Sampling for the European Social Survey Round VII: Principles and Requirements

Guide

2. Version

The ESS Sampling Expert Panel

29th January 2014

http://www.europeansocialsurvey.org/

Summary

The objective of the task *Sampling Design and Sampling Expert Panel* in work package *Fieldwork Preparation* is the „design and implementation of workable and equivalent sampling strategies in all participating countries“. This concept stands for random (probability) samples with estimates of comparable precision. From the statistical point of view, full coverage of the population, non-response reduction, and consideration of design effects are prerequisites for the comparability of unbiased or at least minimum biased estimates. In the following we shortly want to

- describe the theoretical background for these requirements,
- show some examples, how the requirements can be kept in the practices of the individual countries and
- explain, which information the Sampling Expert Panel needs from the NCs to evaluate their proposed sampling schemes.

Important Changes to Round VI

1. There are two operationalisations for the target population depending on the definition of *being aged 15 or older.* One in case eligible individuals can directly be selected from a register and the other if they have first to be identified in the field.
2. Over-sampling in anticipation of expected unequal response rates must not be used. Any proposals for over-sampling due to other reasons must be authorised by the ESS Director.
3. If possible, i.e. if there exists data from previous ESS rounds, the SEP will provide values for the assumed design effect.
4. The information requirements to fill in the Sign Off Forms will not have to be necessarily included into the tender that is put out for application for survey organisations. However, survey organisations, in cooperation with the NC, are still being required to provide the necessary information to the SEP.

Contents

Chapter 1

Principles of Sampling in the ESS

1.1 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 the availability of frames, experience, and of course also the costs in different countries. If, after the survey has been conducted, adequate estimators are chosen, the resulting values are comparable. To ensure this comparability, design weights have to be computed for each country. For this, the inclusion probabilities of every respondent at each stage of selection must be known and recorded in the *Sample Design Data File* (SDDF). Furthermore, also for non-respondents the inclusion probabilities are to be recorded in the SDDF for every stage where the information to calculate them is available.

1.2 Discussion of standards set in the Technical Annex - Specification for participating countries

Only probability samples provide a theoretical basis which allows to infer from the sample to the whole target population or sub-sets of it. As design based inference is one important goal in the project, *probability samples* are required. This, however, is related to other requirements:

- full *coverage* of the target population,
- high *response rates* (ESS: target minimum response rate: 70%),
- the same *minimum effective sample* sizes (n_{eff}) in participating country ($n_{eff} = 1500$ or $n_{eff} = 800$ for countries with an ESS populations (aged 15+) of less than 2 million).

These requirements can only be sensibly discussed in the context of *random samples*. They form a theoretical system that in the end ensures equivalence. The crucial point, however, is that the practical implementation does comply with theoretical framework.

1.2.1 Coverage of the Target Population

An important step in planning a survey is the operationalisations of the population under study. In the case of the ESS it contains persons 15 years or older who are resident within private households, regardless of nationality and citizenship or language. 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 persons 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 field a persons is treated as 15 or older if she or he is 15 at the day the interviewer does the listing of household members.

The more completely the frame covers the persons belonging to the target population,the higher the quality of the sample will be. However, the quality of the sampling frames – e.g. coverage, updating intervals and accessibility – may differ from country to country. Therefore, frames will be evaluated carefully by the responsible sampling expert together with the National Coordinator (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:

a) countries with reliable lists of residents that are available for social research, such as Norway, Sweden, Switzerland or Denmark

b) countries with reliable lists of the households or addresses, that are available for social research, such as Netherlands or the U.K.

c) 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 c). 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 following sampling stages. When no sampling frame for sampling units (e.g. streets, addresses, or households) is available, there are mainly two ways to make the selection:

Listing and selection: Staff from the fieldwork agency lists all units within a certain area, which is defined by the sampling design. The sampling units are then drawn from these lists and given to the interviewer. The enumeration or listing of sampling units and the interviews must be done by different persons.

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.

The question is to which extent *random route techniques* can be judged to be „strictly random“. In Lyberg's (see Lyberg, 2000) view these techniques do result in non-probability samples and should thus be *avoided whenever possible*. At the very least, the following questions have to be answered in order to minimize the interviewer's influence on the selection of respondents:

- How are the algorithms for random routes defined?
- What experience do interviewers have with random walks?
- How can the whole random walk process be controlled?

