Karel van den Bosch

The SHARE sample in Belgium: design, history, results

WORKING PAPER
No. 25/05
April 2025

University of Antwerp
Herman Deleeck Centre for Social Policy
https://www.uantwerpen.be/en/research-groups/csb/

The SHARE sample in Belgium: design, history, results

Karel van den Bosch[1 2*]

[1] Federal Planning Bureau
[2] Centre for Social Policy (University of Antwerp)

Working Paper No. 25/05

April 2025

Abstract

The Survey of Health, Ageing, and Retirement in Europe (SHARE), or "50+ in Europe", is a study that focuses on the European population aged 50 and over. SHARE is a panel (i.e. longitudinal survey). For Belgium it is executed by teams at the University of Liège (for the French-speaking community) and at the University of Antwerp (for the Dutch-speaking community). This report explains the design of the Belgian SHARE sample, discusses how the sample was regularly refreshed by additional samples and looks at some indicators of the representativeness of the SHARE samples for the population of older people in the French-speaking and Dutch-speaking communities of Belgium. A worrying development is that the age group 55-64 is increasingly underrepresented in the SHARE samples for both communities.

Keywords: SHARE-project, sampling, Belgium

Jel classification: H55, I32

Note: This report is the first deliverable of the subcontract agreement of the University of Antwerp with the Federal Planning Bureau. This subcontract agreement is part of the Contractor Agreement between the University of Antwerp and BELSPO regarding the support for the valorisation of federal components of distributed and virtual ESFRI research infrastructures 2021-2026 (contract nr EF/211/CO-SHARE) with Federal Public Planning Service Science Policy within the framework of the project entitled "Co-SHARE ESFRI FED". The goal of this subcontract is to provide support and advice to the SHARE team at the University of Antwerp.

This is a revised version of the report submitted in November 2024, taking account of remarks by the SHARE team at the University of Antwerp.

* Corresponding author: Karel van den Bosch, kvdb@plan.be

Table of contents

List of tables

List of graphs

Executive summary

This report is the first deliverable of the subcontract agreement of the University of Antwerp with the Federal Planning Bureau. This subcontract agreement is part of the Contractor Agreement between the University of Antwerp and BELSPO regarding the support for the valorisation of federal components of distributed and virtual ESFRI research infrastructures 2021-2026 (contract nr EF/211/CO-SHARE) with Federal Public Planning Service Science Policy within the framework of the project entitled “Co-SHARE ESFRI FED”. The goal of this subcontract is to provide support and advice to the SHARE team at the University of Antwerp.

The Survey of Health, Ageing, and Retirement in Europe (SHARE), or "50+ in Europe", is a study that focuses on the European population aged 50 and over. Data are collected every two years in the areas of health (both mental and physical), socio-economic status, and social networks. The study started in 2004 and has since undergone nine rounds of data collection. Belgium is one of the few countries that participated in all nine waves. Despite the panel set-up, the sample needs to be refreshed regularly (i.e. new respondents need to be found).

For reasons of organization (two languages) and of funding (partly by regional authorities) the Belgian participation in SHARE is coordinated by two teams: one at the University of Liège (for the French-speaking community) and one at the University of Antwerp (for the Dutch-speaking community). The two teams work closely together for sampling, questionnaire development and valorisation of the SHARE data.

After a brief introduction in section 1, the second section of the report explains the design of the Belgian SHARE sample. The same design has been used both for the initial sample for the first wave and for the refreshment samples in later waves. It is a two-stage sample with stratification and clustering of municipalities in the first stage, and random sampling of persons in the second stage. The goal is to approximate an EPSEM (“equal probability selection method”) design as closely as possible. The sample is clustered in particular municipalities to facilitate the fieldwork. The national register is the sampling frame for the second stage.

The third section discusses how the sample was regularly refreshed by additional samples, and how well the survey teams were able to retain panel respondents. Refreshment samples are of two kinds. Refreshment ample A is a sample of people who have become part of the target population, i.e. have become 50, since the last wave when a refreshment sample was drawn. Refreshment sample B covers the rest of the target population to compensate for attrition of the panel due to non-response and death. Some refreshment samples were extra large to extend the total sample. The number of respondents who came into the sample through various channels as well as the number of people who exit the panel through death or non-response or who skipped waves is charted for each wave. An important result of the

analysis is that the retention rate (i.e. the proportion of respondents in a particular wave who participate also in the next wave) is consistently high. The average value for the French-speaking community is 78% and for the Dutch-speaking community 83%. Response to the two special COVID-19 surveys in 2020 and 2021 is also discussed.

The fourth section looks at some indicators of the representativeness of the SHARE samples for the population of older people in the French-speaking and Dutch-speaking communities of Belgium. Considering the distribution by age category and comparing with population figures, it is found that the age group 55-64 is increasingly underrepresented in the SHARE samples for both communities. On the other hand, the proportion of the very old (85plus) shows a clear increase in the SHARE samples so that in wave 9 this proportion has caught up to the figures for the population. A second important characteristic is living in a care home. SHARE includes these people and the proportion of the sample living in such an institution has increased substantially across the waves. Third, compared to EU-SILC the SHARE sample contains a larger proportion of older people who are severely limited in their activities due to health problems (an important indicator of the health status). This suggests that SHARE is better than EU-SILC at including older people with serious health problems in the survey. Fourth, it appears that older people with at most lower secondary education are underrepresented in the SHARE samples relative to the population, while those with tertiary education are overrepresented.

1. Introduction

The Survey of Health, Ageing, and Retirement in Europe (SHARE), or "50+ in Europe", is a study that focuses on the European population aged 50 and over.[1] Data are collected every two years in the areas of health (mental, physical and cognitive), socio-economic status, and social networks. The study started in 2004 and has since undergone nine rounds of data collection. In the first SHARE wave in 2004, 11 European countries participated in the study, including Belgium, and by 2024 information has been collected from 140,000 people aged 50 and over in 27 European countries and Israel. SHARE thus enables researchers to gain insight into the living situation of people aged 50 and over in Belgium – also separately for Flanders on the one hand and Wallonia and Brussels on the other – and to compare this with the living situation of their peers in other European countries and Israel.[2] For reasons of organization (two languages) and of funding (partly by regional authorities) the Belgian participation in SHARE is coordinated by two teams: one at the University of Antwerp (Flanders) and one at the University of Liège (Wallonia and Brussels).[3] The two teams work closely together for sampling, questionnaire development and valorisation of the SHARE data.

SHARE is a panel survey, or longitudinal survey. This means that the same group of people (the panel) is followed over time and interviewed every two years. Because of this set-up, it is possible to map changes over time and reactions to these changes. This set-up also means that the sample needs to be refreshed regularly (i.e. new respondents need to be found). First, those who have become 50 or over since the last time a new sample was drawn need to be covered. Second, there is attrition of the panel

---

1 This paper uses data from SHARE Waves 1, 2, 3, 4, 5, 6, 7, 8 and 9 (DOIs: 10.6103/SHARE.w1.900, 10.6103/SHARE.w2.900, 10.6103/SHARE.w3.900, 10.6103/SHARE.w4.900, 10.6103/SHARE.w5.900, 10.6103/SHARE.w6.900, 10.6103/SHARE.w6.DBS.100, 10.6103/SHARE.w7.900, 10.6103/SHARE.w8.900, 10.6103/SHARE.w8ca.900, 10.6103/SHARE.w9.900, 10.6103/SHARE.w9ca900, 10.6103/SHARE.HCAP.0) see Börsch-Supan et al. (2013) for methodological details.(1)
The SHARE data collection has been funded by the European Commission, DG RTD through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982, DASISH: GA N°283646) and Horizon 2020 (SHARE-DEV3: GA N°676536, SHARE-COHESION: GA N°870628, SERISS: GA N°654221, SSHOC: GA N°823782, SHARE-COVID19: GA N°101015924) and by DG Employment, Social Affairs & Inclusion through VS 2015/0195, VS 2016/0135, VS 2018/0285, VS 2019/0332, VS 2020/0313, SHARE-EUCOV: GA N°101052589 and EUCOVII: GA N°101102412. Additional funding from the German Federal Ministry of Education and Research (01UW1301, 01UW1801, 01UW2202), the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, BSR12-04, R01_AG052527-02, R01_AG056329-02, R01_AG063944, HHSN271201300071C, RAG052527A) and from various national funding sources is gratefully acknowledged (see www.share-eric.eu).

2 See https://share-eric.eu for more information on SHARE.

3 See https://www.share-project.be/main/N/default.htm (NL) or https://www.share-project.be/main/F/default.htm (FR) for sources of funding in Belgium and https://share-eric.eu/infrastructure/funding for sources of funding for the international coordination.

mostly because of non-response and to maintain the size of the sample sometimes a refreshment sample covering all ages must be selected.[4]

Note that all results in this report are unweighted, as the focus is on the (realized, i.e. interviewed) sample. Weights are calculated by the central coordination of SHARE, using information provided by the national teams. See the SHARE release guide[5] and De Luca et al. (2021).

The second section of this report explains the design of the Belgian SHARE sample. The third section discusses how the sample was regularly refreshed by additional samples, and how well the survey teams were able to retain panel respondents. The number of respondents who came into the sample through various channels as well as the number of people who exit the panel through death or non-response or who skipped waves is charted for each wave. The fourth section looks at some indicators of the representativeness of the SHARE samples for the population of older people in the French-speaking and Dutch-speaking communities of Belgium: the distribution by age category, the number of respondents in care homes and the proportion of older people who are severely limited in their activities due to health problems.

