# SHARE WAVE 9 METHODOLOGY:

From the SHARE Corona Survey 2 to
the SHARE Main Wave 9 Interview

SHARE
SURVEY OF HEALTH, AGEING
AND RETIREMENT IN EUROPE

# SHARE WAVE 9 METHODOLOGY:

From the SHARE Corona Survey 2 to
the SHARE Main Wave 9 Interview

# SHARE Wave 9 Methodology:

## From the SHARE Corona Survey 2 to the SHARE Main Wave 9 Interview

**Edited by:**
Michael Bergmann
Melanie Wagner
Axel Börsch-Supan

**Authors:**
Michael Bergmann
Arne Bethmann
Axel Börsch-Supan
Paolo Li Donni
Salima Douhou
Theresa Fabel
Tessa-Virginia Hannemann
Magdalena Hecher
Giuseppe De Luca
Marcela Otero
Yuri Pettinicchi
Jeny Tony Philip
Magdalena Quezada
Anna Rieckmann
Gregor Sand
Karin Schuller
Elena Sommer
Barbara Thumann
Iggy van der Wielen
Melanie Wagner
Yasemin Yilmaz

**Published by:**
SHARE-ERIC
Leopoldstrasse 139
80804 Munich
https://share-eric.eu

© SHARE-ERIC, 2024

**Layout and printing by:**
Valentum Kommunikation GmbH
Bischof-von-Henle-Str. 2b
93051 Regensburg

www.valentum-kommunikation.de

**Picture credits:**
SkyLine – stock.adobe.com

**Suggested citation:**
Bergmann, M., M. Wagner and A. Börsch-Supan (2024). SHARE Wave 9 Methodology:
From the SHARE Corona Survey 2 to the SHARE Main Wave 9 Interview. Munich: SHARE-ERIC.

DOI: 10.6103/mv.w09

**Sponsored by:**

MAX-PLANCK-GESELLSCHAFT

This project has received funding from the European Union under grant agreement No 101102412 and the European Union's Horizon 2020 research and innovation programme under grant agreements No 870628, No 101015924.

Federal Ministry of Education and Research

NIH National Institute on Aging

# TABLE OF CONTENTS

# CHAPTER 8

Weights and Imputations in SHARE Wave 9

08

# 8 WEIGHTS AND IMPUTATIONS IN SHARE WAVE 9

*Giuseppe De Luca and Paolo Li Donni*

## 8.1 Introduction

This chapter describes the weighting and imputation strategies used for dealing with problems of unit nonresponse, sample attrition, and item nonresponse in the most recent SHARE studies: the ninth regular wave of SHARE and the second wave of the SHARE Corona Survey. The remainder of the chapter is organized as follows. Section 8.2 provides a brief overview of the key features of these two studies that are relevant for the purposes of our weights and imputation strategies. Section 8.3 focuses on the construction of calibrated survey weights that attempt to compensate for the potential selection effects generated by unit nonresponse and attrition, while Section 8.4 focuses on the construction of (multiple) imputations for the missing values due to item nonresponse errors.

## 8.2 Overview of SHARE Main Wave 9 and SHARE Corona Survey 2

### 8.2.1 SHARE Main Wave 9

The ninth regular wave of SHARE was fielded between October 2021 and September 2022 by means of a Computer-Assisted Personal Interview (CAPI) administered in the same 28 countries that had already participated in the eighth regular wave of the SHARE panel. Ignoring the End-of-Life interviews, SHARE Wave 9 collected data from 69,154 individual interviews in 47,957 households. The sample size available in each country ranges from a minimum of 731 observations for Cyprus and a maximum of 4,802 observations for Poland.

As discussed in Chapter 2, the uncertainty generated by the ongoing COVID-19 pandemic in 2021 prevented the drawing of new refreshment samples. Hence, the gross sample of Wave 9 can be viewed as a follow-up of the sample originally drawn in Wave 8. Note that, in addition to the longitudinal samples from previous waves and the national refreshment samples from batches that were already fielded in Wave 8, it also includes national refreshment samples from batches that were not fielded before the suspension of the Wave 8 fieldwork due to the COVID-19 outbreak in spring 2020.

In total, there are 18 countries that have drawn a refreshment sample in Wave 8: Austria, Belgium, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Israel, Latvia, Poland, Portugal, Slovenia, Spain, Sweden, and Switzerland. The target population of Wave 9 has been defined as the 50+ population in 2019 that survives up to 2021 (i.e., the beginning of the data collection process in Wave 9) because the national gross samples of all these countries cover cohorts of people born in 1969 or earlier who were already age-eligible at the time of the latest refreshment samples in Wave 8. As for other longitudinal studies, this definition of the target population accounts for the sizeable effects of mortality between Wave 8 and Wave 9. It excludes cohorts of people born in 1970 and 1971, who were aged respectively 50 and 51 years in 2021, that are not covered by the national gross samples due to the lack of new refreshment samples in Wave 9. The representativeness of the cohorts of people born in 1968 and 1969 remains problematic for the ten countries (Bulgaria, Cyprus, Greece, Italy, Lithuania, Luxembourg, Malta, the Netherlands, Romania, and Slovakia) that have not drawn refreshment samples in Wave 8. Issues related to the coverage of these cohorts will be addressed in the refreshment samples of Wave 10.

In Section 8.3, we shall see that these survey design features have important implications on the calibrated weights of Wave 9. For example, unlike the other regular waves of SHARE, the calibrated cross-sectional weights of Wave 9 and the calibrated longitudinal weights of the wave combination 8 – 9 aim to reproduce the same target population. These two different sets of calibrated weights differ only in relation to their subsamples of respondents and their sets of population margins.

### 8.2.2 SHARE Corona Survey 2

The second wave of the SHARE Corona Survey was designed to study the long-term impact of the COVID-19 pandemic. It was fielded about one year later than the first wave, between June and August 2021, by means of a Computer-Assisted Telephone Interview (CATI) administered in the same 28 countries that had already participated in the first wave of this study.

The first wave of this study collected data from 57,560 individual interviews in 38,960 households, while its second wave involved 49,254 individual interviews in 33,109 households.[16] The cross-country average household attrition rate is 15 percent, with the lowest retention rate of 68 percent in Sweden and the highest retention rate of 94 percent in Lithuania. About 2 percent of the interviews in the second wave refer to new entries such as new spouses/partners of age-eligible respondents and nonresponding spouses/partners from the first wave that were eligible for the second wave. The balanced sample of respondents who have participated in both waves of the study includes 48,357 individuals.

By design, the first wave was administered to the longitudinal sample of Wave 8, but not to the refreshment sample. The sample of the second SHARE Corona Survey is a follow-up of those households that participated in the first SHARE Corona Survey, without refreshment samples in any of the participating countries. Unlike release 8.0.0, the target population of the first wave has been re-defined as the 50+ population in 2016 (i.e., the time of the latest baseline/refreshment samples drawn in Wave 7) that survives up to 2020 (i.e., the beginning of the data collection process in the first SHARE Corona Survey). Similarly, the target population of the second SHARE Corona Survey is defined as the 50+ population in 2016 that survives up to 2021.

## 8.3 Weighting Strategies

In the ideal situation of complete responses, design weights may allow one to account for the randomness of the sampling process by compensating for the unequal selection probabilities of the various sampling units. Unfortunately, the properties of inferential procedures based on the design weights depend strongly on this ideal assumption, which is almost never satisfied in practice. SHARE is not an exception to this common situation: the baseline/refreshment samples of each wave suffer from problems of unit nonresponse and the longitudinal part of the sample is also subject to problems of attrition. From this viewpoint, it is important to stress that design weights are included in the SHARE release 9.0.0 only to allow the comparison and development of alternative procedures for dealing with unit nonresponse and attrition errors, but we usually discourage users to rely on these weights for standard analyses of the SHARE data.

The basic strategy adopted by SHARE for handling problems of unit nonresponse and sample attrition is the calibration approach of Deville and Särndal (1992), which is summarized in the appendix. This choice is primarily motivated by the fact that, in addition to external auxiliary information on the target population of interest, this approach requires the availability of design weights and auxiliary variables only for the subsample of respondents (but not for the nonresponding units). Moreover, it allows aligning the sample and population marginal distributions of some benchmark variables without specifying an explicit model for the response process. Under the standard missing-at-random assumption, calibrated weights may help reduce the potential selection effects due to different sources of sampling and nonsampling errors. This is therefore the set of weights that we generally recommend using in standard analyses of the SHARE data.

