Sampling and Estimation of Multinational Surveys with Examples from the European Social Survey

Seppo Laaksonen
University or Helsinki and Statistics Finland
Seppo.Laaksonen@Helsinki.Fi

Ventspils, Latvia, Seppo Laaksonen

Multinational surveys 1

- Most Eurostat surveys, the main purpose being to collect data for statistics
  - Harmonized as well as possible using a specific regulation if the survey is regular
  - Subsidiarity principles

- Other EU surveys
  - Eurobarometer is maybe most common
  - Consumer barometer

- Global surveys
  - PISA survey (OECD school achievement survey), very strictly coordinated
  - Gender and Generations Survey (GGS), coordinated by the UN, not many countries yet conducted (will be a panel, three times every two years)

Multinational surveys 2

- Other surveys, more or less volunteer but still coordinated to some extent

- Worlds value survey were designed to provide a comprehensive measurement of all major areas of human concern, from religion to politics to economic and social life and two dimensions dominate the picture:
(1) Traditional/ Secular-rational and
(2) Survival/Self-expression values.
More than 80 societies, on all six inhabited continents.
- International Social Survey Programme (ISSP) is a continuing annual programme of cross-national collaboration on surveys covering topics important for social science research. 44 countries as members.
- International Crime Victimisation Survey (ICVS), approximately 60 countries worldwide. It is coordinated through the UN Office on Drugs and Crime with national coordinators overseeing the project in each participating country. Five cycles of the ICVS have been conducted: 1989, 1992, 1996, 2000 and 2004
- The European Social Survey (ESS)

Co-Workers in the sampling group of the ESS:
Peter Lynn (Essex),
Sabine Häder (Mannheim) and
Siegfried Gabler (Mannheim)

Ventspils, Latvia, Seppo Laaksonen

Starting targets:

High quality, better than in any other social science surveys.
I think that this has been achieved but e.g. the PISA is better but it is not basically a social science survey. PISA is however easier to coordinate well, at least within schools.

Advice from an experienced survey methodologist:
E.g. Kish (1994, 173) writes, "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."

ESS, General Aspects 1

The ESS is an academically-driven social survey designed to chart and explain the attitudes, beliefs and behaviour patterns of Europe´s diverse populations. In parallel with its substantive aims, it aims also to provide a model of best practice in methodology and to contribute towards improvement in methodological standards (further details: www.europeansocialsurvey.org). The ESS is funded via the European Commission's Framework Programmes, with supplementary funds from the European Science Foundation. In each participating nation, the cost of data collection and the appointment of a national co-ordinator (NC) is funded by the national research council or equivalent body. An important principal of the survey is that the data are made freely available: no-one involved in the survey has advance access and there are no restrictions on access. Data can be downloaded from http://ess.nsd.uib.no.

There is a core questionnaire that is administered in every round, along with modules of questions that will change from round to round.

ESS, General Aspects 2

Participation of countries in the ESS 2002-2007

Country	*Round 1*	*Round 2*	*Round 3*	*Country*	*Round 1*	*Round 2*	*Round 3*
Austria	Yes	Yes	Yes	Italy	Yes	No	?
Belgium (Flemish)	Yes	Yes	Yes	Luxembourg	Yes	Yes	No
Belgium (Francophone)	Yes	Yes	Yes	Netherlands	Yes	Yes	Yes
Bulgaria	No	No	Yes	Norway	Yes	Yes	Yes
Cyprus	No	No	Yes	Poland	Yes	Yes	Yes
Czech Republic	Yes	Yes	?	Portugal	Yes	Yes	Yes
Denmark	Yes	Yes	Yes	Romania	No	No	Yes
Estonia	No	Yes	Yes	Russia	No	No	Yes
Finland	Yes	Yes	Yes	Slovak Republic	Yes	No	Yes
France	Yes	No	Yes	Slovenia	Yes	Yes	Yes
Germany	Yes	Yes	Yes	Spain	Yes	Yes	Yes
Greece	Yes	Yes	Yes	Sweden	Yes	Yes	yes
Hungary	Yes	Yes	Yes	Switzerland	Yes	Yes	Yes
Iceland	No	Yes	?	Turkey	No	?	?
Ireland	Yes	Yes	Yes	Ukraine	No	Yes	?
Israel	Yes	No	No	United Kingdom	Yes	Yes	Yes

