European Social Survey (2012)

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

Mannheim, European Social Survey, GESIS

Guide

Final Version

The Sampling Expert Panel of the ESS

January 4, 2012

Summary: The objective of the work package *Sampling and Fieldwork Coordination* 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 National Coordinators to evaluate their proposed sampling schemes.

1 Basic principles for sampling in cross-cultural surveys

Kish (1994, p.173) provides the starting point of the sampling expert panel's work:

> "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-cultural 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 sample member at each stage of selection must be known and recorded in the *Sample Design Data File* (SDDF).

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 countries (ESS: $n_{\text{eff}} = 1,500$ or $n_{\text{eff}} = 800$ where population is smaller than two million inhabitants).

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 works.

2.1 Full coverage of the target population

An important step in planning a survey is the definition 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. In countries in which any minority language is spoken as a first language by 5% or more of the population, the questionnaire has to be translated into that language. This definition applies to all participating countries and thus every person with the defined characteristics must have a non-zero chance of being selected. Thus, 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. The results of these evaluations have to be documented and taken into account when the data are 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/addresses that are available for social research such as, Netherlands or the U.K.

c) countries without reliable and/or available lists such as Portugal or Bulgaria

Drawing a sample is more complicated if no registers (lists) are available (group c). In this case, *multi-stage sample designs* are usually applied, in which the selection of so called *primary sampling units* (PSUs) forms the first stage and the selection of, for example, households within selected PSUs follows at the second stage of sampling.

When no sampling frames of second stage units (e.g. households) are available, there are mainly two ways to make the selection:

Listing and selection : Staff from the fieldwork agency lists all addresses within a certain PSU. The target households are then drawn from these lists and given to the interviewer.

Random route techniques : fieldwork persons select, starting from a randomly selected address, a number of second stage units (e.g. households) which are then being contacted by the interviewer.

The question is to which extent *random route techniques* can be judged to be “strictly random”. In Lyberg’s 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 rules for random routes defined in the countries?
- 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 addresses and then transferring them to the survey office before another person begins contacting the selected addresses.

Even in countries where reliable frames exist, we have to expect pitfalls in the sampling process. It is, for example, difficult to fully cover people with illegal status. Such systematic losses due to undercoverage cannot be ruled out in practice but must be documented carefully.

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. The fixed goal of 70% response rate in the ESS is particularly challenging for some countries where response rates of 50 percent or even less are common. Nevertheless, all efforts should be made to avoid non-response because it increases the danger of biased samples, which should clearly be avoided.

If any reliable information about the expected amount of non-response is available, it should be used in the sample design. 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 should be increased. If the gross sample size in a PSU is 10 on average, the overall expected response rate is 70% but only 40% in big cities, then the gross sample size of PSUs located in big cities should be $10 \cdot \left(\frac{.70}{.40}\right) = 17.5 \approx 18$.

To sum up, 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. Bias, however, is less likely if the response rates are fairly high and appropriate auxiliary data is collected to aid weighting.

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 V of the ESS and can also be expected to be used in round VI.

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” (Shackman, 2001), depending, on the one hand, on the degree of homogeneity in the data, measured by the so called *intraclass correlation coefficient* (ρ).

It is essential that National Coordinators and the fieldwork organizations analyze the data from round I to V to calculate appropriate intraclass correlation coefficients for the sample designs used in their countries. 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 $n_{\mathit{eff}} = 1,500$. In that sense, a large number of PSUs with only a few interviews conducted in each should be the goal.

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, 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 less interviews have to be conducted to reach the required effective sample size of $n_{eff} = 1,500$. Hence, sample designs with small variation in inclusion probabilities are favourable compared to those with larger variation.

In the end, the National Coordinator needs to know the net and gross sample sizes which are required to reach the ESS goals of equal effective sample sizes. Therefore, the sampling expert panel has designed so called *sampling sign-off forms* which capture central aspects of the sample design. In these sign-off forms, 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:

1. prediction of design effects : 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 a) a three-stage sample design and b) a simple random sample design.

Design effect		three-stage sample	simple random sample
$deff_p$	=	1.20	1.00
$deff_c$	=	$1 + (\bar{b} - 1) \cdot \rho$	
	=	$1 + (10 - 1) \cdot 0.05$	
	=	1.45	1.00
$deff$	=	$deff_p \cdot deff_c$	
	=	$1.20 \cdot 1.45$	$1.00 \cdot 1.00$
	=	1.74	1.00

2. prediction of n_{net} : The required net sample size is being calculated as follows

Sample size		three-stage sample	simple random sample
n_{net}	=	$1,500 \cdot deff$	
	=	$1,500 \cdot 1.74$	$1,500 \cdot 1.00$
	=	$2,610$	$1,500$

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}	=	$\frac{n_{net}}{rr \cdot (1-ri)}$	
	=	$\frac{2,610}{0.70 \cdot (1-0.03)}$	$\frac{1,500}{0.70 \cdot (1-0.03)}$
	$\approx$	$3,844$	$2,210$

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!

