SHARE ERIC
SURVEY OF HEALTH, AGEING AND RETIREMENT IN EUROPE

SHARE WORKING PAPER SERIES

SHARE Sampling Guide – Wave 8

Version 8.0.1

Arne Bethmann, Michael Bergmann, Annette Scherpenzeel

Working Paper Series 33-2019

SHARE-ERIC | Amalienstr. 33 | 80799 Munich | Germany | share-eric.eu

mea | MAX-PLANCK-INSTITUT FÜR SOZIALRECHT UND SOZIALPOLITIK
Munich Center for the Economics of Aging
Horizon 2020
Federal Ministry of Education and Research
DFG
NIH National Institute on Aging

1. Introduction

In order to produce high quality data, the Survey of Health, Ageing and Retirement in Europe (SHARE) employs processes at each step of the survey that help to minimize the total survey error (cf. Biemer, 2010). This means that in addition to a well-designed survey instrument, proper field procedures, and diligent data management, it is vital to use sound probability sampling methods to draw a representative sample in each country. This does not mean that there is a single sampling procedure, which is perfect for all countries however (cf. Kish, 1995, p. 173). There is rather a set of basic considerations that should be adhered to in order to optimize the sampling design for each specific country. We hope this document provides you with guidelines helping to devise such a sampling design.

1.1 Overview of the sampling process

Each wave of the SHARE study consists of the main data collection as well as two rounds of preparatory data collection (i.e. pretest and field rehearsal) roughly 6 and 12 months before that. Producing a gross sample that is fielded in the respective data collection can be described as a four step procedure with some specific features in each case (see Figure 3 in the appendix). Firstly, in the development phase, the respective Country Team and Field Agency develop a specific sampling design in close cooperation with SHARE Central. This incorporates deciding on a sampling frame and specifying the sampling procedure. Secondly, the sample must be drawn by the Country Team / Field Agency (or through a third party, e.g. institution hosting a national register) and subsequently processed to produce a gross sample file. In a third step SHARE Central checks that the gross sample conforms to SHARE standards. In the last step SHARE Central and CentERdata prepare and upload the gross sample data for distribution through the Sample Control software (Sample CTRL; formerly: Sample Distributor, SD), which are then combined with the respective addresses uploaded by the Country Team.

1.2 Structure of the document

The following sections will describe the specific tasks of the whole process in more detail and provide some background information. In section 2 we will discuss theoretical considerations for the sampling design. We define the target population in SHARE, discuss possible sampling frames and connected issues like under and over coverage and give our recommendations. The section will also be concerned with the probability sampling procedure as such and provides a very brief introduction of the main technical features commonly encountered in SHARE samples (especially stratification and multi-stage sampling). We will then walk you through the sampling procedure step by step in section 3 using different sampling scenarios, including baseline and refreshment samples, which might apply to your country. The documentation and delivery requirements will also be discussed here. Lastly, the specifics of pre-test and field rehearsal samples will be covered.

2. Sampling design theoretical considerations

- Find most appropriate sampling frame for the target population in your country
- Consider what coverage issues to expect
- Think about regional clustering and an appropriate single or multi-stage design
- See what variables in the sample frame you can stratify on

2.1. Target population

The SHARE target population – the one we want to be able to make representative statements about and hence draw the sample from (Kish, 1965, p. 6; Lohr, 2010, p. 3) consists of all persons aged 50 years and over at the time of sampling who have their regular domicile in the respective SHARE country. For practical reasons – due to the different sampling frames and designs used across the SHARE countries – the age eligibility rule employed in SHARE is: year of birth <= year of sampling - 50, e.g. if the year of sampling is 2018 all birth cohorts older or equal 1968 are age eligible (see next section for other eligibility issues).

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. As these statuses only become known during fieldwork they do not need to be considered during the sampling phase. Residents of retirement or nursing homes on the other hand are considered eligible due to their importance to substantive research on ageing, meaning that they should be sampled in countries where drawing a probability sample for this population is possible.

2.2 Sampling frame and coverage

Probability based sampling requires a list of all sampling units (individuals in this case) in the target population called a sampling frame (Ibid.). Ideally this list can be acquired from national or regional registers of the residential population. If these registers include a date of birth or similar age information it can be easily restricted to the age eligible population. If that is not the case we have a sampling frame with a large amount of over-coverage as it includes all persons under the age of 50 who are not part of the target population ("Not eligible for survey" area in Figure 1). Consequently age-eligibility has to be established during fieldwork in these cases via screening of all members of the sampled household in the SHARE Cover Screen (CV). While this might not be a major issue for the calculation of sampling weights and subsequent data analysis, it will trigger a large increase in survey costs due to the extra number of households that have to be sampled, contacted, and screened in order to achieve the target number of survey interviews.[1]

[1] See section 3 for more information on response rate calculations as well as the difference between gross and net sample.

