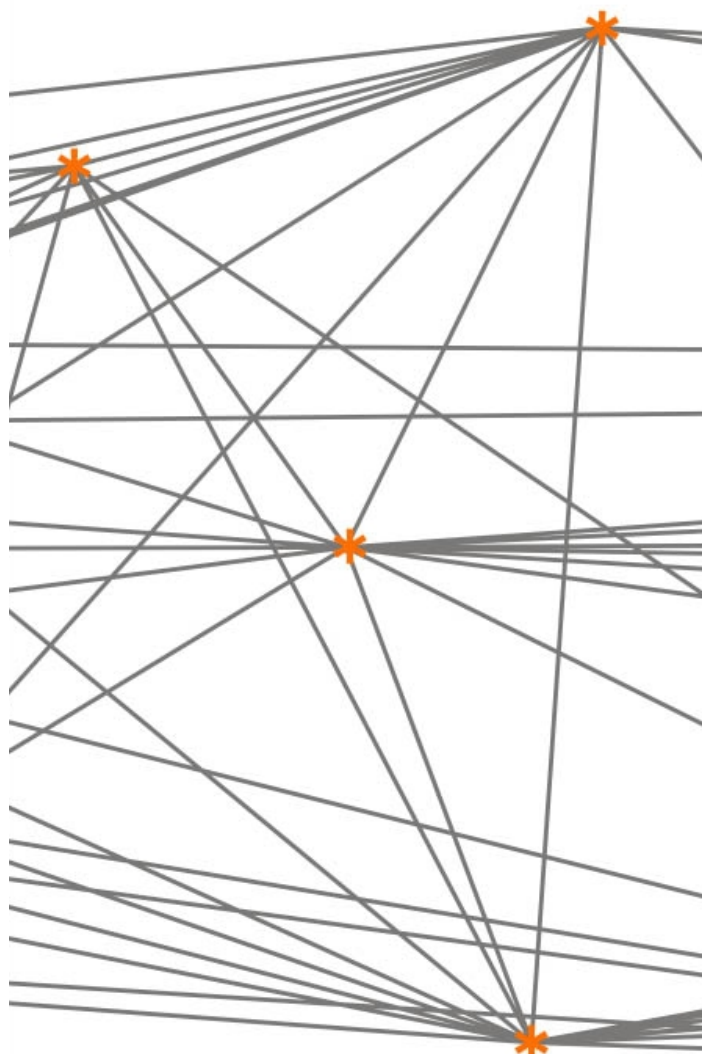




# Guide to *easy*SHARE

release 8.0.0



February 10<sup>th</sup>, 2022

*Please note*

*easySHARE* is intended for student training and teaching purposes. For scientific publications we recommend using the main data set of SHARE, or to carefully study this documentation and the Stata program that extracts and generates *easySHARE* from the main release of SHARE.

*easySHARE* release 8.0.0 is registered as doi: 10.6103/SHARE.easy.800

## Content

1. Introduction .....	3
2. Access .....	3
3. Selection of variables .....	4
4. Structure of the data .....	5
5. Description of variables .....	7
Identifiers and basic demographic information .....	8
Household composition .....	12
Social Support & Network .....	13
Childhood Conditions .....	15
Health .....	18
Cognitive Function Indices .....	23
Behavioral Risks .....	24
Personality .....	25
Work and Money .....	26
6. Examples of how to analyze <i>easySHARE</i> .....	29
... using Stata .....	29
... using SPSS .....	34
... using R .....	38
7. Appendix A: List of variables .....	43

## 1. Introduction

The multitude of countries participating in SHARE with their institutional variety and different languages, combined with a large interdisciplinary set of variables stored in about 25 modules per wave, and the differentiation into individual-, couple- and household-level data, make SHARE a complex data set. The easySHARE-file is a simplified data set for researchers who are less experienced in the quantitative analysis of complex panel data. Its main purpose is student training and teaching.

easySHARE includes the same number of observations as the main release of SHARE but is restricted to a subset of variables. This allows storing easySHARE as one single file without the need for complex merging of waves and modules. easySHARE contains all waves of SHARE and includes variables out of a variety of the SHARE CAPI-modules (CAPI stands for *Computer Assisted Personal Interview*) and in some instances the drop-off questionnaires that collect additional information via self-completion of a paper & pencil questionnaire. We also incorporated variables out of the generated variable data sets or created additional ones. Furthermore, adaptations in easySHARE make the file more comparable with the US Health and Retirement Study (HRS).

To facilitate analyzing the data we transferred the information that was collected on couple- and household-level to the individual level. Several generated variables, such as indexes and recoded health, demographic, social and economic measures allow direct analyses without the need for extensive data preparations. Furthermore, we added a series of additional missing value codes. These can help understanding how missing information evolved. E.g. we added extra missing value codes to tag observations that are missing due to questionnaire filtering.

This documentation gives a quick overview on how we constructed easySHARE and describes the included variables. It also contains examples of how to work with and analyze easySHARE.

## 2. Access

After completing a user statement, every person with scientific affiliation can download the data for free as long as the data are used for no other than scientific purposes. For teaching purposes teachers are asked to complete a teacher statement that is available at [http://www.share-project.org/fileadmin/pdf\\_documentation/easySHARE\\_Teacher\\_Statement.pdf](http://www.share-project.org/fileadmin/pdf_documentation/easySHARE_Teacher_Statement.pdf).

For more information on data access and data download please see <http://www.share-project.org/data-access/user-registration.html>

In case easySHARE data is being used for theses or other scientific publications, besides the mandatory SHARE acknowledgement, the following additional acknowledgement has to be included:

"This paper uses data from the generated *easySHARE* data set (DOI: 10.6103/SHARE.easy.800), see Gruber et al. (2014) for methodological details. The *easySHARE* release 8.8.0 is based on SHARE Waves 1, 2, 3, 4, 5, 6, 7 and 8 (DOIs: 10.6103/SHARE.w1.800, 10.6103/SHARE.w2.800, 10.6103/SHARE.w3.800, 10.6103/SHARE.w4.800, 10.6103/SHARE.w5.800, 10.6103/SHARE.w6.800, 10.6103/SHARE.w7.800, 10.6103/SHARE.w8.800)."

Please cite the dataset in your references as follows:

Börsch-Supan, A. & S. Gruber (2022): *easySHARE*. Release version: 8.0.0. SHARE-ERIC. Dataset. doi: 10.6103/SHARE.easy.800.

Further information on how to expand the *easySHARE* dataset with other variables from the scientific release of SHARE is available in the [SHARE Working Paper 37-2019](#) (Gruber 2019).

### 3. Selection of variables

Central variables have been identified based on the existing modules and comparisons of SHARE and RAND-HRS. This renders variables suitable for comparisons and cross national analysis, particularly with HRS. Variables have been selected to cover each module and topic of SHARE. Problems and requests by researchers gathered from symposiums and meetings, collaborations with researchers and user questions served as a basis for the selection and recoding of variables. Variables have been selected according to the following guidelines:

- low rate of missing information
- cross-country comparability
- maximum cross-wave comparability, i.e. longitudinal measures
- indices instead of variables
- no complex filtering or documentation of filters in target variable

The following topics and variables have been selected for *easySHARE*:

- 1) **Demographics:** age, gender, country of birth, citizenship, education, religion, marital status, age and gender of partner
- 2) **Household composition:** living with partner in the same household, household size, children living in the household
- 3) **Social support & network:** mother/father alive, number of children, residential proximity of children, number of grandchildren, number of living siblings, received and given social support
- 4) **Childhood conditions:** number of books at age ten, relative mathematical skills at age ten, relative language skills at age ten, childhood health status, being vaccinated during childhood
- 5) **Health and health behavior:** self-perceived health, number of chronic diseases, mental health variables, depression scale EURO-D, CASP-12 index for quality of life and well-being, health care utilization, grip strength, body mass index, smoking and drinking behavior, vigorous activities/sports

- 6) **Functional limitation indices:** mobility index, large muscle index, activities of daily living index, gross motor skills index, fine motor skills index, instrumental activities of daily living index, cognitive functions
- 7) **Work & money:** current job situation, term of main job, working hours per week, satisfaction with main job, early retirement plans, able to make ends meet, imputed household net income, wave specific household income percentiles

#### 4. Structure of the data

Panel data like *easySHARE* can be displayed in long or in wide format. The *easySHARE* data are stored in long format, i.e. observations in lines are the respondents, whereas respondents are included as often as they participated in the currently available five waves of SHARE. The variables or columns store reported information. If there are five observations with the same respondent identifier (`mergeid`), this respondent took part in all five waves of SHARE. One advantage of long format is that this form is generally expected for panel analysis.<sup>1</sup>

Table 1 shows the structure of the data for two example respondents. The first respondent was interviewed in Austria which can be derived from the prefix "AT" in `mergeid` or from the country variables (`country` and `country_mod`). This respondent took part in all seven waves as represented by one line per wave. The second respondent who was interviewed in Sweden (prefix "SE") participated in wave 1, did not participate in wave 2 and wave 3, and then took part again in wave 4 and 5, but not in the later waves. For this respondent the data contain only three observations. The information on respondents' wave participation is stored in the variable `wavepart`. This variable can be used to select a balanced panel, e.g. of respondents who participated in all waves.