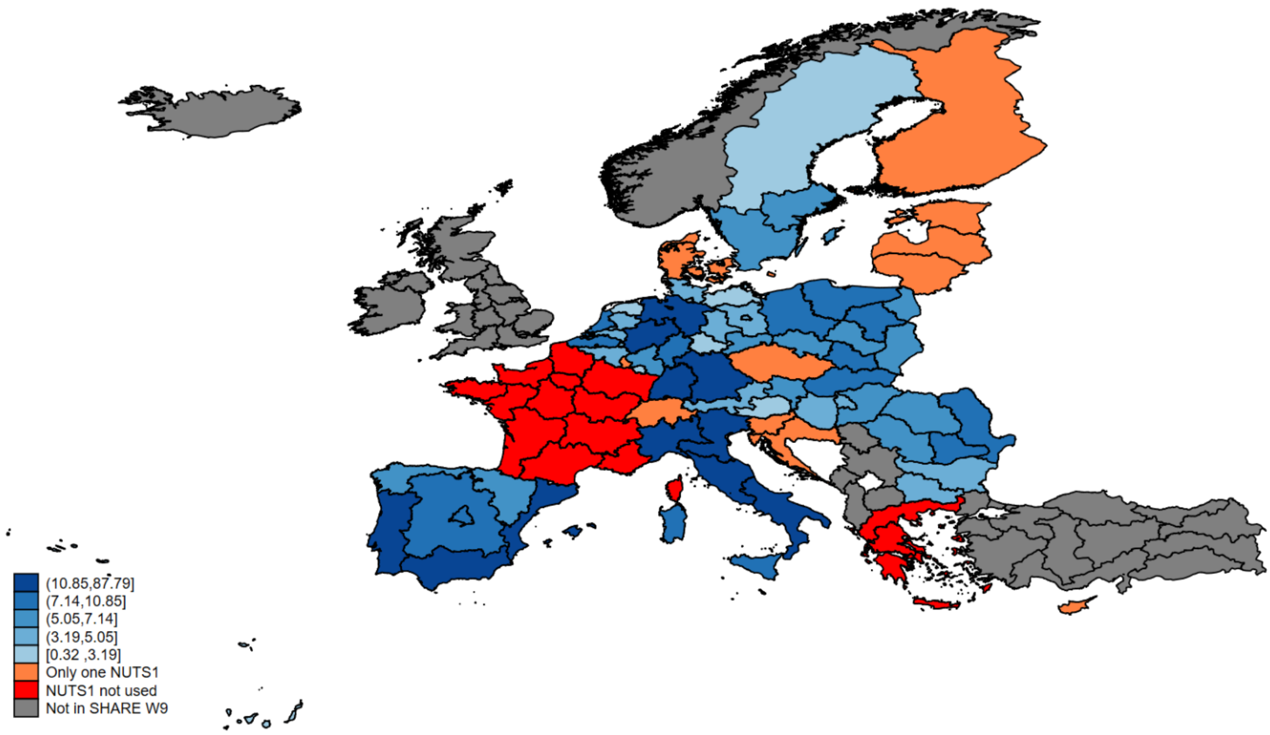
The next subsections provide further information on the calibrated weights available in the SHARE release 9.0.0. Specifically, Section 8.3.1 describes the calibrated cross-section weights of SHARE Wave 9, while Section 8.3.2 describes the calibrated longitudinal weights for selected wave combinations of the SHARE panel. Section 8.3.3 focuses on the calibrated weights for the first two waves of the SHARE Corona Survey, while finally, Section 8.3.4 presents a few additional remarks on the supplementary material for the calibrated weights.

### 8.3.1 Calibrated Cross-sectional Weights of SHARE Wave 9

The calibrated cross-sectional weights of SHARE wave 9 were computed separately by country to match the size of the national 50+ populations in 2019 that survive up to 2021. In each country, we used a logit specification of the calibration function $F(\cdot)$ and a set of population margins for gender-age groups (i.e., males and females in the age classes ([50 – 59], [60 – 69], [70 – 79], [80+]). Mortality of the target population was taken into account by subtracting from each population margin the corresponding number of deaths between 2019 and 2021. Table 8.1 in the appendix shows the resulting set population margins separately by country.

16 The second SHARE Corona Survey also includes 1,216 End-of-Life interviews which are ignored in the construction of weights and imputations.

*Figure 8.1: NUTS1 Population Margins for the Calibrated Cross-Sectional Weights of Wave 9*




In 12 countries (Austria, Belgium, Bulgaria, Germany, Hungary, Italy, the Netherlands, Poland, Portugal, Romania, Spain, and Sweden), we included an additional set of population margins for the 2016 NUTS1 regional areas as illustrated in Figure 1 (Israel is excluded from the figure). This additional set of calibration margins was ineffective in all countries containing only one NUTS1 region.[17] In Greece, NUTS1 calibration margins were excluded because of unsolved inconsistencies in the recoding of NUTS1 codes over time. In Israel, where no NUTS nomenclature is available, we used instead an additional set of calibration margins for three population groups: Jewish Israeli, Arab Israeli, and immigrants from the former USSR. Population data about the calibration margins come from the Central Bureau of Statistics for Israel and the EUROSTAT regional database for all other countries.

As usual, calibrated cross-sectional weights are computed at the individual level for inference to the target population of individuals and at the household level for inference to the target population of households. At the individual level, we assigned an individual-specific weight to each 50+ respondent that depends on the household design weight and the respondent's set of calibration variables (namely, gender, age class, and NUTS1 code). At the household level, we assigned instead a common calibrated weight to all interviewed household members which depends on the household design weight and the set of calibration variables for all 50+ respondents in that household.

By construction, calibrated cross-sectional weights are missing for respondents younger than 50 years (i.e., age-ineligible partners of an age-eligible respondent), for those with missing information on the calibration variables, and for those with missing sampling design weights (i.e., respondents from households for which we do not have sampling frame information). However, the number of these cases is negligible.

### 8.3.2 Calibrated Longitudinal Weights of the SHARE Panel

In addition to calibrated cross-sectional weights, the SHARE release 9.0.0 also includes calibrated longitudinal weights for the purposes of panel data analyses. Although these weights are based on the same calibration procedure, they differ from the cross-sectional weights in two important respects. First, calibrated longitudinal weights are usually computed for the balanced subsample of respondents who have participated in at least two waves of the study. Second,

17 That is the case in Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Latvia, Lithuania, Luxembourg, Malta, Slovakia, Slovenia, and Switzerland.

since mortality is a source of attrition that affects both the sample and the population, calibrated longitudinal weights aim to reproduce the target population at the beginning of a reference period that survives up to the end of the period considered (see, e.g., Lynn, 2009). As discussed in Section 8.2.1, SHARE Wave 9 is somehow an exception. Due to the lack of new refreshment samples in Wave 9, its target population coincides with that reproduced by the calibrated longitudinal weights of the wave combination 8 – 9. However, these two sets of calibrated weights differ in relation to their subsamples of respondents and their sets of population margins.

To simplify the structure of the public release of the data, we still provide calibrated longitudinal weights only for selected wave combinations of the SHARE panel. Those available in the SHARE release 9.0.0 are the 8 possible couples of any two adjacent waves (i.e., the wave combinations 1 – 2, 2 – 3, 3 – 4, 4 – 5, 5 – 6, 6 – 7, 7 – 8, and 8 – 9) and the fully balanced panel (i.e., the wave combination 1 – 2 – 3 – 4 – 5 – 6 – 7 – 8 – 9). The weights of the generic wave combination $t - \cdots - s$ were always computed separately by country to represent the national 50+ populations of Wave *t* that survive up to the interview year of Wave *s*. For example, the wave combination 1 – 2 allows representing the 50+ national populations in 2004 that survive up to 2006, while the fully balanced panel allows representing the national 50+ populations in 2004 that survive up to 2021.

