Sampling for the European Social Survey – Round 4: Principles and requirements

Guide

Final Version

The Sampling Expert Panel of the ESS

February 2008

Summary: The objective of work package 3 is the "design and implementation of workable and equivalent sampling strategies in all participating countries". This concept stands for 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 Co-ordinators 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 practice used in each participating country. The choice of specific design depends on the availability of frames, experience, and of course also the costs in the different countries. If adequate estimators are chosen, the resulting values are comparable. To ensure comparability, design weights have to be computed for each country. For this, the relative selection probabilities of every sample member at each stage of selection must be known and recorded in a *sample design data file*.

This comparability has to be the goal of the sampling strategy and its implementation for the ESS.

2 Discussion of standards set in the Specification for participating countries

Only probability samples provide a theoretical basis which allows us to infer from the sample to the whole target population or sub-sets. 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 (completely responded units) in participating countries (ESS: 1,500 or 800 where population is smaller than two million inhabitants).

These requirements can only be sensibly discussed in the context of probability 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 residential 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, re-

gardless of nationality and citizenship or language[1]. This definition applies to all participating countries and thus every person with the defined characteristics should 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. However, the quality of the frames – e.g. coverage, updating and access – may differ from country to country. Therefore, frames have to be evaluated carefully. 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, Finland

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

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

Drawing a sample is more complicated if no registers (lists) are available (group c). In this instance multi-stage designs are usually applied, in which the selection of municipalities forms the first stage and the selection of households within these municipalities the second stage. Because no sampling frames are available, the crucial problem is the selection of households. There are mainly two ways to go about this. The first is to list all addresses within a certain district of each selected community. The target households are then drawn from these lists. Arguably, it is possible to assess this procedure as one way of drawing a probability sample, even if one which is fairly strongly clustered. Another frequently used way to find target households is the application of random route elements. The question here, however, is the extent to which random routes can be judged to be "strictly random". In Lyberg's view these techniques do result in non-probability samples (see evaluation of the IALS DATE). At the very least, the following questions have to be answered

- 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?

[1] 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.

in order to minimise the interviewer's influence on the selection of respondents. An acceptable method might involve the interviewer doing the complete walk, recording the sampled addresses and transferring them to the survey office before he/she begins contacting any addresses.

Even in countries where reliable frames exist, we have to expect pitfalls in the sampling process. For example, it will be difficult to fully cover people temporarily working abroad. Such systematic losses because of undercoverage cannot be ruled out in practice. However, they must be documented carefully.

2.2 Response rates

Non-response is a second major issue for the representativeness of the target population in the sample. A carefully drawn gross sample from a perfect frame can be worthless if non-contacts and refusals lead to systematic biases. Therefore, it is of essential importance to plan and implement a sufficient number of contacts as well as appropriate field work strategies for the persuasion of the target persons to participate in the survey. However, 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 (see Technical reports of round 1, 2 and 3). Nevertheless, all efforts should be made to avoid non-response because it increases the danger of biased samples, and cell weighting is not such a global mean of "repairing" samples, as some authors argue (Häder and Gabler, 1997).

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 multi-stage stratified and clustered sampling was used in rounds 1, 2 and 3 of the ESS and can also be expected to be used in round 4. For determining the sample sizes, the design effects of the respective country have to be considered to ensure the comparability of estimates with respect to their confidence intervals. The design effect is defined as ratio of the variance of a variable under the actual sampling 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). It is essential that National Coordinators and the fieldwork organizations analyse the data from round 1, 2 and 3 to calculate appropriate intraclass correlation coefficients for the sample designs used in their countries. The cluster size of the selection units also influences the design effect. It should be chosen as small as possible because: The larger the average cluster sizes are, the lower the effective sample size is and the more interviews have to be conducted to reach the minimum size of 1,500. In that sense, a large number of primary selection units should be the goal with only a few interviews in each.