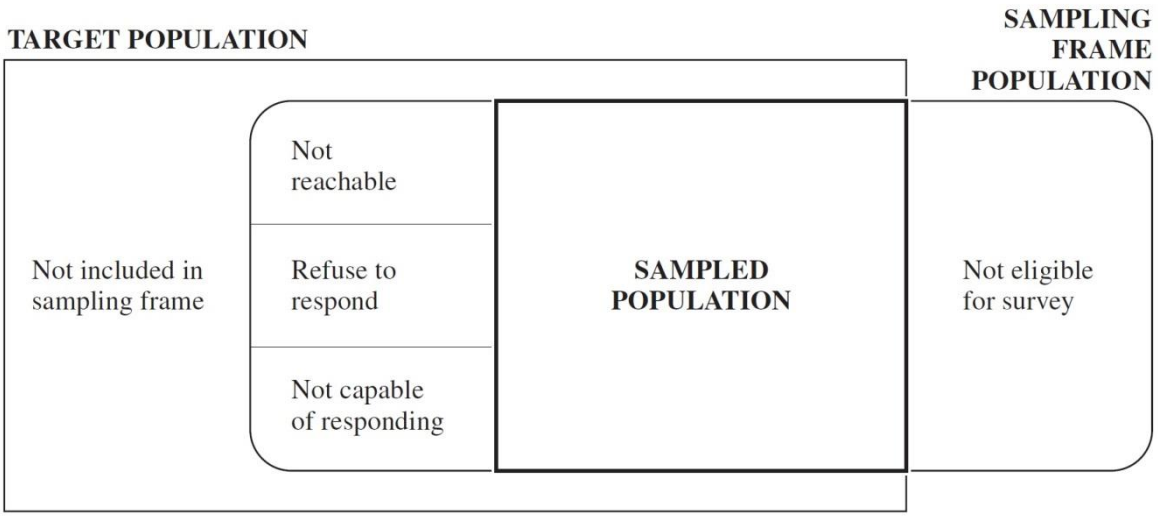

Figure 1: Target population and sampling frame (Lohr, 2010, p. 4)

Depending on the quality and type of register used you may also encounter the more severe problem of under coverage, which means that parts of the target population are not represented in your sampling frame and therefore have a probability of zero to be sampled and subsequently interviewed ("Not included in sampling frame" area in Figure 1). A common example would be the use of a telephone register, which – by definition – does not include households without a phone line. Depending on what percentage of the population does not have a registered telephone line the amount of under coverage can be extensive. This would not matter much if persons in households with a registered telephone line had on average the same characteristics as those without. But in most cases you would probably find that to be an unrealistic assumption and individuals not covered in the register might be quite different e.g. with regard to mobility. In such a situation, analyses based on the sample can be seriously biased.[2]

We would advise to be very careful when selecting your sampling frame and consider potential issues of over and under coverage (see Scherpenzeel et al., 2016). An official person register of the residential population which includes information on individual age is the preferred sampling frame for SHARE as it can be assumed to have the highest possible quality in most cases and is most likely to cover the target population rather precisely, especially when it is updated on a continuous basis. These so called version A samples will also provide the most straight forward sampling and fieldwork process since it is not necessary to conduct age eligibility screening for the residents of each household (see Figure 4 in the appendix). If such a register is not available or accessible in your country you have to draw a version B sample for which we recommend using (in decreasing order of preference):

- official person registers of the residential population without age information,
- official registers identifying households,
- high quality telephone registers,

or other types of register that allow for the identification of individual households. Please note that in all these cases you will have to employ a rather costly screening process. If neither type of register can

[2] For a more thorough treatment of sampling frame problems please refer to Kish (1965, p. 53).

be used for sampling, another, less preferable, solution is to use address or building registers. When using these or similar registers you will also have to do a manual listing of households at each sampled address / building. We will describe the listing procedure in section 3 where we also present area sampling as a backup solution if you cannot use a register at all.

2.3 Random sampling

In order to draw valid inference from samples to target populations for any statistical characteristic (e.g. average health, pension wealth etc.) it is of utmost importance that the sampling procedure is in accordance with the best practices in probability sampling. The key issue here is to be able to calculate inclusion probabilities or the likelihood to be sampled for every unit in the sample. Sample designs that produce an equal probability of selection for each member in the population, so-called EPSEM designs ("equal probability selection method") are often preferable, as these are most efficient in terms of limiting the sampling variance. The most common ones you might come across in SHARE samples would be Simple Random Sampling (SRS), Systematic Sampling (SYS), and multi-stage Probability Proportional to Size (PPS) sampling (see Kish, 1965 and Lohr, 2010).

In a lot of cases you will not be able to strictly follow these approaches due to disproportional stratification (e.g. when oversampling specific birth cohorts in refreshment samples or when defining minimum sample sizes for each region to make sure statistical inference is also possible at a regional level) or specific issues of multi-stage sampling. This means that you will most likely have a non-EPSEM sampling design leading to variation in selection probabilities between units in the sample. In turn design weights have to be calculated to correct for these variations in subsequent analyses. Unequal inclusion probabilities will systematically lead to non-representative samples over-representing units with high inclusion probabilities and under-representing those with low probabilities. Luckily, as long as this systematic mechanism is known and can therefore be modelled, the calculated design weights should provide a perfect correction for any bias that would occur due to the sampling procedure. Therefore, it is important to provide and document all relevant information pertaining to the actual sampling procedure.