An acceptable method involves one fieldwork person doing the complete random walk, recording the sampled units and then transferring them to the survey office before another person begins contacting/visiting the selected units. Furthermore random route techniques have often the drawback that it is very complex to determine the inclusion probabilities of selected sampling units and thus it is often not feasible to calculate them in practice.

Frame Imperfections

Even in countries where reliable sampling frames exist, we have to expect pitfalls in the sampling process. It is often difficult to fully cover the target population in practice. However, there should also be an awareness if persons that do not live in private households, for example students in college dorms, elderly people in retirement homes, military personal in barracks, are part of the sampling frame (over-coverage). Sampling units that are part of the sampling frame and the target population but are not allowed to be contacted by a survey organisation, so called opt-outs, should stay in the sampling frame and be treated as refusals if selected.

1.2.2 Response Rates

If unit non-response is associated with central variables under study, it can introduce severe bias in the survey estimates. Therefore, it is of essential importance to plan and implement a *sufficient number of contact attempts* as well as appropriate field work strategies for the persuasion of the target persons to participate in the survey. All efforts should be made to minimise non-response because it increases the danger of biased estimates, which should clearly be avoided.

Over-sampling 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 balanced response rate across the whole sample.

The transition process from the gross sample to the net sample is of great importance for the quality of the data collected. Comparability of estimates can be achieved only if the net samples are not seriously biased and bias is less likely if the response rates are fairly high.

1.2.3 Design Effects and Effective Sample Size

As already mentioned, a variety of complex sample designs such as stratified random samples, multi-stage sample designs and combinations of them were used in rounds I to VI of the ESS and can also be expected to be used in round VII.

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 defined as ratio of the variance of a variable under the actual sample design to the variance 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), depending, not only, on the degree of homogeneity in the data, measured by the so called *intraclass-correlation coefficient* (ρ). This part of the design effect which is due to selection of clusters of elements instated of elementary units is denoted $deff_c$.

Appropriate interclass correlation coefficients for the 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 the NC and fieldwork organisation will provide in close collaboration with the SEP a presumptive value. Besides the homogeneity, also the size of the PSUs influences the design effect – with n_{net} constant, 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 chosen 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 PSU size, the smaller the design effect and hence the less interviews have to be conducted to reach the required effective sample size of $n_{eff} = 1500$. 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 within a clusters that is present in the population but also by an interviewer effect, e.g. if the same person conducts all the interviews in one cluster.

Another important source which has an effect on the design effect is any departure from equal probability selection methods (epsem), which requires design weighting to correct 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. Hence, 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 inclusion probabilities is due to the selection of a person within a household, $deff_p$ is around 1.2.

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 $n_{eff} = 1500$. Thus, sample designs with small variation in inclusion probabilities are favourable compared to those with larger variation.

In the end, the NC needs to know the net and gross sample sizes which are required to reach the ESS goals of equal effective sample sizes. Therefore, the SEP has designed a *sampling sign-off form* (SoF) which capture central aspects of the sample design. In these SoFs, the calculation of required net and gross sample sizes is documented by the sampling expert responsible for the country. In general, this calculation involves three steps.

For example:

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 predictions for example of a) a threestage sample design and b) 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
	$(1 + (10 - 1) \times 0.05)$	
	1.45	
$deff$	1.2×1.45	1
	1.74	

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}	$1500 \times deff$	$1500 \times deff$
	1500×1.74	1500×1
	2610	

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}	$\frac{n_{net}}{rr(1-ri)}$	$\frac{n_{net}}{rr(1-ri)}$
	$\left[\frac{2610}{0.70 \times (1-0.03)}\right]$	$\left[\frac{1500}{0.70 \times (1-0.03)}\right]$
	3844	2210

NCs and sampling experts 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 1500. 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 1600 to 5376. 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!

1.3 Handling of the Workpackage

In round I to VI we worked with an expert panel on sampling. This panel will continue its work. Its current members are the following sampling specialists:

- Stefan Zins (GESIS – Leibniz Institute for the Social Sciences, Germany)
- Sabine Häder (GESIS – Leibniz Institute for the Social Sciences, Germany)
- Siegfried Gabler (GESIS – Leibniz Institute for the Social Sciences, Germany)
- Seppo Laaksonen (University of Helsinki, Finland)
- Peter Lynn (University of Essex, U.K.)

Depending on the number of participating countries in ESS7 each of the experts will be assigned three to five countries to liaise and support. However, the decision to "sign off" a design will be made together by the whole team.

