Sampling Guidelines: Principles and Implementation for the European Social Survey

The ESS Sampling Expert Panel

17th March 2016

Summary

The document gives an overview on the principles of ESS sampling and provide guidance for the procedure required to approve a sampling design to be used in the ESS.

The Sampling Expert Panel (SEP) is the group of sampling experts tasked by the Core Scientific Team (CST) to evaluate and help implement the sampling design in each of the ESS countries in close cooperation with National Coordinators (NCs). The objective of the SEP is to help to implement the ESS sampling principles as workable sampling designs in all participating countries.

Contents

1. Principles for Sampling in the ESS

1.1 Developing a Sampling Design and Sign-Off Procedure

Since the first ESS round the SEP has worked together with NC teams to implement the ESS sampling specification given the present frameworks for sampling in each country. The panel will continue its work. Its current members are the following sampling specialists:

- Stefan Zins [Chair] (GESIS, Germany)
- Sabine Häder (GESIS, Germany)
- Seppo Laaksonen (University of Helsinki, Finland)
- Peter Lynn (University of Essex, U.K.)
- Siegfried Gabler (GESIS, Germany)

Depending on the number of participating countries in ESS round 8 each of the experts will be assigned to a set of countries to liaise with and provide support when implementing the sampling specification. However, the decision to *sign-off* a design will be made be the entire SEP.

As the main tool to manage the planning of the sampling design, the so-called sampling Sign-off Form (SoF), is used. The SoF documents in detail the sampling design and serves as the point of origin for any other documentation containing information about the sampling design. Once signed off, the SoF describes the sampling design as sanctioned by the SEP on behalf of the CST. All further reports regarding sampling in the participating countries will be based on this document.

The NCs are asked to propose a draft sampling design, which could be the same sampling scheme as used in a previous round. Sampling methods and sampling frames are discussed by the responsible sampling expert and the NC to determine a best practice sampling strategy in a country.

For further planning of the sampling design in a country some key information is needed:

- A benchmark value of the design effect. If a country does not participate for the first time in the ESS and recycles a sampling design of a previews round the SEP will provide the NC with this benchmark. The SEP will calculate such a benchmark value based on the history of estimated design effects of a fix set of variables from the ESS core questionnaire. If this is not possible, e.g. if a country participates in the ESS for the first time or there is a change in the planned sampling design that has a significant effect on design effects, presumed values will be used that the SEP assumes to be reasonable.

- Anticipated response and ineligibles rate. These rates have to be provided by the NC. If applicable, they will be compared with realised response and ineligible rates in previews rounds. The planned response and ineligible rates will also be inspected by ESS fieldwork team.

At this stage, there may be comments or questions of clarification from other panel members

When the responsible panel member regards the sampling design to be ready to be signed off the SoF will be sent to the SEP for evaluation. At this stage, there may be comments or questions of clarification from other panel members. After this review process is concluded, the SEP will provide the NC with a decision regarding the proposed sampling design.

NCs are welcome and encouraged to involve their own sampling experts (and possibly also from the survey organisation involved) in order to clarify details or propose amendments. The knowledge of national sampling experts regarding local practices can be very valuable in the process of designing and implementing ESS sampling procedures in a country. Where necessary, the responsible sampling expert will visit the country to provide local help and support. Consultations by the SEP can continue after the sign-off is completed, but any amendments or alterations to the signed-off sampling design **must** be communicated to the SEP and will be documented in a post-SoF version.

1.2 Basic principles for sampling in cross-national surveys

Kish (1994, p. 173) provides the starting point for the work of the Sampling Expert Panel (SEP):

> *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.*

Following this statement, an optimal sample design for cross-national surveys should consist of the best random sampling practice used in each participating country. The choice of a specific sample design depends on available sampling frames, experience, and of course also the costs of the different sampling designs in different countries. If, after the survey has been conducted, adequate estimators are chosen, the resulting estimates should be comparable. To ensure comparability between the countries the following main principles are set within the ESS regarding sampling:

- The usage of probability samples
- Best possible coverage of the ESS target population
- Similar statistical precision between countries

Only probability samples provides the necessary framework to make inferences from the sample regarding the target population. Every member of the ESS target population in a country should have a larger than zero probability of being included into the sample. To achieve this, the sampling frame(s) should ultimately include all members of the ESS target population. Therefore, it is of great importance to use the best possible administrative sources available to construct the most exhaustive sampling frame for the target population.

