Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

The influence of sampling method and interviewers on sample realization in the European Social Survey

by Natalja Menold

June 2014

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@statcan.gc.ca,

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-877-287-4369

Depository Services Program

• Inquiries line	1-800-635-7943
• Fax line	1-800-565-7757

To access this product

This product, Catalogue no. 12-001-X, is available free in electronic format. To obtain a single issue, visit our website, www.statcan.gc.ca, and browse by “Key resource” > “Publications.”

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under “About us” > “The agency” > “Providing services to Canadians.”

Published by authority of the Minister responsible for Statistics Canada

Cette publication est aussi disponible en français.

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard symbols

The following symbols are used in Statistics Canada publications:

.	not available for any reference period
..	not available for a specific reference period
...	not applicable
0	true zero or a value rounded to zero
0^s	value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p}	preliminary
^{r}	revised
x	suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E}	use with caution
F	too unreliable to be published
*	significantly different from reference category ($p < 0.05$)

Survey Methodology, June 2014
Vol. 40, No. 1, pp. 105-123
Statistics Canada, Catalogue No. 12-001-X

The influence of sampling method and interviewers on sample realization in the European Social Survey

Natalja Menold[1]

Abstract

This article addresses the impact of different sampling procedures on realised sample quality in the case of probability samples. This impact was expected to result from varying degrees of freedom on the part of interviewers to interview easily available or cooperative individuals (thus producing substitutions). The analysis was conducted in a cross-cultural context using data from the first four rounds of the European Social Survey (ESS). Substitutions are measured as deviations from a 50/50 gender ratio in subsamples with heterosexual couples. Significant deviations were found in numerous countries of the ESS. They were also found to be lowest in cases of samples with official registers of residents as sample frame (individual person register samples) if one partner was more difficult to contact than the other. This scope of substitutions did not differ across the ESS rounds and it was weakly correlated with payment and control procedures. It can be concluded from the results that individual person register samples are associated with higher sample quality.

Key Words: Sampling methods; Substitutions by interviewers; Non-observation errors.

1 Introduction

Biases in survey statistics are described by the total survey error models (Groves, Fowler, Couper, Lepkowski, Singer and Tourangeau 2004; Smith 2007). Total survey error results from two types of errors, which are referred to as observation errors and non-observation errors. This article focuses on cross-cultural comparability influenced by factors related to non-observation errors, that is to say the failure of survey statistics to adequately represent the target population. These types of errors – also called representation errors – result from differences between the obtained net sample (number of survey participants, Haeder and Lynn 2007) and the target population.

Previous research in cross-cultural contexts has revealed highly pronounced differences in response rates between countries (Billiet, Phillipsen, Fitzgerald and Stoop 2007; Couper and de Leeuw 2003; de Heer 1999; de Heer and Israis 1992; de Leeuw and de Heer 2002; Hox and de Leeuw 2002; Johnson, O'Rourke, Burris and Owens 2002; Stoop 2005; Symons, Matsuo, Beullens and Billiet 2008), differences in field procedures (Billiet *et al.* 2007; Kohler 2007; Kreuter and Kohler 2009; Smith 2007; Stoop 2005; Symons *et al.* 2008) and differences in sampling methods (Lynn, Haeder, Gabler and Laaksonen 2007). The latter refer to procedures for constructing sampling frames and selecting sample elements. All differences listed may impede cross-cultural comparability. In particular, cross-cultural comparability might be limited due to varying sampling methods to obtain a probability sample since standardising sampling methods is restricted by local availability of sampling frames, by their quality and usability, and by the survey budget (Lynn *et al.* 2007).

