Sampling designs and weighting of the European Social Survey (ESS)

Laaksonen, Seppo

University of Helsinki

Seppo.Laaksonen@Helsinki.Fi

Based much on the material

of European Social Survey Sampling Expert Panel

Responsible institute Gesis (earlier ZUMA), Mannheim; Germany

Chaired by Sabine Häder (2001-2007), Matthias Ganninger (2007-2013) and Stefan Zins (2013+).

The current ordinary members: Siegfried Gabler (Gesis), Sabine Häder (Gesis), Seppo Laaksonen (Helsinki University), Peter Lynn (Essex University)

Content

- References
- Motivation for good and equivalent sampling designs in multinational surveys
- Some ESS sampling principles
- Design effects due to clustering and unequal inclusions
- Effective sample size: minimum 1500 or 800
- Worsening nonresponse and possible anticipation for using this in design
- Sampling design data file, not enough good in any country
- Design weights and post-stratified weights since 2014
- Conclusions and future and problems: mixed-mode, reserve sample

Some references

ESS _ European Social Survey (2014). Documentation of ESS Post-Stratification Weights. http://www.europeansocialsurvey.org/docs/methodology/ESS_post_stratification_weights_documentation.pdf

Holt. D. & Smith, T.M.F. (1979). Post stratification. *Journal of Royal Statistical Society A*, 142, 43-46.

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

Gabler, S., Häder, S. and Lynn, P. (2006): Design Effects for Multiple Design Samples. *Survey Methodology*, 32 (1), pp. 151–120.

Kish, L. (1994). Multipopulation Survey Designs. *International Statistical Review* 62, 167-186.

Kish L. (1995). Methods for design effects. *Journal of Official Statistics* 11, 55-77.

Laaksonen, S. (2015). Sampling design data file. *Survey Statistician* (International Association of Survey Statisticians) 72, 61-66.

Laaksonen, S., Kemppainen, T., Stjernberg, M., Kortteinen, M., Vaattovaara, M. and Lönnqvist, H. (2015). Tackling City-Regional Dynamics in a Survey Using Grid Sampling. *Survey. Research Methods by the European Survey Research Association, Vol 9, No 1, 55-65.* www.surveyresearch.org

Laaksonen, S. & Heiskanen, M. (2014). Comparison of Three Modes for a Crime Victimization Survey, *Journal of Survey Statistics and Methodology 2 (4): 459-483*

It is needed a new version for this that is not up-to-date although still cited:

Lynn, P. & Gabler, S. & Häder, S. & Laaksonen, S. (2007). Methods for Achieving Equivalence of Samples in Cross-National Surveys. *Journal of Official Statistics*, 27, 1, 107-124.

Schouten, B., M. Calinescu, and A. Luiten (2013), "Optimizing Quality of Response Through Adaptive Survey Designs," *Survey Methodology* 39(1), 29-58.

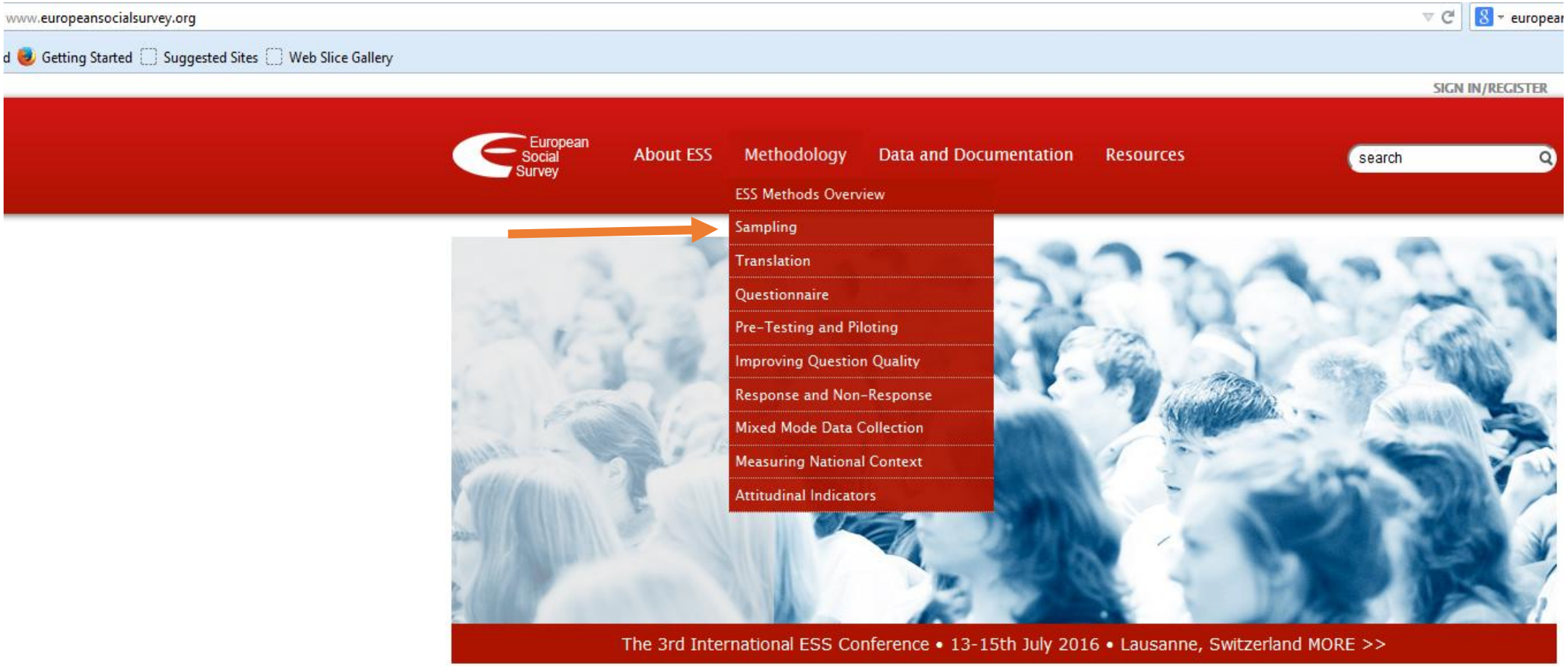
The European Social Survey (ESS) is an academically driven cross-national survey that has been conducted every two years across Europe since 2001. Europe here covers e.g. Israel, Iceland, Turkey and Russia.