Ventspils, Latvia, Seppo Laaksonen

Targets for Sampling (group)

☐ High quality for sampling like for other survey characteristics
☐ As even quality as possible so that the estimates would be as cross-country comparable as possible (equivalence)
☐ Effective and encouraging co-ordination by the sampling expert group:
- The agreed sampling criteria should be satisfied maximally
- The group will 'sign off' the sampling plan finally
- Support for the National ESS co-ordinator (NC) and the survey institute (SI) in all stages.
- Prepare a report together with the NC/SI on the sampling design and the summary report of all country designs.
- Check the sampling file so that this gives opportunity to construct the correct design weights and other sampling weights.
- Evaluate after the fieldwork how well the sampling practice corresponds to the sampling design and take this into account when designing the next ESS round.

Requirements in sampling 1

Note that a sampling design can vary from one country to the next following the best practices of each country. There have been a competition for a SI responsible, and both private and public SI's may have won this fieldwork task.

Only random samples provide a theoretical basis which allows us to infer from the sample to the population or sub-sets of this. As design based inference is one important goal in the project, probability samples are required. However, this is related to other requirements:

- full coverage of the target population
- high response rates (target minimum rate: 70%)
- no substitution

the same minimum effective sample sizes (completely responded units) in participating countries (ESS: 1,500 or 800 where population is smaller than 2 m. inhabitants).

Requirements in sampling 2

Each requirement is demanding and hence all are fulfilled in a few countries only. Too low budget is the main reason for not achieving the requirement of the effective sample size.
The differences between countries are higher than expected.

Target population:

Not big problems, but overcoverage (in-eligibles) varies about 2-15 % but in most cases is less than 5%. Old frames however quite often. Some special regions and groups are not well covered.

Response rates:

Average about 65%, thus not bad. Some countries more than 70%, but some others too low (e.g. CZ, CH). Sampling plan was quite realistic in most countries, in some others too optimistic, maybe one case was too pessimistic. If too optimistic, this leads to a too low number of respondents (net sample size).

Requirements in sampling 3

Substitution was in many private SI's a house style but since not allowed here, many discussions (and more money) were needed to convince a SI to change its fieldwork practice. It is not guaranteed that this would have not been used slightly in some countries.

Close to the substitution problem: much discussion was also needed to find an acceptable random strategy within a psu (primary sampling unit). Fortunately, in some countries, these target persons were possible to draw from a local registry, and randomly, but in many countries not. So, a household or an address was selected first randomly (?) and then using e.g. a Kish-Grid method for choosing a person. Also, the last birthday method was applied.

Requirements in sampling 4

Moreover,
A surprising problem was concerned the definition of a *psu*.
In many countries, fortunately, was a tradition such as to use a census district or something like this as a *psu* for the ESS.
OK!
Consequently, they had some population data on these units.
This also means that *psu*'s were quite small and homogenous.
However, we found countries with very big and small *psu*'s, and demographic data were not good. This may lead to difficulties in getting a satisfactory accuracy and to create correct design weights and other sampling weights.

Requirements in sampling 5

In general, countries differ quite much in drawing a probability sample from its 15+ old population. Even *srs* was used and more often in round two than in round one.

The most common strategy is however,
stratified multi-stage sampling.
Stratification is based
- gender, age and region in some *srs* countries where used (however implicit stratification)
-big regions in other countries
Stages
- maximally four, but two is more common
- *pps* without replacement is like a standard for the first (regional) stage (size = 15+population?)

Requirements in sampling 6

Effective sample size = *neff*

is our 'big' concept, not always simple.

We made our best in order to get this at the minimum level, even taking into account some risk possible to meet in fieldwork.
Budget did not allow however always to achieve this in practice although the accepted sampling plan was based just on a minimum size (strange!).
The example on the next page illustrates which factors we took into account when anticipating a country *neff* for the sampling plan. An important concept here is *DEFF* = Design Effect. We use the two components for it.
DEFF = the variance estimate of the current design divided by the variance estimate of the successful srs design, We had to predict or anticipate it.

Design effect due to unequal selection probabilities (DEFFp)

The ESS guidelines suggested that *DEFFp* should be predicted as follows:

$$D\tilde{E}FF_P = \frac{m\sum_{i=1}^{I} m_i \left(w_i^2\right)}{\left(\sum_{i=1}^{I} m_i w_i\right)^2} \qquad (1)$$

where m_i and w_i denote respectively the number of interviews and the design weight associated with the i^{th} weighting class.