[4] For extensive information on the design and implementation of SHARE see the methodology volumes published for each wave, which are available on https://share-eric.eu/data/data-documentation/methodology-volumes

[5] https://share-eric.eu/data/data-documentation/release-guides

2. Sampling design

2.1. Introduction

The central coordination of SHARE publishes guidelines on the sampling, but these are not very restrictive, as conditions and possibilities vary much across countries (Bethmann and Bergmann, 2023). It is stressed that probability sampling is required, so that the inclusion probabilities or the likelihood to be sampled for every unit in the sample can be calculated. They also mention the advantages of stratification and the need for geographic clustering.

The SHARE target population consists of all persons aged 50 years and over at the time of sampling who have their regular domicile in Belgium. The reference year for sampling purposes in SHARE is the year in which the majority of respondents are age eligible (50+) at start of fieldwork. Other eligibility rules specify that persons are excluded if they are incarcerated, hospitalised or out of the country during the entire survey period, unable to speak the country's language(s) or have moved to an unknown address. The same is true for individuals who have deceased between the time of sampling and fieldwork (Bethmann and Bergmann, 2023, p. 2). These rules can generally only be implemented during the fieldwork.

In Belgium, by and large the same sampling design has been used for the initial sample for the first wave and for the refreshment samples in later waves.[6] It is a two-stage sample with stratification and clustering of municipalities in the first stage, and random sampling of persons in the second stage. The design is such that the drawn sample approximates an EPSEM sample as closely as possible. So-called EPSEM designs ("equal probability selection method") are often preferable, as designs that produce an equal probability of selection for each member in the population are most efficient in terms of limiting the sampling variance (Bethmann and Bergmann, 2023, p. 4). Though separate samples are drawn for Wallonia and Brussels[7] on the one hand, and Flanders on the other hand, the sampling design is similar. The municipalities of the German-speaking community have been excluded, as resources did not permit to translate the questionnaire into German.

2.2. First stage

The first stage involves the selection of municipalities. In order to limit survey costs, in particular travelling costs of interviewers, the sample (outside Brussels) is clustered in 100 municipalities (69 in Flanders and 31 in Wallonia). Municipalities were stratified in both regions. The big cities Antwerp, Charleroi, Gent and Liège each form one stratum. The other municipalities are distributed across three strata

[6] The information in this section comes from various internal documents.

[7] Brussels is of course bilingual, but for practical purposes was included in the Francophone sample. We come back below to the issue of Dutch-speakers in the Brussels sample.

per region, such that these strata have equal size in terms of the target population (i.e. 50 years or over). The criterion for assigning municipalities to one of the three strata was its size in terms of the target population. So the first stratum contains larger towns, the second one middling ones and the third one the smallest municipalities. The strata are constructed in such a way that the total number of people of 50 years or over in each stratum is the same as much as possible. This implies that the number of municipalities in the first stratum is much smaller than in the third one. The stratification ensures that older people from large cities, medium-sized towns and small municipalities are represented proportionally in the sample.

The target sample sizes are distributed across the strata in proportion to the size of the strata in terms of the target population (people age 50plus). This gives the target sample size n_s within each stratum s. The number of municipalities to be selected within each stratum m_s (except for the four large-cities strata) is determined by the formula: $m_s = \text{ROUND}(n_s / 25)$, where 25 is the target cluster size. So in each selected municipality the sample included at least 25 persons. This facilitates the survey fieldwork, as interviewers lose less time in travelling from one interviewee to another. For the four large cities, the target sample size was set proportional to the size of the 50plus population in those cities.

Within each of the strata (except in the four large-cities strata) municipalities were selected proportional to size in terms of the target population, and without replacement. This design implies that in small (large) municipalities, people had a larger (smaller) probability to be in the sample, given that their municipality was selected in the first step. On the other hand, their municipality had a smaller (larger) probability to be sampled. These probabilities cancel out against each other, so that every person in the target population had an equal chance of being selected for the sample.

For the refreshment samples no new municipalities were selected. New SHARE respondents were always selected in the municipalities that were drawn in the first wave.

2.3. Second stage

In the second state, persons were in principle selected by simple random sampling (without replacement). The best sampling frame in Belgium is the national register, which has information on the age, sex and 'relation to the reference person' of all individuals within each household. So when using the national register, the target population could be identified before sampling. Unfortunately, the national register could not be used for the first wave (see below), though it served as sampling frame for the refreshment samples in later waves. The number of persons drawn is considerably higher than the targeted number of interviews, to account for non-response. A conservative estimate for the response of (varying across waves between 40% and 35%) is used in the sampling procedure (i.e. targeted numbers are divided by the assumed response rate to arrive at the number to be sampled).

There is one complication, however. SHARE requires that for all respondents, the spouse, irrespective of age, is also interviewed and so becomes part of the (effective) sample. With simple random sampling of individuals, this would imply that couples where both partners are 50plus have a double probability to be in the sample relative to single persons (and couples where only one partner is 50plus)[8]. Unfortunately, it appeared to be not feasible to group in advance all people 50plus in selected municipalities into those with and those without a spouse who was also 50plus. Instead, after the sample was drawn, the spouses of selected persons were identified, and half of those with a 50plus spouse were removed from the sample by simple random sampling.[9] This procedure ensures that the sample retains its EPSEM character.

For the first wave, the Privacy Commission (a governmental agency charged with protecting the privacy of the personal data of citizens)[10] imposed restrictions on access to the national register that were unacceptable for SHARE. In particular, the National Statistical Institute (which would have taken care of the sampling) was not allowed to send all the sampled addresses directly to the SHARE teams. They first had to send a letter to all sampled households, and only the addresses of people that positively agreed to be interviewed could be communicated to the survey organizations. The reason given for this restriction was the delicate nature of some questions, in particular regarding health and health care. The experience of other surveys which used this procedure indicated that it leads to unacceptably low response rates (at most 35%). For this reason, the telephone directory was used as the sampling frame.[11] This was available on a CD for the whole of Belgium, enabling a (largely) automated extraction of names and addresses. The obvious drawback of the telephone listing is that it does not have information on the age of persons. In order to avoid having to screen a large number of addresses before the actual SHARE interview could be started, the sampled names and addresses were sent to a commercial direct-marketing firm, Wegener DM, which claimed to have age information on about 80% of the Belgian population. Addresses that according to their information contained no people aged 50 or over were removed from the sample. Addresses for which they had no age information were retained and were screened by the interviewers.

---

8 Note that partners below 50 are interviewed, and their data are in the SHARE files, but they do not become part of the research population. Their weights are set to zero or missing. In later waves this status may change when they have passed the age threshold of 50.

9 Formally, the procedure was specified as follows for the National Register to implement:
- within each municipality, persons within the target population are sampled by simple random sampling (without replacement)
- the spouse / partner of each selected person is identified, and his/her age is determined
- if the spouse/partner belongs to the target population, the (original) person is marked as 'target-couple', otherwise she/he is marked as 'target-single'.
- from the group marked as 'target-couple', half are deleted from the sample by simple random sampling
- the selected persons, as well as their spouse/partners, if they belong to the target population, are retained as the final sample.

10 The Privacy Commission was succeeded in 2017 by the Data Protection Authority (French: L'Autorité de protection des données; Dutch: Gegevensbeschermingsautoriteit; see https://www.autoriteprotectiondonnees.be/citoyen / https://www.gegevensbeschermingsautoriteit.be/burger .

11 At that time, almost all households, certainly those with people over 50, had a landline.

2.4. Brussels

Brussels is bilingual, and the preferred language of the inhabitants of Brussels is not registered. The Liège (Francophone) survey team sends both French and Dutch versions of introductory letter to all selected people in Brussels. These letters mention that a Francophone interviewer will contact them in a few weeks. If respondents prefer a Dutch-speaking interviewer, they can inform the survey team by letter, e-mail or telephone. Those respondents would then be transferred to the Antwerp (Dutch-speaking) survey team.[12] In fact, this happened very rarely. In the first wave, only 4 respondents in Brussels were interviewed in Dutch. In later waves there seem to have been even fewer of these, if any.[13]

[12] Note that the SHARE survey is implemented as a CAPI interview on tablets. The central SHARE team responsible for the CAPI programming advised against loading two versions of the CAPI software on the same tablet, as this was likely to produce technical problems. Also, not all interviewers are fluent speakers of both languages.

[13] This cannot be determined with certainty, because the variable identifying the regions (NUTS1) has many missing values in subsequent waves. The four Dutch-speaking respondents from Brussels do not reappear in any subsequent wave. It is possible, however, that they were assigned new identification numbers in the second wave, breaking the links between waves. (At the time in SHARE, the two language communities in Belgium were treated as two different countries by the central SHARE team, and the sample management system could not handle moves between countries.)

3. History

In this section I give an overview of the realized sample, i.e. the people actually interviewed. In a panel survey the greatest concern is always to try to retain as many respondents as possible. Panel attrition occurs when earlier respondents have moved and cannot be traced or when they have moved abroad. In a survey of older people death is also an important cause (see below). However, attrition mostly happens when respondents refuse to participate in the survey in subsequent waves. Attrition represents a loss of information and also in time may reduce the total sample size and affect the representativeness of the sample. Various efforts are made to keep attrition to a minimum, e.g. by sending the most experienced or skilful interviewers to respondents who are hesitating, but it cannot always be avoided. It is of course for the latter reason that refreshment samples were drawn.