As a starting point for the assessment of the sampling designs the sampling expert panel needs the information available from the survey organisations. The NCs should ensure that the questions listed in Chapter 2 can be answered with the help of the survey organisations. Additionally, we ask the NCs to give their comments to the proposed designs, e.g. to evaluate them with the help of their experience. At least the following points should be treated:

1. Is the proposed design good or best practice in the country concerned?

2. Does the survey organisation have experience with the proposed design?

3. Is the proposed response rate realistic?

Item 3 is also be checked with the fieldwork team.

If the information provided by the survey organisation and the additional comments by the NCs is sufficient, the expert panel is enabled to „sign off" the proposals without delay. If the information is not sufficient, the respective expert will start a dialogue with the NC (and possibly the survey organisation involved) in order to clarify details or propose amendments. If necessary, other sampling specialists in the country will be joined in the discussion, so that their knowledge of local practices, arrangements and vocabulary can be drawn on. Similarly, where necessary, the respective sampling expert will visit the country to give help and support. These consultations will be conducted as efficiently as possible to give maximum time for the design to be implemented in good time according to the specification.

Chapter 2

Information to be provided by the Survey Organisation/NCs

Answers to the following questions concerning sampling should be given by the survey organisations and or the NCs.

Description of the target population

- Do the ESS specifications of the sampling 'universe' coincide with the target population definition in the country? It must be noted if any deliberate exclusion of certain eligible persons will be done. For instance if a country has some special jurisdictions that are hard to cover, like overseas territories.

Analytical description of the sample design

Please explain, if applicable, in detail how the following points are to be implemented in your sample design:

- Is the design single- or multi-stage?
- What is the assumed design effect, if the SEPl is not able to estimate it based on data from previous ESS round?
- Does the decision to contact a person only depend on the outcome of the random sampling selection process? That is, is sampling independent of any other random events that occur during fieldwork, e.g. the number of non respondents and ineligibles.
- Are all ultimate sampling units, i.e. selected persons, being contacted/interviewed according to same protocol?

For *each sampling stage* of the design the following specifications are needed:

i) Sampling frame. What is the definition of the sampling units that are listed in the sampling frame. How is the sampling frame of the stage constructed, i.e. which registers or lists are used to enumerate sampling units and what is their information content to construct the sampling frame? Is there any known under- or over-coverage in the used frame?

ii) Sample size. How many units will be selected?

iii) Stratification. Are the sampling units grouped, i.e. stratified, and is sampling performed independently within each stratum? If yes, what are the variables used to stratify the sampling units?

iv) The allocation to strata. If stratification is used then how is the sampling size distributed, i.e. allocated, to the strata?

v) Allocation to the selected preceding sampling units. If the stage is not the first one, how is the sample size of the stage distributed, i.e. allocated, to the sampling units of the stage prior to this stage? How many sampling units are in the selected clusters of the previous stage?

vi) Sampling Method. What method (ideally the algorithm and the software of its implementation) is used to select the sampling units?

Sample size

- How has the effective sample size to be calculated, including estimates of response rates and design effects due to clustering or necessary weighting and eligible rates? This will also be checked by the fieldwork group.

- Will certain population domains (e.g. administrative areas or socio-demographic groups) be over-sampled?. *If yes, any such proposals of over-sampling* ***must*** *be authorised by the ESS Director.*

If multiple sampling designs are used in parallel in a country, i.e. certain domains of the target population are surveyed by different sampling design, for every domain the above information has to be provided. The NCs are responsible for inquiring the survey organizations about these points. As a result, the assigned sampling expert shall be enabled to fill in the Sign Off Form. As an example the following Sign Off Form has been included:

Sampling for the European Social Survey-Round VII

Country:	Name of the country
NC:	Name of the NC and email contact
Other experts:	Names of other relevant experts concerned with sampling issues
Survey Institute:	Name of the Organization entrusted with the fieldwork
Expert:	Member of the SEP assigned to the country
Reference Survey:	If the current sampling design is recycled from a previous round, then which round is it?
Date:	Sign off date

Target Population

All persons 15 years or older at the day of selection who are residents of (country name) within private households, excluding the islands A and B.

Remarks

...

Sampling Design

A 2 stage sampling design.