(This can be expressed equivalently as $1 + cv_w^2$, where cv_w is the coefficient of variation of the weights)

Design effect due to clustering ($DEFF_C$)

The cluster sample size and the intra-class correlation also influence the design effect. Following Kish (1987), the ESS guidelines suggested that $DEFF_C$ should be predicted as follows:

$$\tilde{DEFF}_C = 1 + \left(\bar{b} - 1\right)\rho \qquad (2)$$

where $\bar{b}$ is the mean number of interviews per cluster and ρ is the intra-cluster correlation.

Ventspils, Latvia, Seppo Laaksonen

Combined design effect

The ESS guidelines suggested that the total design effect should be predicted as:

$$D\tilde{E}FF = D\tilde{E}FF_P \times D\tilde{E}FF_C$$

Impact of pre-stratification excluded, often this DEFF below 1

Illustrative example of all factors related to anticipate an ideal gross sample size.

Operation	**Size calculation**
1. Target effective sample size - *neff* (size that can be received with *srs* without missingness).	**1500**
2. Anticipated missingness due to nonresponse (on average , may vary by strata, e.g.)	30% eli 1500/.7 = 2143
3. Anticipated missingness due to overcoverage (on average)	5% eli 2143/.95 = 2256
4. Anticipated cluster effect so that the final cluster size has been anticipated too * and intra-cluster correlation based on earlier experience on similar surveys	$DEFF_c$ = 1+(5.3- 1)*.025 = 1.108 2256*1.108= 2499
5. Anticipated design effect due to unequal inclusion probabilities used in the design*	$DEFF_p$ = 1.25 2499*1.25 = 3125
6. Anticipated risk in fieldwork and then we have the gross sample size (here net sample size = 3150*.7*.95 = 2095)	**3150**

* *should be consistent with figures in points 2 and 3*

Ventspils, Latvia, Seppo Laaksonen

Response rates and realised interviews from round 2 based on the data from April 2006.

	Number of realised interviews	Rate of ineligibles (%)	Response rate (%)	Non-contact rate (%)	Refusal rate (%)
Austria	2256	1.7	62.5	7.8	28.6
Belgium	1778	4.9	61.5	7.1	22.7
Czech Republic	3026	1.3	55.5	n/a	n/a
Denmark	1487	6.4	65.1	5.6	23.9
Estonia	1989	12.7	79.5	5.1	11.4
Finland	2022	1.5	70.8	2.8	21.2
France	1806	7.1	44.2	12.1	39.5
Germany	2870	7.2	52.7	6.2	27.4
Greece	2406	0.1	78.8	3.7	16.4
Hungary	1498	13.5	70.3	6.0	16.0
Iceland	579	5.9	51.3	4.6	39.1
Ireland	2286	8.1	62.5	9.5	22.3
Luxembourg	1635	10.2	52.1	7.7	40.2
Netherlands	1881	3.0	64.5	2.7	28.0
Norway	1760	3.4	66.2	2.1	25.5
Poland	1716	3.8	74.4	2.3	18.2
Portugal	2052	6.4	70.9	2.8	20.0
Slovenia	1442	6.7	70.2	10.2	15.3
Spain	1663	7.8	56.1	13.6	18.6
Sweden	1948	2.3	66.5	4.3	22.6
Switzerland	2141	6.5	47.1	2.9	39.7
United Kingdom	1897	7.9	51.1	8.0	34.0

Ventspils, Latvia, Seppo Laaksonen

DEFF-calculations

We have analysed real *DEFF*'s using round one data.
As mentioned in some countries the definition of *psu*'s was problematic but we found a reasonable solution.

The second aspect is for which variables and for which estimates to calculate these. It is clear that the *DEFF*'s vary from a variable to the next to some extent although they have common features.

Our solution was to take 10 different variables from round one, most indicators or factors, thus based on several initial variables.

Some results on the next page, *median rho* refers to our intra-cluster correlation based on these 10 variables.