3.1. Refreshment samples

Refreshment samples in SHARE are of two kinds. Refreshment sample A is a sample of people who have become part of the target population, i.e. have become 50, since the last wave (or the last wave when a refreshment sample was drawn). Refreshment sample B covers the rest of the target population and is intended to make up for attrition from the panel. Table 1 shows the birth years for the refreshment samples A by wave. Refreshment samples B were selected if and only if a refreshment sample A was drawn. Refreshment samples were not drawn for every wave, and until wave 5, the French and Dutch teams were not working in parallel in this respect. In wave 2, the Dutch team did not take on refreshment samples, because the first wave sample had been much extended (during the fieldwork, in fact) thanks to additional budget from a Flemish funder. For wave 3, the central coordinator of SHARE had advised not to draw refreshment sample because of the special nature of this wave. Called SHARELIFE, the questionnaire was devoted to collecting information about the conditions and events of respondents during their whole lives. Therefore, the refreshment samples A in wave 4 had to cover four birth years for the French-speaking population and six birth years for the Dutch-speaking population. For wave 7, no refreshment samples were drawn, because the central coordinator considered this advisable due to the complicated nature of the survey (new respondents since wave 3 had to answer the SHARELIFE questionnaire, while others took the regular SHARE questionnaire). This necessitated that the refreshment samples A in wave 8 included four birth years. As the fieldwork for wave 8 was interrupted by the COVID pandemic, a big part of this refreshment sample could be contacted only in wave 9.

Table 1. Birth years for refreshment samples A by wave

Wave	BE-fr	BE-nl
2	1955-56	-
3	-	-
4	1957-60	1955-60
5	1961-62	1961-62
6	1963-64	1963-64
7	-	-
8	1965-69	1965-69
9*	1965-69	1965-69

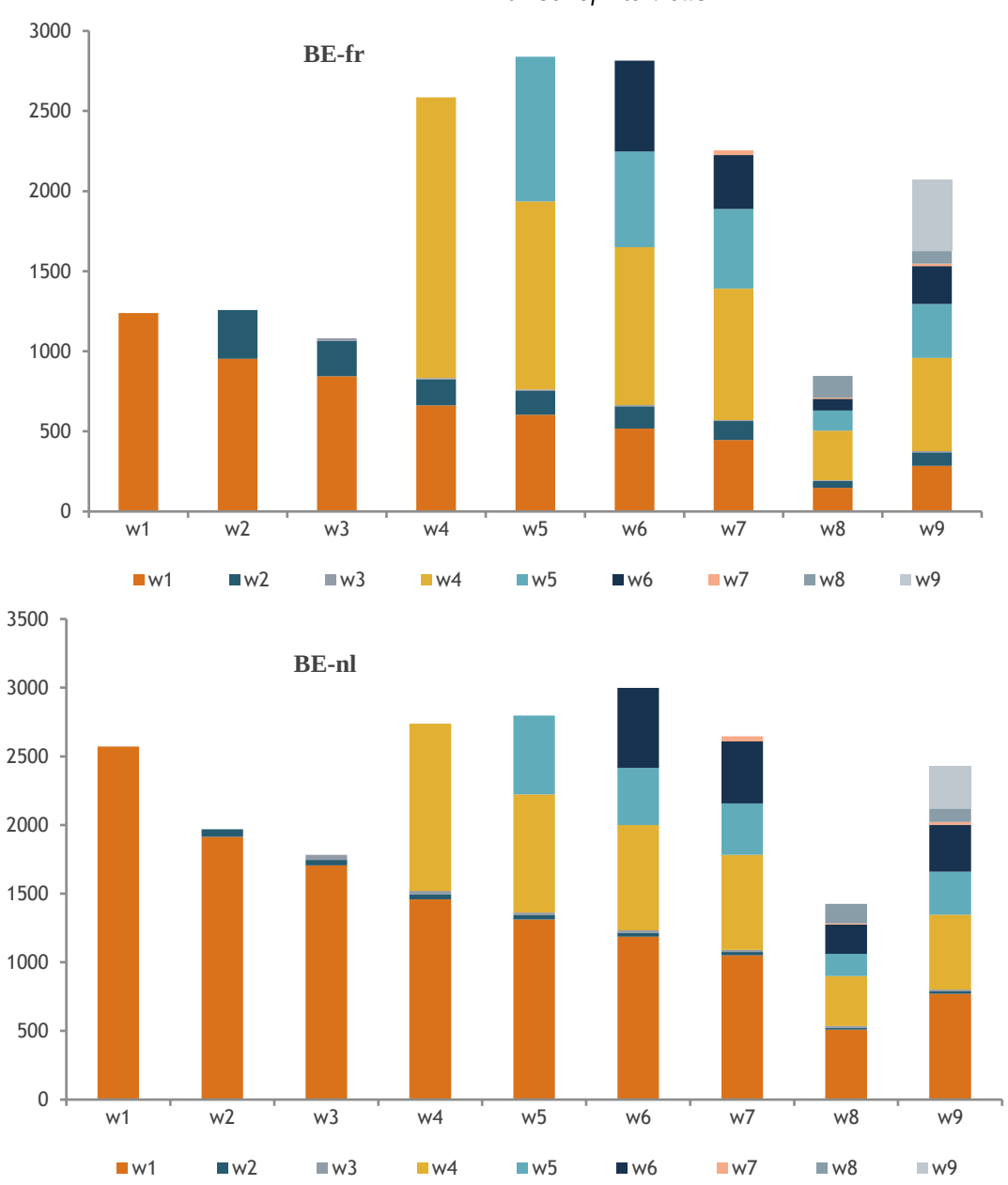
Note: * Fieldwork for wave 8 was interrupted by the COVID pandemic. Therefore, a big part of the refreshment sample for wave 8 was contacted only in wave 9.

Source: SHARE data

3.2. The interviewed sample across waves

What was the size of the sample across the waves, and to what extent did the survey teams manage to retain panel respondents? Figure 1 shows the sample size for each wave, broken down by the wave when these respondents were first interviewed. A number of things stand out. First, in the first wave the size of the sample was much larger on the Dutch-speaking side than for the other language community, due to the additional Flemish funding referred to above. However, the French-speaking team was able to greatly extend the sample in the fourth wave, thanks to extra funding specifically for the French community part of SHARE-Belgium. Second, the bars for wave 8 shows how the fieldwork was cut short by the COVID pandemic. Third, a comparison between subsequent waves of the stacked bars for the same wave of first interview (e.g. the orange bars between waves 1 and 2, or the beige bars between waves 4 and 5) suggests that panel attrition is relatively large between the first wave when respondents enter the sample and the next wave. Once they have participated in two waves, they are more likely to stay in the panel. (It may be seen that even in waves where there were no refreshment samples, as in waves 3 and 7, there were still a small number of new respondents. The reason for this will become clear below.)

Figure 1. Realized sample (i.e. number of interviews) in each wave, by wave of first interview



Note: How to read this graph: The column marked w5 in the top graph shows that in wave 5 at the French-speaking side there were 604 respondents who had participated already in wave 1 (orange bar), 148 who first entered the sample in wave 2 (dark blue bar), 1173 respondents who first took part in wave 3 (beige bar) and 903 respondents who were in the refresher sample for wave 5 (blue bar).

Source: SHARE data

Tables 2A and 2B provide more detail on how these sample numbers by wave came about. These tables represent a kind of bookkeeping exercise on the effective SHARE sample (i.e. sample members with whom an interview was conducted). The column "Total interviewed sample" shows the total number of interviews in each wave; these numbers correspond to the totals of the stacked columns in Figure 1. The

columns to the left labelled "In" provides information on how these respondents were obtained. The column "Interview in previous wave" shows exactly that. Some panel respondents do not participate in one or more waves, but are again recovered in later waves; the number of them (for the wave when they return) is in column "Interview in earlier waves". The columns "Refreshment A" and "Refreshment B" show the number of interviews for refreshment samples A and B, respectively. Finally, the column "Other new entrants" is a kind of residual category: people who had no interview in previous waves, but are also not part of a refreshment sample for that wave. Most of these are people in couples where the other partner was interviewed in an earlier wave, but the person him- or herself did not participate for some reason. A few were new partners of panel respondents.[14]

The four right-most columns of Tables 2A and 2B document the number of respondents in each wave who did not participate in the next wave. Obviously, this could only be ascertained at the time of the next wave. The sum of these columns plus the entry in the column "Interview in previous wave" in the next row corresponds to the "Total interviewed sample". The column "Died, EOL" shows the number of respondents who passed away and for whom an End-Of-Life interview was conducted. SHARE End-Of-Life interviews are conducted (not necessarily during the wave when the death was registered) with people who were close to the deceased, most often the partner or a child, sometimes other relatives and in rare cases non-relatives such as a professional carer. Questions are asked about the cause of death, health and care during the last year of life and the inheritance among other subjects. Obviously, these are sensitive matters and interviewers do not always succeed in obtaining an End-Of-Life interview. The next column, "Died, no EOL", shows the number of respondents who were known to have died but for whom no End-Of-Life interview could be conducted. The column "Skipped wave(s)" lists the number of respondents who did not participate in the next wave, but returned in a later wave. So these correspond to the persons in the column "Interview in earlier wave(s)". Finally, the last column "Non response, no return" shows how many left the panel and did not return (yet). For waves 1 to 6 this is permanent, as households are dropped when no interview was conducted in the last two waves. Non-respondents in waves 7 to 9 might be persuaded to participate in a future wave, unless they refused unequivocally to be interviewed. This column includes not only refusals, but also some people who left the population through emigration or incarceration. (People who moved to a care home are followed, though, see section 4.2.) Moreover, it is not always possible to ascertain whether respondents are still alive or not, and this column may also encompass unobserved deaths.

[14] In the data there is no variable identifying people in refreshment samples A or B. In fact, new respondents were assigned to category "Other new entrants" if they or their partner had been registered in earlier waves either as respondents or non-respondents. (Non-respondents are persons from whom an interview should have been taken, but was not, for whatever reason (no contact, refusal). This must be distinguished from being ineligible for an interview or simply not being in the sample yet.) The remaining new respondents were distributed over the Refresher columns according to their birth year, and following Table 1.)

