# European Social Survey Round 9 Sampling Guidelines: Principles and Implementation

The ESS Sampling and Weighting Expert Panel, 26 January 2018

<u>Contents</u> <u>Page</u>


| | |
|---|---|
| Summary | 2 |
| 1. The ESS Sample Design Process | 3 |
| 1.1 Objectives | 2 |
| 1.2 The Sample Design Process | 2 |
| 1.3 The Sample Design Summary | 4 |
| 2. Principles for Sampling in the ESS | 5 |
| 2.1 Population Coverage | 5 |
| 2.2 Probability Sampling | 6 |
| 2.3 Statistical Precision | 6 |
| 3. Tips for Good Sample Design | 9 |
| 3.1 Sampling Frames | 9 |
| 3.2 Multi-Stage Sampling | 12 |
| 3.3 Stratification | 13 |
| 3.4 Predicting $deff_p$ | 14 |
| 3.5 Predicting $deff_c$ | 17 |
| 4. Calculating the Required Sample Size | 19 |
| Annex: Sample Design Summary | 20 |

**Summary**

The document sets out the principles of ESS sampling and provides guidance on how to produce an effective design that is consistent with these principles. It also explains the procedure required to approve a sampling design to be used in the ESS.

The document has been produced by the ESS Sampling and Weighting Expert Panel (SWEP), a group of experts appointed by the ESS Director to evaluate and help implement the sampling design in each of the ESS countries in close cooperation with National Coordinators (NCs). A core objective of the SWEP is to support NCs in implementing sample designs of the highest possible quality, and consistent with the ESS sampling principles.

**Changes to this Document**

These guidelines have been substantially restructured and rewritten since Round 8. The main changes are:

- The inclusion of explicit tips on how best to handle key aspects of sample design (section 3), including a summary box of "key tips" at the end of each sub-section;
- Worked examples of key calculations ($deff_p$, $deff_c$ and gross sample size);
- Separation of principles (section 2), sample design considerations (section 3), and a description of the process of developing a design and getting it approved (section 1);
- Minor revisions to the "Sign-off Form", which has been renamed the "Sample Design Summary" (Annex).

There are no substantive changes to the 'rules'.

# 1. The ESS Sample Design Process

For the first eight rounds of the ESS, the Sampling Expert Panel (SEP) worked with NC teams to develop the sample design for each country. In June 2017 a new Sampling and Weighting Expert Panel (SWEP) succeeded the SEP. The SWEP will continue to work with NC teams in much the same way that the SEP did previously. In this section we set out the objectives of this process, and how it should work.

Each participating country will be allocated a sampling expert, with whom the NC team should communicate on all matters related to sampling. The experts who make up the panel are:

| | |
|---|---|
| Peter Lynn | (Chair, University of Essex, U.K.) |
| Annette Scherpenzeel | (Munich Centre for the Economics of Aging, Germany) |
| Mārtiņš Liberts | (Central Statistical Bureau of Latvia) |
| Olena Kaminska | (University of Essex, U.K.) |
| Tim Goedemé | (University of Antwerp, Belgium) |

## 1.1 Objectives

The objectives of the sample design process are:

- To ensure that sample designs are consistent with the ESS Round 9 Specifications and are of the highest quality possible;
- To ensure that NC teams are able to identify the most cost-effective sample design parameters;
- To ensure that all relevant details of sample designs are fully documented in the Sample Design Summary (SDS);
- To ensure that all relevant sample design indicators are collected and provided in the Sample Design Data File (SDDF).

To achieve these objectives, the SWEP aims to provide support and guidance to NC teams where it is needed. This document constitutes the core of this guidance, but your allocated sampling expert is also available to answer queries and provide help.

## 1.2 The Sample Design Process

The basic steps of the sample design process are the following. Each of these is discussed in the text which follows:

1. NC team complete the Sample Design Summary (SDS);
2. Sampling expert gives feedback;
3. SDS (possibly revised) is reviewed by the whole panel;

4. Design is "signed off" by the panel;
5. Sample is selected and fieldwork proceeds;
6. SDS is reviewed and amended to reflect any changes since sign-off;
7. Sample Design Data File is deposited

*Step 1: First draft of SDS*. This should be done as early as possible in the process, in order to allow enough time for discussions and possible revisions. Ideally, this should be at least one month before the sample selection process needs to begin.

*Step 2: Sampling expert feedback*. The expert may make suggestions for changes, either to the proposed design, or to the information about the design recorded in the form. Many suggested designs are uncontroversial, being essentially a repeat of a design that has worked well previously. In such cases, there may be no suggested changes, or only minor ones. Sometimes, the feedback will lead to further discussion between the NC team (and perhaps the survey agency) and the expert. The feedback/ discussion, if any, will result in a revised version of the SDS.

*Step 3: Review of SDS by SWEP*. When the expert is satisfied with the proposed design, he or she will circulate it to the other SWEP members for comment. It is possible that this will result in further feedback or queries for the NC team. Alternatively, the panel might agree to "sign-off" the design without further comment. A decision to either sign-off the design or raise further queries will be made within one week of the design being circulated to the full SWEP.

*Step 4: Sign-off*. The expert will inform the NC team that the design has been signed off. This constitutes authorisation to proceed and draw the sample. The expert will also inform the country contact and the CST and will archive the signed off version of the SDS in the SWEP document store.

*Step 5: Sample selection*. Sample selection should follow exactly the procedures and parameters documented in the signed-off SDS. If there is a reason to change any aspect of the design subsequent to sign-off, the sampling expert should be informed immediately.