Until now 36 countries have participated, for one round at least.

The data of six first rounds are ready to be used by everyone.

The next data release is expected to be late October 2015.

Everything important is public and found from the website. Here you see something including Sampling. Naturally, all data are free and can be downloaded by everyone. The only question is how to use it well. This is not easy without good knowledge in survey methodology.



ESS Sampling, Questionnaire Designing, Pre-Testing, Fieldwork with Data Collection, Reduction of Bias and the Reliability of Questions are input harmonized, that is, there are the rules for each and before going to implement them, the plan should be approved or signed off. This signing-off in the case of sampling is made by the Sampling Expert Panel (SEP). We could be more demanding than have been. Very often happens that a country team has promised to do something better in a next round but this has not been the case, unfortunately. Note that Eurostat's surveys are output harmonized.

So, there are many gaps in data of all countries, but the quality of the ESS data is still best in Europe as far as the entire population surveys is concerned. The target population covers all residents of a country who are 15 years old or more. This is not usual, since the age limit is often 64, 69, 74 or 79

The focus on this presentation is in gaps and problems relating to sampling and weighting. It is always hoped that they will be solved in future but since the survey climate is coming more and more difficult, many countries do not try to improve anything but to maintain a basic = minimum quality.

Administration at country level:

The responsible person = National Coordinator (NC), from a academic institution.
- He/she is not in most cases any expert in sampling and hence a subcontractor (either a market survey institute or a statistical office) is responsible for this and fieldwork as well.

Large-scale, cross-national surveys such as the International Social Survey Programme, the European Value Study or the Eurobarometer have become more and more common in the last decades. However, too little attention was given to the topic of sample design for cross-national surveys. But in this century, the situation has changed considerably. Researchers (how is Eurostat?) seem to have realised that it is very important for multinational surveys to select an equivalent sample in each country because lack of equivalence in the samples can undermine the central objective of cross-national comparison.

The European Social Survey was the first survey programme where a whole work-package was dedicated to the “design and implementation of workable and equivalent sampling strategies in all participating countries”.

To explain the basic principles for sampling in cross-cultural surveys Kish provided a starting point: “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.” (Kish 1994) Following this, an optimal sampling design for cross-cultural surveys should consist of the best probability sampling practice used in each participating country.

The key requirements in the Specification for Participating countries:

- Samples must 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
- Individuals are selected by strict random probability methods at every stage
- Sampling frames of individuals, households and addresses may be used
- All countries must aim for a minimum 'effective achieved sample size' of 1,500 or 800 in countries with ESS populations of less than 2 million after discounting for design effects
- Quota sampling is not permitted at any stage
- Substitution of non-responding households or individuals (whether 'refusals', 'non-contacts' or 'ineligibles') is not permitted at any stage

Coverage

In round 7 the following kinds of frames are found:

(i) countries with reliable lists of residents that are available for social research such as Estonia (not well up-to-date in any register country),
(ii) countries with reliable lists of households that are available for social research such as the Netherlands,
(iii) countries with reliable lists of addresses that are available for social research such as Czech Republic,
(iv) countries without reliable and/or available lists such as Portugal.