In addition to having to use design weights when analysing the data there is a second drawback of having unequal sampling probabilities. They can lead to increased variance in the design weights which then produce larger sampling variability, reducing the precision of the estimates (i.e. larger standard errors, confidence intervals etc.).[3] For practical purposes this can be expressed as a reduced effective sample size. We therefore advice you to use sampling designs that strike a balance between getting as close to equal sampling probabilities as possible, and expected costs and other arguments for unequal probabilities like oversampling (see section 2.5).

[3] There are specific sampling approaches that make use of varying sampling probabilities in order to increase precision, depending on domain specific nonresponse, subpopulations with higher variance or specific domains of interest. These approaches can become pretty involved though and should only be considered if they are expected to produce substantive improvements in precision or would alleviate severe sampling problems. Please contact SHARE Central if this applies in your case.

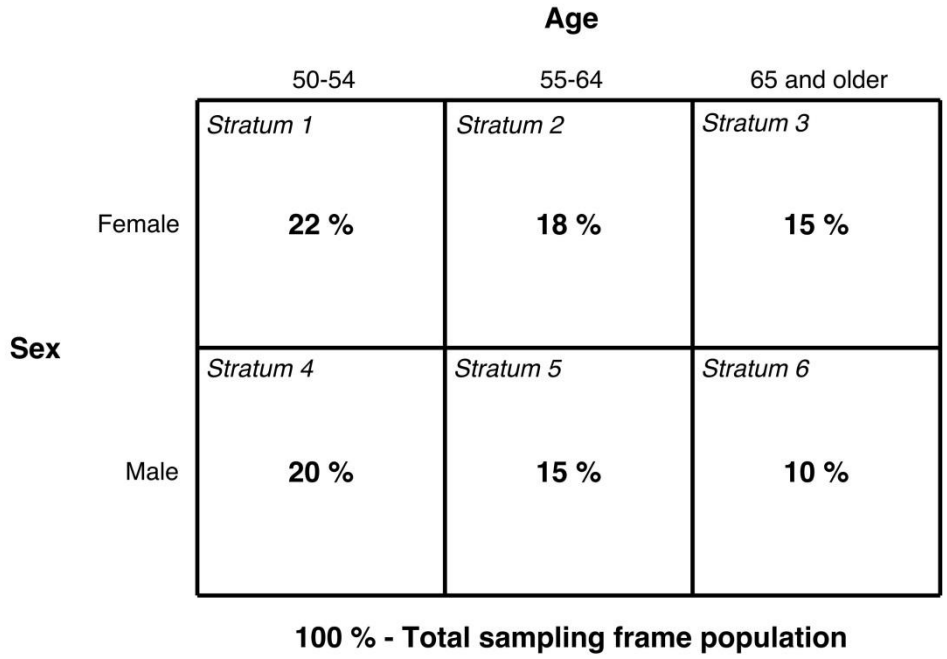

Figure 2: Example for dividing the sampling frame into strata

2.4. Stratification

An important tool for controlling the sampling procedure is stratification. The basic idea is that you slice your sampling frame into separate parts or strata along one or more variables that are included in the sampling frame (e.g. region, gender, age groups etc.) and then draw independent probability samples within these strata (see Figure 2 for an example).[4] Proportional stratification means that the proportion of the gross sample that you draw in each stratum is given by the observed proportion of the respective stratum in the sampling frame. For example if you want to draw a gross sample of 2,000 individuals, 15 % – or 300 – of those should be drawn from stratum 5 (i.e. men between the age of 55 and 64). The advantage of this type of stratification is that you can be certain, that your gross sample will represent the target population with regard to the categories you have stratified on, whereas in a simple random sample without stratification the corresponding percentages would randomly deviate from the distribution. The statistical result of that is an increase in precision that might offset some of the losses in effective sample size due to unequal selection probabilities. The benefit of this approach depends on the specific variables you have available and how strongly they correlate with variables of interest in the survey. Also there are other stratification approaches that might be better suited for certain sampling situations (see Kish, 1965, chapters 3 and 4). So if you have additional variables in your sampling frame we strongly advise you to consider stratifying your sample. Feel free to contact SHARE Central to discuss specifics.

2.5 Oversampling

Oversampling, on the other hand, is useful when you want to increase the proportion of cases from specific strata that end up in the sample, for example when drawing refreshment samples (see also section 3.3). Here you want to get a substantive number of interviews with individuals that became age-eligible since the last sample (baseline or refreshment). Depending on the size of your refreshment

[4] Stratification can be done on any level of a multi stage sampling procedure as discussed in section 2.6, e.g. clusters, households, or individuals.

sample proportional stratification might lead to a relatively low number of cases in these birth cohorts in comparison to older cohorts that would consist of the refreshment and a substantial number of panel cases as well. That is why you might decide to draw a larger proportion of individuals in the strata for the younger cohorts.[5] While this approach helps to fine-tune the age distribution in the gross – and consequently net – sample, you should keep in mind that it will also increase the variability in selection probabilities for the units. The best allocation of refreshment cases depends on different factors like the size of the refreshment sample and prior sampling designs. You should contact SHARE Central to discuss which approach would be best in your situation.