Similar statistical precision between countries is sought to ensure that estimates based on ESS data are of similar quality across the different countries. For this purpose sampling designs are required to have the same *minimum effective sample* sizes (n_{eff}) in all participating country ($n_{eff} = 1500$ or $n_{eff} = 800$ for countries with an ESS populations (aged 15+) of less than 2 million). (For a description of the concept of the effective sample please see Appendix B.1)

1.3 Sampling

The sample is to be selected by strict random probability methods. 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 a another family member volunteers to do the interview instead, this is considered 'substitution'. This is not permitted in the ESS under any circumstance.

Random route techniques: starting from a randomly selected unit (or coordinate) fieldwork persons select, using a specific routing (algorithm), a number of sampling units which are then being contacted/visited by the interviewer.

Random route techniques can in principle approximate a strict random probability design, but we do not encourage their use. The reason for this is that it is hardly ever possible to calculate the exact inclusion probabilities of the persons selected. Random route selection should only be used as a last resort. Either if no sampling frame is available at all, or if the frame is so flawed that not knowing the exact inclusion probabilities can be considered a lesser source of errors than the potential coverage error.

- How is the algorithm for random route defined?
- How can the implementation of random walk process be controlled/monitored?
- What experience do interviewers have with random walks?
- How can the inclusion probabilities of sampled units be approximated?

The task of selecting the sampling units, that is, doing the random walk, and contacting target respondents should be done by different persons. For instance, person A selects the addresses by a random walk, records the addresses and transfers them to the survey agency. Person B will use the resulting addresses from person A's random walk to contact, select and do the interview of persons living at the selected addresses.

Stratification: The SEP strongly recommends using stratification of the sample to achieve a desirable distribution of the persons with respected to important socio-demographic variables (e.g. age, sex, and geographic regions).

1.4 Coverage of the Target Population

Definition. *The target population of the ESS in round 8 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.*

Definition. *As a working definition of a private household for the purpose of selecting a respondent from it, 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.*

Living in a dwelling unit means that this accommodation is associated with the centre of one's life.

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

It excludes: People who have been away for 6 months or more, students away at university or college; temporary visitors and people living in institutions.

Definition. *The definition of being 15 year or older may vary depending on the sampling 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 at the 1st of September in the year in which the survey is conducted.*
- *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 at the day the interviewer does the listing of household members.*

The quality of the sample will be higher, the more complete the sampling frame covers the target population. However, the quality of the sampling frames - e.g. regarding coverage, updating intervals and accessibility - may differ. Therefore, frames will be evaluated carefully by the responsible sampling expert together with the NC. The results of these evaluations have to be documented and taken into account when the data is analysed.

The following differences in frames can be expected:

1. Countries with reliable lists of residents that are available for social research, such as Norway, Sweden, Denmark, or Finland.

2. Countries with reliable lists of buildings or addresses, that are available for social research, such as the Netherlands or the U.K.

3. Countries without reliable and/or available lists of either addresses or households, such as Portugal or Bulgaria

Drawing a sample is more complicated if no registers (lists) are available (group 3). In these cases, multi-stage sample designs are usually applied, in which a list for the selection of the so called *primary sampling units* (PSUs) may exist, but maybe not for some of the subsequent sampling stages. When no sampling frame for sampling units (e.g. streets, addresses, or households) is available random route techniques can be used (see Section 1.3)

Frame Imperfections: Even in countries where reliable sampling frames exist, we have to expect pitfalls in the sampling process. In practice, it is often difficult to fully cover the target population in practice. Also one should be aware of persons that do not live in private households, for example students in college dorms, elderly people in retirement homes, military personal in barracks. These persons are part of the sampling frame but do not belong to the target population (so-called over-coverage). The quality of a sampling frame might also be undermined be the presence of *opt-outs*. These are persons that can be part of the target population but must not be contacted for survey as the ESS. Such exceptions might be due to legal reasons or if persons can make a request by their authorities to not be contacted by survey agencies. In any case, these opt-outs should stay in the sampling frame and be treated as refusals if sampled.