For the calibrated longitudinal weights of two adjacent waves, we used a logit specification of the calibration function $F(\cdot)$ and a set of calibration margins for the size of the target population across eight gender-age groups (i.e., males and females with age at the time of the starting wave in the four classes [50 – 59], [60 – 69], [70 – 79] and [80+]). Compared to the cross-sectional weights of previous waves, we did not control for the 2016 NUTS1 calibration margins due to the smaller number of observations usually available in the national longitudinal subsamples. Moreover, we always accounted for the mortality of the target population by subtracting from each calibration margin the corresponding number of deaths between Waves *t* and *s*. The gender-age population margins of the wave combination 8 – 9 coincides with those presented in Table 8.1 (see appendix). Those of the other wave combinations can be found in the SHARE methodology books of previous waves (available at the SHARE-ERIC website www.share-eric.eu/).

For the calibrated longitudinal weights of the fully balanced panel, we further restricted the set of calibration margins to the six gender-age groups reported in Table 8.2 (i.e., males and females with age in 2004 in the three classes [50 – 59], [60 – 69], and [70+]; see appendix).

As with the calibrated cross-sectional weights, calibrated longitudinal weights are available both at the individual level and at the household level. For the individual weights, the balanced sample consists of respondents interviewed in each wave of the selected wave combination. For the household weights, the balanced sample consists of households with at least one eligible member interviewed in each wave of the selected wave combination. These definitions imply that the balanced sample of households is larger than the balanced sample of individuals. For example, couples with one partner participating in Wave 8 and the other partner participating in Wave 9 belong to the balanced sample of households for the wave combination 8 – 9, even though none of the two partners belongs to the corresponding balanced panel of individuals.

### 8.3.3 Calibrated Cross-sectional and Longitudinal Weights of the SHARE Corona Survey

The SHARE release 9.0.0 includes two sets of calibrated cross-sectional weights for the first two waves of the SHARE Corona Survey and a set of calibrated longitudinal weights for the balanced panel of respondents who participated in both waves of the study.

A description of the calibrated cross-sectional weights for the first SHARE Corona Survey can be found in De Luca et al. (2021). As for release 8.0.0, the new release 9.0.0 includes separate sets of calibrated weights for the CAPI, CATI, and CAPI&CATI subsamples. The target population of the last two subsamples has been however redefined as the 50+ population in 2016 that survives up to 2020. As usual, the calibrated cross-sectional weights of each subsample were computed separately by country using a logit specification of the calibration function, a first set of population margins for the gender-age groups (i.e., males and females in the age classes classes [50 – 59], [60 – 69], [70 – 79], [80+]), and a second set of population margins for the 2016 NUTS1 regional areas. The country-specific population margins of the gender-age groups are presented in Table 8.3 (see appendix). The weights of each subsample were also defined at the individual level for inference to the target population of individuals and at the household level for inference to the target population of households.

For the calibrated cross-sectional weights of the second SHARE Corona Survey, we maintained the distinction between individual-level and household-level weights, but not the distinction between the CAPI, CATI, and CAPI&CATI subsamples. These weights were computed for the cross-sectional sample of 49,254 respondents and 33,109 households who participated in the CATI of the second wave, irrespective of whether they also participated in the CATI of the first wave. The population margins are like those of the calibrated cross-section weights of the first wave, but they now refer to the national 50+ populations in 2016 that survive up to 2021 (see Table 8.4 in the appendix).

Calibrated longitudinal weights were computed for the balanced panel of 48,357 respondents and 33,109 households who participated in the first and second SHARE Corona Survey. Compared to the two cross-sectional samples, this sample excludes the 9,203 respondents who participated only in the first and the 897 respondents who participated only in the second SHARE Corona Survey. The target population coincides with that of the second wave, but the calibrated longitudinal weights were constructed by controlling for the population margins of the gender-age groups only.

### 8.3.4 Supplementary Material and User Guide on Calibrated Weights

Since the SHARE panel now consists of nine waves, one can compute many different types of calibrated longitudinal weights depending on the selected combination of waves and the selected unit of analysis (either individuals or households). In addition, one can compute many different types of calibrated cross-sectional weights for specific subsamples of the data collected in each regular wave of the panel or other related studies, such as the SHARELIFE interviews of waves 3 and 7 or the two waves of the SHARE Corona Survey. These considerations make it clear why the strategy of providing all possible calibrated cross-sectional and longitudinal weights is not feasible, especially in the future when additional waves will be available. For cross-sectional studies based on specific subsamples and longitudinal studies based on other wave combinations, users are required to control for the potential selection effects of unit nonresponse and attrition by computing their own calibrated weights or by implementing some alternative correction methods.

To support users in the nontrivial methodological task, we provide a set of Stata do-files and ado-files that illustrate step-by-step how to compute calibrated cross-sectional and longitudinal weights. Our supplementary material on calibrated weights also includes a dataset with updated information on population size and number of deaths by year, gender, age, and NUTS1 code. Registered users can download this supplementary material on calibrated weights from the SHARE data dissemination website, under the link "Generate Calibrated Weights Using Stata (2020)". A discussion of these step-by-step operations can also be found in the accompanying user guide "Computing Calibrated Weights".

## 8.4 Imputations

Let us now consider the imputation strategies employed to deal with the missing values generated by item nonresponse errors. Section 8.4.1 focuses on the imputations of missing values in SHARE Main Wave 9, while Section 8.4.2 focuses on the imputations of missing values in the SHARE Corona Survey 2.

### 8.4.1 Imputations of Missing Values in SHARE Main Wave 9

Imputations of missing values due to item non-response errors in the regular face-to-face interview of Wave 9 were constructed using the same general procedure adopted in the previous regular waves of SHARE (see, e.g., De Luca et al., 2015). However, we adapted the imputation model to the specific features of the Main Wave 9 interview in terms of branching, skip patterns, proxy interviews, country-specific deviations from the generic version of the questionnaire, and availability of partial information from the sequence of unfolding bracket questions. Moreover, we also attempted to preserve as much as possible the comparability of the imputations across different waves of the SHARE panel. The imputation procedure is essentially based on either the hot-deck method or the fully conditional specification (FCS) method, depending on the prevalence of missing values for the variables collected in the Main Wave 9 interview.

**Hot-deck Imputations**
In SHARE, we use the hot-deck method for variables affected by negligible fractions of missing values (usually, much less than 5 percent of the respondents eligible to answer a specific item on the CAPI questionnaire). This method consists of replacing the missing values in one or more variables for a non-respondent (called the recipient) with the observed values in the same variables obtained from a respondent (called the donor) who is "similar" to the recipient according to some metric (see, e.g., Andridge and Little, 2010).

In Main Wave 9, we computed hot-deck imputations in an early stage, separately by country, and according to a convenient order that accounts for branching and skip patterns in the various modules of the CAPI questionnaire. Donors were selected randomly within imputation classes based on observed auxiliary variables. We imputed first basic socio-demographic characteristics such as age and year of education, which contained very small fractions of missing values. These characteristics were then used as auxiliary variables to impute other variables. Our baseline set of auxiliary variables consisted of country, gender, five age classes ([– 49], [50 – 59], [60 – 69], [70 – 79], [80+]), five groups for years of education ([– 5], [6–10], [11–15], [16–20], [21+]), and two groups for self-reported good/bad health. For some variables, we exploited a larger set of auxiliary variables. For example, we also used the number of children to impute the number of grandchildren and an indicator for being hospitalized overnight during the last year to impute other health-related variables. Variables that are known to be logically related, such as respondent's weight, height, and body mass index, were imputed jointly.