2.6 Clustering

So far we have argued under the assumption that you would want to draw a sample from a complete national register covering every region in your country. While this is advisable from a statistical precision standpoint it comes with (at least) two distinct drawbacks that will render it unrealistic for practical purposes in most countries. Firstly, it might be very resource intensive in terms of time and money to create a high-quality sampling frame covering the whole country (think e.g. of regional or community registers that have to be collected individually). Secondly, the survey cost increases substantially the more interviewers have to travel in order to contact and interview the persons in the sample. That is why it will often be more resource-efficient to cluster the sample in specific regional units, leading to reductions in the cost of creating a sampling frame (in countries where no high-quality national register is available) and in the cost of conducting the interviews.

For practical fieldwork purposes the preferred solution is often employing a multi-stage sampling design, except for cases where SRS is possible. Meaning that in the first stage you will sample regional units (primary sampling units or PSUs, often municipalities, zip code areas or similar units) and that only individuals or households residing within these regional clusters will be part of the second-stage sampling frame. This automatically reduces the amount of work that has to be invested in creating the sampling frame and restricts the distances interviewers will have to travel in order to reach all possible respondents. The sampling procedure for the clusters is similar to sampling individuals or households as described above. You will still need a list of all regional units that you can draw a sample from and you are still advised to use stratification on relevant variables available in your list like municipality size, urbanization, region etc. This works similarly to the example for individuals in section 2.4 depending on the population density and hence the expected number of eligible individuals in each regional units you might want to combine or split up regional units before drawing the sample of regional clusters. There are situations where these clusters still cover areas that are too large for efficient field work. In these cases it might be sensible to add a second stage of regional clustering,[6] e.g. zip code areas within municipalities, to further reduce travel costs for interviewers.

There are two caveats though when using these clustered types of sampling. The first is again about the reduced precision due to an increase in the variability of selection probabilities. Take for example a two stage sampling design, i.e. sampling regional clusters in the first stage and then individuals

[5] The case of refreshment samples will be discussed more in section 3. Also, oversampling and calculating the proper sample sizes for specific subgroups is not trivial in a lot of cases. A spread-sheet that will help you with the calculations will be provided by SHARE Central.

[6] For the sake of clarity the smallest regional units in the sampling design will be referred to as “clusters” in the following discussion, regardless whether these are primary or secondary sampling units.

therein in the second stage. Given that you sample the clusters with equal probabilities and have a fixed sample size within each cluster (say: 10 individuals) a person in a small cluster (e.g. rural with 100 residents) will have a much larger sampling probability (10/100 = 10 %) than somebody in a large cluster (e.g. urban with 1,000 residents: 10/1000 = 1%). The common approach to eliminate that problem is to do Probability Proportional to Size (PPS) sampling which produces equal sampling probabilities for the individuals through setting the sampling probabilities of the clusters proportional to their size: if a cluster contains 0.5 % of the country's total residential population 0.5 % will also be the sampling probability for that regional cluster.[7] If you use strata for cluster sampling the percentage of the population within the specific stratum is the relevant number. In case you have more than one stage of regional clustering (e.g. zip codes within municipalities or districts within cities) you would use this PPS approach only in the first stage.

The second caveat cannot be alleviated that easily since it is due to the regional clustering itself and stems from the observation that individuals living in close proximity to one another are subject to similar influencing factors (which might also be the interviewer), are therefore more similar in their answers and in turn contribute less new information to subsequent analyses. The result is a loss of precision in the estimates (called clustering design effect). Its extent can only be assessed after the data are collected (or from other surveys) and is specific to each item of interest (depending on its homogeneity within clusters), which makes it very hard to estimate the reduction in effective sample size and to compensate in the net sample size. Our advice is therefore to try to keep clustering as low as possible, i.e. you should try to have as many clusters as you can afford in your design with, in turn, only few interviews per cluster.

3. Practical sampling design examples

- Discuss sample design and implementation with SHARE Central and document in Sample Design Form (SDF)
- Coordinate with Country Team / Field Agency and draw sample according to SDF
- Slice gross sample into random fieldwork batches
- Prepare gross sample file, including the sample design variables, using provided Gross Sample Template (GST) and deliver to SHARE Central

In this chapter we will give examples of the practical implementation of the gold-standard version A sample and the alternative version B cases (see Figure 4 in the appendix for a quick overview of the different sampling processes). Please keep in mind, that the specific situation in your country will most likely be slightly different and that you need to adapt the corresponding example to reflect that. We advise you to refer to textbooks on sampling for more detailed descriptions (we recommend Kish, 1965 and Lohr, 2010). Also, please feel free to contact the sampling team at SHARE Central for support.

[7] For our example let us assume that 0.5 % is the sampling probability for the urban cluster, then the probability for the rural cluster is 0.05 % since it is only 1/10 of the size (100 vs. 1,000 residents). The sampling probability for individuals is given by the sampling probability of the cluster multiplied by the sampling probability within the cluster. For an individual in the rural cluster that would be 0.05 % x 10 % = 0.005 %, for the urban cluster 0.5 % x 1 % = also 0.005 %.