*Table 4.1: Data structure*

<code>mergeid</code>	<code>wave</code>	<code>birth_country</code>	<code>wavepart</code>
AT-986403-01	1	Austria	12345678
AT-986403-01	2	Austria	12345678
AT-986403-01	3	Austria	12345678
AT-986403-01	4	Austria	12345678
AT-986403-01	5	Austria	12345678
AT-986403-01	6	Austria	12345678
AT-986403-01	7	Austria	12345678
AT-986403-01	8	Austria	12345678
SE-209636-01	1	Sweden	145
SE-209636-01	4	Sweden	145
SE-209636-01	5	Sweden	145

<sup>1</sup> You can easily „reshape“ the data to the wide panel format. Please find code examples in the section on “Examples of how to analyze *easySHARE*”.

## Missing codes

Compared to the SHARE main release we recoded the missing values in *easySHARE* to one of the following (partially new) codes:

- 3: "implausible value/suspected wrong"
- 7: "not yet coded"
- 9: "not applicable filtered"
- 10: "SHARELIFE interview" (only in wave 7)
- 11: "regular interview" (only in wave 7)
- 12: "don't know / refusal"
- 13: "not asked in this wave"
- 14: "not asked in this country"
- 15: "no information"
- 16: "no drop-off (information in drop-off in this wave)"

There is only one combined missing value code for "don't know" and "refusal" in *easySHARE*. The reason is that when combining two or more variables of the main release of SHARE into a new *easySHARE* variable, we sometimes cannot distinguish between the possibly different missing codes of the original variables. For consistency within *easySHARE* we use the combined "don't know / refusal" code in all variables. The code -15 "no information" is used for missing values that are not explained by filtering or questionnaire design. There are multiple reasons for this code, for example respondents may not have completed the whole interview.

Not every question was asked in all available waves of SHARE. For that reason we implemented the missing code -13: "not asked in this wave". Note that the third wave on people's life histories (SHARELIFE) collected very different information compared to the regular waves in structure and content. This is the reason for many variables set to the -13 code in wave 3. The Wave 7 questionnaire contains a SHARELIFE module for all respondents who did not participate in Wave 3, as well as regular modules for all respondents who already answered a SHARELIFE interview. Therefore, many regular panel variables in *easySHARE* contain the missing code -10 "SHARELIFE interview" in wave 7.

## 5. Description of variables

Some variables in *easySHARE* are copies of the respective variables included in the main release of SHARE that were only complemented by recoding the system missing values due to country-/wave-missing patterns and questionnaire filtering. For these variables the variable name in *easySHARE* remains the same as in the main release of SHARE.

Variables that have been modified beyond the above mentioned or have been specifically generated for *easySHARE* receive new or modified variable names. New indices or scores are named according to conventions (i.e. CASP score, IADLA), generated variables based on several other variables are named similarly to the variable names in the main of SHARE. Variables that have been modified, but do not represent a new concept, got an additional "\_mod" identifier at the end of the variable name.

The overview of variables included in *easySHARE* in the next sections is intended to give a brief overview on the variable generation, the coding format and modifications implemented in *easySHARE*. Appendix A provides a short overview table of all variables included.

## Identifiers and basic demographic information

<b><i>mergeid</i></b>	Person identifier	example "AT-123456-01"
-----------------------	-------------------	------------------------

As in the main release of SHARE *mergeid* is the person identifier. It is unique for each respondent and does not change across waves. It consists of the country abbreviation, a six digit household identifier that is assigned when the household is first sampled and a two digit person identifier.

<b><i>hhid</i></b>	Household identifier	example "AT-123456-A"
--------------------	----------------------	-----------------------

Respondents living in the same household can be identified by *hhid*. The format is similar to *mergeid* (and *coupleid*). For respondents may move out of the household they were originally sampled in, the *hhid* of an individual can change across waves. One part of the household that split gets the appendix "-B" instead of "-A".

Note: While in the coverscreen files (*cv\_r*) of the SHARE main release all household members are represented by a data line, the *easySHARE* data set only stores the observations of responding household members.

<b><i>coupleid</i></b>	Wave specific couple identifier	example "AT-123456-01-02"
------------------------	---------------------------------	---------------------------

*coupleid* identifies couples within a household. Two respondents living together in the same household as a couple (regardless of their legal status) get assigned the same *coupleid*.

Note: As non-responding partners are not included in *easySHARE* there is not always a corresponding observation for every respondent with a non-missing *coupleid*. For individuals without partner or spouse in the household *coupleid* is set to plain missing.

<b><i>wave</i></b>	Wave identifier
--------------------	-----------------

*easySHARE* contains all waves of SHARE. If you want to select a specific wave, the *wave* variable is the easiest way to do so. It is also needed to set up the data for longitudinal analyses.

<b><i>wavepart</i></b>	Wave participation pattern
------------------------	----------------------------

*wavepart* stores information on the wave participation pattern of each respondent. A respondent who participated in all eight waves of SHARE gets the value "12345678", while a respondent who skipped waves 2 and 3 would get the *wavepart* value "145678". Hence, this variable can be used to select any kinds of balanced panels, e.g. of respondents who participated in all waves. The same *wavepart* value is assigned to every single line of the respective respondent.



<b><i>int_version</i></b>	Interview version (baseline or longitudinal)
---------------------------	--

*int\_version* contains the information whether a baseline or a longitudinal interview was conducted

<b><i>int_year</i></b>	Year of interview
------------------------	-------------------

*int\_year* contains the year in which the interview took place. In case the year of interview was missing we replaced the value by the country and wave specific modus year.

<b><i>int_month</i></b>	Month of interview
-------------------------	--------------------

This variable contains the month in which the interview took place. In case the month of interview was missing we replaced the value by the country and wave specific modus (in case the modus is not unique, we used the minimum month). The main reason to impute month and year of interview is to allow computing the age at interview variable for as many respondents as possible.

<b><i>country</i></b>	Country identifier
-----------------------	--------------------

All countries that participated in at least one of the seven waves are included in *easySHARE*: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden and Switzerland. The coding of *country* is the same as in the main release of SHARE.

<b><i>country_mod</i></b>	Modified country identifier
---------------------------	-----------------------------

*country\_mod* is another country identifier. In contrast to *country* it is coded according to the ISO code scheme. Using the same ISO codes as in *birth\_country* and *citizenship* allows combining the information more easily.

<b><i>language</i></b>	Language of questionnaire
------------------------	---------------------------

This variable indicates the language in which the interview was conducted. In some of the SHARE countries (like Switzerland or Israel) there is more than one.

<b><i>female</i></b>	Gender of respondent
----------------------	----------------------

*female* is the "dummy"-style coded gender variable. 0: male; 1: female. For very few respondents gender varies between waves. These are mostly cases that have not been fully checked yet. There is no way to determine the correct information, therefore we coded these respondents' gender to -3 "implausible value" in all instances.

<b>dn002_mod</b>	Month of birth
------------------	----------------

dn002\_mod stores the respondents' month of birth. The original variable dn002\_ differs for quite some respondents between waves. Therefore, easySHARE stores the minimum mode month of birth in case of deviations within a respondent.

<b>dn003_mod</b>	Year of birth
------------------	---------------

dn003\_mod indicates the respondents' year of birth. There are very few respondents for whom year of birth differs between waves. These are mostly cases that have not been fully checked yet. For year of birth is a crucial information and there is no way to determine the correct information, we recode dn003\_mod to -3 "implausible value" for these respondents.

<b>dn004_mod</b>	Born in the country of interview
------------------	----------------------------------

dn004\_mod indicates whether the respondent was born in the country of interview. Hereby first generation migrants can be identified.

<b>age</b>	Age at interview
------------	------------------

age at the time of interview is calculated on the basis of the respondent's month and year of birth (dn002\_mod and dn003\_mod) and month and year of interview (int\_year and int\_month). The month-exact age is divided by 12 to produce age in years.

*Note:* To reduce the number of missing values on this often used variable we imputed int\_year and int\_month if it was missing and we also imputed the month of birth variable (dn002\_) if it was missing. However we did not impute missing year of birth information. This explains why there are still few respondents with missing age information.

<b>birth_country</b>	Country of birth
----------------------	------------------

The variable combines information on whether respondents are born in the (current) country (dn004\_ in the SHARE main release) and the follow-up question in which country they were born (dn005c in the SHARE main release). We used the same ISO coding as in the country\_mod variable.

Using the integrated birth\_country variable and country\_mod you can easily generate a variable telling if the individual was born in the country of interview.

<b>citizenship</b>	Citizenship of respondent
--------------------	---------------------------

citizenship is generated based on the question whether the respondent has the citizenship of the country of interview and if not, which citizenship the respondent has (dn007\_ and dn008c in the SHARE main release).