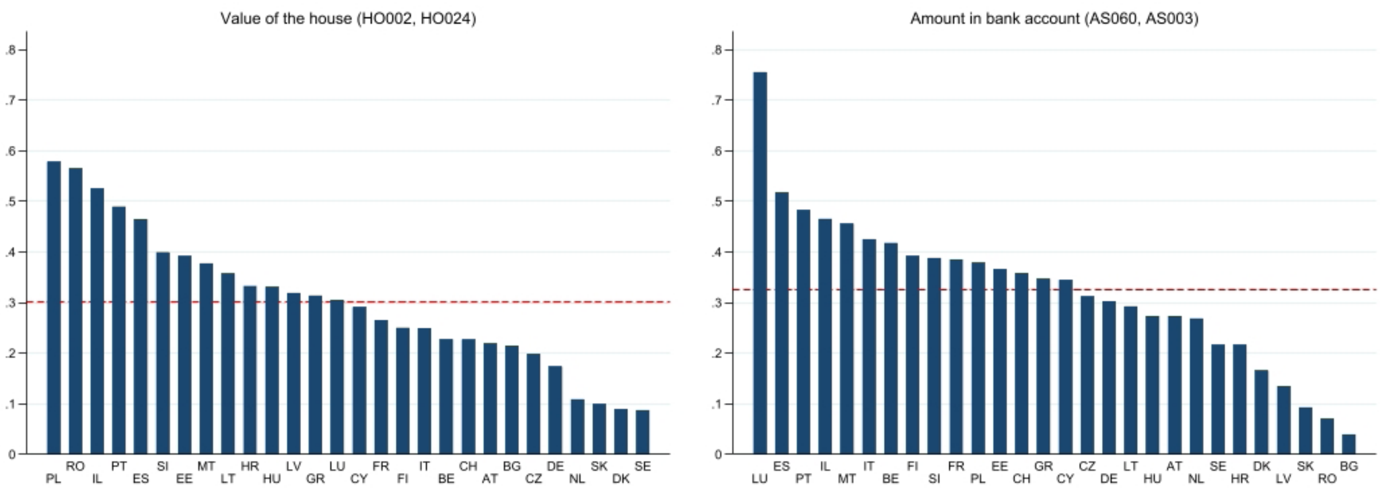
**FCS Imputations**

In the second stage of the imputation procedure, we dealt with the more worrisome issue of item non-response in monetary variables, such as income from various sources, real and financial assets, and consumption expenditures, which were collected by retrospective and open-ended questions that are sensitive and difficult to answer precisely.

Figure 8.2 shows the item non-response rates of two monetary variables: "Value of the house" (HO002, HO024), and "Amount in bank accounts" (AS060, AS003). For the first variable, the percentage of missing values among the eligible respondents ranges from a minimum of 8 percent in Denmark and Sweden to a maximum of 57 percent in Poland (30 percent on average). For the second variable, the item non-response rate ranges from a minimum of 4 percent in Bulgaria to a maximum of 76 percent in Luxembourg (32 percent on average). Similar patterns of item non-response were also observed in the previous waves (see, for example, De Luca et al., 2021). Thus, item non-sampling errors show some degree of persistency both over time and over country.

Since Wave 1, we handled these sizeable fractions of missing values on monetary variables by the FCS method of van Buuren et al. (1999). This method exploits a Gibbs sampling algorithm that imputes a set of variables jointly and iteratively through a sequence of regression models. Assume we want to impute arbitrary patterns of missing values on a set of J variables. At each step of the iterative process, we impute the missing values on the jth variable (j=1,…,J) by drawing from the predictive distribution of a regression model that includes as predictors the most updated imputations of the other J-1 variables (as well as other fully observed predictors). The process is applied sequentially to the whole set of J variables and is repeated in a cyclical manner by overwriting at each iteration the imputed values computed in the previous iteration. Despite a lack of rigorous theoretical justification (see, e.g., Arnold et al., 1999, 2001; van Buuren, 2007), the FCS method is one of the most popular multivariate imputation procedures due to its flexibility in handling complicated data structures and its ability to preserve the correlations of the imputed variables (Raghunathan et al., 2001; van Buuren et al., 2006). Comparisons of the FCS method with other multivariate imputation techniques can be found in Lee and Carlin (2010).

*Figure 8.2: Item Nonresponse Rates for "Value of the House" and "Amount in Bank Accounts" by Country*



In Main Wave 9, we computed FCS imputations separately by country and household type. The household types considered were singles and third respondents (*sample 1*), couples with both partners interviewed (*sample 2*), and all couples with and without a non-responding partner (*sample 3*). The distinction between the first two samples was primarily motivated by the fact of using socio-demographic characteristics of the partner of the designed respondent as additional predictors to impute the missing monetary amounts within couples. The overlapping partition of the last two samples was instead motivated by the need to impute properly total household income in the couples with a non-responding partner.

The set of monetary variables imputed jointly with the Gibbs sampling algorithm was country- and sample-specific as we required a minimum number of donor observations for estimating the regression model associated with each variable.[18] Variables that did not satisfy this requirement were imputed first (either by hot-deck or by regression imputations) and then used as fully observed predictors for computing the FCS imputations of missing values in the other monetary variables.

The imputation of each monetary variable was typically based on a two-stage model that involved a probit model for ownership and a linear regression model for the amount conditional on ownership.[19] Depending on eligibility and ownership, we converted (if needed) non-zero values of monetary variables in annual Euro amounts to avoid modelling differences in the time reference periods of the various variables and the national currencies of non-Euro countries.

In an early stage of the imputation process, we also symmetrically trimmed 2 percent of the complete cases from the country-specific distribution of annual Euro amounts to exclude (and then impute) outliers that may have a large influence on survey statistics. Moreover, we applied logarithm or inverse hyperbolic sine transformations to reduce skewness in the right tails of the conditional distribution of each monetary variable[20].

The set of fully observed predictors was also sample-specific. For singles and third respondents (*sample 1*), our set of predictors consists of gender, age, years of education, self-perceived health, number of children, number of chronic diseases, score of the numeracy test, employment status, and willingness to answer (as perceived by the interviewer in the IV module of the CAPI instrument). For couples with both partners interviewed (*sample 2*), we added a similar set of predictors for the partner of the designed respondent. For couples with a non-responding partner (those remaining in *sample 3* after excluding the couples in *sample 2*), we restricted the additional set of predictors referring to the non-responding partner to age and years of education only.[21]

Imputations of the monetary amounts were always constrained to fall within individual-level bounds that incorporated the partial information available on the missing observations (e.g., country-specific thresholds used to trim outliers in the tails of the observed distribution of each monetary variable, bounds obtained from the sequence of unfolding bracket questions asked by design to non-respondents of open-ended monetary variables and lower bounds based on the observed components of aggregated monetary variables).

As usual, the imputation of total household income received special attention because the CAPI questionnaire provides two alternative measures of this variable. The first measure (*thinc*) can be obtained by a suitable aggregation at the household level of all individual income components, while the second (*thinc2*) can be obtained via the one-shot question on monthly household income (HH017). As argued by De Luca et al. (2015), it is not easy to find strong arguments to prefer one measure over the other. Moreover, the availability of two alternative measures may greatly improve the imputation process because each measure could contribute relevant information on the missing values of the other measure. To avoid understating the first measure of total household income in couples with a non-responding partner, we adopted the following three-stage algorithm:

> **Stage 1.** For singles and third respondents (*sample 1*), we imputed first all monetary variables by the FCS method as discussed above. At the end of each iteration of the Gibbs sampling algorithm, we also computed total household income (*thinc*), household net worth *(hnetw*), and total household expenditure (*thexp*) by suitable aggregations of the imputed income, wealth, and expenditure items. Next, we imputed the second measure of total household income (*thinc2*) using as predictors *thinc*, *hnetw*, *thexp*, and

18 The minimum number of observations was equal to 100 in sample 1 and 150 in samples 2 and 3.

19 For the few variables without an ownership question, such as food at home expenditure (CO002) and total household income (HH017), we used a simple linear regression model.

20 We apply the log transformation to variables with positive support and the inverse hyperbolic sine transformation to variables that may take negative values (e.g., income from self-employment, bank account, and value of own business).

21 In the few cases where the number of donor observations available in the estimation step was lower than 30, we employed a smaller subset of predictors, namely gender, age, years of education, and self-reported health.

the set of socio-demographic characteristics of the household respondent. The imputed values of *thinc2* were constrained to fall in the bounds derived from the sequence of unfolding bracket questions for the variable HH017.

**Stage 2.** For couples with both partners interviewed (*sample 2*), the imputation strategy is similar to the one adopted in stage 1 for the sample of singles and third respondents (*sample 1*). The only difference is that, at each iteration of the Gibbs sampling algorithm, we employed a larger set of predictors that also included the socio-demographic characteristics and the most updated imputations of the monetary variables of the partner of the designed respondent.