3.1 Version A: Individual registers with age information available

The first question for all SHARE sample designs is whether – and if so how – you should use regional clusters. Except in countries were travel costs are not a big issue you would most likely want use a multi-stage stratified cluster sampling design tailored to the specific situation in your country. In countries where a person register with individual age information can be used as the basis for sampling this could be done as follows for a two-stage PPS sample design:

1. **Determine the minimum cluster size** which is feasible to implement in the fieldwork. Let us assume that a minimum of 10 interviewed persons aged 50 or older located geographically near to each other is desirable, to limit the interviewer travel costs. The *net cluster size* would then be 10 persons. The *gross cluster size*, assuming a *response rate* of 40%, is then 25 persons. Usually, municipalities are used as *primary sampling units (PSUs)*.

2. **Make strata** according to the (NUTS) regions in your country, if desired the regions can be crossed with other available criteria such as settlement size (for example: small, medium and large municipalities) to constitute more *strata* and hence reduce sampling variability. In some countries, however, you might have very small municipalities where less than 25 persons live. In that case you should put some of these small municipalities together into one cluster, close in geographical proximity. A good way to do that is to define clusters on the basis of postal codes, for example. In addition, you can divide up large municipalities into multiple clusters (often also by using postal codes, or by neighbourhoods/city-quarters).

3. **Draw a selection of clusters** from each stratum: the *number of clusters to draw in a stratum* is proportional to the size of the 50+ population in the stratum. In total, you draw *k* clusters, where *k* is the desired *gross sample size* divided by 25. For example: Suppose you would need a *gross sample size* of in total 4000 individuals to attain the *net number of interviews* you want to have. Assuming again a *gross cluster size* of 25, the *total number of clusters to draw* then is 4000/25 = 160 (please keep in mind that these numbers are just a hypothetical example). Additionally, the number of clusters (*PSUs*) drawn per stratum must be proportional (see section 2.4) to the share of the stratum in the population aged 50 or older. Also you might want to employ PPS sampling within strata (see section 2.6).

4. **Draw random samples of individuals**, 25 in this example, aged 50 or older from your register within in each of the *PSUs* drawn in step 3. If applicable, we advise you to also use stratification for sampling individuals, e.g. using gender and age as described in chapter 2, in order to reduce sampling variance. This will result in a *gross sample* which is proportionally distributed across strata and not containing any sample points with less than 25 addresses located near each other.

3.2 Version B: Individual registers with age information not available

3.2.1 Household registers and individual registers without age information

If the register available lacks individual information on age it is not possible during sampling to determine whether an eligible person lives in the respective household. The same applies to registers that do not refer to individuals, but households (or entities that households can be derived from, like registered phone numbers). In all these cases fieldwork is more complicated, as household screening must be conducted in order to identify eligible individuals before interviews can be conducted. While being technically easy, since it is implemented in the Case Control software, it is very inefficient due to the

large number of ineligible households that need to be sampled and subsequently contacted and screened.

Apart from the necessary screening procedure during fieldwork, the implementation of version B samples is not very different from version A. You can follow the same four steps for the stratified cluster sample, with one major difference: instead of sampling from the target population (i.e. age 50 or above) you sample from the whole residential population. This has to be reflected when calculating the PPS probabilities for the clusters, as well as when drawing the sample of individuals or households. The restriction to the target population has to be done during fieldwork using eligibility screening of the households. Hence, when calculating *gross sample size* and *gross cluster size* you have to correct for the estimated loss due to screening. That means that you have to divide the *net cluster size* by the assumed *response rate*, as mentioned above, and additionally by the *proportion of households with at least one 50+ individual in the population* (this number can usually be found in national population statistics or Eurostat). For the example above we might assume this proportion to be 50% which would yield a *gross sample size* of 8,000 and a *gross cluster size* of 50 households, while the *net cluster size* and the *total number of clusters* would stay the same (i.e. 10 and 160). The last step of sampling from each selected *PSU* in the register is again similar to the version A sample, except that you cannot stratify on the individual level (e.g. on age or gender).

3.2.2 Address or building registers

In addition to not providing any information on age eligibility, address, building, and similar registers have the drawback that they do not allow for the identification of households at the time of sampling. It is therefore necessary to list all households or dwellings at each address or each building before the final *gross sample* can be prepared. This will produce considerable additional costs and organizational burden. If you have an address or building register that does provide additional information on specific dwellings (e.g. a postal register that has the name and address with something like “second floor, left apartment” or similar) you have basically a household register and can proceed as discussed in section 3.2.1.

In a first step you would draw a sample of addresses/buildings with a similar multi-stage procedure as used in the other types of register samples. In order to determine the necessary *gross sample size* and *gross cluster size* you need to estimate the *proportion of addresses/buildings that contain at least one private household* (i.e. no purely commercial units) and again divide by that quantity. In this respect, it would be helpful to clean the register from commercial addresses/buildings not belonging to the target population before sampling in case this information is available. For example going back to the household register example in 3.2.1 let us assume that 80% of the addresses/buildings contain at least one private household. Again *net cluster size* and *total number of clusters* would stay the same (10 and 160, respectively), but the *gross sample size* would increase to 8,000 / 0.8 = 10,000 and the *gross cluster size* to 50 / 0.8 ≈ 63.