Stage 1

1. The PSUs are the census sections of the 2011 Census. The PSUs do include all citizens registered in the municipal rolls.
2. A number of n PSUs is selected.
3. The PSUs are selected with stratification. The strata have been obtained by crossing two population classification criteria: region of residence (18 regions) and size of habitat (4 brackets). The four brackets are:
   a) target population living in cities with more than 100,000 inhabitants
   b) target population living in cities between 50,001 and 100,000 inhabitants
   c) target population living in municipalities between 10,001 and 50,000 inhabitants
   d) target population living in municipalities with less than 10,001 inhabitants

   64 out of 72 theoretical strata are not empty.
4. The n PSUs are allocated proportionally to the size of the target population within each stratum.
5. The PSUs are selected proportionally to the size of the target population within each PSU using conditional poisson sampling (see, Tillé (2006), Sampling Algorithms).

Remarks

...

Stage 2

1. The ultimate sampling unit is a individual from the target population. The sampling frame is the listing of individuals in the selected PSU.
2. A total of m individuals will be drawn from the selected PSUs.
3. No stratification of individuals within the PSU.
4. The m individuals are allocated equally to the selected PSU, that is from each selected PSUs the same number of individuals is drawn.
5. The individuals will selected by a simple random sample from the selected PSUs.

Remarks

...

Sample size

The design effect has been estimated using the data from the previous ESS round. For the estimation of the mean intra-class correlation coefficient, 23 variables had been used (5 numerical, 10 ordinal and 8 dummy variables).

$$\begin{aligned} \text{Deff}_c &= 1 + (\bar{\text{b}} - 1) \times \rho \\ &= 1 + (4.257 - 1) \times 0.054 \\ &= 1.176 \end{aligned}$$

$$\text{Deff}_p = 1.083$$

$$\begin{aligned} \text{Deff} &= \text{Deff}_c \times \text{Deff}_p \\ &= 1.176 \times 1.083 \\ &= 1.27 \end{aligned}$$

Remarks

...

Target Response Rate & Rate of Ineligibles

rr = 69.1%
ri = 3.6%

Target response rate is 70%, although a safe estimation of 69.1% has been used for the calculation of the sample size. Predicted Eligibility rate: 96.4% This figure has been calculated by the average eligible rates from the last two rounds.

Remarks

...

Effective Sample Size

$$
\begin{aligned}
\text{n}_{\textit{eff}} &= 1500 \\
\\
\text{n}_{\textit{net}} &= \text{n}_{\textit{eff}} \times \text{Deff} \\
&= 1500 \times 1.27 \\
&= 1905 \\
\\
\text{n}_{\textit{gross}} &= \frac{\text{n}_{\textit{net}}}{\text{rr} \times (1 - \text{ri})} \\
&= \frac{\text{n}_{\textit{net}}}{0.691 \times (1 - 0.036)} \\
&= 2860
\end{aligned}
$$

Remarks

...

Analytical Inclusion Probabilities

PROB1

The inclusion probabilities of PSUs.

$$
\begin{aligned}
\text{PROB1} &= \frac{N_{hi}}{N_h} n_h \\
\text{with} &= \\
n_h &= \frac{N_h}{N} ,
\end{aligned}
$$

where N is the total size of the target population, N_h is the size of the target Population in the h-th stratum (of stage one), and N_{hi} the size of the target population of the i-th PSU in the h-th stratum.

PROB2

The conditional inclusion probabilities of individuals given the PSU they belong to has been selected.

$$
\text{PROB2} = \frac{n_{hi}}{N_{hi}} ,
$$

where n_{hi} is the number of individuals selected from each the i-th PSU in the h-th stratum.

General Remarks

Bibliography

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

Lyberg, L. (2000): *Review of IALS - a commentary on the technical report*. In: Measuring Adult Literacy. The International Adult Literacy Survey in the European Context, London: Office for National Statistics.

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

Appendix A

Sampling issues in the „Specification for ESS ERIC Member and Observer countries“ – Round 7 of the ESS

A.1 Population coverage

The ESS will be representative of all persons aged 15 and over (no upper age limit) resident within private households in each country, regardless of their nationality, citizenship or language. Potential under-coverage of certain groups, say because of language problems or sampling frame deficiencies, or for any other reason, must be discussed with the sampling panel prior to deciding on the final sampling method, so that the problem can be remedied if at all possible.

A household is defined as one person living alone or a group of people living at the same address (and have that address as their only or main residence) who either share at least one main meal a day or share the living accommodation (or both). People who are temporarily away (in hospital, on holiday) still belong to the household.

A.2 The sample

The sample is to be selected by strict random probability methods at every stage and respondents are to be interviewed face-to-face (see section 6.2.3 for details on fieldwork). Where a sampling frame of individuals is not available, or lacks sufficient coverage, countries may use a sampling frame of households or of addresses. In these cases, procedures for selecting a household from a multi-household address (where appropriate), and an individual within a household, will be specified and agreed in advance with the sampling panel.