3 Handling of the Workpackage

In round I to V we worked with an expert panel on sampling. This panel will continue its work. Members are the following sampling specialists:

- Matthias Ganninger (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.)

Each of the experts will be assigned about six countries to liase 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 tenders. The National Coordinators should ensure that the questions listed in paragraph 5 can be answered with the help of the tenders. That means that the survey organisations have to be informed by the NCs about these requirements in advance of handing in the tenders. 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:

- Is the proposed design good or best practice in the country concerned?
- Does the survey organisation have experience with the proposed design?
- Is the proposed response rate realistic?

If the information contained in the bidding and the additional comments by the National Coordinators 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 National Coordinator (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 panellist 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.

4 Information need to be contained in the tenders

Answers to the following questions concerning sampling should be given in the tenders from the survey organisations.

Description of the target population

- Are the ESS specifications of the sampling 'universe' adhered to (i.e. all residents aged 15+, regardless of nationality or citizenship, excluding only the homeless and the institutional population)?

Description of the sampling frame

- Is the quality of the proposed sampling frame suited to its proposed purpose (in terms of coverage, updating, access, etc)?

Detailed(!) description of the sample design

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

- How is the sample stratified?
- Is the design single- or multi-stage?

- Which stages are defined?
- How large is the expected degree of homogeneity?
- Is there any oversampling due to expected amount of non-response?
- Are reserve samples planned?

Prediction of design effects

- Prediction of the design effect due to clustering $deff_c$
- Prediction of the design effect due to unequal inclusion probabilities $deff_p$
- Prediction of the overall design effect $deff$

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?
- Will any population subgroups be over-sampled?
- What steps will be taken to achieve the target response rate?

The National Coordinators are responsible for inquiring the survey organizations about these points. As a result, the assigned sampling expert shall be enabled to fill in the following form (as an example see the form of Spain from round V):

Sampling for the European Social Survey-Round V

Country: Spain
NC: Mariano Torcal (mariano.torcal@upf.edu)
Sampling Design: Anna Cuxart (anna.cuxart@upf.edu) and Clara Riba (clara.riba@upf.edu)
Survey Institute: METROSCOPIA, Fernando López (flopez@metroscopia.es)
Expert: Seppo Laaksonen (Seppo.Laaksonen@Helsinki.Fi)

Reference Survey: ESS4
Date:

Target Population, Population Coverage

Persons aged 15 years and over who reside in private households in Spain, including Ceuta and Melilla.

Remarks

None

Sampling Frame

The population census in Spain is structured in census sections taken from the Continuous Census (Padrón Contínuo), which was updated in May 2010, by the Instituto Nacional de Estadística (INE, the public Statistical Office of Spain).
There are 34,600 census sections in Spain. Census section is the most elementary framing unit of eligible voters. The size of sections varies between 500 and 2,000 voters (18+ years old), being the average size of 1,300. At this point, it is necessary to made clear that although census sections are defined with regard to electoral processes, their objective is only to define the boundaries of the administrative units which, on the other hand, are used to carry out sample designs. Therefore, census sections do include all citizens registered in the municipal rolls, regardless of their voting rights.

Remarks

The frame includes all persons living in private houses, whether these being family or collective. This can make that 1,46% individuals be not included in the target population.

Sampling Design

Stratified Two-stage probability sampling.
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:

- First: population aged 15+ living in cities with more than 100,000 inhabitants
- Second: population aged 15+ living in cities between 50,001 and 100,000 inhabitants

- Third: population aged 15+ living in municipalities between 10,001 and 50,000 inhabitants
- Fourth: population aged 15+ living in municipalities with less than 10,001 inhabitants
- 64 of the 72 theoretical strata are not empty.

Stage 1: Selection of PSUs proportionally to population aged 15+.

Stage 2: Random selection of 6 or 7 individuals in each of the PSUs selected in the previous stage (7 in the two first brackets and 6 in the rest).

Remarks

In order to achieve a more accurate sample size for ESS R5, estimations about ineligible rate, response rate and design effect have been obtained as a mean of the results of the last two editions: ESS R3 and R4.
There will not be an overrepresentation of any strata in ESS R5. The analysis of ESS R3 and R4 response rates did not show important differences among the strata.

Design Effects

$$\begin{aligned} DEFF &= DEFF_c = 1.207 \\ DEFF_c &= 1 + (k-1) \cdot p = 1 + (3.973 - 1) \cdot 0.074 = 1.207 \end{aligned}$$

Remarks

A more accurate calculation of the mean intraclass correlation coefficient. A variety of 23 variables had been used (5 numerical, 10 ordinal and 8 dummy variables)

Target Response Rate

70%, although a safe estimation of 66.8% has been handled for the calculation of the sample size.

Remarks

93.9%, this figure has been calculated from the weighted average of R3 and R4 eligible rates. A higher weight has been assigned to the last one.