*Step 6: Final version of SDS*. If the implemented design differs in any way to that documented in the signed-off SDS, the expert should ensure that the SDS is revised to reflect these changes. The expert will then archive the final version of the SDS in the SWEP document store.

*Step 7*: *SDDF deposited*. Almost all of the information to be provided in the SDDF should be captured as a by-product of the sample selection process at the time the sample is selected. This includes variables indicating sampling strata and primary sampling unit (PSU). It is strongly recommended that a draft SDDF is created at the time of sample selection. There are just two variables that can only be added once fieldwork is completed: an indicator ('OUTCOME') of the survey outcome (response, non-response, ineligible) and – in the case of address-based samples – the within-household selection probability ('PROBx').

Please note the importance of step 7. The Sample Design Data File (SDDF) is a key deliverable. Its contents document the sample design and enable the production of design weights, an essential prerequisite for data release. See the ESS Round 9 Specifications and section 5 of the Sample Design Summary (included as an Annex to this document). If you have doubts about the data to provide in the file, please clarify this with your allocated sampling expert.

### 1.3 The Sample Design Summary

A key role in the sign-off process is played by this form (previously known as the ‘sign-off form’). This documents all relevant aspects of the sample design. The Sample Design Summary (SDS) is included as an Annex to this document, showing the information to be entered, with explanatory notes. If in doubt about any of the information required, please ask your assigned sampling expert for advice.

Note that in some countries a different sample design is used in each of two different parts of the country. For example, an unclustered (single-stage) sample may be used in urban areas and a clustered (multi-stage) design in rural areas. We refer to this as a multi-domain design. For multi-domain designs, section 3 of the SDS (‘sample design details’) must be repeated for each domain.

## 2. Principles for Sampling in the ESS

To ensure that ESS samples adequately represent each national population, and provide comparability between countries, the main principles are:

- The use of a sampling frame/method that provides the best possible coverage of the ESS target population;
- The use of probability sampling;
- The use of a design that provides a prescribed level of statistical precision.

Following these principles does not mean that the sample design should be the same in each country. In fact, we follow the approach of Kish[1]:

> *Sample designs may be chosen flexibly and there is no need for similarity of sample designs. Flexibility of choice is particularly advisable for multinational comparisons, because the sampling resources differ greatly between countries. All this flexibility assumes probability selection methods: known probabilities of selection for all population elements.*

Our view is that the ESS should strive to use the best possible random sampling practice in each participating country. The choice of a specific sample design depends on available sampling frames, and population characteristics that influence the costs and practicality of different sample designs in different countries (such as population density and geographic dispersal). The ESS sample designs should enable comparative analysis with useful and estimable precision.

### 2.1 Population Coverage

The target population of the ESS in round 9 is defined as:

> *All persons aged 15 and over (no upper age limit) resident within private households in each country, regardless of their nationality, citizenship or language.*

As a working definition of a private household, it is recommended to follow the definition:

> *One person living alone or a group of people living in the same dwelling unit with its own lockable front door. A dwelling unit is a self-contained place to live that does not require basic facilities such as cooking, washing or toilet facilities to be shared with the occupants of other dwelling units.*

Living in a dwelling unit means that this accommodation is currently the person's main residence.

---

[1] Kish, L. (1994), "Multipopulation survey designs: five types with seven shared aspects," *International Statistical Review* 62, 167-186.

This **includes**: People who are temporarily away for less than 6 months (e.g. on holiday, away working or in hospital); school-age children at boarding school; students sharing private accommodation.

It **excludes**: People who have been, or will be, away for 6 months or more, students away at university or college; temporary visitors (staying for less than 6 months) and people living in institutions.

The definition of being 15 year or older may vary depending on the sample design:

- For designs where persons are sampled directly from a register (given the day of birth is available) a person is treated as 15 or older if she or he is 15 on the 1st of September of the survey year (i.e. 2018, for ESS9).
- For designs where the interviewer has to determine the age of eligible persons in the household a person is treated as 15 or older if she or he is 15 on the day the interviewer does the listing of household members.

## 2.2 Probability Sampling

The sample is to be selected by strict random probability methods at all stages. This means that every member of the ESS target population in a country should have a larger than zero probability of being selected into the sample and that this probability should be known for each person actually selected. The probability of selection for each sampled unit at each stage of the sample design must be recorded, and supplied in the *sample design data file* (SDDF; see the ESS Round 9 Specifications and section 5 of the Sample Design Summary, in the Annex to this document).

Quota sampling is not permitted in any part of the sampling procedure, nor is substitution of non-responding, non-contactable or non-accessible sampling units, be it households, individuals, or even whole apartment buildings. For instance if the selected respondent in a household refuses to participate and another family member volunteers to do the interview instead, this is considered 'substitution'. This is not permitted in the ESS under any circumstance. (Further details about contacting sample members can be found in the document “Interviewer Briefings” on the NC Intranet.)

The use of *random route techniques* is strongly discouraged. The reasons for this are, a) it is rarely possible to implement such techniques in a way that gives all dwellings even approximately equal selection probabilities; b) it is not possible to accurately estimate these probabilities and therefore to obtain unbiased estimates; and c) the method is easily manipulated by interviewers to their advantage, and in ways that are hard to detect. Instead, as a last resort if no better method is possible, we permit the use of *area sampling with field enumeration*. How to do this in a way that is consistent with ESS principles is set out in section 2.1.3.