1.5 Sample Size

For determining the required net and gross sample sizes (n_{net} and n_{gross}, respectively), *design effects* have to be considered to ensure the comparability of estimates. The design effect is a measure for the relative efficiency of an estimator under a studied sampling design. It can tell you how your estimator in combination with your sampling design compares, in terms of accuracy, to same type of estimator under a simple random sample. Its formal definition is the ratio of the variance of an estimator under the studied sample design to the variance of the same estimator computed under the assumption of simple random sampling. The problem is that design effects do not only vary from survey to survey because of different designs but also within one survey from item to item. *In general,* *for a well-designed study, the design effect usually ranges from 1 to 3* (see Shackman, 2001). The driving factor of the design effect is the selection of clusters, instead of directly sampling individual units. For instance, at first municipalities are selected as PSUs, then persons are selected from within the sampled municipalities. The selection of municipalities represents a clustering of the selected persons. Persons within the same municipalities are assumed to be more similar, regarding the surveyed characteristics, as persons across municipalities. A measure for this homogeneity of survey data within clusters is the so-called *intraclass-correlation coefficient*(ρ).

Benchmark values for the design effect of the planned sample designs of participating countries will by provided by the SEP based on the data from previous ESS rounds. If a country participates for the first time in the ESS, the SEP will work in close collaboration with the NC and survey agency to come up with a presumptive value. Also the size of the PSUs influences the design effect. Given the same net sample size n_{net}, a sample design with 15 respondents per PSU will show a larger design effect than a sample design with only 10 respondents per PSU. Hence, the number of respondents per PSU should be as small as possible. Put the other way around: given a certain net sample size, the number of PSUs should be as large as possible.

> **PSU sizes:** The smaller the sample size per PSU, the smaller the design effect and hence the less interviews have to be conducted to reach the required effective sample size of $n_{eff} = 1\ 500$. In that sense, a large number of PSUs with only a few interviews conducted in each should be the goal.

It should be noted that homogeneity within clusters might not only be caused by the similarity of the elements in the population clusters but also by interviewer effects, for example one interviewer conducts all in one cluster.

Another important source which has an effect on the design effect is any departure from equal probability sampling designs, were every person has the same probability of being included into the sample. Samples selected under such designs require weighting of observations to account for different inclusion probabilities. In particular, in countries where the only frames available consist of households/addresses, design effects will be larger than in countries where frames of persons are available. Because, also the design effect due to unequal inclusion probabilities, Deff_p, has to be taken into account when computing the sample sizes. Typically, when the only variation in the inclusion probabilities of persons is due to the selection of a person within a household, Deff_p is around 1.2. (This value depends on the empirical distribution of household sizes.)

> **Variation of inclusion probabilities:** The smaller the variation in inclusion probabilities, the smaller the design effect and hence the fewer interviews have to be conducted to reach the required effective sample size of $n_{eff} = 1\ 500$. Thus, sample designs with small variation in inclusion probabilities are favoured over those with larger variation.

the ESS goal of cross-national comparability, which at this stage of the survey life-cycle is promoted by means of equal effective sample sizes between countries.

In the end, the NC needs to know the net and gross sample sizes which are required to contribute to the ESS goal of cross-national comparability, by means of equal effective sample sizes between countries. The ESS uses a model-based approach to estimate design effects (see Gabler et al., 1999). The following examples shows how a planned value for the net and gross sample size can be calculated:

1. **Prediction of** Deff:
   The responsible sampling expert predicts the expected design effect(s) based on previous rounds and on expected PSU sizes. The following table gives a comparison of the prediction for a three stage sample design and a simple random sample design.