Table 2A. Inflow and outflow from interviewed sample, French-speaking community (BE-fr)

Wave	In:					Total interviewed sample	Out:			
	Refresh-ment A	Refresh-ment B	Other new entrants	Interview in earlier waves	Interview in previous wave		Died, EOL	Died, no EOL	Skipped wave(s)	Non re-sponse, no return
1						1239	24	3	72	186
2	111	156	36	0	954	1257	42	5	58	150
3	0	0	18	62	1002	1082	43	18	92	136
4	427	1315	9	41	793	2585	74	36	179	455
5	191	665	47	94	1841	2838	102	38	150	486
6	154	360	54	185	2062	2815	99	56	139	450
7	0	0	31	153	2071	2255	45	97	873	564
8*	67	65	4	35	676	847	17	16	0	182
9*	260	159	27	993	632	2071				

Notes: EOL: End Of Life interview, see text.

* Fieldwork for wave 8 was interrupted by the COVID pandemic. Therefore, a big part of the sample for wave 8 was contacted only in wave 9.

Columns “out” show what happened to respondents who were not in the next wave. This could only be ascertained in the next wave. E.g. the number 24 in the column “Died, EOL” in the first line of the body of table 2A means that of all respondents in wave 1 (Francophone side) 24 had died at the time of the second wave and End-Of-Life interview was conducted for them.-

Source: SHARE data

Table 2B. Inflow and outflow from interviewed sample, Dutch-speaking community (BE-nl)

Wave	In:					Total interviewed sample	Out:			
	Refresh-ment A	Refresh-ment B	Other new entrants	Interview in earlier waves	Interview in previous wave		Died, EOL	Died, no EOL	Skipped wave(s)	Non re-sponse, no return
1						2571	40	9	165	443
2	0	0	56	0	1914	1970	62	5	98	202
3	0	0	37	143	1603	1783	56	11	83	203
4	536	671	10	90	1430	2737	75	43	143	353
5	135	398	39	101	2123	2796	69	22	133	296
6	163	382	38	139	2276	2998	90	51	86	295
7	0	0	37	132	2476	2645	74	90	856	377
8*	110	26	6	35	1248	1425	23	33	0	172
9*	230	50	25	924	1197	2426				

Notes: EOL: End Of Life interview, see text.

* Fieldwork for wave 8 was interrupted by the COVID pandemic. Therefore, a big part of the sample for wave 8 was contacted only in wave 9.

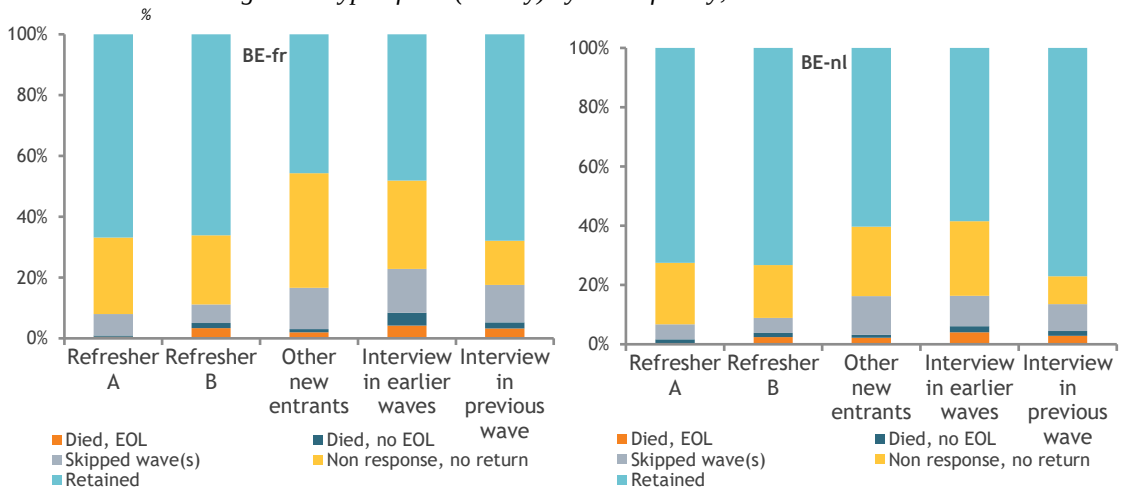
Columns “out” show what happened to respondents who were not in the next wave. This could only be ascertained in the next wave. E.g. the number 24 in the column “Died, EOL” in the first line of the body of table 2A means that of all respondents in wave 1 (Francophone side) 24 had died at the time of the second wave and End-Of-Life interview was conducted for them.-

Source: SHARE data

The tables reveal that the size of the refreshment samples varied quite a lot across waves. The size of sample A is set at such a level that the number of sample cases in the age group concerned relative to the total sample is about the same as the corresponding ratio in the population. It is particularly large when a wider age bracket than normal needed to be included because in previous waves no refreshment samples were drawn. This is the case e.g. in wave 4. Sample B is particularly large in wave 4 for the French-speaking community, because the sample was much extended in that wave (see Figure 1). Otherwise, the size of sample B is mostly determined by the need to compensate for panel attrition and to maintain the size of the total sample, though survey organization and budget considerations can also play a role. Sample B was larger than usual in wave 5 for the French-speaking community because non-response in the previous wave was very high in Brussels. The size of sample B is relatively small in waves 8 and 9 (and there was no refreshment in wave 7), leading to a reduction in the number of total interviewed respondents in those waves. Given that these refreshment cases are interviewed for the first time, it is interesting to observe (Figure 2) that their probability to remain in the interviewed sample in the next wave (67 % in the French-speaking community; 73 % in Flanders) is not much lower than it is for those who were already in the panel in the previous wave. "Permanent" non-response, however, is higher for the refreshment sample as these respondents are less likely to skip waves.

The number of "Other new entrants", mostly through their partner, is small but not negligible. It appears to be more difficult to retain them in the panel (Figure 2). The number of respondents who skipped waves is substantial. These are persons who had an interview in an earlier wave but not the previous one. SHARE makes a big effort (perhaps more than other panel surveys) to interview respondents who were once part of the panel but dropped out in the previous wave. This effort pays out in more longitudinal observations (cases obviously become more useful for research the longer the observation period). However, these respondents are more likely to become "permanent" non-respondents than those who were interviewed in the previous wave. Curiously, they seem also more likely to die than the other groups (the difference is small, but statistically significant). Possibly, this is because illness was the reason why they skipped a wave.

Figure 2. Type of exit (or stay) by kind of entry, across all waves



Source: SHARE data

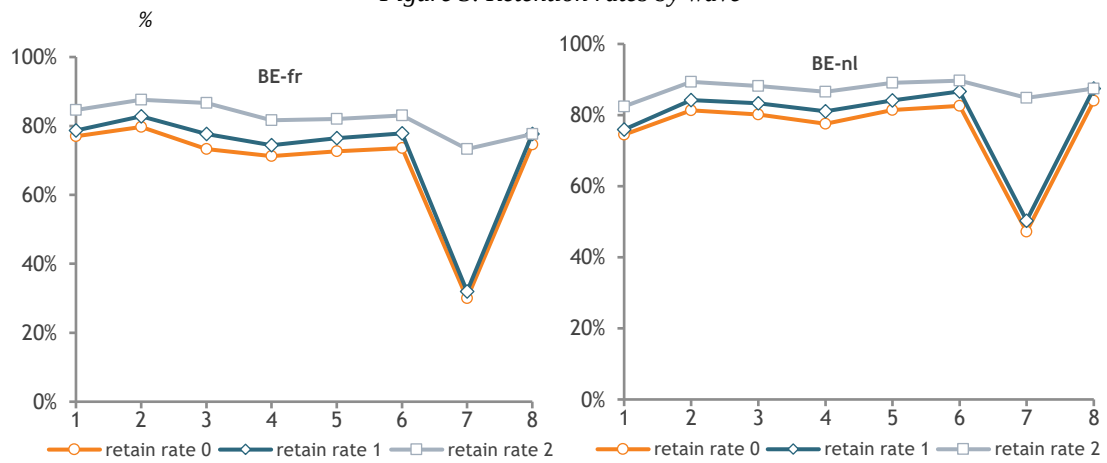
Across waves, the proportion of respondents who have died by the next wave has increased from about 2% between wave 1 and wave 2 to around 6% after wave 7. At the same time, the proportion of deceased respondents for whom an End-Of-Life interview could be conducted has declined from close to 90% in the first two waves to 64% in wave 6 and less than that after wave 7.[15] However, it is likely that both developments are due to a better registration of deaths in the SHARE panel. The national register provided information about deaths for respondents for whom the national register identification number was known (which was not the case for all first-wave respondents).

The column "Skipped waves" shows the number of respondents who were not interviewed in the next wave, but did return in later waves; so these correspond to those in column "Interview in earlier waves" (though not wave-by-wave, as some respondents skipped more than one wave). Finally, the "Non response, no return" column shows non-response that turned out to be permanent (in the earlier waves) or is likely to become so (in waves 6 to 8). Based on these figures, "retention rates" can be calculated indicating how many respondents in a particular wave were retained in the next wave. For a panel survey with the aim of enabling longitudinal research these rates are obviously very important, and should be as high as possible. Figure 3 shows them. Retain rate 0 shows the proportion of respondents in a particular wave that were also interviewed in the next wave. It is affected by deaths or other reasons why respondents are no longer eligible for SHARE as well as by non-response. For retention rate 1, the number of people who died are subtracted from the denominator. For retention rate 2, in addition the number of respondents who skipped waves are added to the numerator. By definition, retention rate 2 is greater than or equal to retention rate 1, which in turn must always be greater than or equal to retention rate 0. For more information on response and non-response in the panel and refreshment samples and

[15] Note that End-Of-Life interviews after wave 7 would be conducted in waves 8 or 9, and were therefore hampered by the COVID pandemic.

other aspects of survey monitoring see Kneip, Malter and Sand (2015) and Sand (2019). These chapters also enable a comparison of the Belgian results with those of other countries.