## 2.3 Statistical Precision

The ESS aims to achieve the same minimum level of precision in each country, as this in turn guarantees a minimum level of precision for comparisons of countries. In practice, the statistical precision of any survey estimate is determined by several factors. Key ones are:

1) Sample size;
2) Distribution of selection probabilities, and their association with the survey variable(s) upon which the estimate is based (see section 2.4);
3) Sample clustering, and the association of the clusters with the survey variables (see section 2.2);
4) Sample stratification, and the association of the strata with the survey variables (see section 2.3);
5) Population variance of the survey variables.

Once a survey is completed, precision can be estimated empirically, provided that indicators of selection probabilities, clusters and strata are available (and weighting variables, if weighting is applied). Precision can, and does, vary between different estimates based on the same sample.

But at the sample design stage, precision must be predicted based on some assumptions, and in a way that provides a single standardised prediction for a sample design (not a separate one for each possible estimate). The ESS uses some simple heuristics to do this. Specifically:

- We are concerned only with precision *relative* to simple random sampling, not with *absolute* precision. This means we do not need to take factor (5) into account;
- We assume only a negligible association between sample strata and survey variables. This means we do not need to take factor (4) into account. In practice, any association is usually modest and has the effect of slightly improving precision. Thus, ignoring this at the sample design stage has the effect of reducing the risk that a sample design will in practice fail to meet the ESS specification of precision;
- We assume no association between selection probabilities and survey variables. This simplifying assumption makes it easier to take factor (2) into account. In practice, associations may improve precision for some estimates and worsen precision for others, so an assumption of no association can be thought of as a kind of 'average' effect.

This leaves us having to take account only of factors (1), (2) and (3). We do this by specifying a minimum *effective sample size* ($n_{eff}$) that should be achieved. This is the size of a simple random sample (i.e. one in which factors (2), (3) and (4) do not apply) that would provide the same precision as the actual design under consideration. This is estimated by adjusting the predicted net sample size (number of interviews achieved, $n$) by the predicted *design effect* (*deff*), a measure of the impact of factors (2) and (3). These factors will always tend to reduce precision, reflected in a value of *deff* greater than 1. Consequently, the greater the variability in selection probabilities, and the larger the cluster sample sizes, the larger the sample size that will be needed to deliver the required effective sample size:

$$n_{eff} = n / deff \qquad (1)$$

The ESS requirement is for a predicted effective sample size of at least 1,500 in each country (with the exception of small countries with a population of fewer than 2 million people aged 15 or over, where the minimum effective sample size is 800). For countries that are able to select an unclustered equal-probability sample of persons from a population register, this translates to a minimum of 1,500 interviews. But for the majority of countries, the minimum number of interviews is considerably larger, due to the effects of clustering and selection probabilities. Calculating the required number of interviews involves predicting *deff* using standard ESS methods that are explained below.

We predict separately the effect of variable selection probabilities ($deff_p$) and the effect of sample clustering ($deff_c$) and then use the product of these two predicted values as our prediction of deff:

$$deff = deff_p \times deff_c$$

The estimation of $deff_p$ requires a prediction of the distribution of overall selection probabilities for sample persons:

$$deff_p = \frac{n\sum_{i=1}^{n} w_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2}$$

where $w_i$ is the weight associated with sample member $i$;

$w_i = {}^{1}/_{p_i}$ , where $p_i$ is the probability of selection of sample member $i$.

Some examples of the estimation of $deff_p$ are presented in section 3.4 below.

The design effect due to clustering - $deff_c$ - is predicted as follows:

$$deff_c = 1 + (\bar{b} - 1)\rho$$

where $\bar{b}$ is the mean number of interviews carried out per primary sampling unit (PSU), and $\rho$ is a measure of the relative homogeneity of a survey measure within the PSU (note that this can be caused both by the relative similarity of people living in the same area and by 'interviewer effects' if the interviews within a PSU tend to be carried out by the same interviewer). Therefore, at the sample design stage it is necessary to predict both $\bar{b}$ and $\rho$. Discussion of how this is done is presented in section 3.5 below.

## 3. Tips for Good Sample Design

### 3.1 Sampling Frames

The quality of the sample will be higher, the more completely the sampling frame covers the target population. The choice of sampling frame will also constrain the extent to which it is possible to control variation in selection probabilities, and hence the likely value of $deff_p$, which will influence the number of interviews that is required. Thus, the choice of sampling frame is of great importance.

The following types of sampling frames are used on the ESS:

1. Lists of residents (population registers). This is generally the preferred type of frame. The use of population registers is discussed further in section 3.1.1 below.
2. Lists of buildings or addresses. This is generally the best option if population registers are not available/accessible, and is discussed in section 3.1.2.
3. Area sampling with field enumeration to create a frame of dwellings. This method can be used when no suitable list of persons, addresses or households exists. How to do this is discussed in section 3.1.3 below.

#### 3.1.1 Population Registers

In recent ESS rounds, around half the participating countries have used population registers as the sampling frame including, for example, Norway, Finland, Slovenia and Belgium. This is generally the preferred type of frame. Some countries have been successful in negotiating for ESS samples to be selected from the national population register, where this was not initially possible. Even if your population register is not usually used for survey sampling, or not outside of the National Statistical Institute, it may be worthwhile exploring under which terms this might be possible.

The main reasons for preferring population registers as a sampling frame are that coverage is good[2] and equal-probability samples can be implemented, leading to $deff_p = 1.0$, which minimises the number of interviews needed in order to meet the effective sample size requirement.