Lynn *et al.* (2007) addressed design effects and the sample sizes necessary to achieve comparability of net samples in the ESS. In doing so, they focused on sample selection prior to the field work stage. However, comparability of samples could also be influenced by interviewers during their work in the

1. Natalja Menold, GESIS - Leibniz Institute for the Social Sciences, Survey Design and Methodology, P.O.Box 12 21 55, D-68072 Mannheim. E-Mail: natalja.menold@gesis.org.

field. Interviewers' degree of freedom in substituting sampled individuals with persons who are not sampled (individuals who are easy to contact and are cooperative) differs based on sampling methods (Hoffmeyer-Zlotnik 2006; Kohler 2007; Sodeur 2007). "Field substitution occurs when a non-responding unit is replaced by a substitute (reserve) unit during the field work stage of the survey process" (Vehovar 1999, page 335). The substitutions addressed by Vehovar are legitimate substitutions that are allowed by protocol. In contrast, this article addresses illegitimate substitutions (referred to below simply as "substitutions") which occur without permission. According to the AAPOR (2003), deliberate substitutions made by interviewers represent a kind of falsification.

The aim of this article is to show whether the effect of interviewers, which is assumed to be associated with substitutions, varies across different sampling methods that are used to obtain probability samples in a cross-cultural context. In addition, it will be shown whether payment, control procedures, the data collector (institution that carries out data collection in the field) or time factors are associated with this interviewer effect. The results may help surveyors when deciding upon sampling methods – a highly relevant cost factor in surveys – and when deciding upon methods to foster interviewer motivation to not substitute. The results are also important for research on interviewer behaviour since they hint at errors associated with interviewer influence on cross-cultural comparability.

The next section (Section 2) provides the theoretical and empirical background of the study presented in this article. The hypotheses are described in Section 3. Section 4 provides information on the procedure and the method used for analysing the influence of the interviewer. The results are presented in Section 5. Finally, Section 6 discusses the results and provides conclusions.

2 Theoretical and empirical background

Substitutions may arise during tasks that interviewers conduct prior to the interview. Interviewers (1) build a sampling frame, for example by generating lists of addresses in surveys; then they (2) elicit cooperation with selected units (addresses, dwellings, households) and they also select individuals for the interview from these units. Finally interviewers (3) elicit cooperation from sampled individuals (Groves *et al.* 2004). In the case of different sampling methods interviewers perform different tasks as described below (figure 2.1).

The first sampling method refers to individual person register samples (denoted below as PRS). Official population registers of individuals are used as sampling frames for PRS. The selection of individuals is conducted prior to the field work stage, thus reducing interviewers' tasks to simply obtaining cooperation from sampled individuals (figure 2.1). In the case of PRS interviewers may influence non-response (*e.g.*, Couper and Groves 1992; de Leeuw and Hox 1996; Durrant, Groves, Staetsky and Steele 2010), but in a theoretical sense they have no influence on the sampling frame or on the selection of sample elements. This level of interviewer impact on non-representation in the case of PRS is shown by the arrow in figure 2.1.

However, as shown for example by Groves *et al.* (2004) selected elements (individuals in the case of PRS) may differ in terms of the probability of being contacted by an interviewer (contactability) and the probability of obtaining survey participation when a contact is given (cooperation). For example, it has been found that people living in urban areas or individuals who are young, single, without children, better educated and socially active are more difficult to contact (Stoop 2004). In contrast, older respondents, women, less-educated people and socially isolated individuals refuse cooperation more often than others (Dohrenwend and Dohrenwend 1968; Stoop 2004; Williams, Irvine, McGinnis, McMurdo and Crombie 2007). If difficulties arise when trying to contact and obtain cooperation from target persons then substitutions may occur. For example, Koch (1995) reported the number of substitutions in a survey in which PRS was used.