Design effect	three-stage sample	simple random sample
Deff_p	1.2	1
Deff_c	$(1+(\bar{b}-1)\times\rho)$	1
	$(1+(10-1)\times 0.05)$	1
	1.45	1
Deff	1.2×1.45	1
	1.74	1

   $\bar{b}$ is the average number of respondents per PSU and ρ the intraclass-correlation coefficient.

2. **Prediction of** n_{net} :
   The required net sample size is calculated in the following examples

Sample size	three-stage sample	simple random sample
n_{net}	1 500 $\times$ Deff	1 500 $\times$ Deff
	1 500 $\times$ 1.74	1 500 $\times$ 1
	2 610	1 500

3. **Prediction of** n_{gross}:
   Based on the expected response rate (rr) and the expected rate of ineligibles (ri) the required gross sample size is calculated as

Sample size	three-stage sample	simple random sample
n_{gross}	$\frac{n_{net}}{rr(1-ri)}$	$\frac{n_{net}}{rr(1-ri)}$
	$\left[\frac{2\ 610}{0.70 \times (1-0.03)}\right]$	$\left[\frac{1\ 500}{0.70 \times (1-0.03)}\right]$
	3 844	2 210

NCs are asked to note that gross sample sizes may have to be larger than usual for similar national or international surveys in order to achieve an effective sample size of 1 500. A sufficient budget therefore needs to be set aside to allow for this. In Round 5, for example, gross sample sizes from all but the smallest country ranged from 1 600 to 5 376. Please discuss this with your sampling expert at the earliest opportunity. If, for any reason, a deviation from this standard is unavoidable, please contact your sampling expert as early as possible!

Response Rates: The transition process from the gross sample to the net sample is of great importance for the quality of the data collected. Realistic estimate of response rates are needed in order to translate needed net sample size into gross sample size. (Please consult the ESS Round 8 Guidelines for enhancing response rates and minimising non-response bias, Stoop et al., 2016.)

Over-sampling: Allocation of the gross sample size to domains (e.g. sampling strata) to counter anticipated low response rates should not be used. Hence, if, for example, empirical evidence suggests that response rates in big cities are much lower than in rural areas, the gross sample size in big cities *must not* be increased because of this. The reason for this is to foster a more balanced response rate across the whole sample.

2. The Sign-off Form

The SoF is structured in the following way:

2.1 Description of the target population

- Any deviation from the definition of the ESS target population must be mentioned here. For instance, the potential under-coverage of certain groups, because of language problems or sampling frame deficiencies, or for any other reason, must be stated here. Often there are groups or areas (e.g. islands or overseas territories) that are deliberately excluded from the sampling frame, even though they are part of the target population.
- The size of the target population should be given.

2.2 Sampling Design

The technical details of sampling design are described here. Each sampling design is structured into one or more sampling domains. Sampling domains are areas or subgroups of the target population to which different sampling designs are applied. For instance for metropolitan areas, with their high population density, cluster sampling is not necessary, because distances between sampled persons are small. Therefore, two sampling domains are used: One for the urban centres of the country and the other for the rural areas, i.e. the rest of the population. Within each sampling domain all sampling stages are described by defining the sampling unit (e.g. addresses or municipalities), the sampling frame of the units (address register or list of municipalities), the number of sampling units selected at that stage, the stratification of sampling frame [if applicable], the allocation of the sample size to the strata [if applicable] at that stage, and final the sampling method used to select the units [within the strata] (e.g. simple random sampling without replacement). If the sampling domain has more than one sampling stage, then for every sampling stage after the first stage it must also be clear how the sample size of that stage is allocated to the selected sampling units of the previous stage.

1. Sampling Domain

1.1. Sampling Stage

unit The definition of the primary sampling units (PSUs).

fram The sampling frame of the primary sampling units (PSUs).

size The number of sampled PSUs.

stra The explicit stratification of the PSUs. If possible a table with the number of PSUs in each stratum should be appended. (Note that PSUs may be individuals.)

allc The allocation method for PSU sample. If possible a table with the number of sampled PSUs in each stratum should be appended.

algo The selection algorithm of the PSUs. If possible the software used to implement the algorithm should be named.