Another important effect is that of departures from equal probability selection methods, which requires design weighting to correct for different inclusion probabilities. In particular, in countries where the only frames available are of households or addresses, design effects will be larger than in countries where frames of persons are available. This fact also has to be taken into account when computing the sample sizes.

3 Summary

Comparability of sampling means that the national surveys must provide minimal biased estimates of comparable precision. These national samples must fully cover equivalent populations (target population). The basic requirement to use probability samples together with the additional requirements discussed in this paper leads theoretically to comparable estimates. However, in the end data quality depends also on the implementation process, e.g. the practical applications. Therefore, this process has to be guided and monitored carefully.

4 Handling of the Workpackage

In round 1, 2 and 3 we worked with an expert panel on sampling. This panel will continue its work. Members are the following sampling specialists:

- Sabine Häder (gesis-zuma, Germany, sabine.haeder@gesis.org)
- Siegfried Gabler (gesis-zuma, Germany, siegfried.gabler@gesis.org)
- Matthias Ganninger (gesis-zuma, Germany, matthias.ganninger@gesis.org)

- Seppo Laaksonen (University of Helsinki, Finland, Seppo.Laaksonen@helsinki.fi)
- Peter Lynn (University of Essex, U.K., plynn@essex.ac.uk)

Each of the experts will be assigned about five 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 join 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.

5 Information required 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

- If lists are to be used, how, if at all, are they to be stratified?
- Is the design single- or multi-stage?
- Which stages are defined?
- How much clustering is proposed?

Sample size

- How has the effective sample size been calculated, including estimates of response rates and design effects due to clustering or necessary weighting?[2]
- Will any population subgroups be over-sampled?
- What steps will be taken to achieve the target response rate?

The National Coordinators are responsible for asking the survey organisations about these points. Once these have been clarified the assigned sampling expert will then be asked to fill in the following form (as an example see the form of the Netherlands from round 3):

[2]For the computation see Appendix 1 of the Specification for participating countries of the ESS.

Sampling for the European Social Survey- Round III

Country:	**The Netherlands**
NC:	Harry B.G. Ganzeboom (h.ganzeboom@hetnet.nl)
Survey Institute:	GfK
Expert:	Siegfried Gabler (gabler@zuma-mannheim.de)
Reference Survey:	ESS Round I+II
Date	13 September 2006

Target Population, Population coverage	Persons aged 15 years and over (no upper age limit) resident in private households in the Netherlands.
Remark	Intra-murals (209,000 = 1.3% of the population of N=16,192,572; Source: Central Bureau of Statistics, Jan 1st 2003) and the sailing and trucking persons are excluded from the gross sample.
Sampling frame	Frame of addresses: Postal delivery points from 'TPG-Afgiftenpuntenbestand' provided by the Dutch Postal Service.
Remark	Most up-to-date and most exhaustive source of postal delivery points out of which P.O. boxes, companies, amusement parks etc. are removed.
Sampling design	Unstratified two-stage probability sampling: • Primary sampling units: Postal delivery points (excluding P.O. boxes and business addresses) selected with equal probability. • Secondary sampling units: Person within a household (using the Last-Birthday-Method: birthday refers to random data on contact form); If more than one household belongs to a postal delivery point up to 5 households are added to the gross sample and within each household one person is selected. For postal delivery points with more than 5 households, 5 households are randomly selected using the Kish table and within each of these households one person is selected.
Remark	Under-representation of persons living in larger households.
Design effects	The sampling design does not contain any clustering. The design effect is only due to differing selection probabilities. $DEFF_C = 1$; $DEFF_P = 1.20$ DEFF = 1.20
Remark	(Round I: $DEFF_P = 1.19$) Round II: $DEFF_P = 1.20$
Target response rate	70%
Remark	(Round I: Response rate = 67.9%) Round II: Response rate = 64.5%

Sample size	Gross sample size = 3,000 Net sample size = (Gross sample size – 3 % ineligible) × Target response rate = 3,000*0.97*0.65 = 1,892 Effective sample size = Net sample size / DEFF = 1,892/1.2 = 1,580
Remark	(Round I: ineligibility rate = 2.4%) Round II: ineligibility rate = 3.0% Claimed effective sample size of 1,500 exceeded

References

HÄDER, S., AND S. GABLER (1997): *Eurobarometer - Measurements for opinions in Europe*chap. Deviations from the population and optimal weights, no. 2 in ZUMA-Nachrichten Spezial. ZUMA.

