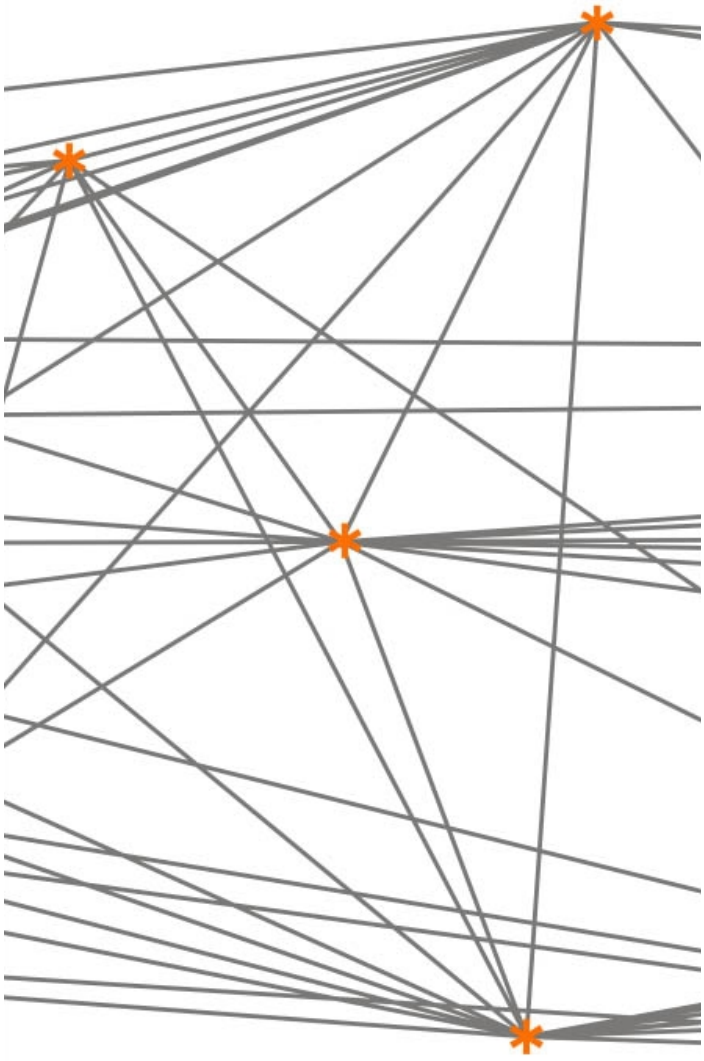
SHARE

Survey of Health, Ageing and Retirement in Europe

SHARE Corona Survey

Release 8.0.0



mea

February 10, 2022

Table of Contents

1 About this release

This release guide describes background and content of the SHARE Corona dataset containing data collected via computer-assisted telephone interviews (CATI) in the two rounds of the SHARE Corona Survey (SCS) between June and August 2020 (1st SCS) and one year later between June and August 2021 (2nd SCS). For further information on the regular SHARE CAPI data, please see Release Guide 8.0.0.

2 New in Release 8.0.0

- Data from the 2nd SHARE Corona Survey available for the first time
- New imputations for the 1st SHARE Corona Survey
- Latest State of sample cleaning

3 About SHARE: General Overview

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of individuals aged 50 or older. SHARE started in 2004 with representative samples of individuals aged 50+ in 11 European countries as a reaction to the growing challenges of population ageing. To date, SHARE conducted eight waves of data collection and covers all continental EU countries plus Switzerland and Israel. SHARE explores this cross-country setting as 'natural laboratory' across scientific disciplines and over time in order to turn the challenges of population ageing into opportunities and provide policy makers with reliable information for evidence based policies.

SHARE applies a concept of ex-ante harmonisation: there is one common generic questionnaire that our country teams translate into the national languages (in some countries more than one language) using an internet based translation tool. However, some internationally highly diverse variables require country-specific measurements and ex-post harmonisation, for example in the areas of education (ISCED) or occupation (ISCO, NACE).