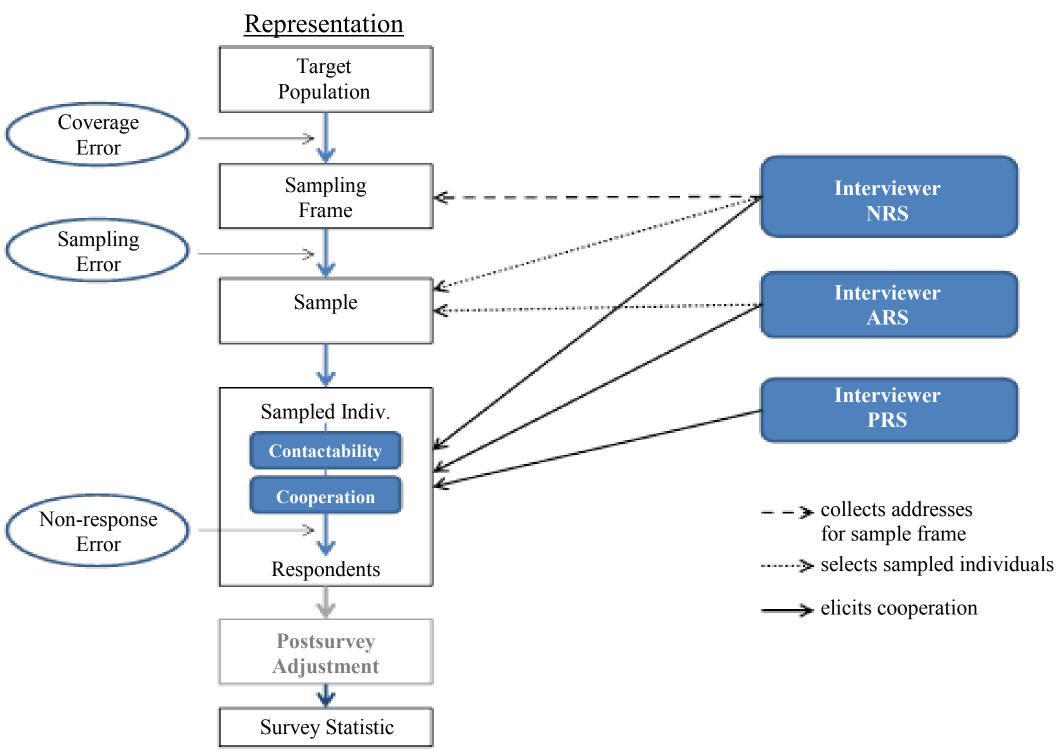

Figure 2.1 Interviewers' tasks in different sampling methods related to coverage, sampling and non-response errors. The path related with representation is adapted from Groves *et al.* 2004, page 48

The next sampling methods discussed are address/household register samples (ARS). In the case of ARS lists of households or addresses are employed as sampling frames. Households or addresses are selected by survey offices prior to the field work stage. In this case interviewers perform tasks two and three (see above): they contact selected units and select individuals for the interview if more than one eligible individual is living at one unit. Interviewers may deliberately deviate from the random selection rules and in this way they can have a negative impact on the selected sample (figure 2.1). Since interviewers have more freedom in selecting sampled individuals with ARS than with PRS, it is assumed that interviewer impact associated with substitutions is higher for ARS than for PRS (figure 2.2). Moreover, in the case of ARS the result of sample selection is not known beforehand and is therefore harder to control than in the case of PRS.

Non-register samples (NRS), in which neither lists of individuals nor lists of addresses/households are available as sampling frames, are described as the third sampling method. These include Random Route Samples (*e.g.*, Arber 2002; ESS Sampling Plans), and Address Listing and Sampling (ALS). In the case of

NRS interviewers themselves generate a sampling frame by listing or collecting addresses within a randomly selected geographical area. Interviewers have to strictly follow instructions concerning the procedures for collection of addresses. Interviewers perform this task in addition to selecting individuals at one address, as described for ARS, and to contacting and obtaining cooperation, as just described for both PRS and ARS (figure 2.1). With NRS interviewers can influence not only sample selection but also the sampling frame. An interviewer can deviate from instructions and chose only addresses where he or she expects to contact the target person and obtain cooperation. Substitutions are particularly likely to occur when using a procedure (Random Route) in which the interviewer conducts the interview at an address which they choose in an area following prescriptions regarding collecting of addresses and the direction of the route through the area. Another type of NRS is more restricted since the interviewer lists the addresses in a geographical area, but the actual selection is conducted by a coordination team (Address Listing and Sampling, ALS). The selected addresses are subsequently assigned to a different interviewer who then conducts the interviews. The degree of interviewer freedom in the case of ALS appears to be similar to that of ARS. However, the instructions for listing or collecting addresses can be ambiguous in the case of both types of NRS (Schnell, Hill and Esser 2011). Therefore, interviewers have more freedom to substitute with NRS than with ARS (figure 2.2).