Figure 3. Retention rates by wave



Notes: The wave number refers to the wave for which the numbers are shown in Table 2. So for wave 1 the retain rates show the proportion of wave 1 respondents that were retained in wave 2 (or later). ..
Retention rate 0: Interview in next wave / Interview in this wave
Retention rate 1: Interview in next wave / (Interview in this wave – Deaths)
Retention rate 2: (Interview in next wave + Skipped waves) / (Interview in this wave – Deaths)
Source: SHARE data

The retention rates move mostly in parallel, suggesting that "permanent" non-response is driving these rates. They increased between the first and second waves (for both Communities), maybe because respondents are less likely to drop out after two interviews than after only one. However, they went down again between waves 3 and 4 and between waves 4 and 5. For the latter, this may be related to the large number of refreshment cases in wave 4. In wave 3, there were no refreshment samples, though. The sharp decline for wave 7 reflects the COVID pandemic during the fieldwork for wave 8, of course. Most of the respondents who could not be contacted in wave 8 were retained in wave 9, but not all, as the decline in retention rate 2 shows.

The most important result to appear from Figure 3 is, obviously, that these retention rates are consistently high. If we take retention rate 1 as the most relevant and excluding wave 7, the average value for the French-speaking community is 78% and for the Dutch-speaking community 83%. The main reason for the lower figure for the French-speaking community is that it is much more difficult to find and re-interview respondents in Brussels than in the other regions of Belgium.

3.3. The COVID-19 surveys

The fieldwork for wave 8 had to be stopped prematurely in March 2020, due to the COVID-19 epidemic and the consequent lockdown. This was also the case in all other countries participating in SHARE. The international SHARE research consortium decided to respond to the unprecedented crisis by initiating a

special COVID-19 survey among the SHARE respondents. The aim was to understand the non-intended effects of the lockdown on health and health behaviours, labour market behaviour and social relationships, among other aspects (Bergmann, Wagner, Yilmaz et al., 2024). Interviews were conducted by CATI (Computer Assisted Telephone Interviewing). The first COVID-19 survey between June and August 2020 was followed by a second one in June-July 2021. Only panel respondents were contacted (so not the refreshment sample), both those who had already been interviewed in wave 8 and those who had not. These interviews were short (about 15 minutes) and conducted by interviewers with SHARE experience.[16] If the respondent had died during the period since the last wave, an End-Of-Life interview was attempted with a survivor.

Table 3. Results for COVID-19 Survey 1, by response (individual) in wave 8

Results for COVID-19 Survey 1	BE-fr			BE-nl		
	Interview in W8			Interview in W8		
	No	Yes	Total	No	Yes	Total
numbers						
No interview	854	131	985	887	128	1,015
SHARE COVID-19 interview	1,108	580	1,688	1,056	1,155	2,211
End-Of-Life interview	52	0	52	56	0	56
Total	2,014	711	2,725	1,999	1,283	3,282
%						
No interview	42.4%	18.4%	36.1%	44.4%	10.0%	30.9%
SHARE COVID-19 interview	55.0%	81.6%	61.9%	52.8%	90.0%	67.4%
End-Of-Life interview	2.6%	0.0%	1.9%	2.8%	0.0%	1.7%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Note: Only people in the longitudinal sample were eligible for the COVID-19 Survey 1, i.e. people who had been interviewed in wave 7 or earlier waves.

Source: SHARE data

[16] https://www.share-project.be/main/N/content.htm

Table 4. Results for COVID-19 Survey 2, by response in COVID-19 Survey 1

Results for COVID-19 Survey 2	BE-fr			BE-nl		
	Interview in COVID-19 survey 1			Interview in COVID-19 survey 1		
	No	Yes	Total	No	Yes	Total
numbers						
Not applicable: out of sample	148	14	162	158	3	161
Not applicable: not listed as household member	13	4	17	2	5	7
Not applicable: household did not participate in this wave	377	229	606	371	148	519
No interview	421	27	448	435	52	487
SHARE COVID-19 interview	25	1,391	1,416	41	1,966	2,007
End-Of-Life interview	1	23	24	8	37	45
Total	985	1,688	2,673	1,015	2,211	3,226
%	No	Yes	Total	No	Yes	Total
Not applicable: out of sample	15.0%	0.8%	6.1%	15.6%	0.1%	5.0%
Not applicable: not listed as household member	1.3%	0.2%	0.6%	0.2%	0.2%	0.2%
Not applicable: household did not participate in this wave	38.3%	13.6%	22.7%	36.6%	6.7%	16.1%
No interview	42.7%	1.6%	16.8%	42.9%	2.4%	15.1%
SHARE COVID-19 interview	2.5%	82.4%	53.0%	4.0%	88.9%	62.2%
End-Of-Life interview	0.1%	1.4%	0.9%	0.8%	1.7%	1.4%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0 %

Source: SHARE data

The results in Table 3 show that the response in the first COVID-19 survey was quite acceptable for a telephone survey, with an overall response of 62 % in the French-speaking community and 67 % in the

Dutch-speaking one. The response was higher among those who had already been interviewed in wave 8, but still more than 50 % in the part of the longitudinal sample who had not. In the French-speaking part of SHARE-Belgium (where wave 8 had not proceeded as far as in the Dutch-speaking part), about two-thirds of people who responded to the first COVID-19 survey had not yet been interviewed in wave 8.

The second SHARE Corona Survey was a follow-up of the first SHARE Corona Survey. "All panel households that participated in the first SHARE Corona Survey and were still eligible were part of the gross sample for the second SHARE Corona Survey."[17] The response was lower in this survey, in particular because – unsurprisingly – very few of the non-respondents in the first COVID-19 survey cooperated in the second one. The code 'Not applicable: household member did not participate in this wave' means that the household was approached but did not participate in the second COVID-19 survey.

How was the response to the second COVID-19 survey associated with participation in the regular SHARE wave 9, which started a few months later in November 2021? Participation was nearly complete (96 % or more) for people who had had an interview in the second COVID-19 survey, irrespective of whether they had been interviewed in wave 8 or not. Interestingly, this was also the case for people who had been coded 'not eligible' for the second COVID-19 survey, provided they had participated in wave 8. Among people who had not had an interview in wave 8, nor in the second COVID-19 survey, the response was virtually nil (4%). Among panel respondents who had participated in wave 8 and in the first COVID-19 survey but not in the second COVID-19 survey, the response in wave 9 was relatively low: 46 % in BE-fr and 64 % in BE-nl. The latter finding could suggest that a refusal to take part in the telephone survey had a negative effect on the response in the later in-person interview.

[17] SHARE Wave 8 Methodology https://share-eric.eu/fileadmin/user_upload/Bilder_Newsredaktion/SHARE_Methodenband_WEB__1_.pdf p. 151.

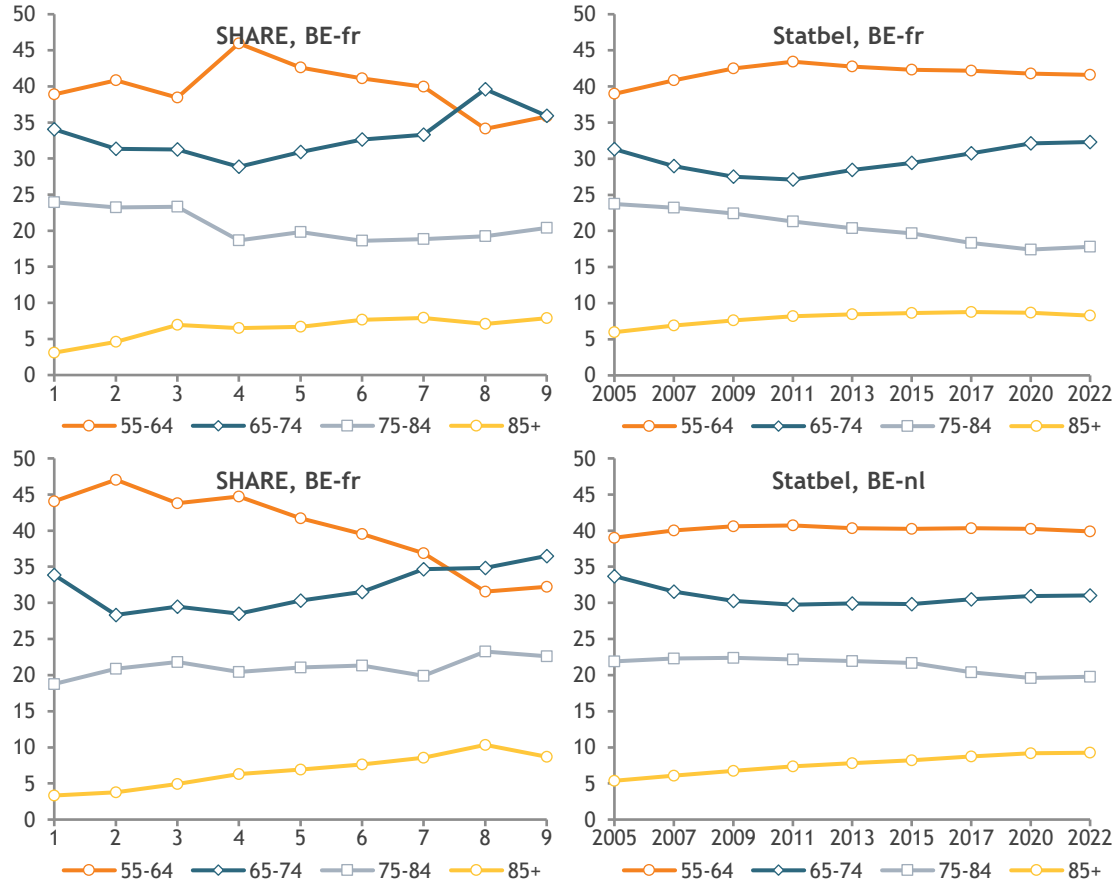
4. Results

In this section I provide some indications about the evolution of the SHARE sample, where possible comparing with external data from populations statistics from EU-SILC. Note that these results are un-weighted; the interest here is to what extent the SHARE sample as such was and remains representative for the population of older people in Belgium. If panel attrition and initial non-response was selective, the unweighted sample could become more biased in time. Variables considered are age, living in care homes, limitations in activity and education level.