Usually, SHARE data collection is based on computer-assisted personal interviewing (CAPI) because it makes the execution of physical tests possible. The interviewers conduct face-to-face interviews using a laptop on which the CAPI instrument is installed.

For further general information on SHARE and regular updates, please visit the project website at http://www.share-project.org

For a detailed description of the previous SHARE waves, eligibility and how to use the data, please see Release Guide 8.0.0.

4 The SHARE Corona Survey

4.1 Switch to telephone interviewing during the COVID-19-pandemic

The outbreak of COVID-19 hit SHARE in the middle of its 8th Wave of data collection and the fieldwork had to be suspended in all participating countries in March 2020. At this point in time, about 70 percent of all expected interviews in the panel sample across countries had been conducted.

To resume fieldwork, a switch to telephone administered interviews (CATI) was decided and a specific questionnaire was developed covering the same topics as the regular SHARE questionnaire - but considerably shortened and targeted to the COVID-19 living situation of people who are 50 years and older. Based on methodological considerations in connection with the health protection of respondents and interviewers, the use of CATI was the preferred alternative to the previous face-to-face-interviewing. For a more detailed description see:

Scherpenzeel, A., Axt, K., Bergmann, M., Douhou, S., Oepen, A., Sand, G., Schuller, K., Stuck, S., Wagner, M., & Börsch-Supan, A. (2020). Collecting survey data among the 50+ population during the COVID-19 outbreak: The Survey of Health, Ageing and Retirement in Europe (SHARE). *Survey Research Methods*, 14(2), 217-221.

4.2 Content of the SHARE Corona data

The generic and the country-specific questionnaires of the first SHARE Corona Survey are available here. The slightly different questionnaire of the second SHARE Corona Survey is available here. The questionnaires cover the most important life domains of the target population and asks specific questions about infections and changes in life during the pandemic:

Health and health behaviour
General health before and after the COVID-19 outbreak, practice of safety measures (e.g. social distancing, wearing a mask)

Mental health
Anxiety, depression, sleeping problems, and loneliness before and after the COVID-19 outbreak

Infections and healthcare
COVID-19 related symptoms, SARS-CoV-2 testing and hospitalization, forgone medical treatment, satisfaction with treatments, vaccination status (only in 2nd SCS available)

Changes in work and economic situation
Unemployment, business closures, working from home, changes in working hours and income, financial support

Social networks
Changes in personal contacts with family and friends, help given and received, personal care given and received.

The naming of items in the SHARE Corona Survey is based on the following principles:

- Items that are identical to previous waves of data collection have the same label plus prefix "ca"
  ➔ Example: *cadn003_* on year of birth
- New items related to a COVID-19 infection are followed by a "c" after the prefix
  ➔ Example: *cac002_* on COVID-19 symptoms
- Items that refer to existing items that needed to be adapted to the specific circumstances during the pandemic can be identified by an "8" in the front of the numbering indicating that the item was introduced in Wave 8
  ➔ Example: *caep805_* on employment status after the outbreak of Corona

Additional to basic household information provided in the Coverscreen module and information covered by the SHARE Corona Questionnaire, the release includes weights as part of the *gv_weights_ca* modules for handling unit nonresponse and attrition errors.

Another module is *sharew8_rel8-0-0_ca_at* containing the Austrian SHARE Corona Survey data conducted later than in the countries. The fieldwork period in Austria was between end of July and end of September 2020 whereas all other countries finished fieldwork before the mid of August (see Table 1). Furthermore, the *sharew8_rel8-0-0_xc* module includes additional pandemic related data conducted via telephone interviews in Austria and the Czech Republic between September 2020 and January 2021.