If an area sample based on a random-route procedure is applied to a sample of addresses or dwellings, it must be ensured that a pre-listing of at least twice as many dwellings as needed for the gross sample is performed from which the required number of dwellings will need to be selected by the survey agency. The person who produces the pre-listing (the enumerator) *should under no circumstances be the same person as the interviewer*. Enumeration should take place before the start of fieldwork.

In any event, the relative selection probabilities of every sample member must be known and recorded, as should any remaining systematic non-coverage problems. *Quota sampling is not permitted at any stage, nor is substitution of non-responding households or individuals (whether "refusals", "non-contacts" or "ineligibles")*. Over-sampling of certain groups (e.g. sample units in low response areas) is not recommended (any efforts to do so MUST be discussed and agreed in advance). The SEP strongly recommends using stratification at least of primary sampling units (PSUs).

A.3 Effective sample size

The minimum "effective achieved sample size" (n_{eff}) must be 1.500, or 800 in countries with ESS populations (aged 15+) of less than 2 million after discounting for design effects. With the help of the Sampling Expert Panel (SEP), each country should determine the appropriate size of its initial issued sample by taking into account the realistic predicted impact of clustering, variation in inclusion probabilities (if applicable), eligibility rates (where appropriate), and response rate. The sampling expert panel will assist in the calculation of the gross sample size required in order to achieve an effective sample size of 1.500 (800) interviews. Note that in some cases (e.g. complex multi-stage sampling design, low expected response rate, etc) this might require a larger gross sample and this needs to be considered when setting the budget for the survey. After the SEP has agreed on and signed off a sampling design, the number of cases in the gross sample must be issued without amendment.

A.4 Documentation of sampling procedures

The precise sampling procedures to be employed in each country, and their implications for representativeness, must be documented in full and submitted in advance of fieldwork to the expert panel for "signing off". This form must subsequently be sent to the HQ-CST for future reference. This precaution is to ensure that all countries within the ESS have defensible (and equivalent) national probability samples of their resident (aged 15 and over) populations.

The following details will be required before the sampling expert panel can "sign off" a country's sample design:

- a description of the target population and of any systematic exclusions due to frame imperfections
- a description of the sampling frame and of the units it comprises at all stages of the design (including information on units that might be used either to stratify the sample or to vary probabilities of selection for certain subgroups, and estimates of any likely under-coverage, over-coverage and ineligibles)

- for those using multi-stage samples a description of how the units at each stage will be selected to result in a random sample of individuals, plus the inclusion probabilities of units at each stage of selection
- details of whether and how the sample is to be clustered geographically, and how the initial clusters are to be selected
- full details of any stratification to be employed
- the calculations on which the predicted effective sample size has been based
- realistic and reliable predictions of design effects (due to clustering and due to unequal inclusion probabilities); response rates; the rate of ineligibles and the required number of interviews as well as the required number of elements to draw the initial sample (gross).

A sample design data file (SDDF) must be produced by each country and delivered to the HQ-CST. It must contain all information about the sample design, such as inclusion probabilities of each stage, information on clustering and stratification. A full and detailed specification of the SDDF is provided in the ESS Data Protocol. *Failure to deliver the sample design data file (SDDF) will be considered an irreparable compromise to quality (see section 10).*

The final sample design will also be fully documented by each NC. This documentation will be translated into one or more variables within the national data file to indicate the relative selection probabilities of cases and to enable appropriate weighting strategies to be calculated. See section 1.4 for information about data protection assured by the ESS Archive at NSD.

Appendix B

Rules for estimating design effects

B.1 Effective Sample Size

The effective sample size (n_{eff}) is the size of a simple random sample which would produce the same precision (standard errors) as the design actually used. Typically, n_{eff} is less than the actual number of achieved interviews, n_{net}, as certain aspects of survey design – for example, clustering or the use of differing selection probabilities – tend to reduce the precision of estimates. The reduction of precision is known as the design effect ($deff$):

$$deff = \frac{\text{Actual sampling variance}}{\text{Sampling variance with simple random samples of same size}} \tag{B.1}$$

$$= \frac{n_{net}}{n_{eff}} \tag{B.2}$$

hence $n_{eff} = \frac{n_{net}}{deff}$

We therefore need to be able to predict the value of $deff$ for a proposed sample design, in order to determine how many interviews should be achieved so as to produce a particular value of n_{eff}. We suggest that two components of $deff$ should be taken into account at the design stage – the design effect arising from differing selection probabilities ($deff_p$) and the design effect arising from clustering ($deff_c$). Then $deff = deff_p deff_c$. We then also need to predict the survey response rate (and the proportion of ineligibles on the sampling frame, if relevant) in order to determine the size of the gross sample (n_{gross}) required in order to achieve approximately n_{net} interviews.