1.2. Sampling Stage

unit The definition of the secondary sampling units (SSUs).

fram The sampling frame of the secondary sampling units (SSUs).

size The number of sampled SSUs.

calc The allocation method of the SSU sample to the selected PSUs. If possible a table with the number of SSUs in each PSU should be appended.

stra The stratification of SSUs within selected PSUs. If possible a table with the number of SSUs in each stratum of each selected PSU should be appended.

allc The allocation method for SSU sample within each PSU. If possible a table with the number of sampled PSUs in each stratum should be appended.

algo The selection algorithm of the SSUs. If possible the software used to implement the algorithm should be named.

1.# The description of any further sampling stage is analogue to the second.

The description of any further sampling domain is analogue to the first.

2.3 Sample Size

Here the sample size is calculated based on the used design effect benchmark and the assumed response and ineligible rates.

Design Effect

$$\begin{aligned} \mathrm{Deff}_c = & \; 1 + (\bar{b} - 1) \cdot \rho \\ \mathrm{Deff} = & \; \mathrm{Deff}_c \cdot \mathrm{Deff}_p \end{aligned}$$

Target Response Rate & Rate of Ineligibles

$$\begin{aligned} \text{Response Rate} = & \; \text{rr} \\ \text{Ineligable Rate} = & \; \text{ri} \end{aligned}$$

Gross Sample Size

$$
\begin{aligned}
\mathrm{n}_{net} &= \mathrm{n}_{\mathit{eff}} \times \mathrm{Deff} \\
\mathrm{n}_{gross} &= \frac{\mathrm{n}_{net}}{\mathrm{rr} \times (1 - \mathrm{ri})}
\end{aligned}
$$

2.4 SDDF

The Sampling Design Data File (SDDF) is a dataset containing information on the sampling design itself. The file consists of the units of the gross sample, including those of a possible reserve sample. The description of the SDDF for ESS round 7, which can used as a template for round 8, can be found in the Annex Sample design data file (pages 93-96) of the ESS7 2014 Data Protocol.

2.5 Variables to be included in the SDDF

Here the variables should be listed that for the described design have to be included in the SDDF. Besides the mandatory variables, like `IDNO`, `PSU` (also for single stage designs) `OUTCOME` and the necessary `PROB` variables and any other variables that may become relevant have to be listed. A complete list of all potentially relevant variables of the SDDF can be found in the *Data Protocol* available by June 2016 on the NC intranet.

2.6 Auxiliary variables in the SDDF

Any auxiliary variable (micro or macro) that the NC agrees to include into the SDDF should be listed here.

The NC and other relevant experts should attempt to include as many relevant auxiliary variables as possible into the SDDF (that is for the gross sample) that might help the CST in the analysis of non-response processes. The auxiliary variables can for instance include information on age, gender, education, civil status geographical region, and employment. We distinguish between two types of auxiliary variables, micro and macro data. The micro data may be available from a central or local population register. The macro data on the same auxiliary variables might be available on the level of PSUs, or the strata of the first sampling stage (if applicable). Macro data is possible to acquire from different sources, including the same registers from which the micro data was gathered and public statistics. All countries should commit themselves to include some micro and macro variables into their SDDF.

2.7 Analytical Inclusion Probabilities

For each sampling domain and sampling stage within domains a mathematical formula must be given that will be used to calculate the `PROB` variables that are required for the sampling design data file.

1. Sampling Domain
   1.1. Formula for `PROB1` in the 1st domain.
   1.2. Formula for `PROB2` in the 1st domain.
   ...
2. Sampling Domain
   2.1. Formula for `PROB1` in the 2nd domain.
   2.2. Formula for `PROB2` in the 2nd domain.
   ...

...

Bibliography

Gabler, S., **Häder, S.** and **Lahiri, P.** (**1999**): *A model based justification of Kish's formula for design effects for weighting and clustering*. Survey Methodology, 25 (1), pp. 105–106.