<b><i>iv009_mod</i></b>	Area of location
-------------------------	------------------

*iv009\_mod* contains information on the area of the building where the interview took place. It is filtered if the interview was not conducted in the respondents' home (information taken from *ho001\_*). Additionally, information was transferred from the household respondent to the other household members.

- 1: A big city
- 2: The suburbs or outskirts of a big city
- 3: A large town
- 4: A small town
- 5: A rural area or village

<b><i>q34_re</i></b>	Religious denomination [only included in wave 1]
----------------------	--

The information about the respondents' religious denomination was part of the drop off questionnaire in wave 1. This question was not included in all later waves.

- 1: Protestant
- 2: Catholic
- 3: Orthodox
- 4: Jewish
- 5: Muslim
- 7: other
- 8: none

<b><i>isced1997_r</i></b>	ISCED classification
---------------------------	----------------------

*isced1997\_r* stores the ISCED-97 coding of education. The values are:

- 0: none
- 1: Level 1 – Primary education or first stage of basic education
- 2: Level 2 – Lower secondary or second stage of basic education
- 3: Level 3 – (Upper) secondary education
- 4: Level 4 – Post-secondary non-tertiary education
- 5: Level 5 – First stage of tertiary education
- 6: Level 6 – Second stage of tertiary education
- 95: still in school
- 96: other

<b><i>eduyears_mod</i></b>	Years of education
----------------------------	--------------------

Years of education are only asked in the baseline interview of each respondent starting in wave 2. Because this question was not asked in wave 1 and 3 the answers collected in waves 2, 4, 5, 6, 7, and 8 were assigned. This leads to relatively high amounts of missing values especially for respondents who were not re-interviewed after their wave 1 interview.

<b><i>mar_stat</i></b>	Marital status [not included in wave 3]
------------------------	---

The marital status is only asked in the baseline interview of each respondent (dn014\_). In later regular waves the respondents are asked if their status has changed or not (dn044). If it has not changed we forwarded the information of dn014\_ to later waves. If it has changed, respondents are asked again for their current status.

- 0: none
- 1: married and living together with spouse
- 2: registered partnership
- 3: married, living separated from spouse
- 4: never married
- 5: divorced
- 6: widowed

## Household composition

<b><i>hhsiz</i></b>	Household size
---------------------	----------------

*hhsiz* contains information about the number of people living in the respondents' household.

<b><i>partnerinh</i></b>	Living with spouse/partner
--------------------------	----------------------------

- 1: living with a spouse/partner in household
- 3: living without spouse/partner in household
- 97: other

<b><i>int_partner</i></b>	Interview of partner available
---------------------------	--------------------------------

*int\_partner* indicates whether an interview of the respondent's partner is available. The variable is filtered when the respondent is living as single or when the partner does not live in the same household.

- 1: yes
- 5: no

<b><i>age_partner</i></b>	Age at interview of respondent's partner
---------------------------	--

*age\_partner* is assigned from the age of a respondent's partner using the wave specific *coupleid*. If they have an interview we use the partners self-report, otherwise *age\_partner* is based on the *cv\_r* information. Hence, for the age of the partner the same imputations were applied as for *age* of respondent, if it is based on self-report information. The variable is filtered when the respondent is living as single or when the partner does not live in the same household.

<b><i>gender_partner</i></b>	Gender of respondent's partner
------------------------------	--------------------------------

*gender\_partner* is based on coverscreen information. The variable is filtered when the respondent is living as single or when the partner does not live in the same household.

- 0: male
- 1: female

## Social Support & Network [not included in wave3]

<b><i>mother_alive</i></b>	Is natural parent alive: mother
----------------------------	---------------------------------

*mother\_alive* is based on *dn026\_1* of the SHARE main release and includes information on whether the respondent's mother is still alive. If available we combined it with information from previous waves and – in waves 4 and 6 – information from the social networks module. Latter is based on the assumption that persons belonging to the respondent's social network are still alive.

- 1: yes
- 5: no

<b><i>father_alive</i></b>	Is natural parent alive: father
----------------------------	---------------------------------

*father\_alive* is based on *dn026\_2* of the SHARE main release and includes information on whether the respondent's father is still alive. If available we combined it with information from previous waves and – in waves 4 and 6 – information from the social networks module. Latter is based on the assumption that persons belonging to the respondent's social network are still alive.

- 1: yes
- 5: no

<b><i>siblings_alive</i></b>	Number of siblings alive
------------------------------	--------------------------

*siblings\_alive* counts the total number of siblings alive (sum of *dn036\_* and *dn037\_* in the SHARE main release). It is zero for respondents who had siblings but none of them is still alive and is filtered if a respondent never had siblings.

<b><i>ch001_</i></b>	Number of children
----------------------	--------------------

*ch001\_* contains the number of children that are still alive including natural children, fostered, adopted and stepchildren. We do not integrate the SHARELIFE information on children from waves 3 and 7, for the concept of collecting information is considerably different. In SHARE, information on children is asked to one respondent per couple only, but transferred to partners/spouses in the easySHARE data set.

<b>ch021_mod</b>	Number of grandchildren
------------------	-------------------------

ch021\_mod contains the number of the respondent's grandchildren. Grandchildren of spouse or partner from previous relationships are included in this variable.

<b>ch007_hh</b>	At least one child in same household
-----------------	--------------------------------------

ch007\_hh indicates whether at least one child lives in the same household or the same building as the respondent (based on ch007\_1 to ch007\_16 and ch526\_1 to ch526\_20 in the SHARE main release; in waves 4 and 6 sn006\_1 to sn006\_7 were additionally used). It is filtered if respondents do not have children.

- 1: yes
- 5: no

<b>ch007_km</b>	Residential proximity of children
-----------------	-----------------------------------

ch007\_km indicates whether at least one child lives less than one kilometer away from the respondent's household (based on ch007\_1 to ch007\_16 and ch526\_1 to ch526\_20 in the SHARE main release; in waves 4, 6 and 8 sn006\_1 to sn006\_7 were additionally used).

- 1: yes
- 5: no

<b>sp002_mod</b>	Received help from outside the household
------------------	--

sp002\_mod contains information on whether the respondents (or if applicable their partners) have received help from any family member from outside the household, any friend or neighbor. Help in this context incorporates personal care (e.g. help with dressing, bathing, eating, getting out of bed, using the toilet), practical household help and help with paper work such as settling financial or legal matters. The question was only answered by the family respondent. We assigned the information also to the partners.

- 1: yes
- 5: no

<b>sp003_1_mod - sp003_3_mod</b>	Who gave help
----------------------------------	---------------

Respondents can name up to three persons outside the household from whom they (or their partners) received help. sp003\_1\_mod contains the first person mentioned by the respondent, sp003\_2\_mod the second and sp003\_3\_mod the third.

<b>sp008_</b>	Given help to others outside the household
---------------	--

sp008\_ contains information on whether the respondent has personally given help to any family member from outside the household, any friend or neighbor. Help in this context incorporates personal care (e.g. help with dressing, bathing, eating, getting out of bed, using the toilet), practical household help and help with paper work such as settling financial or legal matters.

- 1: yes
- 5: no

<b>sp009_1_mod - sp009_3_mod</b>	To whom did you give help
----------------------------------	---------------------------

Respondents can name up to three persons outside the household to whom they gave help. sp009\_1\_mod contains the first person mentioned by the respondent, sp009\_2\_mod the second and sp009\_3\_mod the third. In contrast to received help, the family respondent did not answer the questions on given help on behalf of the couple from wave 4 onwards. This explains the higher amount of missings.

### Childhood Conditions [wave 3 and wave 5 only]

<b>books_age10</b>	Number of books when ten
--------------------	--------------------------

This variable includes information about the number of books that existed in the place the respondents lived in when they were ten years old. Magazines, newspapers, or school books are excluded.

- 1: none or very few (0-10 books)
- 2: enough to fill one shelf (11-25 books)
- 3: enough to fill one bookcase (26-100 books)
- 4: enough to fill two bookcases (101-200 books)
- 5: enough to fill two or more bookcases (more than 200 books)

<b>maths_age10</b>	Relative mathematical skills when ten
--------------------	---------------------------------------

maths\_age10 is about the performance in mathematics compared to other children when the respondent was at the age of ten. It is filtered if the respondent did not go to school.

- 1: much better
- 2: better
- 3: about the same
- 4: worse
- 5: much worse

<b><i>language_age10</i></b>	Relative language skills when ten
------------------------------	-----------------------------------

language\_age10 asks for relative language skills compared to other children when the respondent was at the age of ten. It is filtered if the respondent did not go to school.

- 1: much better
- 2: better
- 3: about the same
- 4: worse
- 5: much worse

<b><i>vaccinated</i></b>	Vaccinations during childhood
--------------------------	-------------------------------

The question "During your childhood, that is, before you turned 16, have you received any vaccinations?" was part of waves 3, 5 and 7.