In all other cases than in (i), there are in fact more than one sampling frame. It often starts from an areal cluster that can be municipality, electoral area or census district.

Some applied selection methods

(i) One stage countries with reliable lists of residents.

- Implicit stratification e.g. in Finland and Sweden in each round (used like srs)
- Proper srs, e.g. in Denmark in Round 7

(ii) Two stage countries with reliable lists of households:

- 1st stage pps, 2nd stage srs using register, e.g. Spain

(iii) Three stage countries with reliable lists of addresses:

- 1st stage pps, 2nd srs, 3rd = individuals e.g. Kish-Grid, (random) last birthday method

(iv) Countries without reliable and/or available lists:

- Locality psu using last census e.g. and pps selection, then more locally collected data even telephone catalogues and GPS, the last stage by last birthday method.

Special sampling practices

In most countries, one design is used for the whole country but there are exceptions that we call two-domain design:

Poland has used from a beginning such a design that

- Simple random sampling has been used in big cities where face-to-face interviewers dot not need to travel far
- Two-stage cluster sampling in rural or less populated areas.

Norway changed the design two times (two-stage cluster, SRS) but in rounds 6 an 7 they have basically the same two-domain design as Poland. The sample allocation is proportional at gross sample level while in Poland it aims at getting equal proportions of respondents at both domains.

I have suggested a two-domain design to be used in f2f surveys also in Finland and Sweden.

A key target thus is to achieve the minimum effective sample size either n_{eff} =1500 or n_{eff} = 800. For this purpose we use design effects that indicate the precision of estimates.

The first design effect that has to be predicted in the design stage is the design effect due to differing selection probabilities (the question about nonresponse not solved).

The second kind of design effects to be predicted is the design effect due to clustering since the cluster size of the selection units and the intra-class correlation also influence the design effect.

To determine the effective sample size of a sampling design its so called design effect (Deff) has to be specified. The design effect quantifies the loss or gain in efficiency (e.g. the sampling variance) of a sampling design in comparison to a SRS. It is always defined with respect to a strategy composed of a sampling design p and an estimator $\hat{\theta}$.

Under the model-based approach Deff$(p, \overline{y}_w)$ is the product of two components, Deff_c, design effect attributed to clustering and Deff_p, the design effect due to unequal survey weights. Thus we have

$$\text{Deff}(p, \overline{y}_w) = \text{Deff}_p \times \text{Deff}_c$$

$$\text{Deff}_p = n\frac{\sum_{j \in s} w_j^2}{(\sum_{j \in s} w_j)^2}$$

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

Here
b^*=net average cluster size
ρ=rho=intra-class correlation

For the estimation of the intra-class correlation (RHO) an ANOVA estimator is used of the following form

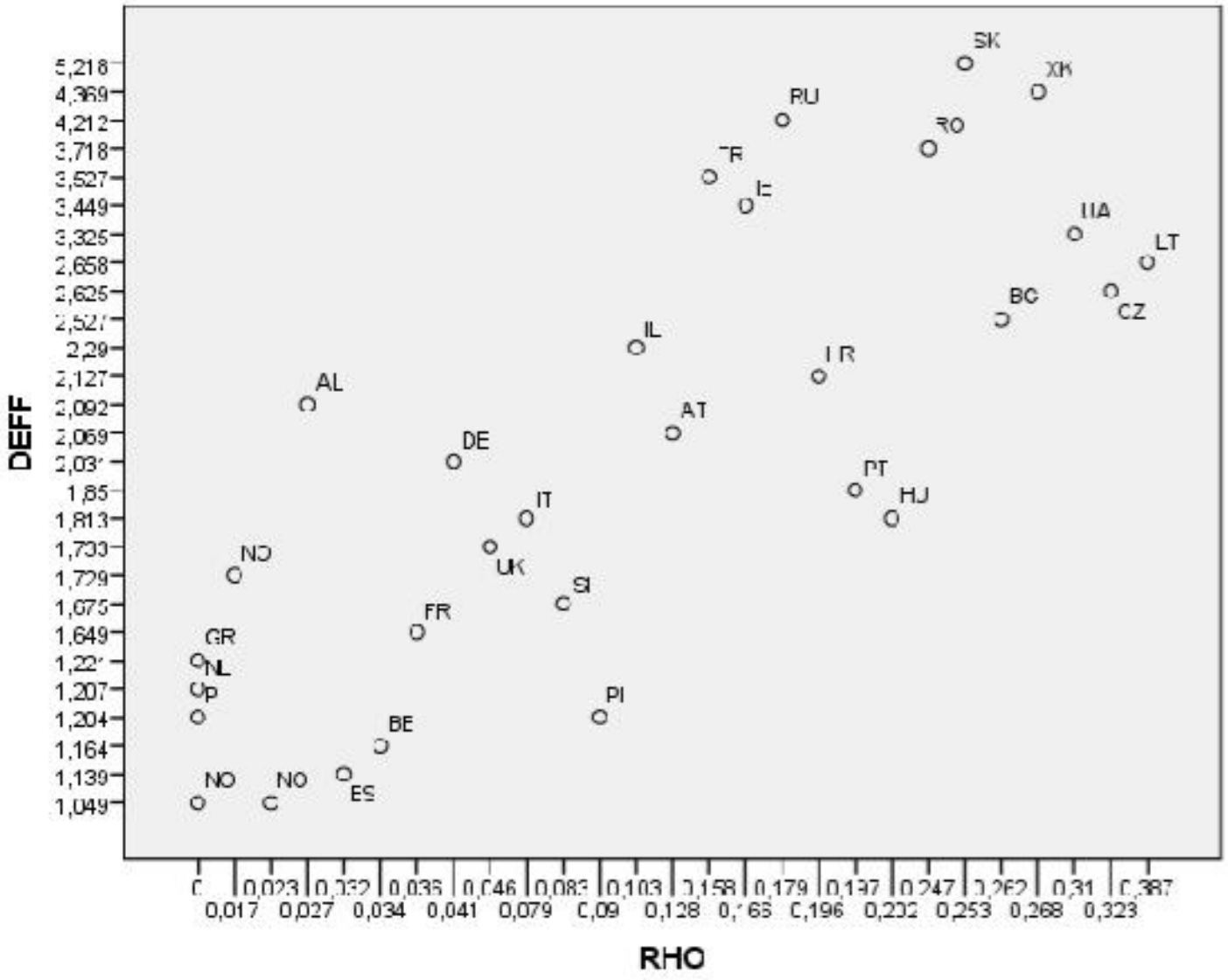
$$\hat{\rho} = \frac{MSB - MSW}{MSB + (K-1)MSW}$$