Sample Size

2,865
The basic values for the estimation of the gross sample size are: valid cases (93.9%), mean response rate (66.8%) and design effect (1.207).

$$\begin{aligned} n_{eff} &= 1,500 \\ n_{net} &= 1,500 \cdot 1.207 = 1,811 \\ n_{gross} &= 1,811/(0.939 \cdot 0.668) = 2,886 \end{aligned}$$

The basic values for the estimation of the gross sample size are: valid cases (93.9%), mean response rate (66.8%) and design effect (1.207).

Remarks

None

Special Features of the Design

The sample design ensures equal probability of individual selection for all the individuals in the same stratum.

Remarks

Two pre-test samples will be selected in Madrid and Barcelona with the same sampling design than the ordinary sample. In order to raise the specification of 50 completed interviews, 6 sections and 42 cases in each city will be selected.

References

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

LYBERG, L. (2000): *Measuring Adult Literacy*chap. Review of IALS – a commentary on the technical report. Office for National Statistics.

SHACKMAN, G. (2001): "Sample size and design effect," `http://www.albany.edu/~areilly/albany_asa/confweb01/abstract/Download/shackman.pdf`.

A Sampling issues in the "Specifications for participating countries" – Round 6 of the ESS

6.1 Population coverage

The survey 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 undercoverage 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.

6.2 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 7 for details on fieldwork). Where a sample frame of individuals is not available, or lacks sufficient coverage, countries may use a sample 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. 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' or 'non-contacts')**. Over-sampling of certain subgroups must be discussed and agreed in advance with the sampling panel.

Effective sample size

The minimum 'effective achieved sample size' should be 1,500, after discounting for design effects (see appendix II), or 800 in countries with populations of less than 2 million. Thus, with the help of the sampling panel, each country should determine the appropriate size of its initial issued sample by taking into account the realistic estimated impact of clustering, eligibility rates (where appropriate), over+sampling and response rate. The sampling panel will help to calculate the gross sample size required in order to achieve an effective sample size of 1,500 interviews. Note that in some cases this might require a large gross sample and needs to be considered when setting the budget for the survey. The gross sample agreed should then be reflected without change in the number of cases issued to the field.

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 to the expert panel for 'signing off' and subsequently to the CCT for 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 panel can 'sign off' a country's sample design:

- a description of the sampling frame and of the units it comprises (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 undercoverage, duplication 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 survey 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.

A sample design data file needs to be produced by each country and delivered to the CCT. 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 about this is provided by the sampling panel. **It will not be possible to include the national data in the integrated data file without the provision of the sample design data.**

The final sample design will also be fully documented by each national team in the national technical report of the survey. 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 sections 8.3 and 8.4

for information about data protection assured by the ESS data archive.

B Rules for estimating design effects

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*):

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

$$deff = \frac{n_{net}}{n_{\text{eff}}}, \quad \text{so } n_{\text{eff}} = \frac{n_{net}}{deff}$$

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 $deff = deff_p \times deff_c$. 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.

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. Another reason why differing selection probabilities might be used is if important minority groups were to be over-sampled.

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

$$deff_p = \frac{m \sum m_i w_i^2}{(\sum m_i w_i)^2}$$

where there are m_i 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.

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:

$$\mathit{deff}_c = 1 + \left(\bar{b} - 1\right) \cdot \rho,$$

where $\bar{b}$ is the mean number of respondents per cluster and ρ is the intra class 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 (see Kish, 1994, Survey Sampling, pp. 161-164 (New York: Wiley and Sons, Inc.)). 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.

Example: How to determine the size of the gross sample

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

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

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:

   $$\mathit{deff}_c = 1 + (15 - 1) \times 0.02 = 1.28.$$

   [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	$0.35 \cdot n_{net}$	1	$0.35 \cdot n_{net}$	$0.35 \cdot n_{net}$
2	0.45	$0.45 \cdot n_{net}$	2	$0.90 \cdot n_{net}$	$1.80 \cdot n_{net}$
3	0.12	$0.12 \cdot n_{net}$	3	$0.36 \cdot n_{net}$	$1.08 \cdot n_{net}$
4	0.06	$0.06 \cdot n_{net}$	4	$0.24 \cdot n_{net}$	$0.96 \cdot n_{net}$
5	0.02	$0.02 \cdot n_{net}$	5	$0.10 \cdot n_{net}$	$0.50 \cdot n_{net}$
				$1.95 \cdot n_{net}$	$4.69 \cdot n_{net}$

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

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

3. Thus, we predict $deff = 1.28 \times 1.23 = 1.57$. Consequently, to achieve n_{eff} > 1,500 with this design, we need $n_{net} > 1,500 \times 1.57 = 2,355$.

4. The final stage is to calculate the gross sample size in order to achieve around 2,355 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:

$$n_{gross} = \frac{\left(\frac{n_{net}}{0.80}\right)}{0.95} = \frac{\left(\frac{2,355}{0.80}\right)}{0.95} \approx 3,098$$

So we would select a gross sample of at least 3,100 addresses.