Once you have drawn the respective samples in each cluster you have to send out “listers” who go to each address and write down the number of private households/dwellings, as well as the specific building layout (cf. Kish, 1965, section 9.6 “Listing dwellings”), so households in the sample can be identified

and hence contacted by an interviewer during fieldwork later on. After collecting and safely storing[8] all these lists you should select a single household via Simple Random Sampling at each address. Only these households would then constitute the *gross sample*.

3.2.3 No register available, area sampling

In the most unfortunate situation where you do not have any register available you would have to resort to area sampling, which is the most expensive, burdensome and least controllable of the sampling procedures employed in SHARE samples. In a first stage you could, again, use the regional clustering process discussed for the version A sample if you have at least a list of all available regional units (most often called PSUs) with corresponding information for stratification.[9]

Since we strongly advise to make a full listing of all households in each PSU it is even more important to keep these as small as possible. Therefore, municipalities could still be too large and geographical segments like neighborhoods, streets, building blocks or postal code areas might be a better option. The listing process for the addresses/buildings within each PSU is similar to the process described above. The aim is to produce a database of all addresses/buildings and consequently households/dwellings for each of the sampled PSUs that can be used for the sampling of households much like in register samples.

In general, we do not advise to use random walk procedures instead of the full listing since it further complicates the listing procedure and makes it more error prone. It was also shown in simulation studies that random walk procedures violate the assumption of equal selection probabilities (i.e. lead to biased selections; cf. Bauer, 2016), rendering it impossible to calculate proper design weights for data analysis and hence hindering valid statistical inference. This situation would be further aggravated when random walk procedures are not only used for listing households, but are integrated with interview attempts. In this case you have even less control about the household selection process, since it will be hard to ensure that interviewers comply with the selection protocol. Using random walk procedures is therefore the least desirable sampling approach by far and can only be done in close cooperation with SHARE Central.

3.3 Refreshment samples

Refreshment samples are drawn in order to account for sample size reduction due to panel attrition (natural mortality as well as longitudinal unit non-response). If feasible you should try to replicate the sampling design that was used in prior waves. While largely similar to a baseline sample, refreshment samples have a specific issue with regard to handling the population that became age eligible between the last and the current sample. If you have individual information on age (sample version A) you have several options: First, you can restrict your sample to only the birth cohorts that became age eligible since drawing the last sample. Imagine you drew the last sample in 2010, then the youngest eligible birth cohort would have been 1960. If you wanted to sample again in 2018 all the cohorts younger than 1960 up to and including 1968 would be age eligible. Second, you can oversample these youngest cohorts while simultaneously doing a proportional sample for the older birth cohorts, since the former cases would otherwise be underrepresented in your unweighted sample leading to low statistical power for analyses in this specific subgroup. A third option would be to just draw a simple proportional

[8] This information will be needed for fieldwork as well as weighting later on.
[9] If not even a list of PSUs is available you could use geographical segments from maps or aerial photographs.

sample over the entire eligible range of birth cohorts. This option is also the only possibility regarding sampling frames that do not provide age information on the individual level (sample version B). In this case you draw your sample as described in section 3. Which of these approaches you use should be discussed with SHARE Central. In case you plan to restrict your sample to the youngest cohorts or you want to oversample these cohorts, we also provide a spread-sheet template to help you with calculating the proper numbers.

3.4 Documentation of the sampling design

When designing a new baseline or refreshment sample it is required to document the specific details in the Sample Design Form (SDF, which will be provided by SHARE Central). This serves several purposes. Firstly, it is necessary that SHARE Central understands and approves the design to ensure that the sample complies with comparable quality standards across all participating countries, while considering the specific situation in each country. Therefore, the SDF is a tool to communicate and discuss the development and implementation of a sampling design between the Country Team/Survey Agency and SHARE Central. Because this is a serious task and hence should take some time, we strongly advise to start the sampling design process as early as possible. This applies in particular for countries where the application process for ordering a probability sample is known to be difficult and tedious. Secondly, The SDF is used to inform the process of quality checking the gross sample file provided by the Country Teams/Survey Agencies, before releasing it for data collection via the new Sample CTRL software. After fieldwork has ended the SDF and the gross sample are the basis for understanding the specific sampling procedure in order to calculate design weights for data analysis. And lastly, it will also provide information for end user documentation on sample design elements necessary to conduct proper analyses (e.g. correcting standard errors for cluster design effects).

3.5 Preparation of the gross sample file

For conducting the actual fieldwork, a gross sample file is needed that includes all sampled individuals or households which are loaded into the Case Control system via the Sample Control. In order for this process to work you need to deliver the gross sample file to SHARE Central at least four weeks before your national training sessions are scheduled (see Figure 3 in the appendix for a timeline of the sampling process). The file format and necessary columns are defined in the Gross Sample Template (GST, see appendix for a detailed description) that will be provided by SHARE Central and might differ depending on your specific sampling design. After receiving the gross sample file, SHARE Central will verify that it contains the necessary information for all sampled cases. Also some double checks of the sampling process will be conducted as far as that is possible based on GST and SDF. Please keep in mind that for data protection reasons addresses and last names must not be included in the gross sample file delivered to SHARE Central. First names should be provided for samples from registers of individuals.