Kish, L. (**1994**): *Multipopulation Survey Designs: Five Types with Seven Shared Aspects*. International Statistical Review, (62), pp. 167–186.

Shackman, G. (**2001**): *Sample size and design effect*. Technical report, Albany Chapter of American Statistical Association.
URL http://faculty.smu.edu/slstokes/stat6380/deff%20doc.pdf

Stoop, I., **Koch, A.**, **Halbherr, V.**, **Loosveldt, G.** and **Fitzgerald, R.** (**2016**): *Field Procedures in the European Social Survey Round 8: Guidelines for Enhancing Response rates and minimising nonresponse bias*. Technical report, London: ESS ERIC Headquarters.

Appendices

A. Sampling issues in *Round 8 Survey Specification for ESS ERIC Member, Observer and Guest countries*

On sampling see Section 8 pages 23-25 in Round 8 Survey Specification for ESS ERIC Member, Observer and Guest countries.

B. Estimator Used for the Design Effect and Effective Sample Size

B.1 Design Effect and Effective Sample Size

We want to compare the quality of statistics estimated from different sample surveys, where by *quality* we mean the *sampling variance* or *sampling error.* So we need an instrument to plan the different samples to insure that estimates of the same statistic are of similar quality across the countries. For the ESS this is done by the so-called *effective sample size* (n_{eff}). The n_{eff} is the number of elements required by simple random sample (SRS) to yield the same precision as under the actual, possibly complex, sampling design p. For instance, if a sampling design has an effective sample size of 1 500, then this implies that the design is as precise as a SRS of size 1 500. Formally the n_{eff} is defined as the ratio

$$\mathrm{n}_{eff} = \frac{\mathrm{n}_{net}}{\mathrm{Deff}}$$

where Deff is the so-called design effect. The design effect expresses how well a design fares in comparison to reference design SRS, in terms of sampling variance. The definition of Deff is

$$\mathrm{Deff} = \frac{\mathrm{V}\left(\hat{\theta}\right)_p}{\mathrm{V}\left(\hat{\theta}\right)_{\mathrm{SRS}}} ,$$

where $\mathrm{V}\left(\hat{\theta}\right)_p$ is the variance of estimator $\hat{\theta}$ under sampling design p and $\mathrm{V}\left(\hat{\theta}\right)_{\mathrm{SRS}}$ the variance of $\hat{\theta}$ under SRS. Sampling design p is the actual sampling design of the survey. If

- Deff > 1, then precision is lost by not using SRS and if
- Deff < 1, then precision is gained by not using SRS.

Depending on sampling design p, Deff can be difficult to estimat, if no assumptions about the distribution of the study variables are made. (The sample was never selected by a SRS, so it is particular difficult to find suitable estimators for the denominator in Deff.) For this reason a model based approach is used to estimat design effects for the ESS. The used model is intended to describe the effect of a sampling design involving clustering on the model variance of $\hat{\theta}$. The following sampling scheme is considered:

At first a sample of clusters $s_I = \{1, \ldots, n_I\}$ is selected from a population of clusters $\mathcal{U}_I = \{1, \ldots, N_I\}$. A single cluster $\mathcal{U}_i$, $i \in \mathcal{U}_I$ contains multiple elements of the target population, i.e. $U_i = \{1_i, \ldots, N_i\}$. At a second stage from each cluster $\mathcal{U}_i$ $i \in s_I$ an independent sample of elements $s_i \subset U_i$ of size n_i is selected. The total sample size is $\sum_{i \in s_I} n_I = n$.

The estimator $\hat{\theta}$ for which the design effects are estimated is $\overline{y}_w$, an estimator for the population mean of the variable of interest $\mathcal{Y}$. Estimator $\overline{y}_w$ is given by

$$\overline{y}_w = \frac{\sum_{i \in s_I} \sum_{k \in s_i} w_{ik} y_{ik}}{\sum_{i \in s_I} \sum_{k \in s_i} w_{ik}} ,$$

where w_{ik} is the survey weight associated with the k-th observation in the i-th cluster, and y_{ik} is the value of the variable of interest $\mathcal{Y}$ of the k-th observation in the i-th cluster.