$$K = (M-1)^{-1}\left(n - \sum_{i=1}^{M} \frac{n_i^2}{n}\right), \text{M} = \text{the number of clusters}$$

The RHO has a big role for the complete DEFF. This formula can give even negative values If MSB<MSW, although the simplest solution does not give. We here do not go to all details but we see e.g. that the DEFFc is lower if the net average cluster size is smaller. It means that the RHO is declining by the response rate (not satisfactory). On the other hand, the RHO decreases when the number of the clusters increases. We have no requirement for the minimum cluster counts as in the PISA where it is 150. This number is below this only in a few countries.

Sampling design based design effects and intra-class correlations (RHO) by country in the round when participated most recently; if the design has been changed, a second round is included. The results are based on the median of the 38 ordered scaled variables that are mostly used in survey analysis. The countries are sorted by their DEFF's. NA = not available.

COUNTRY	ROUND	RHO	DEFF	
IS	6	-0,007	1,000	
CH	6	NA	1,000	
DK	6	NA	1,000	
EE	6	NA	1,000	
FI	6	NA	1,000	
SE	6	NA	1,000	
EE	4	NA	1,024	
NO	6	0,000	1,049	SRS Domain
NO	6	0,023	1,049	Cluster Domain
ES	6	0,032	1,139	
LU	2	NA	1,152	
BE	6	0,034	1,164	
CY	6	NA	1,201	
PL	6	0,000	1,204	SRS Domain
PL	6	0,090	1,204	Cluster Domain
NL	6	0,000	1,207	
GR	5	0,000	1,221	
CH	4	NA	1,240	
LV	4	NA	1,311	
FR	6	0,036	1,649	
SI	6	0,083	1,675	
NO	1	0,017	1,729	
UK	6	0,046	1,733	
HU	6	0,202	1,813	
IT	6	0,079	1,813	
PT	6	0,197	1,850	
DE	6	0,041	2,031	
AT	3	0,128	2,069	
AL	6	0,027	2,092	
HR	5	0,196	2,127	
IL	6	0,103	2,290	
BG	6	0,262	2,527	
CZ	6	0,323	2,625	
LT	6	0,387	2,658	
UA	6	0,310	3,325	
IE	6	0,166	3,449	
TR	4	0,158	3,527	
RO	4	0,247	3,718	
RU	6	0,179	4,212	
XK	6	0,268	4,369	
SK	6	0,253	5,218	



Scatter plot:
RHO versus DEFF;
the countries with NA not included

Summary of the required anticipations (predictions)

The basic idea is to take advantage of the most recent similar surveys that are often the latest ESS surveys. Hence the sampling team chair has computed such figures to country persons unless they cannot do it themselves.
We thus anticipate:

- The intra-class correlation
- The response rate that should as realistic as possible
- The ineligibility rate is often difficult. Unfortunate, they are often unbelievably low.
- Using the three previous points we can anticipate DEFFc
- DEFFp is anticipated respectively, this depends on the sampling design.
  - It should be basically easy in two-stage design without the anticipation of the regional or other response rate
  - The same question is concerned all other designs
  - We can very easily anticipate the DEFFp of the selection of an individual of the household or the address (about 1.2-1.25)

Due to worsening financial problems the anticipation figures seem to be too optimistic and the effective sample size is not achieved.