We also require you to use subsamples or fieldwork batches. The sample batches must be random subsamples (within each cluster/PSU if applicable, so each one is represented appropriately in each batch) of the entire gross sample considering the specific sample design and are introduced sequentially depending on progress of fieldwork. Only after a batch has been exhausted, the next batch will be opened. The reasons for using batches are two-fold: First, it gives you more control over fieldwork progress and hence stimulates good performance as batches that are not started can be excluded from

the denominator of the response rate formula[10]. Second, it preserves your sample quality in case the fieldwork results are better than expected and the target number of interviews is attained before the end of fieldwork. If you do not have fieldwork batches, you are obliged to fully exhaust the complete gross sample, since stopping halfway would result in a nonrandom selection. Using batches allows you to stop fieldwork before the complete sample has been exhausted, thus saving costs.

The batches must be random subsamples (without replacement) of the final list of units (individuals or households) in the gross sample. The size of the first batch should be large enough that it would suffice to reach the planned number of net interviews under optimistic assumptions with regard to response rates.[11] The rest of the gross sample should be divided into additional batches depending on the field-work conditions in your specific country. In order to ensure proper randomization the subsamples should be taken within each cluster (if you use a multi stage sample).

It is important that the country team leader has full control over when a new batch is started, after checking the number of successful interviews and the response rate achieved so far. The decision to put a new batch in the field or not should be taken on the basis of

a) the number of interviews achieved in the first batch as compared to the desired total number of interviews;
b) the possibility of still improving the response rate / number of interviews in the first batch by extra efforts of the agency for the most difficult cases;
c) the fieldwork time left.

Activate new batches only after all households in the previous batch have been assigned a final code (=interview, refusal, or non-contact[12]). In addition, make sure that still enough fieldwork time is left to also have at least 6 contact attempts for all households in the new batch: make a projection of the time it took to finalize all units in the first batch to determine how many additional batches to release. Note that the addresses of the second (and any further) batch must be put in the field in all regions and clusters at the same time: A region that runs out of first batch addresses earlier than others will have to wait. There must not be made partial, regional-selective releases of batches. To activate a batch you will need to contact SHARE Central. In case of an agreement CentERdata will release the additional batch(es) through the Sample Control.

3.6 Pretest and field rehearsal samples

For pretest (Pre) and field rehearsal (FiRe) following the same probability sampling procedure as for the main data collection can often be a challenge or even prohibitively costly. While we encourage using the main sampling design for Pre and FiRe if doing so is possible without major problems, we

[10] Response rate definitions in SHARE pertain to: The American Association for Public Opinion Research. 2016. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition*. AAPOR. URL (02 May 2018): http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169thedition-final.pdf

[11] Whereas the total gross sample should be drawn large enough to suffice even under worst case conditions. Please feel free to contact SHARE Central to discuss realistic scenarios for the gross sample size.

[12] SHARE standards require that at least 6 contact attempts are made in each household (from which 2 must be done on weekday daytime, 2 on weekday evening, and 2 at the weekend) before the final code "non-contact" can be assigned (see Annex 1 "Specifications" of SHARE model contract.

also acknowledge that in many cases a less strict, non-probability approach might be sufficient. The main goal of Pre and FiRe samples is to be able to thoroughly test the instrument and fieldwork processes in as many different scenarios as possible. Apart from a large enough net sample size this means that all relevant combinations of respondent characteristics to test routing, different questionnaire modules, and variants of the fieldwork process should occur in the sample. A fully representative sample, on the other hand, is not mandatory for proper testing. Therefore, we advise to draw stratified or even quota samples based on age groups, gender, working status, and language (if there is more than one in your survey). A convenient way could be a commercial access panel or similar data source. Additionally, you can draw the samples for Pre and FiRe at the same time to save costs and time.

The number of net interviews in Pre and FiRe has to be at least 100 each. Countries which already participated in SHARE in previous waves should use the respective panel samples for testing. In addition to that you should draw refreshment samples that yield at least 50 net interviews if you plan to draw a refreshment sample for the main data collection in the current wave and 25 if you do not. You need to draw larger refreshment samples if you anticipate a need to top up the panel sample in order to reach 100 interviews. New countries that participate in SHARE for the first time need to draw sufficiently large samples for Pre and FiRe to reach the mandatory 100 interviews. Preparation of the gross sample file for Pre and FiRe has also to be based on the Gross Sample Template. It must also be delivered to SHARE Central sufficiently in advance of the corresponding fieldwork (at least 4 weeks before the national training sessions are scheduled), so it can be checked and transferred to the Sample Control.

References and recommended reading

- Bauer, J. (2016). Biases in Random Route Surveys. *Journal of Survey Statistics and Methodology*, 4(2), pp. 263–287.
- Biemer, P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly*, 74(5), pp. 817–848.
- Kish, L. (1965). *Survey sampling*. New York: John Wiley & Sons.
- Lohr, S. (2010). *Sampling: Design and analysis* (2 ed.). Brooks/Cole: Nelson Education.
- Scherpenzeel, A., Maineri, A. M., Bristle, J., Pflüger, S.-M., Mindarova, I., Butt, S., Zins, S., Emery, T., Luijkx, R. (2016). Report on the use of sampling frames in European studies. Deliverable 2.1 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221.