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 4 of the ESS

5 Specification for the Survey

Sampling (see 5.1 to 5.4)

- Full coverage of the residential population
- Use of strict random methods at all stages
- Minimum 'effective' sample size of 1,500 (or 800 where population is under 2 million) (i.e. to obtain an effective sample size of 1,500, the actual number of interviews will be greater than this)
- Target minimum response rate of 70% and a target maximum non-contact rate of 3%

5.1 *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[1]. 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.

5.2 *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 5.12). Where a sample frame of individuals is not available, 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.

5.3 *Effective sample size*

The **minimum 'effective achieved sample size' should be 1,500**, after discounting for design effects (see Appendix 1), 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 actual gross achieved sample size required in order to achieve an effective sample size of 1,500 interviews.

[1] Please note that questionnaires are to be available in all languages spoken as a first language by 5 per cent or more of the population and interviewers must be available to administer them (see 5.12). For speakers of certain minority languages (spoken by fewer than 5 per cent of the population), however, it may be possible to adapt the questionnaire produced by another participating country. If National Coordinators wish to offer translated questionnaires to these smaller minority language groups, they should refer to the CCT for advice. Countries are not, however, required to interview language minorities under the 5% cut-off and must never allow interviewers to perform their own 'oral' translations for this purpose.

5.4 *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 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 under-coverage, 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.

The final sample design will be fully documented by each national team in the national technical report of the survey. Furthermore, a sample design data file has to be produced by each country and then delivered to the sampling panel. 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.

The final sample design will 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.

B Rules for estimating design effects

Effective Sample Size

The effective sample size (neff) is the size of a simple random sample which would produce the same precision (standard errors) as the design actually used. Typically, neff is less than the actual number of achieved interviews, m, 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 = Actual sampling variance / Sampling variance with simple random samples of same size;

DEFF = m/neff, so neff = m/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 neff. 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 initial sample (n) required in order to achieve approximately m 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

$$DEFF_p = \frac{m(\sum_i m_i w_i^2)}{(\sum_i m_i w_i)^2}$$

where there are m_i respondents in the i^{th} selection probability class, each receiving a weight of w_i, where
where α means 'proportional to'

$$w_i \alpha \frac{N_i}{m_i}$$

(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:

$DEFF_c = 1 + (b\text{-}1)\ \rho$

where b is the mean number of respondents per cluster and ρ is the intra-cluster 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 issued sample

We have prescribed neff > 1500.

To determine m, we must first estimate $DEFF = DEFF_p \times DEFF_c$

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 (adjusted if necessary to allow for multiple households at some addresses) to estimate the number of respondents in each selection probability class, thus:

No. of persons aged 15+ at address i	Proportion of addresses in population 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	0.35m	1	0.35m	0.35m
2	0.45	0.45m	2	0.90m	1.80m
3	0.12	0.12m	3	0.36m	1.08m
4	0.06	0.06m	4	0.24m	0.96m
5+	0.02	0.02m	5	0.10m	0.50m
				1.95m	4.69m

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 $DEFF_p$:

$$DEFF_p = m \times 4.69m/(1.95m)^2 = 4.69/1.95^2 = 1.23$$

3. Thus, we predict DEFF = 1.28 x 1.23 = 1.57. Consequently, to achieve neff > 1,500 with this design, we would need m > 1,500 x 1.57 = 2,355.

4. The final stage is to calculate the sample size to select initially in order to be likely to achieve around 2,355 interviews. Suppose we anticipate a response rate of 80% and that 5% of the sampling frame units will be ineligible (e.g. addresses which do not contain a resident household), then:

n = (m / 0.80) / 0.95 = 3,098

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