- 1: Yes
- 5: No

<b><i>childhood_health</i></b>	Childhood health status
--------------------------------	-------------------------

This variable on childhood health was part of waves 3, 5 and 7. It does not refer to a specific age but the general health status throughout the entire childhood.

- 1: Excellent
- 2: Very Good
- 3: Good
- 4: Fair
- 5: Poor
- 6: Health varied a great deal

## Health: Physical Health

<b><i>sphus</i></b>	Self-perceived health (US version)
---------------------	------------------------------------

In wave 1 respondents were randomized to answer the self-perceived health item either at the beginning (ph003\_) or at the end of the physical health questionnaire module (ph052\_). sphus combines both in one variable.

- 1: excellent
- 2: very good
- 3: good
- 4: fair
- 5: poor



<b><i>chronic_mod</i></b>	Number of chronic diseases
---------------------------	----------------------------

This variable is based on the multiple answer question ph006\_ that asks which of the listed chronic conditions the respondents had according to their doctors (“Has a doctor ever told you that you had ...”). The question was asked in waves 1, 2, 4, 5, 6, 7, and 8. However, the answer options have been modified between waves. Therefore, we created a new index called `chronic_mod` that only counts the conditions that were included as answer options in each wave. For the list of answer options varies, the “other condition” option does change its meaning between waves and hence is not taken into account for `chronic_mod`. The following list shows the conditions included in the index:

1. A heart attack
2. High blood pressure or hypertension
3. High blood cholesterol
4. A stroke or cerebral vascular disease
5. Diabetes or high blood sugar
6. Chronic lung disease (...)
10. Cancer or malignant tumor (...)
11. Stomach or duodenal ulcer, peptic ulcer
12. Parkinson disease
13. Cataracts
14. Hip fracture or femoral fracture

The following answer options are *not* included in `chronic_mod`:

- |                             |                           |               |
|-----------------------------|---------------------------|---------------|
| 7. Asthma                   | : only available in waves | 1, 2          |
| 8. Arthritis                | : only available in waves | 1, 2, 4       |
| 9. Osteoporosis             | : only available in waves | 1, 2          |
| 15. Other fractures         | : only available in wave  | 4, 5, 6, 7, 8 |
| 16. Alzheimer’s (...)       | : only available in wave  | 4, 5, 6, 7, 8 |
| 17. Benign tumor (...)      | : only available in wave  | 2             |
| 18. Other affective (...)   | : only available in wave  | 5, 6, 7, 8    |
| 19. Rheumatoid Arthritis    | : only available in wave  | 5, 6, 7, 8    |
| 20. Osteoarthritis (...)    | : only available in wave  | 5, 6, 7, 8    |
| 21. Chronic kidney disease: | only available in wave    | 6, 7, 8       |

*Difference to similar variables provided in SHARE main release:* The variables `chronicw1`, `chronicw2`, `chronicw4`, `chronicw5`, `chronicw6`, `chronicw7`, and `chronicw8` in the main release are computed based on all answer options provided in the respective wave, including the “other condition” option.

## Health: Mental health variables

<b><i>casp</i></b>	CASP-12 score (Quality of life)
--------------------	---------------------------------

The CASP-12 score measures quality of life and is based on four subscales on control, autonomy, pleasure and self-realization. The CASP score is the sum of these four subscales and ranges from 12 to 48. The items were asked in the drop off questionnaire in wave 1 and then moved into the CAPI from wave 2 onwards. For drop off questionnaires not being filled in by all respondents there is a higher fraction of missing observations in wave 1.

<b><i>euro1</i></b>	Depression (part of EURO-D)
---------------------	-----------------------------

*euro1* is taken from the *gv\_health* module. It is based on *mh002\_* in the mental health module in the SHARE main release. It indicates whether the respondent has been sad or depressed in the last month.

- 0: not selected
- 1: selected

<b><i>euro2</i></b>	Pessimism (part of EURO-D)
---------------------	----------------------------

*euro2* refers to the respondent's hopes for the future. It is based on *mh003\_*.

- 0: not selected
- 1: selected

<b><i>euro3</i></b>	Suicidality (part of EURO-D)
---------------------	------------------------------

This variable gives information on suicidal feelings based on *mh004\_*. The question is: "In the last month, have you felt that you would rather be dead?"

- 0: not selected
- 1: selected

<b><i>euro4</i></b>	Guilt (part of EURO-D)
---------------------	------------------------

*euro4* is based on *mh005\_* in the mental health module in the SHARE main release. It indicates whether the respondents tend to blame themselves or feel guilty about anything.

- 0: not selected
- 1: selected

<b><i>euro5</i></b>	Sleep (part of EURO-D)
---------------------	------------------------

Whether a person has trouble with sleeping is reported in *euro5* respectively *mh007\_*.

- 0: not selected
- 1: selected

<b>euro6</b>	Interest (part of EURO-D)
--------------	---------------------------

euro6 shows changes in the general interest in things (mh008\_ in the mental health module).

- 0: not selected
- 1: selected

<b>euro7</b>	Irritability (part of EURO-D)
--------------	-------------------------------

"Have you been irritable recently" is the question behind euro7 (mh010\_ in the mental health module).

- 0: not selected
- 1: selected

<b>euro8</b>	Appetite (part of EURO-D)
--------------	---------------------------

Changes in the respondent's appetite are shown in this variable that is based on mh011\_.

- 0: not selected
- 1: selected

<b>euro9</b>	Fatigue (part of EURO-D)
--------------	--------------------------

This variable is based on mh013\_ that asks whether a respondent had too little energy to do the things she/he wanted to do in the previous month.

- 0: not selected
- 1: selected

<b>euro10</b>	Concentration (part of EURO-D)
---------------	--------------------------------

Information on difficulties with the concentration on a television program, film, radio program or reading is given in euro10. This variable combines mh014\_ and mh015\_ of the mental health module in the SHARE main release.

- 0: not selected
- 1: selected

<b>euro11</b>	Enjoyment (part of EURO-D)
---------------	----------------------------

"What have you enjoyed doing recently?" is the question behind mh016\_ which is the basis for euro11.

- 0: not selected
- 1: selected

<b>euro12</b>	Tearfulness (part of EURO-D)
---------------	------------------------------

mh017\_ asks whether the respondent has cried at all in the last month? Taken from the gv\_health module this information is stored in euro12.

- 0: not selected
- 1: selected

<b>eurod</b>	Depression scale EURO-D
--------------	-------------------------

The EURO-D symptom scale measures the current depression and is constructed from questions in the mental health module (mh002\_ – mh017\_) as a composite index of twelve items: depressed mood, pessimism, suicidality, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness. The scale ranges from 0 “not depressed” to 12 “very depressed”.

Health: Health care

<b>hc002_mod</b>	Doctor visits
------------------	---------------

hc002\_mod contains the number of doctor visits within the past year. The question behind this variable is: “During the last twelve months, about how many times in total have you seen or talked to a medical doctor about your health?” (question text of wave 4 questionnaire). Please note that hc002\_mod is based on hc602\_ for wave 6 due to a slightly different question text.

<b>hc012_</b>	Hospital stay
---------------	---------------

hc012\_ indicates whether respondents have been in a medical, surgical, psychiatric or any other specialized hospital overnight during the last twelve months.

- 1: yes
- 5: no

<b>hc029_</b>	Nursing home
---------------	--------------

Overnight stays in a nursing home during the last twelve months are indicated by hc029\_.

- 1: yes, temporarily
- 3: yes, permanently
- 5: no

## Health: Physical measurements

<b>maxgrip</b>	Maximum of grip strength measure (from gv_health)
----------------	---

According to instructions two grip strength measurements on each hand were recorded using a so-called dynamometer. The variable `maxgrip` is defined as the maximum grip strength measurement of both hands (2x2) or of one hand (1x2).

Valid measurements are defined when the two measurements of one hand do not differ by more than 20kg. If the difference was above that limit (>20kg) the measurements for that hand were recoded as MISSING. If grip strength was only measured once on one hand, this measurement has also been recoded as MISSING. However, if there were two measurements on the other hand, these measurements have been included. Grip strength measurements of zero "0" or above 100 kg ( $\geq 100$  kg) have been recoded as MISSING.

## Health: Functional limitation indices

Although SHARE provides a huge variety of indices on physical and mental health as well as health behavior, some of them are not directly comparable with US-measures and those in RAND HRS. Since one of the aims of easySHARE is to increase comparability with HRS and the use of SHARE outside of Europe, some new indices on health were recoded for easySHARE. If desired, additional indices provided in the SHARE main release files can be merged.

<b>adlwa</b>	Activities of Daily Living w&h Index
--------------	--------------------------------------

`adlwa` is the sum of the three tasks: dressing (`ph049d1`), bathing or showering (`ph049d3`) and eating/cutting up food (`ph049d4`) (*Wallace and Herzog*). The higher the index is the more difficulties with these activities and the lower the respondent's mobility. `adlwa` ranges from 0 to 3.