**Stage 3.** Imputations of all monetary variables for the subsample of couples with both partners interviewed were obtained in stage 2. In stage 3, these couples were included in the imputation sample only as donor observations to impute the missing values in monetary variables for the remaining subsample of couples with a non-responding partner. As before, we imputed first all monetary variables for the responding partners using the FCS method. Unlike stage 2, the predictors referring to the non-responding partner now consisted, however, of age and years of education only. At the end of each iteration of the Gibbs sampling algorithm, we also imputed the *thinc2* using *hnetw*, *thexp*, and socio-demographic characteristics of the responding partner as predictors and the bounds obtained from the sequence of unfolding bracket questions for the variable HH017. Finally, we imputed thinc using *thinc2*, *hnetw*, *thexp*, and the set of socio-demographic characteristics of the responding partner as predictors, couples with two partners interviewed as donors, and the sum of imputed individual income sources of the responding partner as a lower bound.

To account for the additional variability generated by the imputation process, we provide five imputations of the missing values by independent replications of the hot-deck and FCS methods. Notice that neglecting this additional source of uncertainty by selecting only one of the five available replicates in the generated dataset of imputations may result in misleadingly precise estimates. After an initial set of burn-in iterations, convergence of the Gibbs sampling algorithm for FCS imputations was assessed by the Gelman-Rubin criterion (see, e.g., Gelman and Rubin, 1992, and Gelman et al., 2004) applied to the mean, the median, and the 90th percentile of the five imputed distributions of each monetary variable.

### 8.4.2 Imputations of Missing Values in the SHARE Corona Survey 2

Since item nonresponse rates in the CATI data of the Second SHARE Corona Survey were generally much less than 5 percent, most variables were imputed by the hot-deck method. We used the FCS method only for 15 variables collected in Section E (Economic situation) and Section W (Work) of the questionnaire administered in the second wave. As for SCS1, the variables collected in these two sections suffer from somewhat larger amounts of item nonresponse. Moreover, Section E contains missing data by design due to the presence of a filter in the routing (see De Luca et al., 2021). Regarding possible issues of data comparability across the first and the second SHARE Corona Survey, we note that seemingly similar questions may present relevant differences in terms of question wording, answer categories, time-reference period, branching, and skip patterns. To mark these differences within the generated dataset of imputations, we assigned slightly different variable names to items whose comparability is more doubtful.

**Hot-deck Imputations**

We first computed 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 of the second wave. The imputation classes for this method were generally 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 ([– 5], [6–10], [11–15], [16–20], [21+]), a binary indicator for good self-perceived health, and a binary indicator for changes in the self-perceived health status during the last three months.[22] 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. For some variables, we employed a larger set of auxiliary variables. For example, we used one additional binary indicator for keeping distance from others in public when imputing several variables included in Section H (Health and health behaviour), Section C (Corona-related infection), and Section Q (Quality of healthcare) of the CATI questionnaire of the second SHARE Corona Survey. Furthermore, we jointly imputed missing values of the variables that are logically related. For example, we jointly imputed variables related to illness or health conditions since the last interview (CAH004) in Section H, those related to the COVID-19 symptoms (CAC102, CAC103) in Section C, and those related to forwent medical treatment

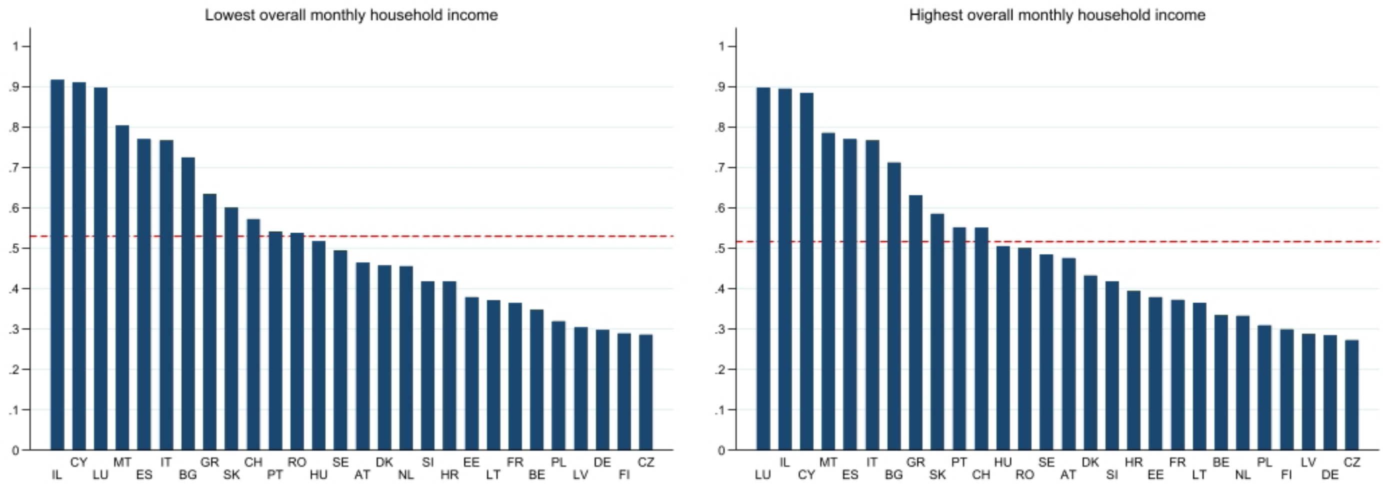
22 The information on years of education was obtained from the most recent CAPI data collected in the regular waves of SHARE.

since the outbreak (CAQ105 and CAQ106) in Section Q. In total, we imputed sequentially about 200 variables. As for the hot-deck imputations of the CAPI data collected in the regular SHARE waves, the imputation databases of the first and the second SHARE Corona Survey contain five multiple imputations of the missing values and a flag variable for each imputed variable, which allows users to identify the imputed observations.

**FCS Imputations**

After hot-deck imputations, we constructed FCS imputations for fifteen variables: four of them related to changes in hours of work (namely CAW121, CAW122, CAW124, and CAW125), and the other eleven related to changes in the financial situation of the household (namely CAE100, CAE105, CAE107, CACO107, CAE111, CAE112, CAE103, and CAE104). As shown in Figure 8.3, the two most worrisome variables are the lowest (CAE107) and the highest (CAE105) overall amounts of monthly household income after taxes and contributions. In particular, the first respondent of each household was first asked whether monthly household income had been the same every month since the last interview (CAE100). Respondents who provided a negative answer to this question were then asked to report the lowest and the highest overall amount of monthly household income. The unweighted cross-country average of the item nonresponse rates for these two variables are 53 and 51 percent, respectively. In Luxembourg, Israel, Cyprus, and Malta, where the item nonresponse rates are around 90 percent, we adopted a country-pooling strategy to increase the extremely low number of donors.[23]

*Figure 8.3: Item Nonresponse Rates for Lowest and Highest Overall Monthly Household Income by Country*




23 Specifically, we increased the number of donors by pooling Malta with Italy, Cyprus with Greece, Luxembourg with both Belgium and the Netherlands, and Israel with all other European countries.

Except for these more problematic cases, FCS imputations were constructed separately by country. At each iteration of the Gibb sampling algorithm, we used a linear regression model for the continuous variables (CAE105 and CAE107), a simple hot-deck method for the lowest and the highest hours of work (CAW122 and CAW125), a logit model for five binary variables (CAW121, CAW124, CAE100, CAE111, and CAE112), a multinomial logit model for the categorical variable CACO107, and a multivariate hot-deck method for the six binary indicators related to financial support received since the outbreak of the pandemic (CAE103 and CAE104). For the variables CAE105 and CAE107, 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. In addition to the variables imputed jointly within the Gibb sampling, our baseline set of observed predictors consists of age, years of education, and binary indicators for female respondents, living with a spouse/partner, and good self-perceived health. For all variables of Section E, we also used a binary indicator for being retired. For the variables imputed by either simple or multivariate hot-deck methods, all continuous predictors within the Gibb sampling were discretized to form the imputation classes. In some cases, we imposed a set of country- and item-specific exclusion restrictions to avoid possible problems of collinearity, imprecise estimates, and convergence problems in the context of non-linear models. As for the other types of imputations provided by SHARE, we always provide five multiple imputations of the missing values. After an initial set of 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.