Sampling Design Data File

The creation of the sampling design data file (SDDF) is the last phase in sampling designing. We have developed a template for this purpose. The newest version should have been used since round 6. This template includes the detailed description about this gross sample file. The file should contain one record for each selected address, in the case of address-based samples, or for each selected individual, in the case of individual-based (population register) samples.

In addition to the specification of the statistical units the SDDF template includes fairly detailed requirements for the variables and their meta data. We here summarize these variables into five groups. Unfortunately, the quality thus is not good, and even meta data are not informative in all cases.

(i) Inclusion probabilities of each stage.
(i) Other variables directly relating to sampling design (psu=primary sampling unit that can be a cluster or individual, explicit stratum, implicit stratum)
(ii) Outcome of the survey fieldwork (respondent, ineligible, non-respondent)
(iii) Macro auxiliary variables, statistics for the target population level (cluster psu's, explicit strata, calibration margins) Incl. 1-2 variables in two countries
(iv) Micro auxiliary variables for individuals and their groups, e.g. gender, age, education level, regional or areal codes, language, ethnic or other background, household member data incl. children, civil status, employment status, register income, etc.). A general quality is fairly poor but varying over rounds to some extent.

Example: UK three-stage sampling design of the round 5 of the ESS
NO STRATA

	Sample stage	Sample size	Per cent
Inclusion probability at first stage of sampling (selection of PSUs)	1st stage	4640	100
Inclusion probability at second stage of sampling (selection of addresses)	2nd stage	4640	100
Inclusion of selection of dwelling unit at sampled address	2nd stage	4070	87.7
Inclusion of selection of household at selected dwelling unit	2nd stage	3835	82.7
Probability of selection of adult at selected household	3rd stage	3460	74.6
Unit response	Fieldwork	2366	51.0

Factually, there are five stages in the sampling data file. This is a usual case in countries that they can calculate missingness with a stage.

Example: Three-stage sampling for Russia with 10 regional explicit strata

	Sample stage	Sample size	Per cent
Inclusion probability at first stage of sampling (selection of PSUs)	1st stage	3982	100
Inclusion probability of household	2nd stage	3982	100
Inclusion probability of adult at selected household	3rd stage	2595	65.2
Unit response	Fieldwork	2595	65.2

You see that they did not try to get any detailed data on non-respondent households that is quite common outside register countries

As always, the response rates vary and this is good to take into account in sample allocation as was made in Russia. Today, too many countries do not take care of this problem but hopefully in a future.

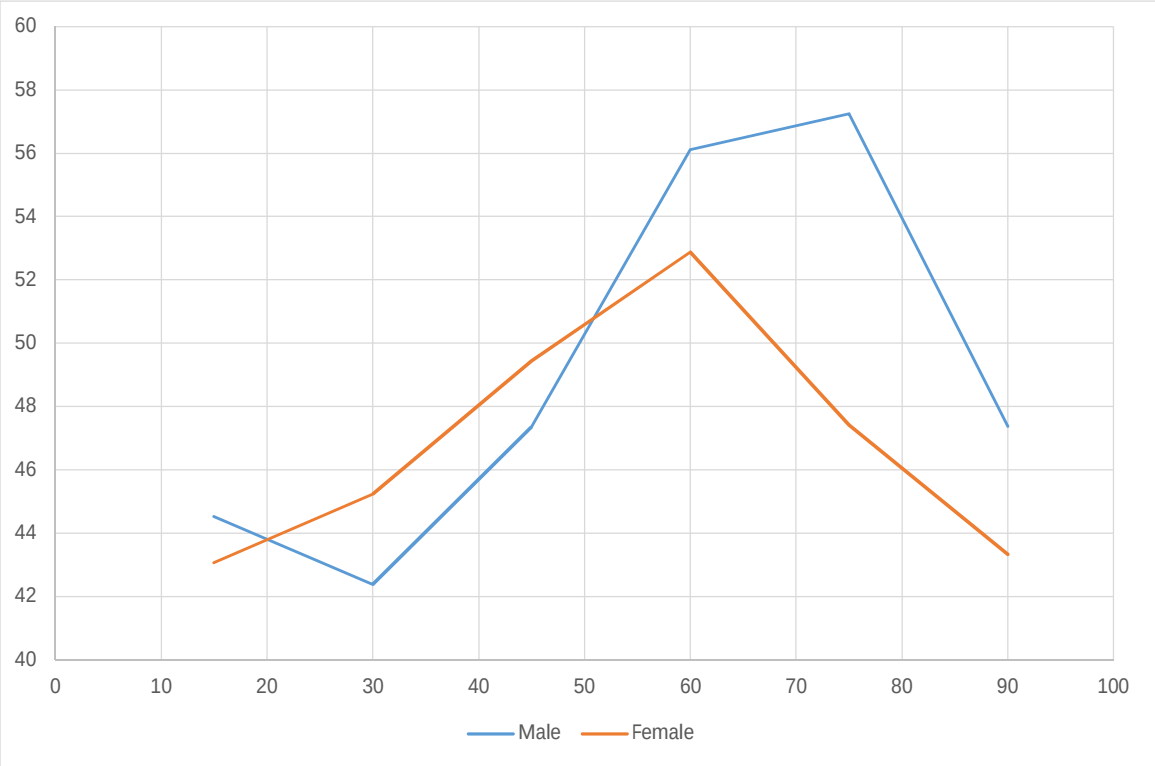
Response rate in Russia by strata, the 5th round of the ESS		
REGION OF SETTLEMENT = Explicit Stratum	Gross sample size	Per cent
1	451	51.7
2	989	57.0
3	190	81.1
4	199	74.4
5	436	71.1
6	439	78.6
7	491	72.5
8	396	62.6
9	222	58.6
10	169	63.3