4.3 Sample and fieldwork design

For the instrument of the SHARE Corona Survey, the sample included 1) panel members who had not been interviewed in Wave 8 before the suspension of fieldwork and 2) panel members who had already been interviewed in Wave 8. Both respondent groups received the same questionnaire; the only difference is that the panel members who had not been interviewed face-to-face in Wave 8 were asked questions on changes in the household composition since their last interview (Coverscreen). Respondents who had already been asked in Wave 8 did not have to answer these questions again. However, the coverscreen (cv_r) module of the 1st SCS includes information on household composition for all households that participated in the first SHARE Corona Survey. Users need to be aware, that this information might be collected during the regular fieldwork of Wave 8 and thus not at the same time as the SHARE Corona data. The variable *cvdate* as part of the *cv_r* module informs the user in which field phase the information was collected. The *update_ca* variables indicate when specific cv_r variables were subject to change in the second field phase of the interview.

Regarding the fieldwork design, several aspects of the normal SHARE fieldwork design had to be adapted as a consequence of the mode switch from CAPI to CATI. However, SHARE's principle of ex-ante harmonization by providing the same software tools and programmed questionnaire to all survey agencies was also followed for the CATI.

4.4 Participating countries

Table 1: Countries & Fieldwork Times

Country ID	Country	Fieldwork time 1st SCS in 2020	Fieldwork time 2nd SCS in 2021
11	Austria[1]	July 20th – September 30th	June 22nd – August 5th
12	Germany	June 19th – August 3rd	June 29th – August 3rd
13	Sweden	June 17th – August 14th	June 28th – August 13th
14	Netherlands	June 19th – July 31st	June 10th – August 3rd
15	Spain	June 11th – August 10th	June 9th – August 14th
16	Italy	June 9th – July 31st	June 9th – August 6th
17	France	June 16th – July 31st	June 8th – July 30th
18	Denmark	June 10th – August 7th	June 20th – August 2nd
19	Greece	June 12th – August 7th	June 21th – August 10th
20	Switzerland	June 9th – August 6th	June 10th – August 5th
23	Belgium	June 8th – August 10th	June 8th – August 1st
25	Israel	June 4th – August 5th	June 2nd – August 4th
28	Czech Republic	June 8th – August 6th	June 3rd – August 4th
29	Poland	June 8th – August 5th	June 9th – August 10th
31	Luxembourg	June 25th – August 5th	June 14th – July 28th
32	Hungary	June 18th – August 11th	June 16th – August 9th
33	Portugal	June 11th – August 10th	June 5th – August 4th
34	Slovenia	June 8th – August 12th	June 4th – July 26th
35	Estonia	June 8th – July 27th	June 10th – August 1st
47	Croatia	June 15th – August 9th	June 17th – August 1st
48	Lithuania	June 13th – July 31st	June 8th – July 8th
51	Bulgaria	July 2nd – August 14th	June 9th – August 12th
53	Cyprus	June 11th – August 10th	June 8th – August 14th
55	Finland	June 12th – August 10th	June 16th – August 4th
57	Latvia	June 24th – August 11th	June 11th – August 2nd
59	Malta	June 11th – August 10th	June 21st – August 13th
61	Romania	June 9th – August 11th	June 13th – August 1st
63	Slovakia	June 12th – July 30th	June 10th – July 12th

[1] Due to the later fieldwork period, the Austrian SHARE Corona Survey data are provided as separate module

4.5 Weights

Giuseppe De Luca & Paolo Li Donni – University of Palermo

Based on the participation in the two survey instruments of wave 8, one can distinguish three subsamples in the 1st SCS of primary interest: CAPI, CATI and CAPI & CATI. In contrast to that, a differentiation of subsamples is not necessary in the 2nd SCS as there was only one survey instrument.