***References***

*Andridge, R. and Little, R. (2010). A review of hot deck imputation for survey non-response. International Statistical Review, 78: 40-64.*

*Arnold, B.C., Castillo, E. and Sarabia, J.M. (1999). Conditional specification of statistical models. New York: Springer.*

*Arnold, B.C., Castillo, E. and Sarabia, J.M. (2001). Conditionally specified distributions: An introduction. Statistical Science, 16: 249-274.*

*De Luca, G., Celidoni, M. and Trevisan, E. (2015). Item nonresponse and imputation strategies in SHARE Wave 5. In Malter, F. and Börsch-Supan, A. (Eds.), SHARE Wave 5: Innovations & Methodology, Munich (pp. 85-100): MEA, Max Planck Institute for Social Law and Social Policy.*

*De Luca, G., Li Donni, P. and Rashidi, M. (2021). Weights and imputations in the first SHARE Corona Survey. In M. Bergmann & A. Börsch-Supan (Eds.). SHARE Wave 8 Methodology: Collecting cross-national survey data in times of COVID-19. (pp. 175-178). MEA, Max Planck Institute for Social Law and Social Policy.*

*Deville, J.C. and Särndal C.E. (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association, 87: 376-382.*

*Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7: 457-511.*

*Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004). Bayesian data analysis, 2nd ed. Boca Raton, FL: Chapman and Hall.*

*Haziza, D. and Lesage, E. (2016). A discussion of weighting procedures for unit nonresponse. Journal of Official Statistics, 32: 129–145.*

*Lee, K.J. and Carlin, J.B. (2010). Multiple imputation for missing data: Fully conditional specification versus multivariate normal imputation. American Journal of Epidemiology, 171: 624-632.*

*Lundström, S. and Särndal, C.E. (1999). Calibration as a standard method for treatment of nonresponse. Journal of Official Statistics, 15: 305-327.*

*Lynn, P. (2009). Methods for longitudinal surveys. In P. Lynn (Ed.). Methodology of longitudinal surveys (pp. 1-19). Chichester: Wiley.*

*Raghunathan, T.E., Lepkowski, J.M., Van Hoewyk, J. and Solenberger P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. Survey Methodology, 27: 85–95.*

*Van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. Statistical Methods in Medical Research, 16: 219-242.*

*Van Buuren, S., Boshuizen, H.C. and Knook, D.L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. Statistics in Medicine, 18: 681–694.*

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

# APPENDIX

*Table 8.1: Gender-age Population Margins for the Calibrated Cross-sectional Weights of Wave 9 and the Longitudinal Weights of Waves 8-9*

| Country | Men | | | | Women | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | [50-59] | [60-69] | [70-79] | [80+] | [50-59] | [70-79] | [60-69] | [80+] | |
| AT | 685,340 | 461,937 | 325,998 | 123,942 | 686,610 | 506,881 | 412,500 | 222,857 | 3,426,065 |
| BE | 793,962 | 630,772 | 387,164 | 183,509 | 787,172 | 666,687 | 467,759 | 326,934 | 4,243,959 |
| BG | 457,078 | 405,776 | 238,029 | 81,483 | 470,241 | 505,399 | 378,070 | 164,265 | 2,700,341 |
| CH | 641,999 | 448,773 | 310,744 | 132,226 | 632,659 | 468,383 | 364,876 | 224,335 | 3,223,995 |
| CY | 53,082 | 44,783 | 27,965 | 10,353 | 54,592 | 47,222 | 32,988 | 14,863 | 285,848 |
| CZ | 656,982 | 617,225 | 375,945 | 105,560 | 648,816 | 701,720 | 528,147 | 221,220 | 3,855,615 |
| DE | 6,731,208 | 4,914,177 | 3,384,703 | 1,807,698 | 6,685,991 | 5,273,764 | 4,098,271 | 3,050,914 | 35,946,726 |
| DK | 397,067 | 316,838 | 245,718 | 80,616 | 394,397 | 331,083 | 277,040 | 129,108 | 2,171,867 |
| EE | 80,707 | 66,245 | 35,416 | 14,635 | 88,723 | 91,223 | 67,289 | 45,067 | 489,305 |
| ES | 3,392,398 | 2,445,652 | 1,626,734 | 836,157 | 3,468,783 | 2,673,512 | 2,020,060 | 1,486,861 | 17,950,157 |
| FI | 362,575 | 342,088 | 230,952 | 84,904 | 364,457 | 368,454 | 283,069 | 158,951 | 2,195,450 |
| FR | 4,245,922 | 3,687,740 | 2,339,133 | 1,157,572 | 4,476,548 | 4,155,613 | 2,875,030 | 2,181,218 | 25,118,776 |
| GR | 705,218 | 583,989 | 421,097 | 246,435 | 782,048 | 667,962 | 524,756 | 357,557 | 4,289,062 |
| HR | 279,379 | 252,297 | 132,707 | 52,278 | 295,085 | 290,652 | 198,331 | 112,325 | 1,613,054 |
| HU | 575,738 | 544,451 | 287,483 | 92,341 | 620,259 | 711,482 | 480,236 | 233,180 | 3,545,170 |
| IL | 403,051 | 339,665 | 198,837 | 87,840 | 424,112 | 383,298 | 243,479 | 134,756 | 2,215,038 |
| IT | 4,514,206 | 3,410,822 | 2,543,552 | 1,236,337 | 4,713,506 | 3,751,864 | 3,102,241 | 2,190,314 | 25,462,842 |
| LT | 192,243 | 134,684 | 71,227 | 31,897 | 225,036 | 195,438 | 145,808 | 92,876 | 1,089,209 |
| LU | 45,320 | 29,436 | 16,515 | 7,000 | 41,875 | 29,554 | 18,998 | 12,445 | 201,143 |
| LV | 121,961 | 94,058 | 50,755 | 19,700 | 143,723 | 136,922 | 106,464 | 63,077 | 736,660 |
| MT | 29,859 | 28,862 | 18,794 | 6,244 | 28,711 | 29,410 | 21,902 | 10,633 | 174,415 |
| PL | 2,262,420 | 2,239,510 | 960,991 | 381,485 | 2,385,769 | 2,727,782 | 1,486,890 | 906,503 | 13,351,350 |
| PT | 1,206,983 | 1,084,184 | 528,781 | 229,121 | 1,220,036 | 1,352,443 | 817,677 | 451,340 | 6,890,565 |
| RO | 647,257 | 542,678 | 440,297 | 160,803 | 631,981 | 552,696 | 478,621 | 253,024 | 3,707,357 |
| SE | 151,920 | 131,963 | 69,625 | 27,806 | 149,092 | 138,066 | 91,484 | 59,247 | 819,203 |
| SI | 342,438 | 299,902 | 132,279 | 40,281 | 355,856 | 362,136 | 213,564 | 96,928 | 1,843,384 |
| SK | 685,340 | 461,937 | 325,998 | 123,942 | 686,610 | 506,881 | 412,500 | 222,857 | 3,426,065 |

*Table 8.2: Gender-age Population Margins for the Longitudinal Weights of the Fully Balanced Panel (Waves 1-9)*

| Country | Men | | | Women | | | Total |
|---|---|---|---|---|---|---|---|
| | [50-59] | [60-69] | [70+] | [50-59] | [60-69] | [70+] | |
| AT | 379,050 | 237,429 | 43,660 | 433,645 | 327,882 | 103,694 | 1,525,360 |
| BE | 533,138 | 270,010 | 62,093 | 587,102 | 369,793 | 149,900 | 1,972,036 |
| CH | 410,876 | 223,289 | 49,922 | 438,326 | 284,517 | 110,289 | 1,517,219 |
| DE | 4,047,658 | 3,106,488 | 623,620 | 4,462,982 | 4,062,234 | 1,300,948 | 17,603,930 |
| DK | 303,041 | 148,036 | 22,104 | 325,012 | 185,908 | 52,476 | 1,036,577 |
| ES | 1,961,861 | 1,104,283 | 281,005 | 2,290,760 | 1,593,952 | 658,554 | 7,890,415 |
| FR | 3,112,718 | 1,548,742 | 461,584 | 3,614,349 | 2,154,995 | 1,114,092 | 12,006,480 |
| IT | 2,994,456 | 1,882,995 | 381,023 | 3,384,385 | 2,593,872 | 934,061 | 12,170,792 |
| SE | 528,640 | 276,188 | 53,184 | 550,267 | 331,073 | 115,289 | 1,854,641 |