A further advantage of population registers is that individual-level auxiliary variables including age and gender are typically available for use in stratification. This tends to increase sample precision to a greater extent than stratification only by higher-level auxiliary variables such as small area characteristics. Stratification by frame variables such as age, gender and region is strongly encouraged.

With population registers one should be aware that some persons on the frame may not belong to the target population (so-called over-coverage). This would include those who do not live in private

[2] Though this is usually the case, it is not universally true. Some population registers may be updated infrequently and can therefore suffer from being out-of-date.

households (for example, students in college dorms, elderly people in nursing homes, military personal in barracks) and those who are not currently resident in the country though still registered (e.g. working or studying abroad for a period exceeding six months). There may also be under-coverage, for example of illegal immigrants or recent immigrants.

The quality of a population register as a sampling frame might also be undermined be the presence of *opt-outs*. These are persons who are rightly part of the target population but must not be contacted for survey research. This can occur due to legal reasons or because persons can make a request not to be contacted for research or marketing purposes. These opt-outs should stay in the sampling frame and be treated as refusals if sampled.

### 3.1.2 Address Lists

Common lists of this kind are lists of addresses held by the postal delivery service and lists of dwellings held for land registration or taxation purposes. This type of frame has been used in recent rounds by United Kingdom, Portugal and Ireland. Such lists tend to have the advantage of good population coverage. However, a disadvantage is that it is not usually possible to select equal-probability samples of persons, so design effects tend to be higher than with population register samples, leading to the need to carry out a larger number of interviews.

Generally (unless an indicator of the likely number of residents is available), the most efficient design possible with an address list is to select an equal-probability sample of addresses. At each address it will then be necessary for the interviewer in the field to implement a procedure to randomly select one person to interview at the address. This within-address selection is what causes a loss of statistical precision ($deff_p$ > 1.0) as the overall selection probabilities of persons will be inversely-proportional to the number of persons aged 15 or over residing at the address. Selection of one person per address/household tends to lead to values of $deff_p$ in the range 1.2 to 1.3. The value can be well predicted from knowledge of the distribution of household size.

There are two common and acceptable types of procedures for randomly selecting one person at a sample address: *Kish grid methods*, and *birthday methods*.

Kish grid methods (Kish, 1948) are based upon the idea of listing eligible persons in a predetermined order (for example, ascending order of age, or alphabetical order of given name) and then using a random number to identify which person on the list should be selected. The random numbers are generated in advance for each sample address, so that the interviewer has no control over which number should be used. The predetermined rule for the order in which persons should be listed is important as this is what permits the correct application of the method by interviewers to be checked. The rule should not contain ambiguities. Alphabetical order is more ambiguous than age, for example, as people can be known by more than one version of their name (Johannes or Hans; William or Bill).

Variants of the birthday method include next-birthday, last-birthday, and nearest-birthday (which can be either in the past or in the future). The selection rule is therefore determined by the relationship between the date on which the interviewer is making the selection and the dates on which household members have their birthdays. These methods have the advantage of being less intrusive than the Kish grid method: the interviewer does not need to ask for the names or ages of residents. For this reason there is a tendency to obtain lower refusal rates with birthday selection methods. On the other hand, the methods are more error-prone. It can sometimes be difficult to work out which birthday is nearest (which is why next- or last-birthday methods are usually preferred to nearest-birthday). And the household member answering the selection questions may deliberately nominate someone who they know not to be the correct person. This may even be done in collusion with the interviewer. Interviewers are typically not asked to check the selection (as that would undermine the simple and non-intrusive nature of the procedure), but data collected later, in the interview, can often be used to identify whether the correct selection was made.

As Kish grid methods and birthday methods both have their own advantages, the preferred method differs between countries, between survey organisations and between researchers. We prefer Kish grid methods, but either type of method is acceptable for the ESS, provided it is well implemented, with some form of quality control.

### 3.1.3 Area Sampling with Field Enumeration

Area sampling designs involve at least three stages of sample selection: small areas, addresses/dwellings, and persons. The first stage in such a design is to select a probability sample of small areas such as villages, grid squares, streets or city blocks. The frame of areas may come from an existing list (administrative areas, census enumeration areas, postal areas, street directory) or may be created for the purpose of sample selection (e.g. identifying and listing areas on maps). If it is not possible to access or create a list of such areas, it may be necessary to resort instead to random route methods. Please consult your sampling expert if this is to be considered.

An efficient multi-stage design (see section 3.2 below) involves selecting the first-stage units with probability proportional to the number of households or – preferably – persons that they contain. Thus, if possible the frame of small areas should include some indicator of the number of households or persons in the area, to be used for this purpose. Use of an approximate indicator is still likely to be more efficient than selecting areas with equal probabilities (for example, if the areas are Census enumeration areas and the only size measure available is from a Census carried out several years ago).

At the second stage, an enumerator must make a complete listing of the dwellings in the area, from observation. To meet the ESS requirements for complete coverage and known probabilities of selection, it is important that the boundaries of each area are clearly defined and that the enumerator is able to identify those boundaries on the ground. The list is then returned to the central field office, where a random selection of dwellings is made to constitute the survey sample. The list of sampled dwellings is then passed to an interviewer. The enumerator and interviewer should not be the same person. This is important, to ensure that interviewer subjectivity cannot influence the sample selection (for example, if the interviewer excludes from the list certain dwellings that he or she would not like to have to visit). For the same reason, it is even more important that the selection of dwellings should be made in the central field office rather than by the enumerator or interviewer.