The baseline strategy adopted by SHARE to deal with problems of unit nonresponse and panel attrition relies again on the calibration approach of Deville and Särndal (1992). For each subsample we always provide calibrated cross-sectional weights, as well as calibrated weights at individual level for inference to the target population of individuals and calibrated weights at the household level for inference to the target population of households. For the purposes of longitudinal analyses, we also provide calibrated longitudinal weights for the balanced sample of respondents who have participated in the first two waves of the SCS. As for cross-sectional calibrated weights, the longitudinal calibrated weights are defined both at the individual and at the household levels. The calibration procedure for calibrated cross-sectional and longitudinal weights coincides with that described in chapter 15.2 of Release Guide 8.0.0 for the CAPI subsample. There are only two major differences. First, the weights are now calculated on two different subsamples of respondents using a possibly different specification calibration function. Second, while the population calibration margins for the calibrated cross-sectional weights of the CATI and CAPI&CATI subsamples of the 1st SCS coincide with those of the CAPI subsample, the population calibration margins of the 2nd SCS refer to the population of people aged 50+ in 2021. Similarly, the calibrated longitudinal weights aim to represent the population of people aged 50+ in 2019 that survive up to 2021. Since official statistics are not yet available, data about population size and number of deaths are based on projections provided by the Central Bureau of Statistics for Israel and from the EUROSTAT regional database for all other countries.

4.6 Imputations

So far, we have produced imputations of the missing values due to item non-response in the CATI data of the 1st SCS only. A similar set of imputations for the 2nd SCS will be available in the future releases. Imputations for the CATI data of the 1st SCS were constructed separately from the imputations for the CAPI data. Since the fraction of missing values in the CATI data was generally much less than 3 per cent, the imputation procedure for this data set draws mainly on the hot-deck method. Two exceptions worth noting are the monetary variables on overall monthly household income before the outbreak (CAHH017) and the lowest overall monthly household income since the outbreak (CAE005). As for the monetary variables collected in the standard CAPI questionnaire, these two open-ended questions are very sensitive and difficult to answer precisely, especially in a CATI mode. In addition, unlike the CAPI questionnaire, the CATI questionnaire does not include sequences of unfolding bracket

questions that may allow valuable interval data on the missing monetary amounts to be recovered.

In addition to missing data due to "Don't know" and "Refusal" answers, for some variables we also imputed other types of data inconsistencies due to routing errors in Section E (Economic situation) and measurement errors in Section W (Work) of the CATI questionnaire. Specifically, Section E depends on a filter variable, *CAE001*, which controls in turn all other questions included in this section (i.e. all variables starting with CAE). The problem is that, by design, this section would have to be asked only to the first respondent in the household. However, the filter variable was not automatically assigned by the interviewing software, but rather it was left open for selection by the respondent/interviewer. Thus, instead of having only one respondent per household, the data contain a set of households who have skipped all questions in the economic section and another set of households with two respondents per household. For the specific purposes of imputations, we adopted the following strategy: in households that skipped the economic section, we imputed the missing values on all variables of this section by selecting the household member with the minimum person identifier (*mergeid*). In households that have two respondents per household, we selected first the respondent with the largest number of valid answers to the remaining variables of the economic section. Depending on the number of valid answers, we then selected the respondent with the minimum person identifier within each household. The CATI imputation database contains an indicator variable (*RESP_E*) that allows the household member who was selected as the eligible respondent for the economic section to be identified.

In Section W, respondents who reported being employed or self-employed at the time when COVID-19 broke out were first asked about their usual working hours per week before the outbreak (*CAW020*). Next, respondents were asked whether they reduced/increased the number of working hours and finally the lowest/highest number of hours of work since the outbreak (*CAW021*, *CAW022*, *CAW024*, *CAW025*). These variables were affected by two types of measurement errors: (i) people who reported a reduction in the number of working hours, but the lowest number of hours of work since the outbreak was in fact greater than or equal to the number of hours of work before the outbreak; (ii) people who reported an increase in the number of working hours, but the highest number of hours of work since the outbreak was in fact smaller than or equal to the number of hours of work before the outbreak. To handle these types of measurement errors, we imputed in this case all data inconsistencies due to misreporting on the sequence of variables *CAW020*, *CAW021*, *CAW022*, *CAW024* and CAW025.