4.1. Age distribution

We first look (in Figure 4) at the distribution of the SHARE sample by broad age category, comparing this with population statistics from Statbel. Only those aged 55plus are considered, as the size of the age group 50-54 varies quite a lot across waves due to the irregular drawing of refreshment samples.

Figure 4. Age distribution of SHARE sample (55plus) across waves, compared with population statistics
Cumulative %



Note : The Statbel data refer to the 1st January of the year shown. These years correspond to the SHARE waves. Mostly, the SHARE survey was conducted during the first half of the year, or it straddled the change of year. FR refers to Wallonia and Brussels.
Source: SHARE data, Statbel data

A number of things stand out. First, the proportion of people aged 55-64 increased strongly up to wave 4 (2011) in the SHARE sample for the French-speaking community, and afterwards declined. In this, it followed the population trend. However, the decrease in the sample was stronger than in the population. In the SHARE sample for the Dutch-speaking community, the decrease in the proportion of the 55-64 age group was even more pronounced. As a result in wave 9 (2022) these proportions fall short of the population figures by more than 5 percent-points for the French-speaking community and by nearly 8 percent-points for the Dutch-speaking community. Second, the proportions for the age group 65-74 follow opposite trends, first falling and then rising, both in the SHARE samples and in the population. Again, the trends in the samples are stronger than in the populations, in particular for the Dutch-speaking community. This leads to an overestimate of the proportions of the 65-74 age group in the samples by wave 9 (2022), by about 4 percent points for both language communities. Third, for the next age group 75-84, the French-speaking community SHARE sample follows roughly the same slightly downward trend as is observed in the population figures, and the proportions are similar too. In the SHARE sample for the Dutch-speaking community, this proportion rises slightly across waves, in contrast to the slightly declining trend in the population. Finally, the proportions of the very old (85plus) show a clear increase in the SHARE samples for both communities, an increase which is larger than in the population. While in the first wave these proportions were below those for the population, by wave 9 (2022) the SHARE samples have caught up and the differences are minimal.

The most important and somewhat alarming result from this comparison is that the age group 55-64 is increasingly underrepresented in the SHARE samples for both communities. Further analyses have suggested that this is mainly due to a lower-than-average response in the refreshment sample for the age group 50-54 (mostly in refreshment sample A, but in some waves also in refreshment B).[18] Once they are interviewed at least once, these respondents are as likely to be retained in the panel as older age groups. As they grow older across waves, the initially lower response leads to a growing deficit of cases in the age group 55-64.

4.2. Care homes

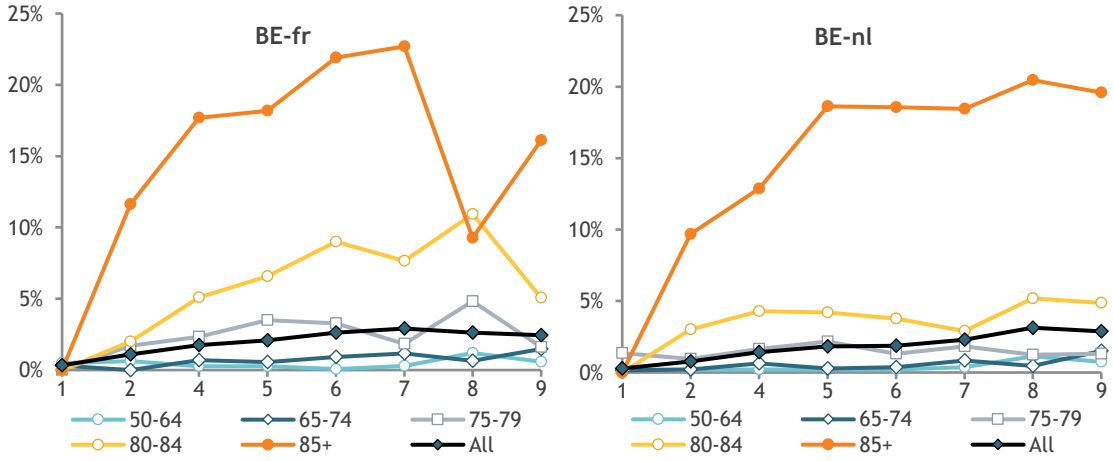
Another result from the comparison of the age distribution of the SHARE samples across waves is the gradual increase of the proportion of the oldest old (aged 85plus). This is related to the fact that SHARE includes people in collective households, i.e. care or nursing homes.[19] In this it differs from EU-SILC.

[18] For the relatively large refreshment samples in waves 4 and 5, the National Register provided the birth years of all sampled persons. Therefore, I could compare the age distribution of the raw sample, including non-response, with the age distribution of those from these samples who were successfully interviewed.

[19] People in collective households were excluded from the sample in the first wave, but not in later waves. More importantly, older people in the panel were followed and interviewed, if possible, when they had moved to a care home. If people were unable to answer some factual questions due to mental health issues, a proxy respondent for those questions was allowed.

Figure 5 shows the proportion of the SHARE samples who are living (and were probably interviewed) in a care home.[20] This proportion increased most in the first six waves. It is largest (up to 22%) among the 85plus, but substantial also in the age group 80-84. Unfortunately, there are no good administrative data on the number of older people in care homes. Statbel collects and publishes the number of people in collective households (based on National Register data), but these figures much underestimate the size of the actual care home population, because many people do not change their official residence when moving to a care home, but keep their old address (see Van den Bosch et al., 2011). The RIZIV used to collect data on the care home population, but this series has been discontinued since the sixth state reform, which transferred responsibility for care homes to the Communities.

Figure 5. Percentage of SHARE samples in care homes across waves, by age category



Source: SHARE data

4.3. Limitations in activities

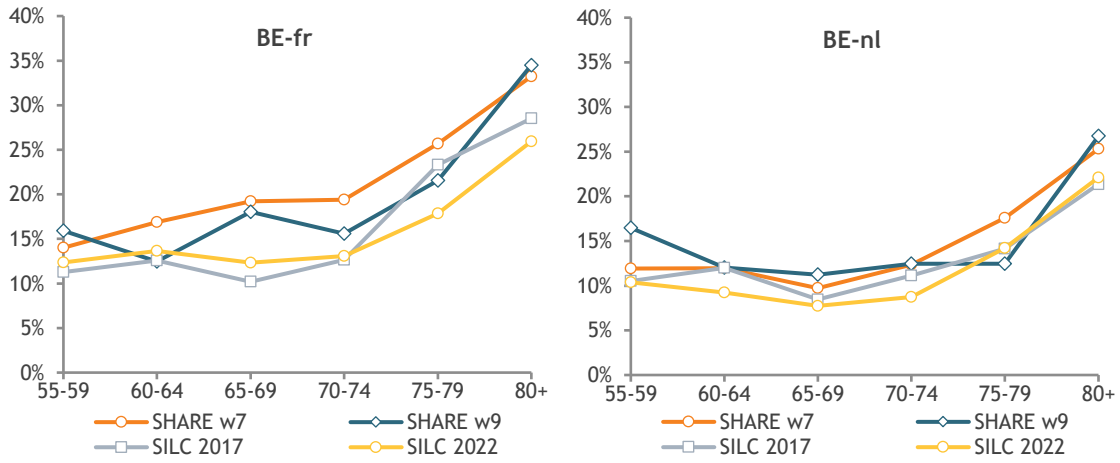
Collecting data on all aspects of health takes up a big part of the SHARE questionnaire. Health is of obvious importance for labour market participation, social participation, well-being and other dimensions of life. Therefore it is important that the SHARE sample is also representative for people with less than good health. People with health problems can be more reluctant to participate in a survey.

The SHARE sample can be compared with the EU-SILC data regarding health status, as both contain the same question about limitations in activities due to health problems. This is an important indicator of the health situation of people. In both surveys the three possible answers are: severely limited, limited, and not limited. Those answering "severely limited" are likely to suffer from serious health problems.

[20] It might be surprising that this proportion is not exactly zero in the first wave. This is because people in care homes often do not have their official residence there, but keep their old address (see Van den Bosch et al., 2011). When these were drawn into the sample, interviewers were sometimes directed by partners or neighbours to the care home.

In Figure 6 the proportions who are severely limited are compared between the last two complete waves of SHARE (waves 7 and 9) and the SILC data for the corresponding years (2017 and 2022) by age category from 55 up. Both SHARE and EU-SILC indicate that the proportion of people who are severely limited increases strongly after age 75. For the French-speaking community the proportions in SHARE are always higher than in EU-SILC for the age categories 65-69 and higher, except for the 75-79 category in wave 9. The SHARE curves are closer to those derived from EU-SILC for the Dutch-speaking community, but for the 80plus the former clearly exceed the latter. I also looked at the proportions of older people who were either limited or severely limited in activities. Those proportions are consistently higher in SHARE than in EU-SILC for all age categories. These results suggest that SHARE is better than EU-SILC at including older people with serious health problems in the survey. This could of course be related to the fact that unlike EU-SILC, SHARE includes older people in care homes.