Once the selection of dwellings has been made, the third selection stage is to select one person to interview at each dwelling. This step is the same as described in section 3.1.2 above.

| **Key Tips on Sampling Frames: Summary** |
| --- |
| Population registers are preferred. It may be possible to negotiate their use for the ESS; |
| Lists of addresses are a suitable alternative, provided they have comprehensive coverage. However, they have some disadvantages; |
| If neither population registers nor address lists are available, area sampling with field enumeration of dwellings can be used. |

## 3.2 Multi-stage Sampling

Multi-stage sample designs are used either to increase the cost-efficiency of the design (as they result in a sample which is *clustered*, usually within relatively small geographical areas, such that each sample cluster forms the workload for one interviewer) or because the constraints on available sampling frames leave no choice (for example, as in the case of area sampling with field enumeration). Examples include:

2-stage. First stage small geographical areas; second stage persons (population register)

3-stage. First stage small geographical areas; second stage addresses; third stage persons (address list or area sampling)

4-stage. First stage small geographical areas; second stage addresses; third stage households; fourth stage persons (address list or area sampling)

4-stage. First stage large geographical areas; second stage small geographical areas; third stage addresses; fourth stage persons (address list or area sampling)

Key aspects of multi-stage designs are the following:

- The overall selection probability of each person is the product of the conditional selection probabilities at each stage of the sample design. Careful control of the relationship between these probabilities is therefore important;
- The predicted design effect due to clustering ($deff_c$) depends on two features of the sample design: the relative homogeneity of the first-stage units (*primary sampling units*, *PSUs*), and the number of interviews conducted in each.

With respect to the control of probabilities, an efficient design is one in which PSUs are selected with probability proportional to the number of addresses/persons in the PSU, and subsequently the same number of addresses/persons is selected in each sampled PSUs. If there are practical reasons for wanting to vary the sample size of addresses/persons per PSU between two or more strata (for example, a smaller sample size per PSU in rural areas than in urban areas), then the first-stage probabilities should be modified to compensate (for example, a larger probability for PSUs in rural areas).

Regarding $deff_c$, the following points should be noted:

Relatively heterogeneous PSUs are desirable (greater precision and therefore fewer interviews required). Typically, larger geographical areas are more heterogeneous than smaller areas. Thus, if possible, larger rather than smaller areas should be used. Even an increase from a mean PSU size of, say, 1,000 households to 2,000 households is likely to be worthwhile, so it is worth considering whether smaller units could be combined to create larger units prior to sample selection;

Smaller *sample* size per PSU is desirable. Thus, to the extent possible, the number of sampled PSUs should be maximised and the number of sampled persons per PSU minimised.

**<u>Key Tips on Multistage Sampling: Summary</u>**

Larger areas are preferred to smaller areas as PSUs;

A larger number of sampled PSUs is preferred to a smaller number;

Large variability in the size of PSUs (within strata) is undesirable;

If possible, PSUs should be sampled with probability proportional to the number of addresses/persons in the PSU, and a fixed number of addresses/persons then selected in each PSU.

### 3.3 Stratification

Proportionate stratified sampling can improve the precision of sample estimates. If, for example, strata are regions, this ensures that the sample distribution by region matches the population distribution: there is no random sampling variation in respect of region.

Stratification can be either explicit or implicit. With *explicit stratification*, the units on the sampling frame are sorted into distinct strata and a sample is selected independently from each stratum. With *implicit stratification* the units on the sampling frame are sorted (ranked) in a meaningful order and a systematic sample (every $n^{th}$) is then selected from the sorted list. Either method is effective at improving precision. More important is the choice of variables to define the stratification.

Stratification is more effective the more strongly associated the stratification variables are with the survey variables (i.e. the ESS questionnaire variables). Individual-level variables such as age and gender (available only with population register frames) therefore tend to be more beneficial than regions or characteristics of small areas such as population density or local unemployment rate.

When sampling from address lists or using area sampling, the PSUs can usually be selected by stratified sampling, where strata are defined by geography (e.g. regions) or by geographically-linked data (e.g. summary Census data for each PSU). This may require a prior step of linking geographically-referenced data to the frame of PSUs using geographical identifiers.

With multi-stage sampling of addresses, the addresses can usually be selected by implicit stratified sampling, where the sorting is done by some indicator of geographical location, such as street name, or the order in which the enumerator recorded the dwellings (area sampling) or indicators such as postal code or geo-location (address lists). Such ordering ensures the sample of addresses within each PSU is spread throughout the entire PSU, which tends to be beneficial.

**Key Tips on Stratification: Summary**

Proportionate stratification is beneficial and is preferred to simple random sampling;

Stratification can be either explicit or implicit;

Choose stratification variables that are correlated with the survey variables;

With multi-stage sampling, PSUs can usually be stratified by geography or by geographically-defined variables;

Addresses within PSUs should be selected by systematic random sampling with implicit stratification, in preference to simple random sampling.

### 3.4 Predicting $deff_p$

To predict $deff_p$ requires a prediction of the distribution of (relative) overall probabilities of selection for survey respondents. Most ESS sample designs are of one of the following two types, for which this prediction is simple:

- Equal-probability sample of persons; $deff_p = 1$ ;
- Equal-probability sample of addresses; one person selected at each address; selection probabilities are inversely proportional to the number of persons aged 15 or over at the address; $deff_p$ depends only on the national distribution of household size. See example 1.