*Table 8.3: Gender-age Population Margins for the Calibrated Cross-sectional Weights of the SHARE Corona Survey 1*

| Country | Men | | | | Women | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | [50-59] | [60-69] | [70-79] | [80+] | [50-59] | [60-69] | [70-79] | [80+] | |
| AT | 637,866 | 417,832 | 281,282 | 86,511 | 647,778 | 470,086 | 368,118 | 172,444 | 3,081,917 |
| BE | 775,927 | 590,542 | 324,454 | 130,053 | 778,267 | 635,176 | 415,517 | 254,297 | 3,904,233 |
| BG | 455,181 | 390,366 | 194,602 | 56,042 | 483,161 | 510,731 | 322,957 | 113,894 | 2,526,934 |
| CH | 606,363 | 422,558 | 264,036 | 93,488 | 595,799 | 449,237 | 323,006 | 172,686 | 2,927,173 |
| CY | 51,520 | 41,531 | 23,019 | 6,510 | 53,722 | 44,629 | 27,709 | 10,060 | 258,700 |
| CZ | 649,368 | 608,454 | 279,703 | 73,414 | 657,256 | 718,311 | 420,401 | 165,254 | 3,572,161 |
| DE | 6,351,644 | 4,304,875 | 3,183,899 | 970,594 | 6,377,015 | 4,748,270 | 4,087,726 | 1,886,272 | 31,910,295 |
| DK | 375,143 | 315,119 | 197,057 | 51,648 | 375,312 | 332,606 | 231,240 | 92,513 | 1,970,638 |
| EE | 79,069 | 59,430 | 30,157 | 9,146 | 91,564 | 87,293 | 63,806 | 31,283 | 451,748 |
| ES | 3,166,767 | 2,239,962 | 1,406,427 | 626,791 | 3,260,050 | 2,492,484 | 1,825,440 | 1,180,199 | 16,198,120 |
| FI | 361,707 | 342,827 | 176,741 | 56,248 | 367,590 | 376,644 | 231,207 | 118,382 | 2,031,346 |
| FR | 4,148,876 | 3,590,836 | 1,867,857 | 843,581 | 4,426,446 | 4,074,812 | 2,418,113 | 1,714,987 | 23,085,508 |
| GR | 670,778 | 557,650 | 379,700 | 171,774 | 752,838 | 642,214 | 493,750 | 253,767 | 3,922,471 |
| HR | 285,171 | 228,636 | 115,448 | 32,748 | 305,874 | 274,353 | 184,851 | 77,543 | 1,504,624 |
| HU | 559,830 | 505,231 | 234,902 | 63,160 | 630,608 | 685,435 | 426,715 | 167,490 | 3,273,371 |
| IL | 378,698 | 318,856 | 155,380 | 60,535 | 405,109 | 364,304 | 197,397 | 97,559 | 1,977,838 |
| IT | 4,250,255 | 3,346,875 | 2,225,989 | 865,784 | 4,484,039 | 3,719,901 | 2,847,202 | 1,689,294 | 23,429,339 |
| LT | 190,228 | 114,689 | 66,478 | 20,836 | 231,282 | 178,040 | 144,672 | 65,367 | 1,011,592 |
| LU | 40,908 | 25,880 | 13,712 | 4,899 | 38,370 | 26,287 | 16,924 | 9,556 | 176,536 |
| LV | 122,877 | 81,928 | 45,631 | 12,066 | 149,653 | 128,742 | 104,832 | 42,842 | 688,571 |
| MT | 29,681 | 27,649 | 14,125 | 3,872 | 29,526 | 29,043 | 17,443 | 7,425 | 158,764 |
| NL | 1,206,766 | 992,292 | 527,420 | 156,221 | 1,203,965 | 1,018,612 | 615,614 | 291,328 | 6,012,218 |
| PL | 2,409,916 | 1,993,489 | 738,169 | 264,958 | 2,605,739 | 2,522,699 | 1,256,881 | 670,998 | 12,462,849 |
| PT | 672,599 | 541,235 | 343,108 | 122,188 | 759,000 | 651,620 | 488,444 | 251,516 | 3,829,710 |
| RO | 1,107,833 | 979,861 | 459,954 | 151,305 | 1,188,438 | 1,266,455 | 761,680 | 301,473 | 6,216,999 |
| SE | 609,338 | 543,045 | 361,950 | 110,367 | 598,517 | 560,211 | 406,527 | 191,559 | 3,381,514 |
| SI | 149,534 | 118,974 | 59,029 | 18,260 | 148,238 | 128,100 | 83,651 | 44,784 | 750,570 |
| SK | 347,489 | 263,160 | 101,331 | 27,080 | 370,518 | 333,425 | 179,838 | 68,457 | 1,691,298 |

*Table 8.4: Gender-age Population Margins for the Calibrated Cross-sectional Weights of the SHARE Corona Survey 2 and the Calibrated Longitudinal Weights of the SHARE Corona Survey 1 and the SHARE Corona Survey 2*

| Country | Men | | | | Women | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | [50-59] | [60-69] | [70-79] | [80+] | [50-59] | [60-69] | [70-79] | [80+] | |
| AT | 633,144 | 409,391 | 267,017 | 71,558 | 645,350 | 464,974 | 356,675 | 146,827 | 2,994,936 |
| BE | 770,086 | 578,734 | 307,017 | 106,992 | 774,901 | 627,990 | 401,297 | 216,740 | 3,783,757 |
| BG | 445,567 | 372,335 | 175,978 | 42,822 | 478,750 | 499,993 | 303,913 | 92,182 | 2,411,540 |
| CH | 603,331 | 416,589 | 253,099 | 78,120 | 594,095 | 445,566 | 314,566 | 149,276 | 2,854,642 |
| CY | 51,151 | 40,889 | 21,982 | 5,419 | 53,548 | 44,267 | 26,906 | 8,514 | 252,676 |
| CZ | 642,244 | 590,359 | 258,976 | 58,017 | 653,935 | 708,239 | 402,041 | 136,776 | 3,450,587 |
| DE | 6,351,644 | 4,304,875 | 3,183,899 | 970,594 | 6,377,015 | 4,748,270 | 4,087,726 | 1,886,272 | 31,910,295 |
| DK | 372,552 | 309,298 | 187,847 | 43,012 | 373,615 | 328,709 | 223,672 | 80,036 | 1,918,741 |
| EE | 77,891 | 57,532 | 28,199 | 7,599 | 91,093 | 86,224 | 61,544 | 27,033 | 437,115 |
| ES | 3,143,113 | 2,198,491 | 1,338,825 | 526,131 | 3,248,419 | 2,472,262 | 1,776,831 | 1,025,173 | 15,729,245 |
| FI | 358,951 | 336,683 | 168,425 | 47,641 | 366,302 | 373,109 | 224,627 | 103,168 | 1,978,906 |
| FR | 4,114,631 | 3,525,947 | 1,786,943 | 715,862 | 4,408,662 | 4,039,762 | 2,358,742 | 1,507,573 | 22,458,122 |
| GR | 664,823 | 546,126 | 361,242 | 145,764 | 749,908 | 636,155 | 478,187 | 215,569 | 3,797,774 |
| HR | 281,563 | 221,336 | 106,940 | 26,009 | 304,296 | 270,445 | 175,770 | 64,058 | 1,450,417 |
| HU | 549,220 | 484,796 | 216,140 | 50,728 | 625,266 | 671,979 | 404,624 | 139,064 | 3,141,817 |
| IL | 424,664 | 352,130 | 237,761 | 114,705 | 441,737 | 391,704 | 280,984 | 167,068 | 2,410,752 |
| IT | 4,223,885 | 3,288,099 | 2,113,838 | 720,360 | 4,468,949 | 3,686,305 | 2,760,311 | 1,453,787 | 22,715,534 |
| LT | 186,445 | 109,665 | 60,996 | 16,705 | 229,761 | 175,220 | 138,315 | 54,877 | 971,984 |
| LU | 40,628 | 25,430 | 13,065 | 4,093 | 38,255 | 26,019 | 16,405 | 8,336 | 172,231 |
| LV | 120,641 | 78,574 | 41,974 | 9,705 | 148,658 | 126,674 | 100,093 | 35,962 | 662,281 |
| MT | 29,477 | 27,174 | 13,449 | 3,271 | 29,412 | 28,786 | 16,923 | 6,432 | 154,924 |
| NL | 1,199,649 | 975,597 | 500,125 | 128,108 | 1,198,865 | 1,007,045 | 594,385 | 248,245 | 5,852,019 |
| PL | 2,370,911 | 1,921,576 | 680,242 | 213,200 | 2,588,726 | 2,481,567 | 1,200,171 | 568,492 | 12,024,885 |
| PT | 666,208 | 530,372 | 325,511 | 100,045 | 756,326 | 645,828 | 473,191 | 215,473 | 3,712,954 |
| RO | 1,084,217 | 937,163 | 417,871 | 119,931 | 1,178,578 | 1,241,689 | 717,539 | 248,849 | 5,945,837 |
| SE | 606,031 | 534,751 | 345,984 | 91,260 | 596,542 | 554,605 | 394,058 | 164,440 | 3,287,671 |
| SI | 148,277 | 116,316 | 55,548 | 14,758 | 147,675 | 126,781 | 80,655 | 37,733 | 727,743 |
| SK | 342,573 | 254,381 | 93,433 | 21,807 | 368,352 | 328,272 | 171,212 | 57,307 | 1,637,337 |