The imputation procedure for these five variables of Section W and other eleven variables of Section E is based on the Fully Conditional Specification (FCS) method. One important difference with respect to the FCS method used to impute the monetary variables of the CAPI questionnaire is that we account for the continuous, binary or categorical nature of the sixteen variables that need to be imputed jointly.

a) Hot-deck imputations

We performed hot-deck imputations separately by country and according to a convenient order of the variables that accounts for branching and skip patterns in the CATI questionnaire. Imputation classes for the implementation of this method were based on the following set of auxiliary variables: country, gender, five age classes ([– 49], [50– 59], [60– 69], [70 – 79], [80+]), a binary indicator for respondents living with a spouse/partner, five groups for years of education, a binary indicator for good self-perceived health before the outbreak, and a binary indicator for changes in the self-perceived health status during the outbreak. The first four auxiliary variables are fully observed, while the last three auxiliary variables contain very small fractions of missing values that were imputed first using only the first four variables. Since the CATI data consist of longitudinal respondents only, the information on years of education was obtained by the most recent CAPI data available for each respondent. For some variables we exploited a larger set of auxiliary variables. For example, we used two additional binary indicators for wearing a face mask in public spaces and keeping distance from others in public when imputing several variables included in Section H (Health and health behavior), Section C (Corona-related infection) and Section Q (Quality of healthcare) of the CATI questionnaire. Furthermore, we jointly imputed missing values on the variables that are logically related. Specifically, in Section H, we jointly imputed variables regarding illness or health conditions since the last interview (*CAH003* and *CAH004*). In Section C, we imputed jointly variables regarding the COVID-19 symptoms (*CAC002* and *CAC003*), whether testing positive (*CAC004* and *CAC005*) or negative (*CAC007* and *CAC008*) for coronavirus, and whether being hospitalised (*CAC010* and *CAC011*) or having died (*CAC013* and *CAC014*) due to an infection from coronavirus. In Section Q, we imputed jointly variables regarding forwent medical treatment since the outbreak (*CAQ005* and *CAQ006*), postponed medical appointment (*CAQ010* and *CAQ011*), denied medical appointment (*CAQ015* and *CAQ016*), whether treated in hospital and the associated level of satisfaction (*CA025* and *CA027*), and whether visited by a doctor and the associated level of satisfaction (CAQ020 and CAQ022).

In total, we imputed sequentially by the hot-deck method about 200 variables. As for the hot-deck imputations of the CAPI data, the CATI imputation database contains five multiple imputations of the missing values and a flag variable associated to each imputed variable which allows the users to identify the imputed observations.

b) FCS imputations

After the hot-deck imputations, we constructed the FCS imputations for the missing values on the five hours of work variables collected in Section W (*CAW020*, *CAW021*, *CAW022*, *CAW024* and *CAW025*) and other eleven variables collected in Section E: overall monthly household income before the outbreak (*CAHH017*), lowest overall monthly household income since the outbreak (*CAE005*), a set of six binary indicators for received financial support (*CAE003* and *CAE004*), household's ability to make ends meet (*CACO007*), a binary indicator for the postponement of regular payments (*CAE011*) and a binary indicator for dipping into savings to cover the necessary day-to-day expenses (*CAE012*). FCS imputations of these sixteen

variables were always constructed separately by country. We do not use separate imputation models for different household types, but we always include a binary indicator for living with a spouse/partner in our set of observed predictors.

At each iteration of the Gibb sampling algorithm, we used a linear regression model for the continuous variables (*CAW020*, *CAW022*, *CAW025*, *CAHH017* and *CAE005*), a logit model for four binary variables (*CAW021*, *CAW024*, *CAE011* and *CAE012*), a multinomial logit model for the categorical variable *CACO007*, and the multivariate hot-deck method for the six binary indicators related to financial support received since the outbreak (CAE003 and CAE004).