Other possible sources of variation in selection probabilities include:

- Oversampling regions or subpopulations of particular analytical interest. Examples include non-Jewish areas in Israel and (until round 8) East Germany. If, for example, 20% of the population lives in Region A and are given twice the selection probability of persons in Region B, then persons in Region A will have a relative weight of 0.5. And persons in Region A will constitute one-third of the sample (because (20%*2P)/((20%*2P)+(80%*P)) = 1/3, where P is the probability of selection in Region B). Thus:

$$deff_p = \frac{n\sum_{i=1}^{n} w_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2}$$

$$= n\frac{(0.333n\times0.5^2)+(0.667n\times1^2)}{\left((0.333n\times0.5)+(0.667n\times1)\right)^2} \qquad = \frac{0.750}{0.833^2} \qquad = 1.08$$

- Selecting PSUs with probability proportional to an approximate size measure, which does not correspond perfectly with the current size measure, which is identified only at the next sampling stage. The design effect will be modest if the two size measures are highly correlated (perhaps in the range 1.01 to 1.05), but will be larger the lower the correlation. An example is the design in Portugal at round 8, where census districts were selected with probability proportional to the number of households registered in the 2011 Census, but addresses were subsequently selected (in 2017) from the electricity company's list of all households in the district.

If two or more different sources of variation in selection probabilities can be assumed to be independent, we can estimate their effects separately and then take the product as our prediction of $deff_p$. For example, suppose the over-sampling of a region, as described above, is combined with a 3-stage design as in example 1. If we had no reason to suppose that the household size distribution differed substantially between the two regions, then:

$$deff_p = 1.08 \times 1.21 \quad = 1.31.$$

**Example 1: Predicting $deff_p$**

| Sampling stage | Design | | Probability |
|---|---|---|---|
| 1 | Select 200 PSUs with probability proportional to number of addresses in the PSU | $PROB1 =$ | $\frac{200 \times N_k}{N}$ |
| 2 | Equal-probability selection of 12 addresses from each PSU | $PROB2 =$ | $\frac{12}{N_k}$ |
| 3 | Equal-probability selection of 1 person from each address | $PROB3 =$ | $\frac{1}{N_{jk}}$ |
| Overall | | $PROB1 \times PROB2 \times PROB3 =$ | $\frac{2400}{N \times N_{jk}}$ |

where $N_{jk}$ is the number of persons aged 15 or over resident at address *j* in PSU *k* , and $N_k$ is the total number of addresses in PSU *k*. Note that $N = \sum_{k=1}^{K} N_k$,where *K* is the total number of PSUs in the population.

With this design, the design weight for person *i* (the person at address *j* in PSU *k*) will be $w_{jk} = \frac{N \times N_{jk}}{2400}$. Suppose the distribution of household size in the country (perhaps estimated from Census data, or from a previous survey) is as follows:

| Number of persons aged 15 or over in the household, $N_{jk}$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Percentage of households in the country | 25 | 52 | 15 | 5 | 2 | 1 |

With this information, we can predict $deff_p$ :

$$deff_p = \frac{n\sum_{i=1}^{n} w_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2} \qquad = \frac{n\sum_{i=1}^{n}\left(\frac{N \times N_{jk}}{2400}\right)^2}{\left(\sum_{i=1}^{n}\frac{N \times N_{jk}}{2400}\right)^2} \qquad = \frac{n\sum_{i=1}^{n}(N_{jk})^2}{\left(\sum_{i=1}^{n} N_{jk}\right)^2}$$

$$= n\frac{(0.25n\times1)+(0.52n\times4)+(0.15n\times9)+(0.05n\times16)+(0.02n\times25)+(0.01n\times36)}{\left((0.25n\times1)+(0.52n\times2)+(0.15n\times3)+(0.05n\times4)+(0.02n\times5)+(0.01n\times6)\right)^2}$$

$$= \frac{5.34}{2.1^2} \qquad = 1.21$$

This is a typical value for the design effect due to selecting one person per household/address. In countries with higher proportions of larger households, the design effect will be larger, but usually in the range 1.2 to 1.3.

### 3.5 Predicting ***deff$_c$***

For single-stage, unclustered, samples, $deff_c = 1$. However, for multi-stage (clustered) samples, it is necessary to predict the design effect due to clustering. To do this, we need predictions of both the mean number of interviews per PSU, $\bar{b}$, and the relative homogeneity of persons living within the PSU, $\rho$. The predicted value of $\bar{b}$ is simply the ratio of the total number of achieved interviews to the number of sample PSUs. However, the required number of achieved interviews is determined by the prediction of *deff*, so the problem is circular and must be solved iteratively.

The intra-cluster correlation coefficient, $\rho$, will in practice vary between survey variables and estimates. However, to determine the required sample size only one value can be used. The SWEP will, after each round of data collection, estimate values of $\rho$ for a standard set of over 100 items (means and proportions) from the core questionnaire. The mean value across the items within a country will be published in the ESS Quality Matrix. If the same, or similar, geographical units are to be used as PSUs in a subsequent round, then this empirical mean from previous rounds should be used as the prediction of $\rho$ for the current round. For a country that has not taken part in ESS previously, or which has not used a clustered design before, the best prediction may be the ESS estimate from a 'similar' country.