Variable of interest $\mathcal{Y}$ is assumed to follow the model

$$\begin{aligned} \mathrm{E}\left(y_{ki}\right) &= \mu \\ \mathrm{V}\left(y_{ki}\right) &= \sigma^2 \\ \mathrm{COV}\left(y_{ki},\, y_{k'i'}\right) &= \begin{cases} \sigma^2 \rho & \text{for } k \neq k', i = i' \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

The design effect of $\overline{y}_w$ and design p can be decomposed into two factors, Deff_p the effect of using unequal weights w_{ik} and Deff_c the effect of using cluster sampling. Thus we have

$$\mathrm{Deff} = \mathrm{Deff}_p \times \mathrm{Deff}_c \, , \tag{B.1}$$

with

$$\mathrm{Deff}_p = n \frac{\sum_{i \in s_I} \sum_{k \in s_i} w_{ik}^2}{\left(\sum_{i \in s_I} \sum_{k \in s_i} w_{ik}\right)^2}$$

and

$$\mathrm{Deff}_c = 1 + (b^* - 1)\rho \, .$$

where, $b^* = \frac{\sum_{i \in s_I} \left(\sum_{k \in s_i} w_{ik}\right)^2}{\sum_{i \in s_I} \sum_{k \in s_i} w_{ik}^2}$ and ρ the Intraclass Correlation Coefficient (ICC). To estimate ρ an ANOVA estimator is used, which has the following form

$$\hat{\rho} = \frac{MSB - MSW}{MSB + (K-1)MSW} \, , \tag{B.2}$$

with

$$MSB = \frac{SSB}{n_I - 1} \, ,$$

where $SSB = \sum_{i \in s_I} n_i (\overline{y}_i - \overline{y})^2$ and

$$MSW = \frac{SSW}{n - n_I} \, ,$$

with $SSW = \sum_{i \in s_I} \sum_{k \in s_i} (y_{ik} - \overline{y}_i)^2$ and

$$K = (n_I - 1)^{-1} \left(n - \sum_{i \in s_I} \frac{n_i^2}{n} \right) \ ,$$

where $\overline{y}_i = \sum_{k \in s_i} y_{ik} {n_i}^{-1}$ the sample mean of the i-th cluster and $\overline{y} = \sum_{i \in s_I} \sum_{k \in s_i} y_{ki} n^{-1}$ the overall sample mean.

For sampling designs that do not use cluster sampling or select only one element in each cluster $b^* = 1$ (i.e. $\rho = 1$), we have $\mathrm{Deff}_c = 1$, i.e. there is no cluster effect. If the survey weights are equal for all elements in the sample $\mathrm{Deff}_c = 1$, else $\mathrm{Deff}_c > 1$, which means that unequal survey weights will increase the estimator of the design effect.

Since the ESS as a multitude of variables the SEP estimates model design effects for a selected number of variables from the ESS core questionnaire. The arithmetic mean of the estimated design effects of these variables is used as the benchmark value for Deff in the planning of the net sample sizes.

C. Example: Estimation of the Design Effect

Table C.1: Sample of PSUs

	PSU	PROB1	PROB2	x_i	$\overline{y}_i$	SW_i	n_{gross_i}	n_{neti}
1	20	1.0397	0.0115	867	8069	3102255	10	8
2	66	0.1643	0.0730	137	942	77901	10	8
3	96	0.2087	0.0575	174	1218	107813	10	10
4	136	0.2614	0.0459	218	1410	158269	10	5
5	100	0.4425	0.0271	369	4529	809714	10	8
6	105	0.2530	0.0474	211	1561	198928	10	6
7	120	0.5624	0.0213	469	7002	1468327	10	7
8	224	0.3250	0.0369	271	2005	239271	10	6
9	228	0.2410	0.0498	201	1629	148593	10	5
10	450	0.2650	0.0452	221	1745	214707	10	8

Suppose we have a two-stage sampling design. At the first stage $n_I = 10$ municipalities, the PSUs, are selected with probability proportional to the size of their target population. Table C.1 contains a sample of PSUs that has been selected by such a procedure. x_i is the population size of i-th PSU and $\overline{y}_i$ is the mean of the variable of interest in the i-th PSU. PROB1 is the expected selection frequency of the PSUs. For PSU 20 in the sample PROB1 is greater than one. This means that this PSU is, because of its large relative size, included with certainty into the sample. It has however the chance of been included a second time. The probability for a second inclusion is equal to 0.0397, the non-integer part of its PROB1 value.