The above minimum requirement of the SDDF thus is fulfilled in each country but the other the two variable groups are not well used since they have not been well available in most countries. However, there should be any reasons not to include macro variables (areal, regional, psu level). We here mention the two main purpose for the SDDF.

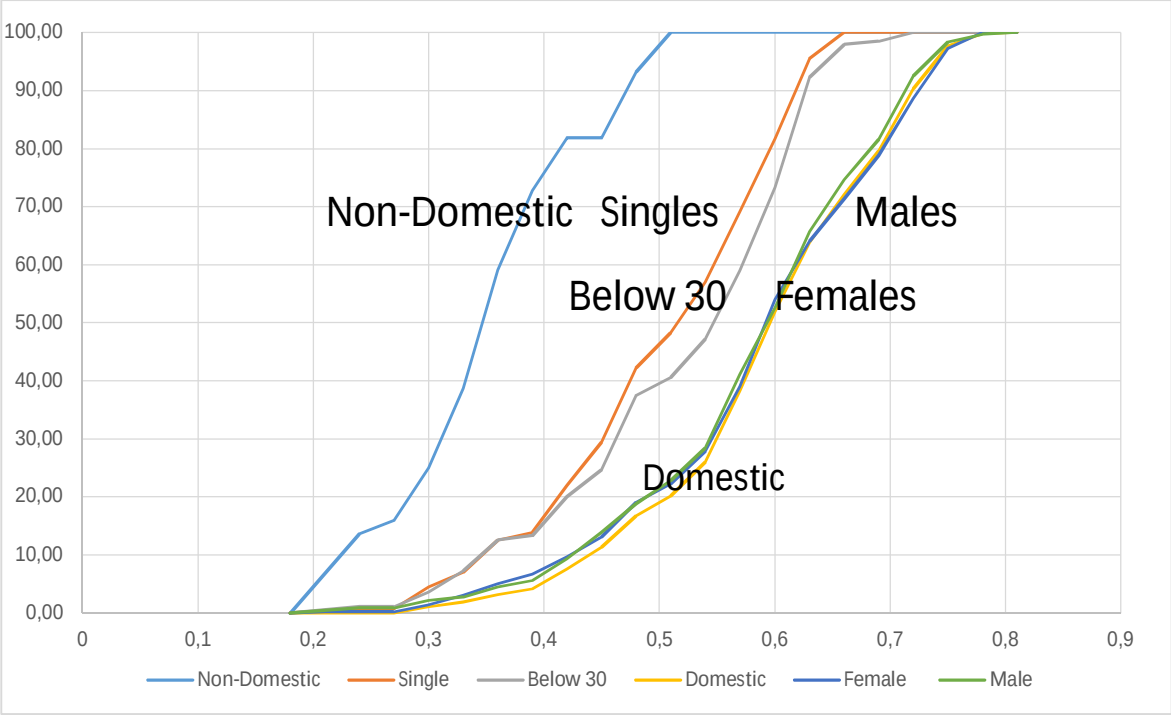
(i) Quality analysis, especially concerning unit nonresponse and ineligibility.
(ii) Nonresponse weighting.

We give some examples about both of them.

The country that improved the quality in Round 7: Sweden. No micro variables earlier, now the two, gender and age group. Now we get the following response rates by their interaction. (age = the lower limit of 15 years old groups).



The country with a good micro auxiliary variables in Round 3 but later nothing: Denmark. The response propensities cumulatively for some population groups.