Figure 6. Proportion who are severely limited in activity, by age category in SHARE and in EU-SILC



Notes: For SHARE: waves 7 and 9. For EU-SILC corresponding years 2017 and 2022.
Sources: SHARE data, EU-SILC data

4.4. Education level

The education level attained is a key variable in many domains of life, including health, labour market participation, income and well-being. Population data about the education level are available in the Census 2021, which is limited to a single year. The best source with data for all years is probably the Labour Force Survey, as its sample is large and participation is compulsory. However, published data (by Eurostat) are limited to the group aged below 65. A comparison for all years and all age groups is possible with EU-SILC. In all comparisons below, three education levels are distinguished: ISCED 0-2 (up to

lower secondary education), ISCED 3-4 (higher secondary education[21]) and ISCED 5+ (tertiary education).

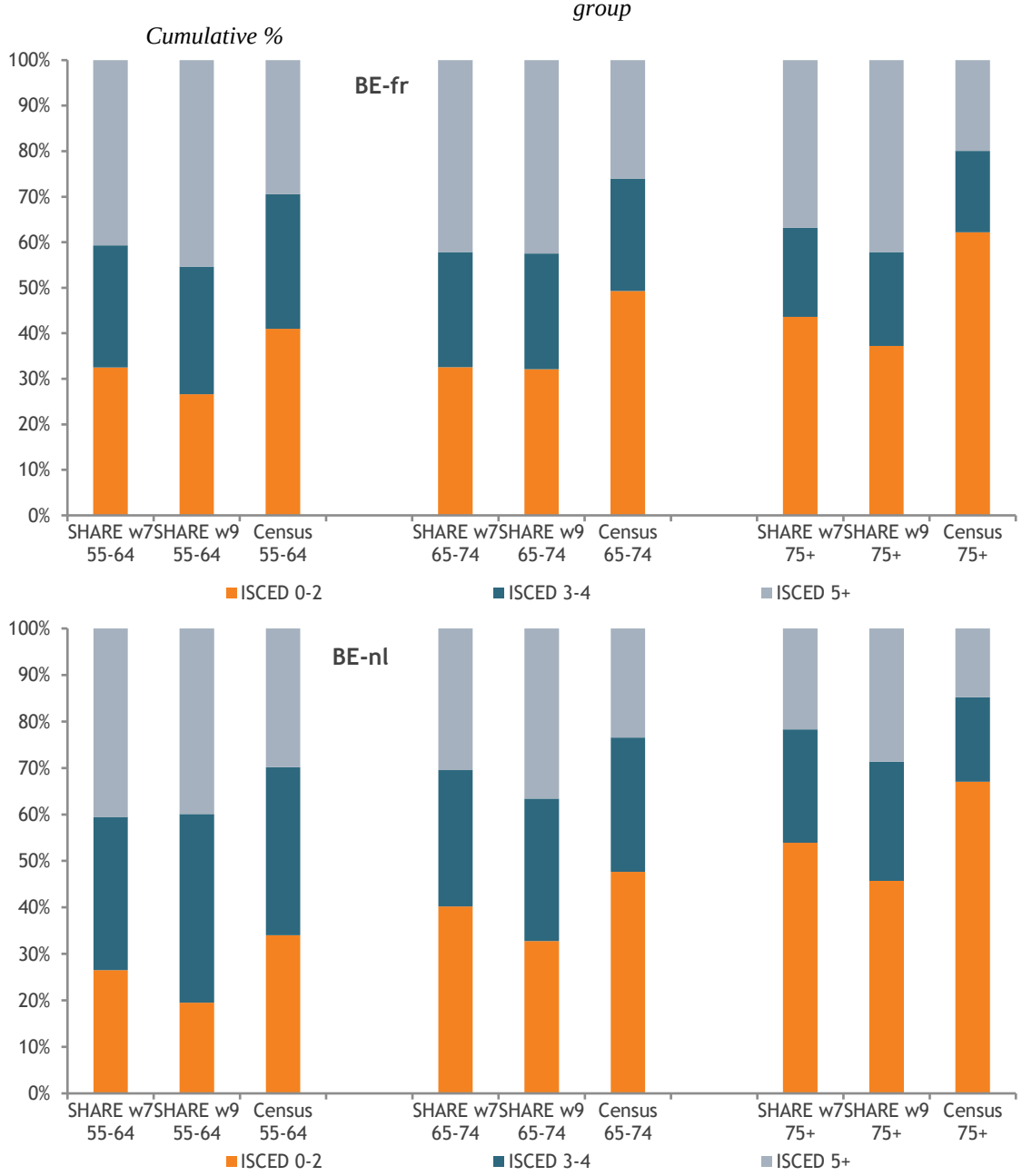
In Figure 7 the Census of 2021 is compared with SHARE wave 9 (2022) and wave 7 (2017)[22]. The latter is included because wave 9 is one year later than the Census, and as the education level attained is clearly higher the younger the birth cohort, even one year can make a difference. Compared with the Census, the proportion of people with at most lower secondary education is much lower in the SHARE sample in all age groups. Conversely, the proportion with tertiary education is higher in SHARE compared to the Census 2021. The differences are slightly smaller for SHARE wave 7 than for wave 9, but follow the same pattern. The contrast with the 2021 Census is larger for the French-speaking sample than for the Dutch-speaking one. The education data in the 2021 Census are derived from a large variety of administrative and survey sources. They have some limitations, including that the education level could not be ascertained for some people, in particular immigrants.[23] (See Statbel, 2024, for more details.)

[21] ISCED 4 corresponds to post-secondary non-tertiary education, but in Belgium very few people have this level of education.

[22] Wave 8 is not used, because the COVID-19 pandemic terminated it prematurely, see above.

[23] Among those aged 55+, the proportion of missing education data in the Census 2021 is 6% in Flanders, 7% in Wallonia and 15% in Brussels (calculated from the published data).

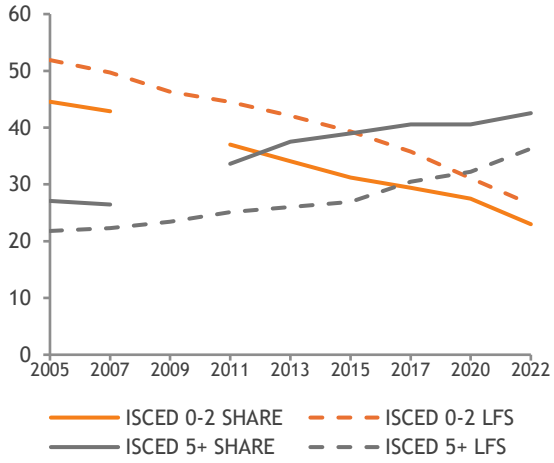
Figure 7. Education level of SHARE respondents in waves 7 and 9 compared with the Census 2021, by age group



Source: Statbel, Census 2021 and SHARE data

Similar differences are found between the Labour Force Survey (LFS) data for the age group 55-64 and SHARE (Figure 8). LFS results by age group are only published for Belgium as a whole, not by region. In both data sources the proportion of people with at most lower secondary education declines dramatically during the period 2005-2022, while the proportion with tertiary education strongly increases. However, the first proportion is always lower in SHARE than in the LFS, while the opposite is true for the proportion with tertiary education. The gaps appear to have narrowed somewhat in the last SHARE waves.

Figure 8. Education level of SHARE respondents aged 55-64 compared with the Labour Force Survey, 2005-2022
Cumulative %



Source: SHARE micro-data and Eurostat database table edat_lfse_03

Figure 9, which presents the comparison with EU-SILC, shows that the double tendency of a fall in the proportion of people with at most lower secondary education and an increase in the percentage of people with tertiary education occurs in all age groups. Interestingly, for the Dutch-speaking samples the SHARE curves and the ones derived from EU-SILC virtually coincide. For the French-speaking the proportion of people with at most lower secondary education is lower in SHARE than in EU-SILC, while the opposite is true for the percentage of people with tertiary education.