<b>adla</b>	Activities of Daily Living Index
-------------	----------------------------------

`adla` is the sum of the five tasks dressing (`ph049d1`), bathing or showering (`ph049d3`), eating, cutting up food (`ph049d4`), walking across a room (`ph049d2`) and getting in or out of bed (`ph049d5`). The higher the index is the more difficulties with these activities and the lower the mobility of the respondent. `adla` ranges from 0 to 5.

<b>iadla</b>	Instrumental Activities of Daily Living Indices
--------------	---

`iadla` is the sum of telephone calls (`ph049d10`), taking medications (`ph049d11`) and managing money (`ph049d13`). The higher the index is the more difficulties

with the activities and the lower the mobility of the respondent. *iadla* ranges from 0 to 3.

<b><i>iadlza</i></b>	Instrumental Activities of Daily Living Indices
----------------------	---

*iadlza* is the sum of telephone calls (*ph049d10*), taking medications (*ph049d11*), managing money (*ph049d13*), shopping for groceries (*ph049d9*) and preparing a hot meal (*ph049d8*). The higher the index is the more difficulties with the activities and the lower the mobility of the respondent. *iadlza* ranges from 0 to 5.

<b><i>mobilityind</i></b>	Mobility Index
---------------------------	----------------

*mobilityind* is the sum of walking 100 meters (*ph048d1*), walking across a room (*ph049d2*), climbing several flights of stairs (*ph048d4*) and climbing one flight of stairs (*ph048d5*). The higher the index, the more difficulties with these activities exist and the lower the mobility of the respondent. It ranges is from 0 to 4.

<b><i>lgmuscle</i></b>	Large Muscle Index
------------------------	--------------------

*lgmuscle* is the sum of sitting two hours (*ph048d2*), getting up from chair (*ph048d3*), stooping, kneeling, crouching (*ph048d6*) and pulling or pushing large objects (*ph048d8*). The higher the index, the more difficulties with the activities and the lower the mobility of the respondent. *lgmuscle* ranges from 0 to 4.

<b><i>grossmotor</i></b>	Gross Motor Skills Index
--------------------------	--------------------------

*grossmotor* is the sum of walking 100 meters (*ph048d1*), walking across a room (*ph049d2*), climbing one flight of stairs (*ph048d5*) and bathing or showering (*ph049d3*). The higher the index is the more difficulties with the activities and the lower the gross motor skills of the respondent. *grossmotor* ranges from 0 to 4.

<b><i>finemotor</i></b>	Fine Motor Skills Index
-------------------------	-------------------------

*finemotor* is the sum of picking up a small coin (*ph048d10*), eating/cutting up food (*ph049d4*) and dressing (*ph049d1*). The higher the index is the more difficulties with the activities and the lower the fine motor skills of the respondent. *finemotor* ranges from 0 to 3.

## Cognitive Function Indices

<b><i>recall_1</i></b>	Recall of words, first trial
------------------------	------------------------------

`recall_1` contains the number of words recalled in the first trial of the word recall task (`cf008tot` in the SHARE main release). It ranges from 0 to 10.

<b><i>recall_2</i></b>	Recall of words, first trial
------------------------	------------------------------

`recall_2` contains the number of words recalled in the delayed word recall task (`cf016tot` in the SHARE main release). It ranges from 0 to 10.

<b><i>orienti</i></b>	Orientation to date
-----------------------	---------------------

Orientation to date, month, year and day of week is also a generated variable from the `gv_health` module on the basis of `cf003` - `cf006`. `orienti` ranges from 0 (good) to 4 (bad). The high amount of missing values in waves 4, 5, 6, 7, and 8 is due to a routing filter: only baseline respondents get to the respective questions.

<b><i>numeracy_1</i></b>	Numeracy Score 1 (percentage)
--------------------------	-------------------------------

`numeracy_1` gives information on the Mathematical performance (percentage calculation) of the respondents. The variable is based on `cf012` - `cf015` respectively the corresponding variables in the `gv_health` module and ranges from 1 (bad) to 5 (good). The high amount of missing values in waves 4, 5, 6, 7, and 8 is due to a routing filter: only baseline respondents get to the respective questions. Respondents who participated before, got a new set of questions (`numeracy_2`).

<b><i>numeracy_2</i></b>	Numeracy Score 2 (subtraction)
--------------------------	--------------------------------

`numeracy_2` is the second test on Mathematical performance (subtraction). It was asked to respondents of waves 4, 5, 6, 7, and 8 who already participated in one of panel waves.

## Behavioral Risks

<b>bmi</b>	Body Mass Index
------------	-----------------

This variable is based on individual weight (`ph012_`) and height (`ph013_`), and is calculated according to the formula:  $BMI = (ph012_ / (ph013_ * 2) * 10\ 000$ . The variable is part of the `gv_health` module in the SHARE main release.

<b>bmi2</b>	Body Mass Index (categorized)
-------------	-------------------------------

In `bmi2` the values of `bmi` are summarized into the following categories:

- 1: below 18.5 – underweight
- 2: 18.5 - 24.9 – normal
- 3: 25-29.9 – overweight
- 4: 30 and above – obese

<b>smoking</b>	Smoke at present time
----------------	-----------------------

This variable is based on `br002_` and the respective filter question `br001_`. `smoking` indicates whether the respondents smoke at present. The high amount of -15 values in wave 6 results from a change in the routing of `br002_`.

- 1: yes
- 5: no

<b>ever_smoked</b>	Ever smoked daily
--------------------	-------------------

`ever_smoked` is based on `br001_` of the SHARE main release and indicates if a respondent ever smoked daily. Please be aware that `br001_` is only asked in the baseline interview. In the variable `ever_smoked` information was transferred to later waves.

- 1: yes
- 5: no

<b>br010_mod</b>	Drinking behavior
------------------	-------------------

`br010_mod` contains information on drinking behavior. Please note that in wave 1 the reference period is the last six months whereas in the subsequent waves the reference period is the last three months. The question is: "During the last six (three) months, how often have you drunk any alcoholic beverages, like beer, cider, wine, spirits or cocktails?"

- 1: not at all
- 2: less than once a month
- 3: once or twice a month
- 4: once or twice a week



- 5: three or four days a week
- 6: five or six days a week
- 7: almost every day

<b>br015_</b>	Vigorous activities
---------------	---------------------

br015\_ gives information on the frequency of doing vigorous activities such as sports, heavy housework, or a job that involves physical labor.

- 1: more than once a week
- 2: once a week
- 3: one to three times a month
- 4: hardly ever, or never

Personality: 10-item Big-Five inventory (BFI-10)

<b>bfi10_extra_mod</b>	Extraversion
------------------------	--------------

This generated variable is based on ac701\_ ("I see myself as someone who is reserved") and ac706\_ ("I see myself as someone who is outgoing, sociable").

- 1: low
- ...
- 5: high

<b>bfi10_agree_mod</b>	Agreeableness
------------------------	---------------

This generated variable is based on ac702\_ ("I see myself as someone who is generally trusting") and ac707\_ ("I see myself as someone who tends to find fault with others").

- 1: low
- ...
- 5: high

<b>bfi10_consc_mod</b>	Conscientiousness
------------------------	-------------------

This generated variable is based on ac703\_ ("I see myself as someone who tends to be lazy") and ac708\_ ("I see myself as someone who does a thorough job").

- 1: low
- ...
- 5: high

<b><i>bfi10_neuro_mod</i></b>	Neuroticism
-------------------------------	-------------

This generated variable is based on ac704\_ (“I see myself as someone who is relaxed, handles stress well”) and ac709\_ (“I see myself as someone who gets nervous easily”).

- 1: low
- ...
- 5: high

<b><i>bfi10_open_mod</i></b>	Openness
------------------------------	----------

This generated variable is based on ac705\_ (“I see myself as someone who has few artistic interests”) and ac710\_ (“I see myself as someone who has an active imagination”).

- 1: low
- ...
- 5: high

## Work and Money: Employment

<b><i>ep005_</i></b>	Current job situation
----------------------	-----------------------

“In general, how would you describe your current situation?” is the question behind ep005\_.

- 1: retired
- 2: employed or self-employed (including working for family business)
- 3: unemployed
- 4: permanently sick or disabled
- 5: homemaker
- 97: other

<b><i>ep009_mod</i></b>	Employee or self-employed in (main) job
-------------------------	---

ep009\_mod specifies the current job situation if the respondent answered in ep005\_ that she/he is employed. Note that the answer options in wave 5 are slightly different: 1. “private sector employee” and 2. “public sector employee”.

- 1: employee
- 2: civil servant
- 3: self-employed

<b>ep011_mod</b>	Term of (main) job
------------------	--------------------

ep011\_mod indicates whether the respondents have a short-term or a permanent contract. Short-term contract means less than 3 years here.

- 1: short-term
- 2: permanent

<b>ep013_mod</b>	Working hours per week
------------------	------------------------

The hours that the respondents usually work during a week are asked in ep013\_mod – regardless of their basic contracted hours. Meal breaks should be excluded and paid or unpaid overtime included.