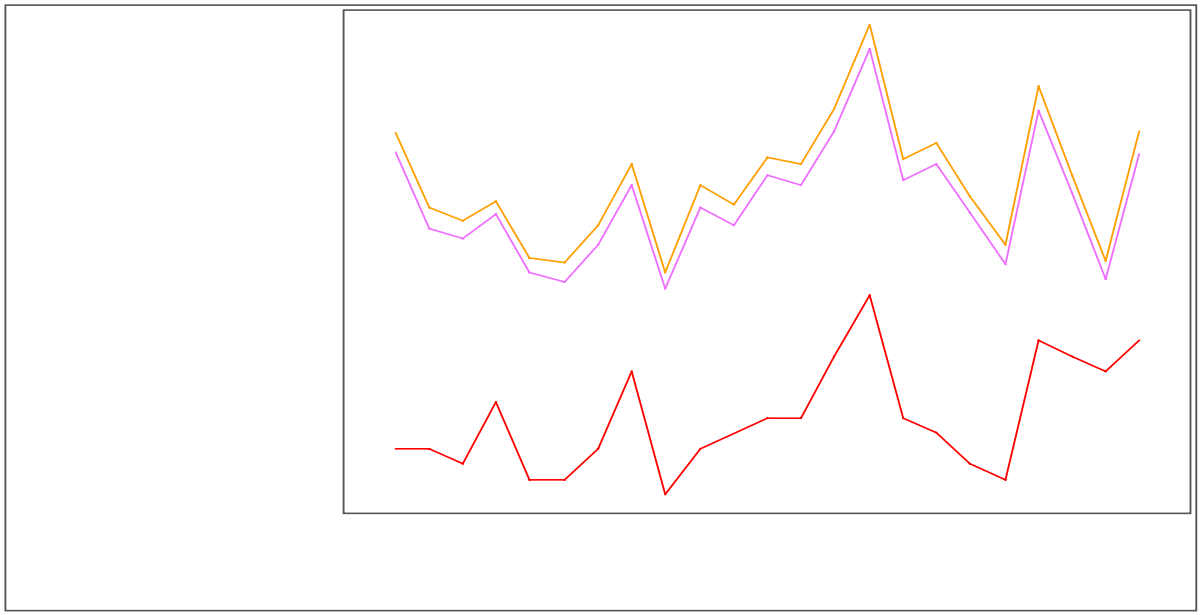
Estimation of design effects for a number of countries

Country	Median ρ	max b^*	*DEFFc*	*DEFFp*	*DEFF*
AT	0.11	6.49	1.61	1.24	2.01
BE	0.04	6.56	1.22	1	1.22
CH	0.03	8.83	1.27	1.21	1.54
CZ	0.15	2.94	1.28	1.25	1.61
DE	0.06	18.85	2.03	1.11	2.26
ES	0.15	4.96	1.60	1.22	1.95
FR	0.05	7.42	1.34	1.23	1.65
UK	0.03	12.06	1.40	1.22	1.69
HU	0.05	8.68	1.36	1	1.36
NL	-	-	1	1.19	1.19
NO	0.01	30.03	1.41	1.43	2.03
PL	0.05	10.07	1.32	1.02	1.35
PT	0.14	5.07	1.57	1.83	2.88
SI	0.03	10.76	1.33	1	1.33