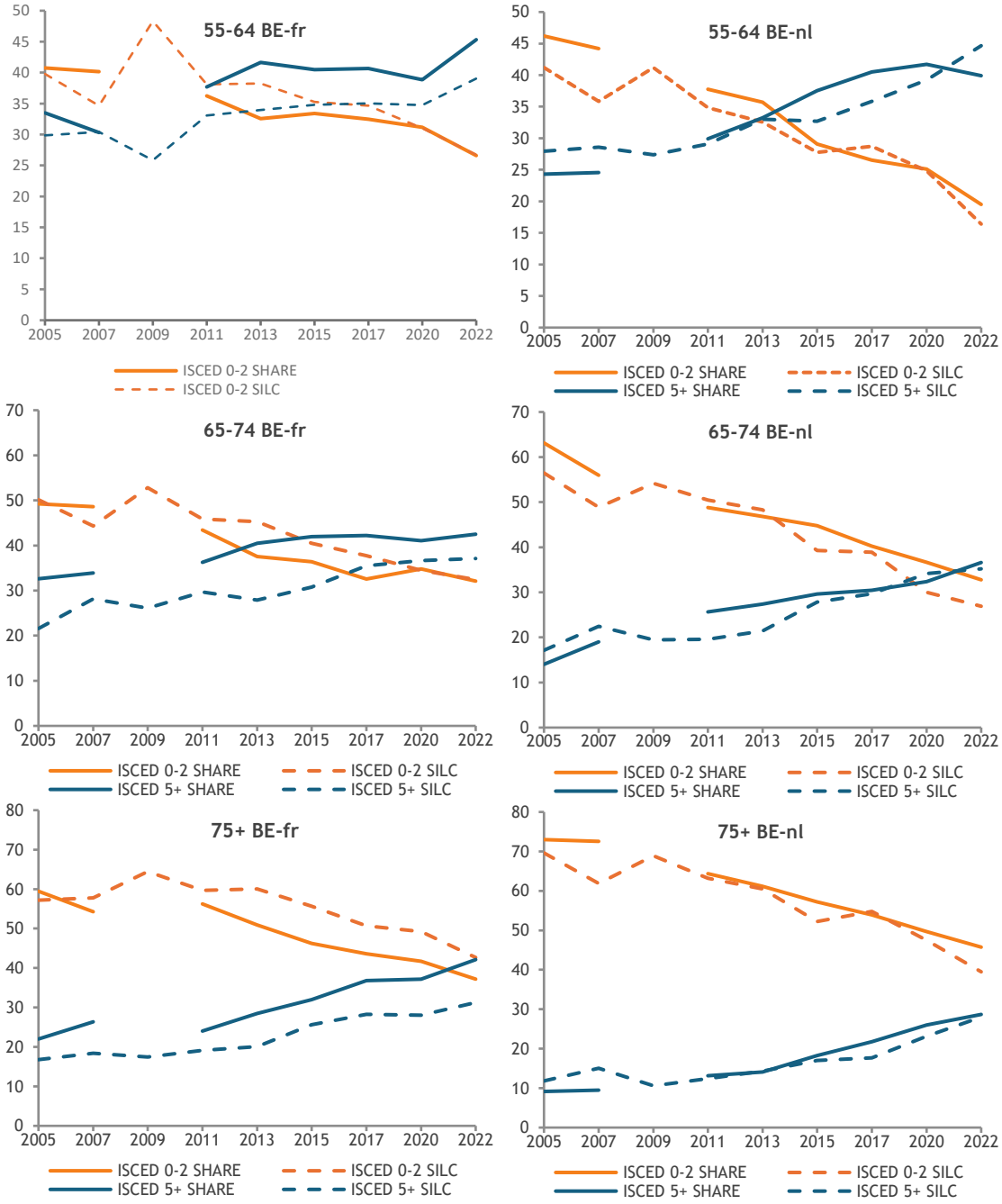
There are two possible interpretations of these findings. First, people with less formal education may be less likely to participate in surveys, including SHARE. Second, people may overstate their level of education in surveys, e.g. saying that they have tertiary education when they had started a university study but never finished it. The available evidence suggests that the second mechanism is probably not very important. The administrative data used in the Census 2021 (i.e. from educational institutions on diplomas they have awarded) are available mostly for those that finished their education fairly recently, while for the majority of the older population the sources were probably the Censuses of 2001 or 2011.[24] If people overstate their education in a survey, there is no reason why they would not do the same thing in response to similar Census questions. Another reason to favour the first interpretation is that within SHARE it was found that respondents with more education are more likely to participate in the next wave.[25]

Summing up, it appears that older people with at most lower secondary education are underrepresented in the SHARE samples relative to the population, while those with tertiary education are overrepresented. The greater correspondence between the proportions in SHARE and in EU-SILC is probably due to EU-SILC suffering from the same problem.

[24] Other sources which do not rely on diplomas awarded include sample surveys like the LFS and EU-SILC and the data collected by labour market organizations from people looking for employment.

[25] In a logistic regression of non-response in the next wave (excluding waves 7 and 9 and those who had died) on education, controlling for wave and age category, in the French-speaking sample the odds-ratios of non-response for those with ISCED 3-4 and ISCED 5+, relative to those with ISCED 0-2 were 0.83 and 0.69 respectively, and in the Dutch-speaking sample they were 0.82 and 0.65 respectively (standard errors between 0.04 and 0.06).

Figure 9. Education level of SHARE respondents compared with EU-SILC, by age group, 2005-2022
Cumulative %



Source: SHARE microdata and EU-SILC microdata

5. Conclusion

The Survey of Health, Ageing, and Retirement in Europe (SHARE), or "50+ in Europe", is a study that focuses on the European population aged 50 and over. Data are collected every two years in the areas of health (both mental and physical), socio-economic status, and social networks. The study started in 2004 and has since undergone nine rounds of data collection. In addition there were two special COVID-19 surveys. Belgium is one of the few countries that participated in all waves. Despite the panel set-up, the sample needs to be refreshed regularly (i.e. new respondents need to be found).

For reasons of organization (two languages) and of funding (partly by regional authorities) the Belgian participation in SHARE is coordinated by two teams: one at the University of Liège (for the French-speaking community) and one at the University of Antwerp (for the Dutch-speaking community). The two teams work closely together for sampling, questionnaire development and valorisation of the SHARE data.

The sample design is a two-stage sample with stratification and clustering of municipalities in the first stage, and random sampling of persons in the second stage. This has been used both for the initial sample for the first wave and for the refreshment samples in later waves. The goal is to approximate an EPSEM ("equal probability selection method") design as closely as possible.

The sample was regularly refreshed by additional samples. These are needed to cover the people who have joined the target population (i.e. have become 50) since the last wave and to compensate for attrition of the panel due to non-response and death. An important result is that the retention rate (i.e. the proportion of respondents in a particular wave who participate also in the next wave) is consistently high. The average value for the French-speaking community is 78% and for the Dutch-speaking community 83%.

Some indicators of the representativeness of the SHARE samples for the population of older people in the two largest communities of Belgium were studied. Considering the distribution by age category and comparing with population figures, it is found that that the age group 55-64 is increasingly underrepresented in the SHARE samples for both communities. This has motivated the Dutch-speaking SHARE team to focus the refreshment sample B in wave 10 on this age group. On the other hand, the proportion of the very old (85plus) shows a clear increase in the SHARE samples so that in wave 9 this proportion has caught up to the figures for the population. Second, SHARE includes older people in care homes and the proportion of the sample living in such an institution has increased substantially across the waves. Third, compared to EU-SILC the SHARE sample contains a larger proportion of older people who are severely limited in their activities due to health problems. This suggest that SHARE is better than EU-SILC at including older people with serious health problems in the survey partly thanks to interviews in care homes. Fourth, it appears that older people with at most lower secondary education are

underrepresented in the SHARE samples relative to the population, while those with tertiary education are overrepresented. The weights provided in the SHARE databases will correct at least some of these deviations of the sample from the populations. Also, there are statistical and econometric tools that allow research to be conducted while controlling for sampling bias, see e.g. Flawinne et al. (2023).

References

Bergmann, M., T. Kneip, G. De Luca and A. Scherpenzeel (2019). *Survey participation in the Survey of Health, Ageing and Retirement in Europe (SHARE), Wave 1-7. Based on Release 7.0.0*. SHARE Working Paper Series: 41-2019. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

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.

Bergmann, M., M. Wagner, Y. Yilmaz, K. Axt, J. Kronschnabl, Y. Pettinicchi, D. Schmidutz, K. Schuller, S. Stuck and A. Börsch-Supan (2024). SHARE Corona Surveys: study profile. *Longitudinal and Life Course Studies* 15 (4) 506-525.

Bethmann, Arne and Michael Bergmann (2023). *SHARE Sampling Guide 10*. SHARE Working Paper Series 87-2023.

Börsch-Supan, A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck and S. Zuber (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, 42(4) 992-1001.

De Luca, Giuseppe, Paolo Li Donni and Moslem Rashidi (2021). Weights and imputations in SHARE Wave 8. In: Bergmann, M. & Börsch-Supan, A. (eds.), *SHARE Wave 8 Methodology: Collecting cross-national survey data in times of COVID-19*. Munich: MEA, Max Planck Institute for Social Law and Social Policy

Flawinne, Xavier , Mathieu Lefebvre, Sergio Perelman, Pierre Pestieau and Jérôme Schoenmaeckers (2023). Nursing homes and mortality in Europe: Uncertain causality. *Health Economics*, 32 134–154.

Kneip, T., Malter, F., & Sand, G. (2015). Fieldwork monitoring and survey participation in fifth wave of SHARE. In F. Malter & A. Börsch-Supan (Eds.), *SHARE Wave 5: Innovations & methodology*. Munich: Munich Center for the Economics of Aging (MEA) at the Max Planck Institute for Social Law and Social Policy (MPISOC). pp. 102-159.

Sand, G. (2019). Fieldwork monitoring and survey participation in the seventh wave of SHARE, in: Bergmann, M., Scherpenzeel, A., & Börsch-Supan, A. (eds.). *SHARE Wave 7 Methodology: Panel innovations and life histories*. Munich: MEA, Max Planck Institute for Social Law and Social Policy, pp.

Scherpenzeel, A., Axt, K., Bergmann, M., Douhou, S., Oepen, A., Sand, G., Schuller, K., Stuck, S., Wagner, M., and 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.

Statbel (2024). Census 2021: *Metadata voor het thema 'Onderwijs.* https://statbel.fgov.be/sites/default/files/files/documents/Census2021/EDU_METADATA_CENSUS_2021_NL.pdf (accessed 22/12/2024).

Van den Bosch K, Willemé P, Geerts J, Breda J, Peeters S, Van De Sande S, Vrijens F, Van de Voorde C, Stordeur S. (2011). *Toekomstige behoefte aan residentiële ouderenzorg in België: Projecties 2011-2025*. Health Services Research (HSR). Brussel: Federaal Kenniscentrum voor de Gezondheidszorg (KCE). KCE Reports 167A. D/2011/10.273/63

Federal Planning Bureau

The Federal Planning Bureau (FPB) is a public agency that carries out, in support of political decision-making, forecasts and studies on economic, social-economic and environmental policy issues and examines their integration into a context of sustainable development. It shares its expertise with the government, parliament, social partners, national and international institutions.

The FPB adopts an approach characterised by independence, transparency and the pursuit of the general interest. It uses high-quality data, scientific methods and empirical validation of analyses. The FPB publishes the results of its studies and, in this way, contributes to the democratic debate.

The Federal Planning Bureau is EMAS-certified and was awarded the Ecodynamic enterprise label (three stars) for its environmental policy.

Rue Belliard 14-18 – Belliardstraat 14-18, 1040 Brussels

+32-2-5077311

contact@plan.be

Responsible publisher: Baudouin Regout