<b>ep026_mod</b>	Satisfaction with (main) job
------------------	------------------------------

The question behind ep026\_mod is: “All things considered I am satisfied with my job. Would you say you strongly agree, agree, disagree or strongly disagree?”

- 1: strongly agree
- 2: agree
- 3: disagree
- 4: strongly disagree

<b>ep036_mod</b>	Look for early retirement in (main) job
------------------	---

The wish for early retirement is indicated by ep036\_mod: “Thinking about your present [main/secondary] job, would you like to retire as early as you can from this job?”

- 1: yes
- 5: no

## Work and Money: Household income

Note that the income variables of the SHARE main release are indicating *gross values* in wave 1 and *net values* in waves 2, 4, 5, 6, 7 and 8. The module gv\_grossnet in the SHARE main release makes it possible to generate net income also for wave 1.

<b>co007_</b>	Household able to make ends meet
---------------	----------------------------------

The question is: “Thinking of your household's total monthly income, would you say that your household is able to make ends meet...” This variable is taken from the consumption module and not from the imputed variables.

- 1: with great difficulty

- 2: with some difficulty
- 3: fairly easily
- 4: easily

<b><i>thinc_m</i></b>	Household net income, imputed
-----------------------	-------------------------------

The variable was generated using `gv_grossnet` for wave 1. For the other waves, it is based on the `thinc` variable of `gv_imputations`. In wave 7, it is only filled for respondents with a regular interview (-10 for SHARELIFE respondents of wave 7).

<b><i>income_pct_w1</i></b>	Household income percentiles for wave 1
-----------------------------	---

`income_pct_w1` contains the household income percentiles for respondents who participated in wave 1. The higher the percentile to which the household belongs, the higher is the household income.

<b><i>income_pct_w2</i></b>	Household income percentiles for wave 2
-----------------------------	---

`income_pct_w2` indicates the household income percentiles for respondents who participated in wave 2. For Ireland, `income_pct_w2` is missing (-7).

<b><i>income_pct_w4</i></b>	Household income percentiles for wave 4
-----------------------------	---

`income_pct_w4` indicates the household income percentiles for respondents who participated in wave 4.

<b><i>income_pct_w5</i></b>	Household income percentiles for wave 5
-----------------------------	---

`income_pct_w5` indicates the household income percentiles for respondents who participated in wave 5.

<b><i>income_pct_w6</i></b>	Household income percentiles for wave 6
-----------------------------	---

`income_pct_w6` indicates the household income percentiles for respondents who participated in wave 6.

<b><i>income_pct_w7</i></b>	Household income percentiles for wave 7
-----------------------------	---

`income_pct_w7` indicates the household income percentiles for respondents who participated in wave 7.

<b><i>income_pct_w8</i></b>	Household income percentiles for wave 8
-----------------------------	---

`income_pct_w8` indicates the household income percentiles for respondents who participated in wave 8.

## 6. Examples of how to analyze *easySHARE*

... using Stata<sup>2</sup>

After downloading and saving *easySHARE* you can load the data set into the Stata memory with the `use` command followed by the storage location of the data.

```
use D:\easySHARE\data\easyshare_rel8-0-0.dta, clear
```

First we reduce our data set to the key variables of the following exemplary analyses using the `keep` command.

```
keep mergeid wave int_year ep005_maxgrip
```

Panel data can be stored in “wide format” or in “long format”. *easySHARE* is provided in long format because this is the common format for panel analyses. In the long format, each observation stands for a respondent at a specific point in time, and the variables are their observed characteristics. Each row represents one point in time (or wave) per respondent. As *easySHARE* contains all waves of SHARE there is a maximum of eight rows per respondent. If a respondent participated in just one wave there is only one data line for this respondent.

In the wide format each observation stands for one respondent and the variables are their characteristics on the respective time points. There is only one row per respondent and each column represents one wave-specific variable.

*Figure 6.1: long vs. wide data format*

Long format			Wide format		
mergeid	wave	int_year	mergeid	int_year1	int_year2
AT-986403-01	1	2004	AT-986403-01	2004	2007
AT-986403-01	2	2007			

To reshape the data from long into wide format we use the `reshape` command. `mergeid` is the identifier variable (*i*) and the variable by which to organize the data is `wave` (*j*).

```
reshape wide int_year ep005_maxgrip, i(mergeid) j(wave)
```

The corresponding Stata output shows that in wide format `wave` is dropped and that the number of observations decreases whereby the number of variables increases. Variables are set automatically to system-missing if respondents did not participate in a wave. Using the wide format often facilitates computing difference scores between time points or to give descriptive information on the analysis sample.

---

<sup>2</sup> The example analysis is based on Stata version 14.

Figure 6.2: Stata output for reshaping the data from long to wide format

Data	long	->	wide
Number of obs.	412110	->	140125
Number of variables	5	->	25
j variable (8 values)	wave	->	(dropped)
xij variables:			
	int_year	->	int_year1 int_year2 ... int_year8
	ep005_	->	ep005_1 ep005_2 ... ep005_8
	maxgrip	->	maxgrip1 maxgrip2 ... maxgrip8

For our example analysis we need the data in long format. So we reshape the data back.

```
reshape long
```

This procedure leads to more observations than in the original data file because every respondent has one data line for each of the eight waves now. The variable values are set to system-missing for waves in which the respondent did not participate.

Figure 6.3: Stata output for reshaping the data from wide to long format

Data	wide	->	long
Number of obs.	140125	->	1.1e+06
Number of variables	25	->	5
j variable (8 values)		->	wave
xij variables:			
	int_year1 int_year2 ... int_year8	->	int_year
	ep005_1 ep005_2 ... ep005_8	->	ep005_
	maxgrip1 maxgrip2 ... maxgrip8	->	maxgrip

To get the original number of observations we drop the lines with missing information for the year of interview. Make sure you use a variable that had no missing values before the reshape for this step.

```
drop if int_year ==.
```

Before starting with the analysis we first look at the central variables. `maxgrip` is the maximum value of the grip strength test. It ranges from 1 to 99.

```
summarize maxgrip if maxgrip > 0
```

 Figure 6.4: Stata output for the summarizing `maxgrip` (missing values excluded)

Variable	Obs	Mean	Std. Dev.	Min	Max
maxgrip	373,276	33.31181	11.80805	1	99

ep005\_ contains information on the self-reported current job situation of the respondents. The output shows that retirement is the most frequent employment status. This is not surprising keeping in mind that only persons age 50 and over (and their partners) are eligible for SHARE.

```
tabulate ep005_
```

Figure 6.5: Stata output for the tabulation of ep005

Current job situation	Freq.	Percent	Cum.
-15. no information	4,776	1.16	1.16
-13. not asked in this wave	28,463	6.91	8.07
-12. don't know / refusal	253	0.06	8.13
1. retired	219,147	53.18	61.30
2. employed or self-employed	96,214	23.35	84.65
3. unemployed	10,098	2.45	87.10
4. permanently sick or disabled	12,744	3.09	90.19
5. homemaker	34,984	8.49	98.68
97. other	5,431	1.32	100.00
Total	412,110	100.00	

Regression analyses require Stata to know what the missing codes are. Using the `mvdecode` command, we can set all easySHARE missing codes to system-missing.

```
mvdecode maxgrip ep005, mv(-1/-15)
```

Afterwards we recode a new variable `Retired` based on `ep005_`. The new variable `Retired` is set to "1" for respondents in retirement and to "0" for all other employment statuses.

```
recode ep005 (2/97=0) (1=1), gen(Retired)
```

Now we estimate a cross-sectional linear regression model for the latest wave 8.

```
regress maxgrip Retired if wave ==8
```

Figure 6.6: Stata output for an exemplary linear regression

Source	SS	df	MS	Number of obs	=	42,170
Model	50089.3606	1	50089.3606	F(1, 42168)	=	404.95
Residual	5215833.91	42,168	123.691755	Prob > F	=	0.0000
				R-squared	=	0.0095
				Adj R-squared	=	0.0095
Total	5265923.27	42,169	124.876646	Root MSE	=	11.122

maxgrip	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Retired	-2.352387	.1168978	-20.12	0.000	-2.581509 -2.123265
_cons	33.67178	.0969635	347.26	0.000	33.48173 33.86184

To fully exploit the *easySHARE* panel data we now estimate a longitudinal model. For this purpose we first declare the data to be panel data. Unfortunately Stata does not accept person IDs in string format for this procedure. But there is a simple way for generating a new numerical ID variable:

```
bysort mergeid: gen mergeid_n = _n ==1
replace mergeid_n = sum(mergeid_n)
```

Using the new person identifier `mergeid_n`, the `xtset` command declares the data to be a panel.

```
xtset mergeid_n wave
```

Then we can estimate any panel model, here we choose a `mle` specification:

```
xtreg maxgrip Retired, mle
```



Figure 6.7: Stata output for an exemplary panel model

```

Random-effects ML regression
Group variable: mergeid_n

Random effects u_i ~ Gaussian

Log likelihood = -1191963.5

Number of obs      =    345,381
Number of groups   =    131,408

Obs per group:
    min =          1
    avg =          2.6
    max =           7

LR chi2(1)        =    5290.22
Prob > chi2       =     0.0000