The following table shows how predictions of $deff_c$ depend on the predicted values of $\bar{b}$ and $\rho$:

| $deff_c$ | $\bar{b} = 4$ | $\bar{b} = 8$ | $\bar{b} = 12$ | $\bar{b} = 16$ | $\bar{b} = 20$ |
|---|---|---|---|---|---|
| $\rho = 0.02$ | 1.06 | 1.14 | 1.22 | 1.3 | 1.38 |
| $\rho = 0.04$ | 1.12 | 1.28 | 1.44 | 1.6 | 1.76 |
| $\rho = 0.06$ | 1.18 | 1.42 | 1.66 | 1.9 | 2.14 |
| $\rho = 0.08$ | 1.24 | 1.56 | 1.88 | 2.2 | 2.52 |

It can be seen that the design effect increases quite rapidly as both $\bar{b}$ and $\rho$ increase. For most ESS countries, $\rho$ is in the range 0.04 to 0.08, so *deff$_c$* can become considerable if $\bar{b}$ exceeds 10.

**Example 2: Effect of *$deff_c$* on Required Number of Interviews**

Suppose that we estimate $\rho = 0.06$ for the proposed PSUs and that the proposed sample design has $\bar{b} = 16$. Then, the predicted value of $deff_c$ (from the table in section 3.5 above) is 1.90. If we further suppose that a separate calculation, similar to that in example 1, has produced a prediction of $deff_p$ = 1.22, then we can now estimate the overall *deff*:

$$deff = deff_p \times deff_c \qquad = 1.22 \times 1.90 \qquad = 2.32$$

We can now estimate the required minimum number of interviews:

$$n = 1{,}500 \times deff \qquad = 3{,}480$$

But this number could be reduced if we change the sample design to have smaller sample sizes per cluster. Reducing $\bar{b}$ to 8 would reduce $deff_c$ to 1.42 (from the table in section 3.5 above) and hence the minimum number of interviews would reduce to 2,599. Achieving this reduction in $\bar{b}$ would involve increasing the number of sample PSUs from 217 to 325. These two designs provide equivalent precision, as does an intermediate design with 3,038 interviews and 253 PSUs. The choice between these designs – set out in the table below – and others of equivalent precision should depend on the associated field costs. The preferred sample design should be the one that maximises precision for a fixed budget or minimises the budget required to deliver a fixed precision.

Sample Designs of Equivalent Precision

| Design | Completed interviews | Sample PSUs | $\bar{b}$ |
|---|---|---|---|
| 1 | 3,477 | 217 | 16 |
| 2 | 3,038 | 253 | 12 |
| 3 | 2,599 | 325 | 8 |

## 4. Calculating the Required Sample Size

The steps in calculating the minimum required gross (initial) sample size are:

i. Predict $deff_p$ (section 3.4);
ii. Predict $deff_c$ (section 3.5) and hence $deff = deff_p \times deff_c$;
iii. Calculate the minimum required number of interviews, $m$: $m = n_{eff} \times deff$
iv. Predict the response rate, *rr*, and the ineligibility rate, *ri*. The predicted response rate should be realistic but should not be lower than the response rate achieved at the previous round. Where possible, methods for improving the response rate should be proposed and agreed with the ESS fieldwork team. In all cases, the predicted response rate should be agreed with the ESS fieldwork team prior to confirming the required gross sample size. The ineligibility rate, *ri*, indicates the proportion of selected sample units (persons or addresses) that are likely to turn out to be ineligible for the survey (for example, persons who have died or moved abroad, or who reside in institutions; or addresses that are vacant, demolished, or non-residential). This can usually be well estimated from other social surveys, perhaps including the previous round of ESS, that have used the same sampling frame.
v. Calculate the minimum required gross (initial) sample size, $n$: $n \geq m/\big(rr \times (1 - ri)\big)$.

This calculation is illustrated in example 3 below.

**Example 3: Sample Size Calculation (Ireland, Round 8)**

A clustered design, with address-based sampling, so $deff_p$ depends on the household size distribution, as in example 1 above.

i. Based on the distribution of household size (persons aged 15 or over) from the latest issue of the Quarterly National Household Survey (http://www.cso.ie/en/qnhs/), $deff_p$ = 1.206.
ii. At ESS Round 7, $\rho$ = 0.10, and proposed design gives 5.45 interviews per PSU, so $deff_c = 1 + (5.45 - 1) \times 0.10 = 1.445$. Thus, $deff = 1.206 \times 1.445 = 1.743$.
iii. $m = 1{,}500 \times 1.743 = 2{,}614$.
iv. Response rate is predicted to be 60%, similar to ESS7. 9.25% of sampled addresses are expected to be vacant, based on 2016 Census returns.
v. $n \geq 2{,}614/(0.60 \times 0.9075) = 4{,}800$

Note: 480 PSUs to be selected, hence gross sample of 10 addresses per PSU and net sample of $10 \times 0.9075 \times 0.60 = 5.45$ interviews per PSU, as in step ii above.