Ventspils, Latvia, Seppo Laaksonen

Quality check

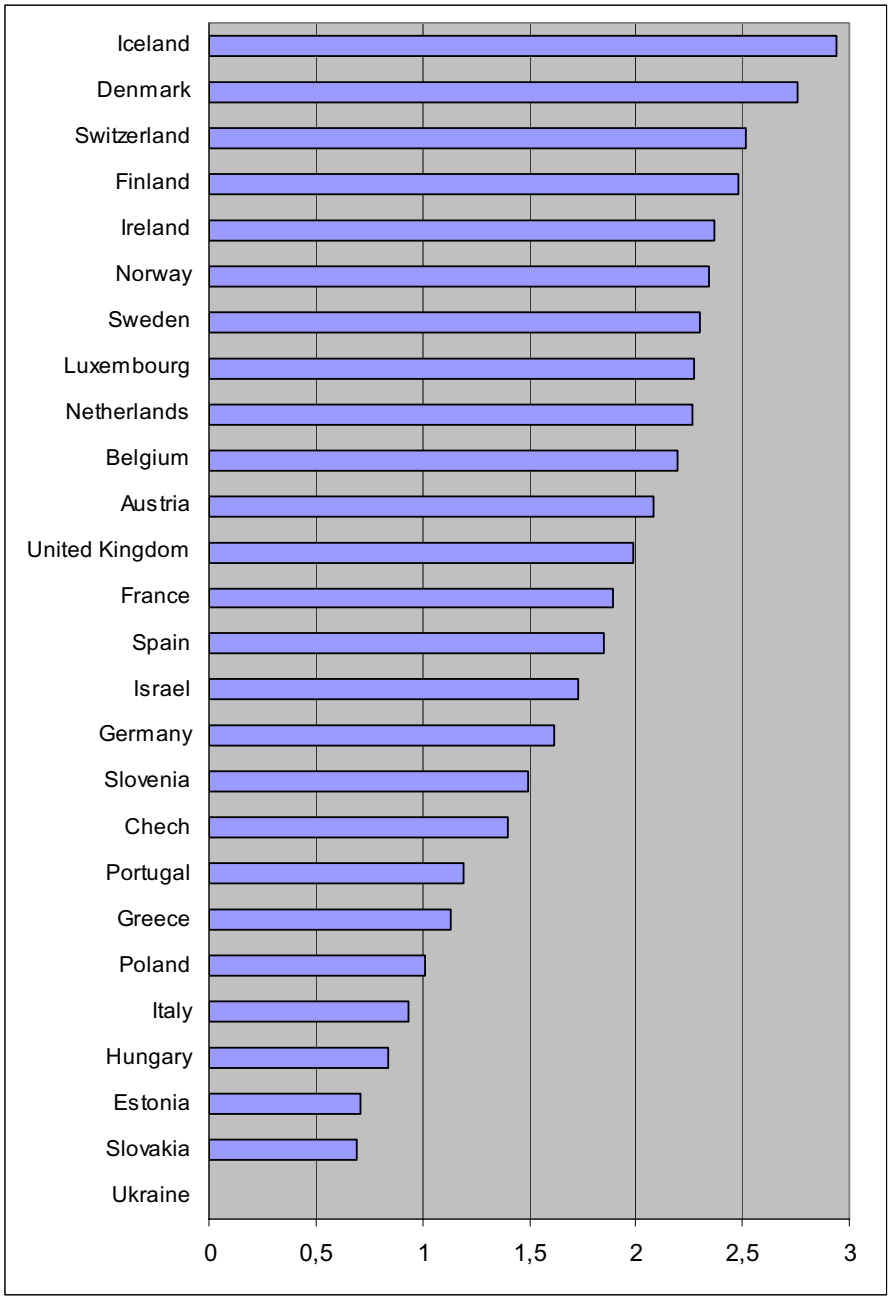
Average household size by country based on the UNECE data from early 2000 (lower line) and from the ESS micro data from 2004-2005 (upper lines with 95% confidence intervals).