SDDF in Banocoss Countries

Countries, I am now coordinating

- Norway: varying and fairly good in rounds 1 to 6, but no auxiliary variables in Round 7 since a new sampling expert did not like to include them even though it is fairly easy Norway
- Estonia: Basic quality, best when Statistics Estonia was survey organization but it has been difficult to get more auxiliary variables

Other countries

- Denmark: best ever in Round 3 when I was coordinating, later thus nothing; a good point is that they can now get all people to participate (earlier opt-out was too easy, being 12-13 %)
- Finland: some micro auxiliary variables, no macro. I have asked to improve but nothing happened.
- Sweden: now first time two auxiliary variables but they can easily get more.

Design weights only until 2014

These are first calculated for a gross sample and then converted to the respondent assuming *the missingness is ignorable*. Hence e.g. these weights are equal to one in the countries using simple random sampling (Finland, Sweden etc.).
Naturally these weights are not at all satisfactory. Fortunately, better weights are available for each round now, they are called post-stratified weights although they are not exactly such weights.

Truncation (robusting) in weights?

In order to avoid too high weights, the largest ones are truncated around the value of 4. This is a subjective choice. It is not well argued for two reasons.

- One point is that if the weight is too high, the reason should be checked and the procedure should be improved if not correct.
- Secondly, the low small weights should also be checked and truncated unless any specific good reason found.

If the weights are truncated, this should be done symmetrically, that is, the minimum weight would be the inverse of the maximum = ¼= 0.25. But there are weights below this limit.

The ESS public data thus include the two weights since 2014, design weight and post-stratified weight, both so that the average is = 1 in each country (called analysis weight or relative weight). Their sum = the number of the respondents.

There exists the third weight as well, called population weight. If the ordinary sampling weight is needed, it can be easily calculated as follows:

$$w_k = 10000\, d_k w_P$$

Here d_k refers either to the design weight or to the post-stratified weight and w_P to the population weight. The target population size is not taken from the sampling design file but from the EU statistics from the survey period. The same figure could also be in the SDDF, as the sum of the inverses of the final inclusion probabilities but this is not the case. These two date bases thus are not consistent but this imbalance is not dramatic.

For example, the Spanish SDDF gives the 15+ population = 42.5 million but the above calculation 3 million less, 39.2 million. I do not know the reason for this difference.

Post-stratified weights (or raking-ratio weights)

are obtained by adjusting the design weights in such a way that they will replicate the distribution of the so called control data. As control data the two marginal population distribution have been used, one for the cross-classification of age (3 age classes plus one missing class), gender (2 plus one), and education (4 plus one), GAE (gender x age x education) and the second one for the variable region. The same variables should be in the survey data with the same codes. There are some differences between rounds when handling education.

The population distributions for those adjusting variables are obtained from the European Union Labour Force Survey (LFS) data. Its advantage is that the quality is fairly equal but how good it is, it is not clear.

Nevertheless, these weights are best possible and they should have been used.

Post-stratified weights (or raking-ratio weights)

The table of next page is illustrative:

- We can wonder how it is possible that the ratio in two countries (Slovakia and Luxembourg) is below 1. Maybe something strange in sampling or in reweights or both,
- Some ratios are close to one (Iceland, Poland, Russia, Lithuania). One reason at least in Poland and Russia is that they are used anticipated response rates in sample allocation by region.
- The ratio is rather high in several countries including Italy, Finland, Germany and Denmark. Until now the ESS has not been allowed to allocate sample by e.g. regional response rates. If this would be the case, these countries should change the design essentially. This will be discussed when going toward Round 8. The strange thing in my opinion and some others but by all is that the proportional gross sample allocation leads to higher standard errors and thus not being good. An interesting point is that Norway's ratio is 1.09 apart from two-domain design but the allocation is about proportional since they follow this strange recommendation.

Design effect due to the two weights by country, sorted by their ratio

Country	Round	DEFF_post_strati-fication	DEFF_Design	Ratio
SK	6	1,52	1,92	0,79
LU	2	1,44	1,52	0,95
IS	6	1,04	1,03	1
PL	6	1,03	1,02	1,01
RU	6	1,35	1,34	1,01
LT	6	1,36	1,33	1,02
GB	6	1,32	1,27	1,04
HU	6	1,05	1	1,04
IE	6	1,28	1,24	1,04
EE	6	1,05	1	1,05
UA	6	1,29	1,23	1,05
CH	6	1,06	1	1,06
SI	6	1,06	1	1,06
GR	5	1,31	1,22	1,07
NL	6	1,29	1,21	1,07
HR	5	1,34	1,24	1,08
IL	6	1,12	1,04	1,08
ES	6	1,1	1,01	1,09
FR	6	1,4	1,28	1,09
NO	6	1,09	1	1,09
AT	3	1,38	1,26	1,1
BE	6	1,1	1	1,1
CZ	6	1,39	1,27	1,1
TR	4	1,36	1,24	1,1
SE	6	1,12	1	1,12
CY	6	1,36	1,2	1,13
PT	6	1,55	1,38	1,13
BG	6	1,35	1,17	1,15
DK	6	1,17	1	1,17
DE	6	1,32	1,11	1,19
FI	6	1,2	1	1,2
IT	6	1,3	1	1,3