B.2 Design Effects due to Differing Selection Probabilities

In some countries which have accessible population registers, it will be possible to select an equal-probability sample from the survey population. In other countries, it will be necessary to select the sample in stages, with the penultimate stage being residential addresses. In this case, each person's selection probability will depend on their household size.

If differing selection probabilities are to be used - for whatever reason - the associated design effect should be predicted. This can be done very simply, using the following formula

$$deff_p = \frac{m\sum_{i=1}^{I} m_i w_i^2}{\sum_{i=1}^{I}(m_i w_i)^2}$$

where I is the number of probability classes and m_i the respondents in the ith selection probability class, each receiving a weight of $w_i \propto \frac{N_i}{m_i}$, where $\propto$ means „proportional to". Note that this formula assumes that the population variance of survey variables will not vary over selection probability classes – a reasonable assumption in most situations.

B.3 Design Effects Due to Clustering

It is anticipated that in most countries it will be efficient to select a multi-stage, clustered, sample. In such situations there will also be a design effect due to clustering:

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

where $\bar{b}$ is the mean number of respondents per cluster and ρ is the intraclass-correlation (or "rate of homogeneity") – a measure of the extent to which persons within a clustering unit are more homogeneous than persons within the population as a whole. This design effect can be estimated, at least crudely, from knowledge of other surveys and/or the nature of the clustering units.

In practice, all elements of the overall design effect, including that due to differing selection probabilities and that due to clustering, will take different values for different survey estimates. For sample design purposes, an average value should be used.

B.4 Example: How to determine the size of the gross sample

We have prescribed $n_{eff} > 1500$.

To determine n_{gross}, we must first determine n_{net} and therefore estimate $deff = deff_p \times deff_p$

1. Suppose the proposed clustering units are administrative areas of around 5.000 households on average and that based on data from other surveys, we expect that for these areas, ρ will take values of around 0.02 for many variables. Then, if we are proposing a design with a mean of 15 interviews per cluster:

   $$deff_c = 1 + (15 - 1) \times 0.02 = 1.28.$$

   [Note: If there is no available empirical evidence at all upon which to base an estimate of ρ, then we suggest that a value of 0.02 should be used.]

2. Suppose that the only available sampling frame is a list of addresses and that these must be selected with equal probabilities. The proposed design is then randomly to select one person to interview at each address. This is the only aspect of the proposed design that involves differing selection probabilities. Then, we can use population statistics on the distribution of household size to estimate the number of respondents in each selection probability class, thus:

No. of persons aged 15+ in household i	Proportion of households in population $\frac{H_i}{H}$	No. of achieved interviews m_i	Relative weight		
			w_i	$m_i w_i$	$m_i w_i^2$
1	0.35	0.35 n_{net}	1	0.35 n_{net}	0.35 n_{net}
2	0.45	0.45 n_{net}	2	0.90 n_{net}	1.80 n_{net}
3	0.12	0.12 n_{net}	3	0.36 n_{net}	1.08 n_{net}
4	0.06	0.06 n_{net}	4	0.24 n_{net}	0.96 n_{net}
5	0.02	0.02 n_{net}	5	0.10 n_{net}	0.50 n_{net}
				1.95 n_{net}	4.69 n_{net}

The population distribution of household size appears in the second column. From this, we can predict that the sample distribution will be as shown in the third column. We can thus predict $deff_p$:

$$deff_p = n_{net} = n_{net}\frac{4.69 \times n_{net}}{(1.95 \times n_{net})^2} = \frac{4.69}{1.95^2} = 1.23$$

3. Thus, we predict $deff = 1.28 \times 1.23 = 1.57$. Consequently, to achieve $n_{eff} > 1500$ with this design, we need $n_{net} > 2355$.

4. The final stage is to calculate the gross sample size in order to achieve around 2355 interviews. If we anticipate a response rate of 80% and 5% of ineligibles (e.g. addresses which do not contain a resident household), we have:

$$n_{gross} = \frac{\frac{n_{net}}{0.80}}{0.95} = \frac{\frac{2355}{0.80}}{0.95} \approx 3099$$

So we would select a gross sample of at least 3099 addresses.