```

maxgrip	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Retired	-2.252197	.0308824	-72.93	0.000	-2.312725	-2.191668
_cons	34.37223	.0357913	960.35	0.000	34.30208	34.44238
/sigma_u	10.84695	.0234125			10.80116	10.89293
/sigma_e	4.7002	.007196			4.686117	4.714325
rho	.8419165	.000729			.8404834	.8433411

```

LR test of sigma_u=0: chibar2(01) = 2.9e+05          Prob >= chibar2 = 0.000

```

... using SPSS<sup>3</sup>

After downloading and saving *easySHARE* you can load the data set into the SPSS memory with the `get` command followed by the storage location of the data.

```
GET
FILE='D:\easySHARE\data\easyshare_rel8-0-0.sav'.
```

First we reduce our data set to the key variables of our analysis using the `SAVE` subcommand `KEEP`. Afterwards we load the reduced data set in the memory.

```
SAVE OUTFILE= 'D:\easySHARE\data\easyshare_reduced.sav'
/KEEP mergeid wave int_year ep005_ maxgrip.
```

```
GET
FILE='D:\easySHARE\data\easyshare_reduced.sav'.
```

Panel data can be stored in “wide format” or in “long format”. *easySHARE* is provided in long format because this is the common format for panel analysis. In the long format, each observation stands for a respondent at a specific point in time, and the variables are their observed characteristics. Each row represents one point in time (or wave) per respondent. As *easySHARE* contains all waves of SHARE there is a maximum of eight rows per respondent. If a respondent participated in just one wave there is only one data line for this respondent.

In the wide format each observation stands for one respondent and the variables are their characteristics on the respective time points. There is only one row per respondent and each column represents one wave-specific variable.

*Figure 6.8: long vs. wide data format*

Long format			Wide format		
mergeid	wave	int_year	mergeid	int_year1	int_year2
AT-986403-01	1	2004	AT-986403-01	2004	2007
AT-986403-01	2	2007			

To reshape the data from long into wide format we use the `CASESTOVARS` command. `mergeid` is the identifier variable and the variable by which to organize the data is `wave`. It seems a good idea to sort the data set (again) in SPSS before running the `CASESTOVARS` command using `SORT CASES`.

```
SORT CASES mergeid wave.

CASESTOVARS
/ID=mergeid
/INDEX = wave.
```

---

<sup>3</sup> The example analysis is based on SPSS version 19.

Figure 6.9: SPSS output for reshaping the data from long to wide format

Cases In	412110
Cases Out	140125
Cases In/Cases Out	2,9
Variables In	5
Variables Out	25
Index Values	8

For our example analysis we need the data in long format. So we reshape the data back.

```
VARSTOCASES
/MAKE int_year from int_year.1 int_year.2 int_year.3 int_year.4 int_year.5
int_year.6 int_year.7 int_year.8
/MAKE ep005_ from ep005_.1 ep005_.2 ep005_.3 ep005_.4 ep005_.5 ep005_.6
ep005_.7 ep005_.8
/MAKE maxgrip from maxgrip.1 maxgrip.2 maxgrip.3 maxgrip.4 maxgrip.5
maxgrip.6 maxgrip.7 maxgrip.8
/INDEX = wave.
```

Regression analyses assume that there are no missing values in the variables. Here we set all *easySHARE* missing codes to system-missing using the `recode` command.

```
RECODE ep005_ maxgrip (-15 thru -1=SYSMIS).
EXECUTE.
```

Before starting with the analysis we first look at the central variables. `maxgrip` is the maximum value of the grip strength test. It ranges from 1 to 99.

```
DESCRIPTIVES VARIABLES = maxgrip
/STATISTICS=MEAN STDDEV MIN MAX.
```

Figure 6.10: SPSS output for the description of `maxgrip`

	N	Minimum	Maximum	Mean	Std. Deviation
Max. of grip strength measure	373276	1	99	33,31	11,808
Gültige Werte (Listenweise)	373276				

ep005\_ contains information on the self-reported current job situation of the respondents. The output shows that retirement is the most frequent employment status. This is not surprising keeping in mind that only persons above the age of 50 are eligible for SHARE.

```
FREQUENCIES VARIABLES= ep005_
/ORDER=ANALYSIS.
```

Figure 6.11: SPSS output for the frequency distribution of ep005

		Current job situation			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	retired	219147	53,2	57,9	57,9
	employed or self-employed	96214	23,3	25,4	83,3
	unemployed	10098	2,5	2,7	86,0
	permanently sick or disabled	12744	3,1	3,4	89,3
	homemaker	34984	8,5	9,2	98,6
	other	5431	1,3	1,4	100,0
	Total	378618	91,9	100,0	
Missing	System	33492	8,1		
Total		412110	100,0		

Afterwards ep005\_ is recoded. The newly generated variable Retired is 1 for respondents in retirement and 0 for all other employment statuses.

```
RECODE ep005_ (1=1) (2 thru 97=0) INTO Retired.
EXECUTE.
```

Now we estimate a cross-sectional linear regression model for wave 8.

```
REGRESSION
/SELECT=wave EQ 8
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT maxgrip
/METHOD=ENTER Retired.
```

Figure 6.12: SPSS output for an exemplary linear regression

Coefficients <sup>a,b</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	33,672	,097		347,262	,000
	Retired	-2,352	,117	-,098	-20,123	,000

a. Dependent Variable: maxgrip.1: max. of grip strength measure  
 b. Selecting only cases for which wave = 8

To fully exploit the *easySHARE* data we now apply a longitudinal model. In contrast to Stata, SPSS does accept person IDs in string format for this procedure. Here we use the MIXED command:

```
MIXED maxgrip WITH Retired
/FIXED=Retired | SSTYPE(3)
/METHOD=ML
/RANDOM=INTERCEPT | SUBJECT(mergeid)
/PRINT=SOLUTION.
```

Figure 6.13: SPSS output for an exemplary panel model

Estimates of Fixed Effects <sup>a</sup>							
Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	34,372229	,035777	190275,778	960,726	,000	34,302106	34,442351
Retired	-2,252197	,030841	303594,543	-73,027	,000	-2,312643	-2,191750

a. Dependent Variable: maxgrip.1: max. of grip strength measure.

... using R

After downloading and saving *easySHARE* you can rename the data set (e.g. *easyshare*) for an easier handling in the R code later on. Then load the data set into the R memory with the `load` command followed by the storage location of the data.

```
load("D:\easySHARE\data\easySHARE_rel8_0_0.rda")
```

First we reduce our *easyshare* data set to the key variables of the following exemplary analyses by creating a character vector and using this for the new defined data set.

```
variables <- c("mergeid", "wave", "int_year", "ep005_", "maxgrip")
easyshare <- easySHARE[variables]
```

Panel data can be stored in “wide format” or in “long format”. *easySHARE* is provided in long format because this is the common format for panel analyses. In the long format, each observation stands for a respondent at a specific point in time, and the variables are their observed characteristics. Each row represents one point in time (or wave) per respondent. As *easySHARE* contains all waves of SHARE there is a maximum of eight rows per respondent. If a respondent participated in just one wave there is only one data line for this respondent.

In the wide format each observation stands for one respondent and the variables are their characteristics on the respective time points. There is only one row per respondent and each column represents one wave-specific variable.

*Figure 6.14: long vs. wide data format*

Long format			Wide format		
mergeid	wave	int_year	mergeid	int_year1	int_year2
AT-986403-01	1	2004	AT-986403-01	2004	2007
AT-986403-01	2	2007			

To reshape the data from long into wide format we use the `reshape` command out of the *stats* package. `mergeid` is the identifier variable and the time variable is `wave`. It is important to arrange the data according to the time variable while reshaping. Otherwise the order of variables could be disarranged in the wide format.

```
install.packages("stats")
library("stats")

easyshare_wide <- reshape(easyshare[order(easyshare$wave),],
  v.names = c("int_year", "ep005_", "maxgrip"),
  idvar = "mergeid", timevar = "wave",
  direction = "wide")
```

For our example analysis, we need the data in long format. After we define a character vector with the variable names that we need for the `varying` argument, we reshape the data back into long format.

```
names <- names(easyshare_wide[2:25])
easyshare <- reshape(easyshare_wide,
  idvar = "mergeid", timevar = "wave",
  times = 1:8
  varying = c(names),
  sep=".",
  direction = "long")
```

This procedure leads to more observations than in the original data file because every respondent has one data line for each of the seven waves now. The variable values are set to NA for waves in which the respondent did not participate.