Appendices

Description of the Gross Sample Template (GST)

- Use the GST to deliver relevant gross sample information to SHARE Central
- Each row corresponds to one sampling unit, i.e. either
    - a person or
    - a household
- **Do NOT include last names or addresses! This would be a severe breach of data protection rules.**
- **Again for data protection reasons, use Cryptshare for all transfers of sample data**
- Only yellow fields have to be filled in for household samples

- Additionally, white fields have to be filled in for person samples. If the birth date is not available the dotted fields "age" or "age_category" can be included instead.
- Grey fields are mandatory (if applicable) in the Main sample only. For Pretest (Pre) and Field Rehearsal (FiRe) samples these can be left empty.

The following list provides some additional information about the GST fields:

pidcom and **hhidcom** – these are the person and household identifiers that will be used for all further internal data processing. It is therefore vital that these are **unique** and **consistent**. The first two characters of both ids are the country code of your country. For pidcom the last two digits are a person counter within the household (i.e. "01" for all new samples). For hhidcom the last character should always be an "A" for all new samples (Main, Pre and FiRe). The 6 digit number in the middle has to be generated to be unique within you country **across all past and current waves and sample types** as well. Please be very diligent with assuring these criteria!

first_name, gender, month, year, age, age_category – these are individual characteristics of a sample from a person register. In order to identify the sampled individual when contacting the household at least **first_name**, **gender** and **year** have to be provided and are therefore **mandatory for person samples**. year may be substituted for age or age_category if it is unavailable.

replicate and **REPLICATEACTIVE** – these indicators define the fieldwork batches that have to be used at least in the main data collection. Please see the section on gross sample preparation for information on how to create the batches. The replicates should be numbered starting with "1" not "0". REPLICATEACTIVE should be set to "1" for the batch that is active at the start of fieldwork.

linkage – this indicates if you will do record linkage in the main data collection. Please only fill in a "1" in this case and "0" otherwise.

nuts3, lau1, lau2 – these regional classifications are only mandatory in the Main sample. lau1 can only be provided if this classification is available in your country.

psu, **psu_stratum**, **psu_popsize**, **ssu**, **ssu_stratum**, **ssu_popsize**, **final_stratum –** describe the probability sampling design. Since probability sampling is only mandatory in the Main sample, these fields only need to be filled in Main, too. psu and ssu are indicators for the first and second sampling stage clusters respectively. Depending on your specific sampling design you might only have one clustering stage (or even none), in which case the ssu specific fields would stay empty. If you used stratification at one or both cluster sampling stages you must indicate the stratum of the specific psu or ssu in psu_stratum and/or ssu_stratum. You also need to provide the size of the target population within each psu and/or ssu if it is relevant for your sampling design. If you used stratification in the final sampling stage, individual persons or households within clusters, you need to indicate the stratum in the final_stratum field. These fields should cover the information necessary to describe most sampling designs encountered in SHARE. Please provide any additional fields necessary if your sampling design cannot be described with these standard GST fields, but do not delete columns.

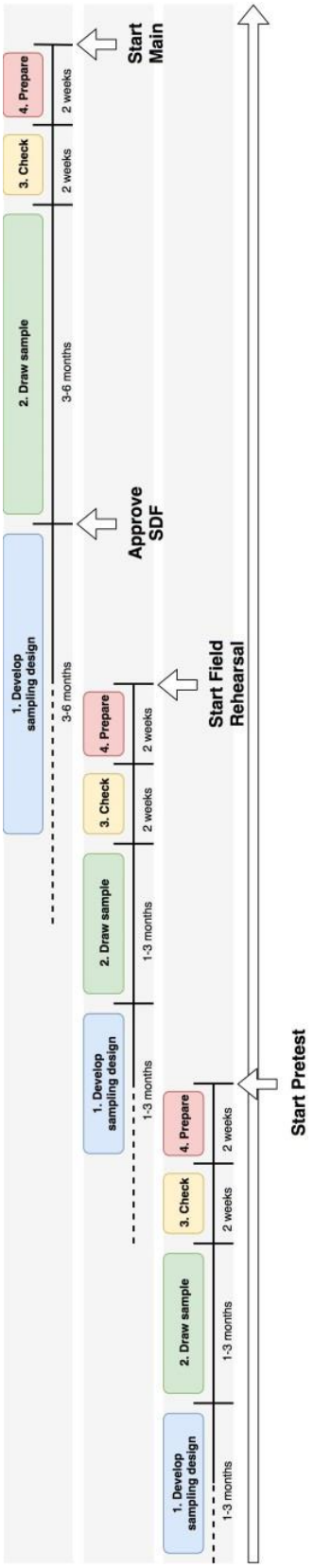

Figure 3: Sampling process time line

Figure 4: SHARE sampling decision tree

SHARE Sampling Decision Tree