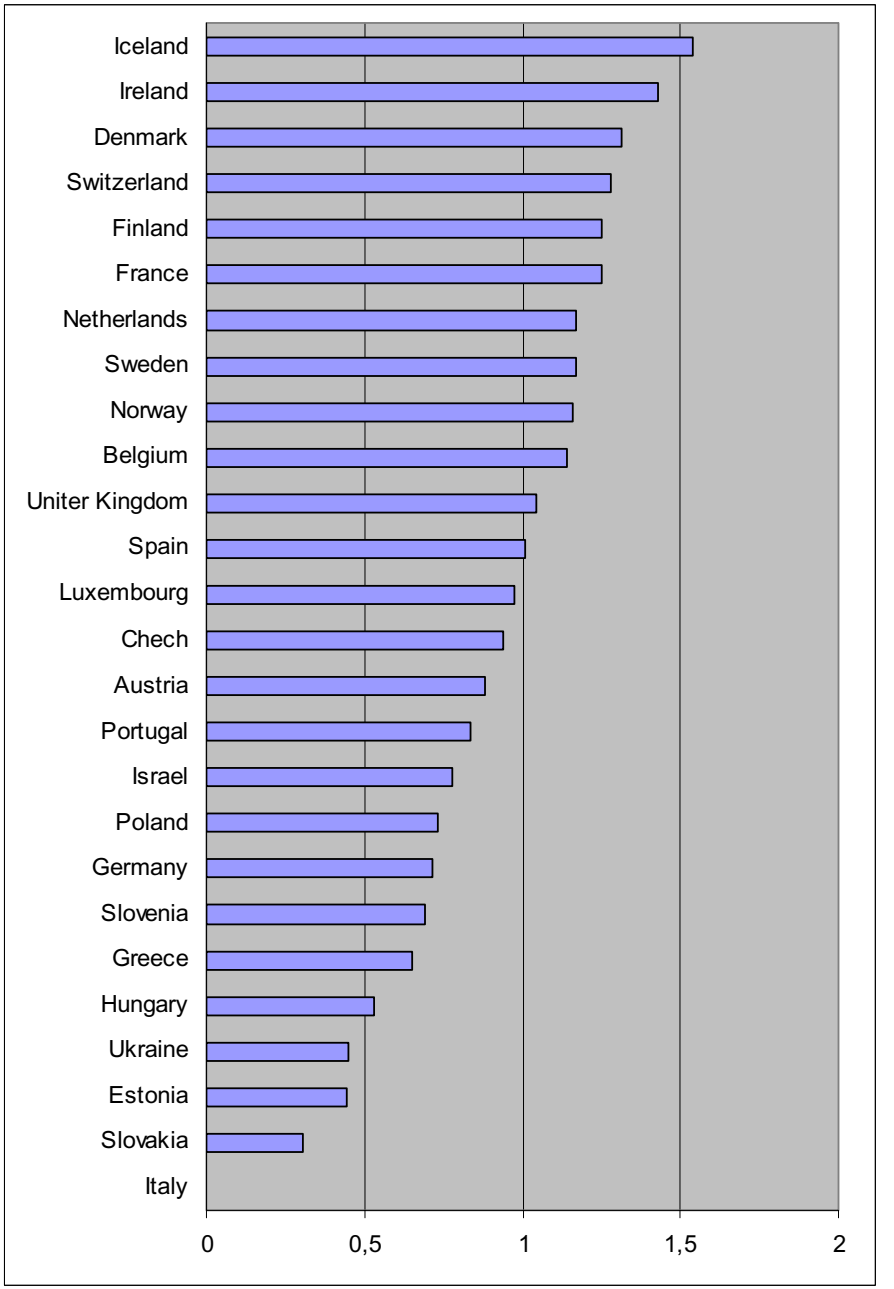


Analysis: Example on Happiness

Taking all things together, how happy would you say you are? Please use this card.

Extremely unhappy										Extremely happy	(Don't know)
00	01	02	03	04	05	06	07	08	09	10	88





Ventspils, Latvia, Seppo Laaksonen

Critics and future 1

Oversampling has been used in some countries and also so that the expected differences in nonresponse/ overcoverage between regions have been taken into account. But this could be exploited much more, also in *srs*-countries where it is well-known that response rates vary much by region and other domain. So, pre-stratification would be *my recommendation* for these countries too, and consequently leading varying weights if the anticipation is not complete.

But this does not belong to our sampling group.

Critics and future 2

My second recommendation is to insert the new sampling weights into the ESS archive data files in addition to the current design weights (DWEiGHT).

And in the first stage these weights should be required for the *srs* countries that have always the weights equal to one in the current integrated archive file.
This is not even difficult since the countries have already created such weights, based on simple post-stratification or other calibration.

Later, we should require all countries to add information for nonresponse adjustments. For example, all countries are able to add to a sampling file some variables for nonrespondents, That could be used easily for adjustments.

Conclusion

The quality of the ESS could be improved
for example using my proposals.

Nevertheless, the quality of this survey
is one of the best ones in the world if the
demanding multinational surveys are concerned.

Thank you

Kish, L. (1992). Weighting for unequal P_i. *Journal of Official Statistics*, 8:2, 183-200.

Kish, L. (1994). Multipopulation survey designs: five types with seven shared aspects. International Statistical Review, 62, 167-186.

Kish, L. (1995). Methods for design effects. *Journal of Official Statistics*, 11:1, 55-77.

Lynn, P., Gabler, S. & Häder, S. & Laaksonen, S. (2006). Methods for achieving equivalence of samples in gross-national surveys. *Journal of Official Statistics*. (in print)

Philippens, M. and Billiet, J. (2006). Nonresponse in cross-national surveys. Results of the European Social Survey. *ESS Working Paper*.