At the second sampling stage 10 persons are selected from a municipality by a simple random sample for each time it gets selected into the sample. n_{neti} is the number of respondents in the i-th PSU and $\text{PROB2}_i = \frac{10}{x_i}$ is the inclusion probability of a person in the i-th PSU, given the PSU has been selected. The net sample size is $\text{n}_{net} = \sum_{i \in s_I} n_{\text{net}_i} = 71$.

The design weights are all equal to one for this design, because the inclusion probabilities of all persons is the same in this setting. This is easily shown by

$$\text{PROB1}_k = \frac{10 x_i}{\sum_{i \in \mathcal{U}_I} x_i},$$

$$\text{PROB2}_k = \frac{10}{x_i}, \text{ and}$$

$$\text{PROB1}_k \text{PROB2}_k = \frac{100}{\sum_{i \in \mathcal{U}_I} x_i} \ ,$$

where PROB1_k and PROB1_k are the values of PROB1 and PROB1 for the k-th respondent, respectively. Thus, for this particular design we have $\text{Deff}_c = 1$. Which means that this design has no weigthing effect.

To estimate the ICC ρ we compute MSW (mean squares within) and MSB (mean squares between). Table C.1 contains $\text{SW}_i = \sum_{k \in s_{\text{net}_i}} (y_{ik} - \overline{y}_i)^2$, which are the squared differences between the all observations of $\mathcal{Y}$ from the i-th PSU and the PSU mean $\overline{y}_i$, for $i = 1, \ldots, n_I$. Then

$$SSW = \sum_{i \in s_I} SW_i = 6.5257783 \times 10^6 \ .$$

The mean squares within MSW are then calculated as

$$MSW = \frac{SSW}{n_I - 1} = \frac{6.5257783 \times 10^6}{9} = 7.2508648 \times 10^5 \ .$$

The overall sample mean $\overline{y} = 3099.51$, thus the SSB is given by

$$SSB = \sum_{i \in s_I} (\overline{y}_i - 3099.51)^2 = 6.0580131 \times 10^7 \ ,$$

and the MSB by

$$MSB = \frac{SSB}{\text{n}_{net} - n_I} = \frac{6.0580131 \times 10^7}{61} = 9.931169 \times 10^5 \ .$$

Now we estimate ρ by $\hat{\rho}$ where

$$\begin{aligned}
\hat{\rho} &= \frac{MSB - MSW}{MSB + (K-1)MSW} \\
&= \frac{9.931169 \times 10^5 - 7.2508648 \times 10^5}{9.931169 \times 10^5 + (7.0642 - 1) \times 7.2508648 \times 10^5} \\
&= 0.0497 \ .
\end{aligned}$$

To determine Deff_c we need also to calculate b^*. Because the weights are all equal to one for this design we have

$$b^* = \frac{\sum_{i \in s_I} n_{\text{net}_i}^2}{\text{n}_{net}} = \frac{527}{71} = 7.4225,$$

thus

$$\text{Deff}_c = 1 + (7.4225 - 1) \times 0.0497 = 1.32 \ .$$

Finally, our estimate for the overall design effect $\text{Deff} = 1.32$, as $\text{Deff}_p = 1$. Since the net sample size is 71 the estimated effective sample is $\text{n}_{\mathit{eff}} = \frac{\text{n}_{net}}{\text{Deff}} = 54$. This can be interpreted in the following way: The sample mean $\overline{y}$, under the considered sampling design, is as efficient as under a simple random sample of size 54.