For the *CAHH017* and *CAE005* variables, we symmetrically trimmed 2 percent of the complete cases from the country-specific distribution of each variable to exclude (and then impute) outliers that may have a large influence on survey statistics. Furthermore, we transformed all continuous variables using either the logarithm (*CAW020*, *CAW025* and *CAHH017*) or the inverse hyperbolic sine (*CAW022* and *CAE005*) transformations to reduce skewness in the left tails of their distributions. In addition to the variables imputed jointly in the Gibb sampling, we used as observed predictors a binary indicator for female respondents, a quadratic polynomial in age, years of education, a binary indicator for living with a spouse/partner and its interaction with age of the spouse/partner, a binary indicator for good self-perceived health and a binary indicator for changes in the self-perceived health status during the outbreak. In the linear regression models for *CAHH017* and *CAE005* and in the multinomial model for *CACO007*, we also included as observed predictors a binary indicator for being employed before the outbreak. For the multivariate hot-deck imputations of the six binary indicators related to financial support received since the outbreak we used instead a similar set of observed predictors (properly discretized to form the imputation classes) plus the quantiles of *CAE005* computed at each iteration of the Gibb sampling algorithm. The final FCS imputation model adopted in each country was subject to an accurate fine-tuning for the choice of the predictors. Specifically, we had to impose a set of country and item-specific exclusion restrictions to avoid possible problems of collinearity, very imprecise estimates, as well as problems of convergence and perfect prediction in the context of non-linear models. As for the other types of imputations in SHARE, we always provide five multiple imputations of the missing values. After an initial set of 15 burn-in iterations, convergence of the Gibbs sampling algorithm was assessed by the Gelman-Rubin criterion applied to the mean, median and 90th percentile of the distribution of each continuous variable and the mean of the distribution of each discrete variable. In all countries, convergence was achieved before the 50th iteration.

4.7 Data Access

Access to the data is provided free of charge for scientists worldwide. After registration as a SHARE user, the SHARE users can easily log in to the SHARE Research Data Center and download the SHARE data sets. Please find more information here: http://www.share-project.org/data-access.html

In addition to the interview month and year of the COVID-19 interviews, which are included in the cv_r module, registered SHARE users can apply for access to an additional data set that includes the date of the COVID-19 interviews. To be granted access to this data, users have to fill in and sign a special user statement in which they are requested to provide a comprehensible justification why access to this data is needed in order to carry out their scientific research. After submission of the statement to info@share-project.org, SHARE Central will review the application. In the case of successful applications, the respective user will be informed that they can access the data via their SHARE user accounts and download the COVID-19 Survey Interview Date data from the SHARE Research Data Center.

4.8 Citation Requirements

Please cite the SHARE Corona Survey 1 data set as:

Börsch-Supan, A. (2022). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 8. COVID-19 Survey 1. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w8ca.800

Please cite the SHARE Corona Survey 2 data set as:

Börsch-Supan, A. (2022). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 9. COVID-19 Survey 2. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w9ca.800

Please cite the SHARE Corona Survey 1 interview date data sets as:

Börsch-Supan, A. (2022). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 8. COVID-19 Survey 1 Interview Date. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w8caintd.800

Please cite the SHARE Corona Survey 2 interview date data sets as:

Börsch-Supan, A. (2022). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 9. COVID-19 Survey 2 Interview Date. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w9caintd.800

And please do not forget to report all your publications (not only journal articles but also books, book chapters, working papers, discussion papers, theses, etc.; also non-English) to SHARE Central via info@share-project.org

Responsible for database management and preparation of the releases:

Stephanie Stuck, Sabrina Zuber, Fabio Franzese, Stefan Gruber,
Tim Birkenbach, Senta-Melissa Pflüger, Josefine Atzendorf,
Carolina Brändle, Magdalena Gerum

and all SHARE country team operators.

Contact: info[at]share-project.org