**The calibration approach of Deville and Särndal (1992)**

Let $U = \{1, \ldots, i, \ldots, N\}$ be a finite population of $N$ elements, from which a probability sample $s = \{1, \ldots, i, \ldots, n\} \subseteq U$ of size $n \leq N$ is drawn according to a probability-based sampling design. Unless otherwise specified, we shall assume that the inclusion probability $\pi_i = \Pr(i \in s)$ is known and strictly positive for all population units. To describe the basic ideas and the key properties of the calibration approach, we consider first the ideal situation of complete response where all units in the sample $s$ agree to participate to the survey. Then, we relax this ideal setup to describe the key implications of non-response errors on the properties of this weighting method.

The sampling design weights $w_i = \pi_i^{-1}$ are typically used to account for the randomness of the sampling process and the variability of the inclusion probabilities across sample units due to stratification and clustering strategies (additional details can be found in Chapter I.1). For example, one can estimate the population total $t_y = \sum_{i \in U} y_i$ of a variable of interest $y$ by the Horvitz-Thompson estimator:

$$\hat{t}_y = \sum_{i \in s} w_i y_i. \tag{1}$$

Under the ideal setup of complete response, this estimator is known to be design unbiased, that is $E_p(\hat{t}_y) = t_y$, where $E_p(\cdot)$ denotes the expectation with respect to the sampling design.

Let us assume now that the sampling frame or other external sources such as census data and administrative archives provide supplementary data on a q-vector of categorical auxiliary variables $x_i = (x_{i1}, \ldots, x_{iq})^\top$ with known population totals $t_x = \sum_{i \in U} x_i$. We shall refer to the auxiliary variables $x_i$ as calibration variables and to their population totals $t_x$ as calibration margins. The basic idea of the calibration approach is to determine a set of *calibrated weights* $w_i^*$ that are as close as possible to the design weights $w_i$ and that satisfy the constraints

$$\sum_{i \in s} w_i^* x_i = t_x. \tag{2}$$

Thus, given a distance function $G(w_i^*, w_i)$ and the availability of survey data on $(w_i, x_i^\top : i = 1, \ldots, n)$ and population data on the calibration margins $t_x$, the aim of the procedure is to determine the calibrated weights $w_i^*$ by minimizing the aggregate distance $\sum_{i \in s} G(w_i^*, w_i)$ with respect to $w_i^*$ subject to the q equality constraints in (2). Under some regularity conditions on the distance function $G(w_i^*, w_i)$ (see Deville and Särndal 1992), the solution of this constrained optimization problem exists, is unique and can be written as

$$w_i^* = w_i F(\eta_i), \qquad i = 1, \ldots, n, \tag{3}$$

where $\eta_i = x_i^\top \lambda$ is a linear combination of the calibration variables $x_i$, $\lambda = (\lambda_1, \ldots, \lambda_q)^\top$ is the q-vector of Lagrangian multipliers associated with the constraints (2), and $F(\cdot)$ is a calibration function, which is uniquely determined by the distance function $G(w_i^*, w_i)$.

A key feature of the calibration approach is that many traditional re-weighting methods such as post-stratification, raking, and generalized linear regression (GREG) correspond to special cases of the calibration estimator

$$\hat{t}_y^* = \sum_{i \in s} w_i^* y_i \tag{5}$$

for particular choices of the calibration function $F(\cdot)$ (or, equivalently, of the distance function $G(\cdot,\cdot)$). Table 1 in Deville and Särndal (1992) presents various functional forms for $G(w_i^*, w_i)$ and $F(\eta_i)$. The chi-square distance function $G(w_i^*, w_i) = (\mathrm{w}_i^* - \mathrm{w}_i)^2 / 2w_i$, which leads to the widely used GREG estimator, has the advantage of ensuring a closed form solution for the calibrated weights $w_i^*$. However, this distance function is unbounded and depending on the chosen set of calibration variables it may also lead to negative weights. Different specifications of the calibration function may avoid these issues, but the underlying optimization problems may not admit a solution and the Lagrange multipliers must be computed numerically. In SHARE, we rely on the logit specification of the distance function

$$G(w_i^*, w_i) \propto \left(\frac{\mathrm{w}_i^*}{\mathrm{w}_i} - l\right) \ln\left(\frac{\mathrm{w}_i^*/\mathrm{w}_i - l}{1 - l}\right) + \left(u - \frac{\mathrm{w}_i^*}{\mathrm{w}_i}\right) \ln\left(\frac{\mathrm{u} - \mathrm{w}_i^*/\mathrm{w}_i}{u - 1}\right),$$

which leads to a calibration function of the form

$$F(\eta_i; u, l) = \frac{l(u-1) + u(1-l)\exp(a\eta_i)}{\mathrm{u} - 1 + (1 - \mathrm{l})\exp(a\eta_i)},$$

where $a = [(1-l)(u-1)]^{-1}(u-l)$. Unlike other distance functions, these functional forms restrict in advance the range of feasible values for the calibrated weights by suitable choices of the lower bound *l* and the upper bound *u*. Specifically, if a solution exists, then it must satisfy the restriction $\mathrm{w_i l} \leq \mathrm{w}_i^* \leq w_i u$.

As discussed in Deville and Särndal (1992), effectiveness of the calibrated weights depends crucially on the correlation between the study variable *y* and the calibration variables *x*. In the extreme case when *y* can be expressed as a linear combination of *x*, it is clear that the calibrated estimator $\hat{t}_y^*$ gives an exact estimate of $t_y$ for every realized sample *s*. Under suitable regularity conditions, the class of calibration estimators $\hat{t}_y^*$ satisfies other desirable asymptotic properties. For example, the estimators obtained by alternative specifications of the distance function are asymptotically equivalent to the GREG estimator based on a chi-squared distance function. Thus, in large samples, calibrated weights are robust to arbitrary choices of the calibration function $F(\cdot)$.

Unfortunately, this property does not necessarily extend to the more realistic cases where survey data are affected by nonresponse errors. Previous studies by Lundström and Särndal (1999) and Haziza and Lesage (2016) suggest that in these cases alternative specifications of the calibration function $F(\cdot)$ correspond in practice to imposing different parameterization of the relationship between response and calibration variables. Moreover, statistical properties of calibration estimators depend as usual on the validity of the missing at random assumption.