CONCLUSION AND SOME FUTURE VIEWS

The ESS should be based on high-quality, comparable, probability samples in each country. It was also recognized that to achieve this would be a significant challenge and would require a well-resourced central infrastructure of a kind not previously implemented on any cross-national survey. Initial experiences were broadly favorable.

Our experience is that sample designs have indeed improved over the rounds in some countries. However, there are also countries in which the quality of the sample design has reduced over time. In some cases, this has occurred because of a change in the national team commissioned to be responsible for the survey and/or a change in fieldwork agency. Such changes can be associated with a change in access to sampling frames, or simply with different institutional practices or different understanding of the survey requirements.

We have encountered cases in which a country has changed the sample design in a negative way, perhaps due to a need to reduce costs, or due to internal reorganization of fieldwork operations. These changes could involve introducing a clustered design, increasing cluster sample sizes, reducing intervals in systematic sampling, or using a master sample approach, for example. We suggest that all participating countries always require close attention.

Though not part of sample design, maintaining good response rates has also proved difficult for the ESS. This has had implications for the objective of achieving a certain minimum effective sample size, as this depends not only on the sample design parameters but also on the achieved sample size.

ESS – like many surveys – has experienced substantial variations in response rates between regions at sub-national level. Some countries have responded to this by introducing into the survey design what can be thought of as an adaptive design feature. This involves applying a different sampling fraction in different geographical strata, inversely proportional to the anticipated response rate, such that the responding sample, rather than the selected sample, should approximate the population distribution over strata.

This practice was consistent with the ESS sampling specification but was banned by the ESS Central Scientific Team prior to round 7, reflecting concerns that it could lead to field agencies making less effort than they might in the urban areas. I hope that this decision will be changed.

Also in the vein of adaptive or responsive designs is the idea of selecting a "reserve" sample, to be deployed only in the eventuality of substantially lower-than-anticipated response rates. This has been done in just two countries (Germany and France in Round 7), despite some resistance from the SEP. The arguments for and against this practice have close parallels with those for the variable sampling fraction approach discussed in the previous paragraph. It is a design feature that was not on the agenda when the ESS began but has been introduced at the request of a survey agency and has consequently generated considerable debate amongst the central teams of the ESS. This strategy is basically equal to the previous strategy but not as well controlled.

As far as the sampling design file in Round 7 was concerned, it is not possible to know in the case of Germany this reserve sample has been used. Instead, the French file is more precise as the figure on next page shows.

SDDF FILE RESULT, FRANCE ROUND 7

Respondent

Non-respondent

In-eligible

Unknown

Frequency Percent Col Pct	Table of OUTCOME by RESSAM		
OUTCOME(Field outcome)	RESSAM(Reserve sample identification)		
	1	3	Total
1	1799 35.60 42.80	118 2.33 13.87	1917 37.93
2	1617 31.99 38.47	182 3.60 21.39	1799 35.60
3	405 8.01 9.64	52 1.03 6.11	457 9.04
9	382 7.56 9.09	499 9.87 58.64	881 17.43
Total	4203 83.16	851 16.84	5054 100.00

Initial sample

Reserve

17.4 % of the gross sample is not used

One major change that may occur in the future is a shift to mixed-mode data collection for the ESS. Should this occur, we do not foresee any associated major change to the survey's approach to sample design. The requirements will remain the same, as will the process for developing and agreeing the design. Indirect effects on sampling may be the introduction of even greater uncertainty regarding likely response rates, and hence the size of sample that should be selected initially.

In conclusion, it seems that the ESS has broadly succeeded (so far) in introducing and maintaining the use of high-quality comparable samples.

Notable achievements have included the introduction of good sample designs, and design features, in places where these were not previously known, as well as incremental improvements in other places, partly building on the cumulative experience of ESS participation. On the other hand, the need to commit sustained effort to maintaining good practice was perhaps underestimated. Nothing should be taken for granted: success last time is no guarantee of success next time!

I think that the ESS data base (including its integrated and cumulative versions) is the best European data source when comparing countries in several social science phenomena. E.g. Eurobarometer, Worlds Value Survey and EU-SILC are not as good than the ESS. However. I have to say that the PISA is even better in assessment variables but not in ordinary survey questions with high item nonresponse. On the other hand, its target population is very limited.

I hope that we are happier in future with the ESS as well.
Thank you for your attention