To get the original number of observations we drop the lines with missing information for the Variable `int_year`. Make sure you use a variable that had no missing values before the reshape for this step.

```
easyshare <- easyshare[!is.na(easyshare$int_year),]
```

Now we order our data set by `mergeid` and drop the `row.names`.

```
easyshare <- with(easyshare, easyshare[order(mergeid), ])
row.names(easyshare) <- NULL
```

Before starting with the analysis, we look at the central variables first. `maxgrip` is the maximum value of the grip strength test. It ranges from 1 to 99.

```
summary((subset(easyshare, maxgrip > 0))$maxgrip)
```

Figure 6.15: R output for the summarizing `maxgrip` (missing values excluded)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	25.00	32.00	33.31	41.00	99.00

`ep005_` contains information on the self-reported current job situation of the respondents. The output shows that retirement (category 1) is the most frequent employment status followed by employed or self-employed (category 2). This is not surprising keeping in mind that only persons age 50 and over (and their partners) are eligible for SHARE.

```
library("catspec")
ctab(table(easyshare$ep005_))
```

Figure 6.16: R output for the tabulation of ep005

Var1	Count	Total %
- 15	4776.00	1.16
- 13	28463.00	6.91
- 12	253.00	0.06
1	219147.00	53.18
2	96214.00	23.35
3	10098.00	2.45
4	12744.00	3.09
5	34984.00	8.49
97	5431.00	1.32

Regression analyses require R to know what the missing codes are. Here we set all easySHARE missing codes to NA.

```
easyshare$ep005_[easyshare$ep005_ < 0] <- NA
easyshare$maxgrip[easyshare$maxgrip < 0] <- NA
```

Afterwards we recode a new variable Retired based on ep005\_. The new variable Retired is set to "1" for respondents in retirement and to "0" for all other employment statuses.

```
install.packages("car")
library("car")
easyshare$Retired <- recode(easyshare$ep005_,
                           "c(2, 3, 4, 5, 97)=0; 1=1; else=NA")
```

Now we estimate a cross-sectional linear regression model, e.g. for wave 8

```
lm <- lm(maxgrip ~ Retired,
         data = subset(easyshare, (wave == 8),
                       na.action = na.exclude)
summary(lm)
```



Figure 6.17: R output for an exemplary linear regression

```

Call:
lm(formula = maxgrip ~ Retired, data = subset(easyshare, (wave ==
      8)), na.action = na.exclude)

Residuals:
    Min       1Q   Median       3Q      Max
-32.672  -8.319  -1.672   7.681  51.328

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  33.67178    0.09696   347.26  <2e-16 ***
Retired      -2.35239    0.11690   -20.12  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.12 on 42168 degrees of freedom
(4563 observations deleted due to missingness)
Multiple R-squared:  0.009512, Adjusted R-squared:  0.009488
F-statistic:  405 on 1 and 42168 DF, p-value: < 2.2e-16

```

To fully exploit the *easySHARE* panel data we now estimate a longitudinal model. For this purpose we first declare the data to be panel data. Using the package `plm` with the command `pdata.frame`, the data can be declared to be panel data.

```

install.packages("plm")
library("plm")
easyshare_panel <- pdata.frame(easyshare,
                               index = c("mergeid", "wave"))

```

Then we can estimate any panel model. We choose a model with random effects here:

```

plm <- plm(maxgrip ~ Retired, data = easyshare_panel,
           index = c("mergeid", "wave"), model = "random")
summary(plm)

```

Figure 6.18: output for an exemplary panel model

```

Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)

Call:
plm(formula = maxgrip ~ Retired, data = easyshare_panel, model = "random",
     index = c("mergeid", "wave"))

Unbalanced Panel: n = 131408, T = 1-7, N = 345381

Effects:
              var std. dev share
idiosyncratic 22.033  4.694 0.162
individual    113.984 10.676 0.838
theta:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.5975 0.7031  0.7853  0.7515  0.8071  0.8361

Residuals:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-37.410 -2.977  -0.293   0.034   2.864  67.946

Coefficients:
              Estimate Std. Error z-value Pr(>|z|)
(Intercept) 34.377590   0.035574 966.368 < 2.2e-16 ***
Retired     -2.260775   0.030914 -73.131 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    9527700
Residual Sum of Squares: 7707700
R-Squared:               0.19154
Adj. R-Squared:         0.19153
Chi sq: 81553.9 on 1 DF, p-value: < 2.22e-16

```

## 7. Appendix A: List of variables

<b>Variable</b>	<b>Description</b>
mergeid	Person identifier (fix across modules and waves)
hhid	Household identifier (wave specific)
coupleid	Couple identifier (wave specific)
wave	Wave of data collection
wavepart	Wave participation pattern
int_version	Interview version (baseline or longitudinal)
int_year	Interview year
int_month	Interview month
country	Country identifier
country_mod	Country identifier (ISO coded)
language	Language of questionnaire
female	Gender of respondent
dn002_mod	Month of birth
dn003_mod	Year of birth
dn004_mod	Born in the country of interview
age	Age at interview (in years)
birth_country	Country of birth (ISO coded)
citizenship	Citizenship of respondent (ISO coded)
iv009_mod	Area of location
q34_re	What religion do you belong to or feel attached to mostly?
isced_r	Education of respondent in ISCED-97 Coding
eduyears_mod	Years of education
mar_stat	Marital status
hhsize	Household size
partnerinh	Living with spouse/partner
int_partner	Interview of partner available
age_partner	Age at interview of partner

Variable	Description
gender_partner	Gender of partner
mother_alive	Is natural parent still alive: mother
father_alive	Is natural parent still alive: father
siblings_alive	Number of siblings alive (based on: dn036_, dn037_, dn034_)
ch001_	Number of children
ch021_mod	Number of grandchildren
ch007_hh	At least one child lives in household/building (based on ch007_1-ch007_16)
ch007_km	At least one child lives less than 1 km away (based on ch007_1-ch007_16)
sp002_mod	Received help from outside the household
sp003_1_mod	Who gave help: person 1
sp003_2_mod	Who gave help: person 2
sp003_3_mod	Who gave help: person 3
sp008_	Given help to others outside the household
sp009_1_mod	To whom did you give help: person 1
sp009_2_mod	To whom did you give help: person 2
sp009_3_mod	To whom did you give help: person 3
books_age10	Number of books when ten
maths_age10	Relative position to others mathematically when ten
language_age10	Relative position to others language when ten
vaccinated	Vaccinations during childhood
childhood_health	Childhood health status
sphus	Self-perceived health – us version
chronic_mod	Number of chronic diseases (easySHARE version)
casp	CASP: quality of life and well-being index
euro1	Depression (part of EURO-D)
euro2	Pessimism (part of EURO-D)
euro3	Suicidality (part of EURO-D)
euro4	Guilt (part of EURO-D)

Variable	Description
euro5	Sleep (part of EURO-D)
euro6	Interest (part of EURO-D)
euro7	Irritability (part of EURO-D)
euro8	Appetite (part of EURO-D)
euro9	Fatigue (part of EURO-D)
euro10	Concentration (part of EURO-D)
euro11	Enjoyment (part of EURO-D)
euro12	Tearfulness (part of EURO-D)
eurod	Depression scale EURO-D – high is depressed
bfi10_extra_mod	Extraversion
bfi10_agree_mod	Agreeableness
bfi10_consc_mod	Conscientiousness
bfi10_neuro_mod	Neuroticism
bfi10_open_mod	Openness
hc002_mod	How often seen or talked to medical doctor last 12 months
hc012_	In hospital last 12 month
hc029_	In a nursing home
maxgrip	Max. of grip strength measure
adlwa	Activities of daily living w&h index (high: has difficulties)
adla	Activities of daily living index (high: has difficulties)
iadla	Instrumental activities of daily living index (high: has difficulties)
iadlza	Instrumental activities of daily living index 2 (high: has difficulties)
mobilityind	Mobility index (high: has difficulties)
lgmuscle	Large muscle index (high: has difficulties)
grossmotor	Gross motor skills index (high: has difficulties)
finemotor	Fine motor skills index (high: has difficulties)
recall_1	Recall of words, first trial (based on c008tot)
recall_2	Recall of words, delayed (based on cf016tot)
orienti	Orientation to date, month, year and day of week

Variable	Description
numeracy_1	Score of first numeracy test (percentage)
numeracy_2	Score of second numeracy test (subtraction)
bmi	Body mass index
bmi2	Body mass index categories
smoking	Smoke at the present time
ever_smoked	Ever smoked daily
br010_mod	Days a week consumed alcohol last 3 month
br015_	Sports or activities that are vigorous
ep005_	Current job situation
ep009_mod	Employee or self-employed in (main) job
ep011_mod	Term of (main) job
ep013_mod	Total hours worked per week (main) job
ep026_mod	Satisfied with (main) job
ep036_mod	Look for early retirement in (main) job
co007_	Is household able to make ends meet
thinc_m	Household net income, imputed
income_pct_w1	Hhd income percentiles wave 1
income_pct_w2	Hhd income percentiles wave 2
income_pct_w4	Hhd income percentiles wave 4
income_pct_w5	Hhd income percentiles wave 5
income_pct_w6	Hhd income percentiles wave 6
income_pct_w7	Hhd income percentiles wave 7
income_pct_w8	Hhd income percentiles wave 8