# Sample Design Summary: ESS Round 9

| | |
|---|---|
| **Country:** | **<country>** (<abbreviation>) |
| **NC:** | <name> (<email address>) |
| **Other Experts:** | <name> (<email address>) |
| **Survey Institute:** | <institute name> |
| **Sampling Expert:** | <name> (<email address>) |
| **Country Contact:** | <name> (<email address>) |
| **Reference Survey:** | <> |
| **Date:** | <date> |
| **Status:** | ☐ Pre sign-off<br>☐ Signed off<br>☐ Post sign-off amendment<br>☐ Final (post-field work) |

## 1.1 Target Population

| | |
|---|---|
| **Number of residents aged 15 or older in the country:** | <number> |
| **Source and reference date:** | <details> |

## 1.2 Population Coverage

*<Describe here any population subgroups not covered by the sample design. Include an estimate of the proportion of the total population that each subgroup accounts for>*

## 2. Summary of the Sample Design

*<Provide an overview of the sample design in one or two paragraphs. Outline the sampling frame, the source of any other data used in the design, the stratification to be used, and the clustering to be used (number and nature of primary sampling units), if any.>*

*For multi-domain designs, repeat this section for each domain*

# 3. Sample Design Details

## First Sampling Stage

**unit:** *<State the units to be selected, e.g. municipalities, electoral divisions, postal sectors, addresses, persons, etc>*

**frame:** *<Describe the sampling frame of these first-stage units>*

**size:** *<Number of units to be selected>*

**strata:** *<Describe how the units are stratified prior to selection. If the stratification is explicit, state how many strata there are and how they are defined>*

**allocation:** *<Describe how the number of units to select from each stratum is determined (if applicable)>*

**algorithm:** *<Describe how it is determined which units to select (in each stratum). For example, simple random sampling, systematic sampling; with equal probabilities or with probability proportional to size; etc>*

## Second Sampling Stage

**unit:** *<State the units to be selected, e.g. addresses, persons, etc>*

**frame:** *<Describe the sampling frame of these second-stage units>*

**size:** *<Number of units to be selected within each sampled first-stage unit>*

**strata:** *<Describe how the units are stratified prior to selection. If the stratification is explicit, state how many strata there are and how they are defined>*

**allocation:** *<Describe how the number of second-stage units to select from each first-stage unit is determined>*

**algorithm:** *<Describe how it is determined which second-stage units to select within each first-stage unit. For example, simple random sampling, systematic sampling; with equal probabilities or with probability proportional to size; etc>*

## Remarks

*<An optional space to provide any further comments or explanations about the sample design>*

# 4. Planning the Sample Size

## History of Planned and Realised Values

*<This section will be pre-filled by the SWEP with predicted and actual values from previous rounds for* $\rho$, $\bar{b}$, $deff_p$, $deff_c$, $deff$, *response rate,* $n_{gross}$, $n_{net}$ *and* $n_{eff}$*>*

## Parameters of the Planned Gross Sample Size

| Achieved interviews per cluster ($\bar{b}$) | Intraclass Correlation Coefficient ($\rho$) | Design Effect due to Selection Probabilities ($\text{Deff}_p$) | Response Rate (rr) | Ineligible Rate (ri) | Effective Sample Size ($n_{eff}$) |
|---|---|---|---|---|---|
| <> | <> | <> | <> | <> | <> |

## Design Effect

$$\text{Deff}_c = 1 + (\bar{b} - 1) \times \rho$$

$$= 1 + (<> -1) \times <>$$

$$= <>^{*}$$

------------- --- ---------------------------------------

$$\text{Deff}_p = <>^{*}$$

------------- --- ---------------------------------------

$$\text{Deff} = \text{Deff}_p \times \text{Deff}_c$$

$$= <>^{*}$$

**results have been rounded to 3 d.p.s*

## Gross Sample Size

$$\text{Min. } n_{net} = \text{Deff} \cdot n_{eff}$$

$$= <> \times <>$$

$$= <>^{**}$$

------------- --- ---------------------------------------

$$\text{Target } n_{net} = <>^{**}$$

------------- --- ---------------------------------------

$$n_{gross} = \frac{n_{net}}{\text{rr} \times (1 - \text{ri})}$$

$$= \frac{<>}{<> \times (1 - <>)}$$

$$= <>^{**}$$

------------- --- ---------------------------------------

** *results have been rounded to 0 d.p.s.*

## Remarks

*<An optional space to provide any further comments or explanations about parameters of the sample design, including assumptions about ineligibility rates and response rates>*

# 5. Sampling Design Data File (SDDF)

## Variables to be included in the SDDF

*Edit this list and include details as appropriate, e.g. of stratification variables, value ranges, etc.*

| Variable | Description |
|---|---|
| idno | Personal identifier |
| prob1 | Probability of selection at first stage of sampling |
| prob2 | Conditional probability of selection at second stage of sampling |
| prob3 | Conditional probability of selection at third stage of sampling |
| stratex1 | Indicator of explicit stratum at first stage of sampling |
| stratim1 | Order of selection of PSU |
| stratim2 | Order of selection of person within PSU |
| strtval1 | Value of the first variable used to stratify PSUs |
| strtval2 | Value of the second variable used to stratify PSUs |
| psu | PSU identifier |
| samppoin | Sampling point identifier |
| outcome | Final outcome |
| frame1 | Information from sampling frame: <> |
| frame2 | Information from sampling frame: <> |
| frame3 | Information from sampling frame: <> |

## Probabilities of Selection

*<Define the values of the PROB variables that will be included in the SDDF. For example, for probability proportional to size selection of municipalities as PSUs, using a population register count as the size measure, we might have* $PROB1_i = n1\frac{N_i}{N}$*, where* $n1$ *is the number of PSUs to be sampled,* $N_i$ *is the population count for the* $i^{th}$ *municipality and* $N$ *is the total population count for all municipalities on the frame; etc.>*

## Remarks

*<Please provide further comments or explanations about variables to be included in the SDDF, for example the coding of strata or the source/definition of additional variables from the sampling frame>*

# Appendix

*<Supplementary material such as tables of PSUs by strata, or population counts by strata>*