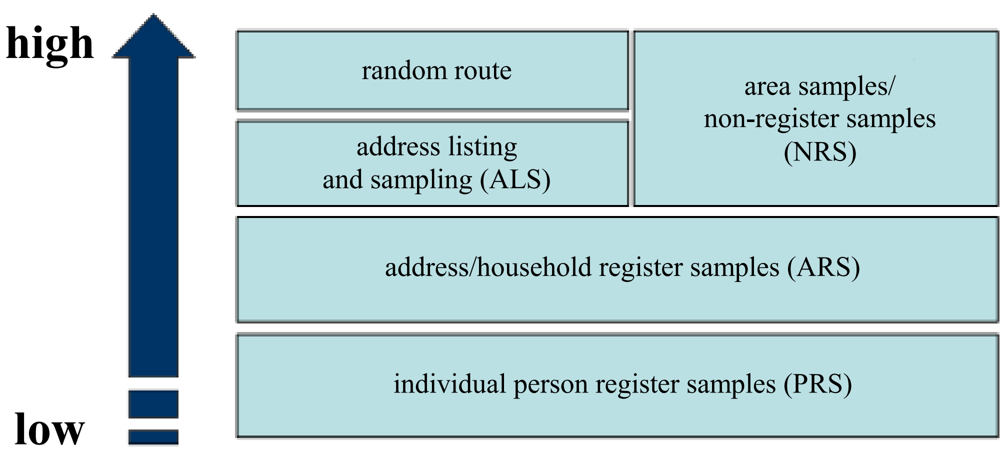

Figure 2.2 Degree of interviewers' freedom to substitute in various sampling methods

Deviations in obtained net samples, which are associated with deviations from the rules of random sample selection (*e.g.*, substitutions), can be empirically analysed using a method developed by Sodeur (1997). This method works by defining a subpopulation with one fixed and known parameter, then the statistics representing this parameter are observed in a subsample that is defined in a corresponding manner. The more the observed statistic deviates from the population parameter, the stronger the error from non-observation is. This article considers the gender ratio of heterosexual couples, which is known to be a 50/50 population parameter. Within the limits of random fluctuation any sample from the population of heterosexual couples should produce a proportion of males of around 50%. Significant deviations from this level of 50% indicate deviations from sample selection standards, such as through substitutions (see section 4.2 for details).

Using this method Sodeur (2007) and Hoffmeyer-Zlotnik (2006) found that deviations from the true parameter of a 50/50 gender ratio vary between different rounds of the German General Social Survey (ALLBUS), which also involved different sampling methods. The authors found that males who are difficult to contact are interviewed less often than females (since the males are the breadwinners in households with children). Aside from contactability, differences in cooperation between the partners can play a role (Hoffmeyer-Zlotnik 2006). If partners are retired they have comparable levels of contactability but they may differ in terms of cooperation. The retired man, now at home, feels responsible for providing information to an interviewer about the household (as its “head”). The woman can refuse to participate since the man likes to cooperate. An interviewer who contacts such households may interview men instead of women in order to avoid refusals (Hoffmeyer-Zlotnik 2006).

Kohler (2007) found larger deviations from the parameter of a 50/50 gender ratio in NRS samples as compared to other sampling methods in six cross-cultural surveys (Eurobarometer 62.1, European Quality of Life Survey EQLS’03, ESS 2002, ESS 2004, European Value Study 1999, International Social Survey Program, ISSP 2002). Unfortunately, the sampling method effect obtained by Kohler (2007) was survey-specific. The most poorly designed samples – area samples with a NRS – were used predominantly in one survey (EQLS). The differences that Kohler found between a random route and other sampling methods could thus be due to differences between EQLS and other surveys. Other researchers (Hoffmeyer-Zlotnik 2006; Souder 1997) addressed the effect of sampling methods on interviewer impact associated with substitutions while considering only a single German national survey; the results of this research are not applicable to cross-cultural contexts. Therefore, it is important to address the question regarding the relationship between sampling methods and interviewer impact associated with substitutions in cross-cultural surveys. Additionally, it is important to consider other explanatory factors which can affect substitutions. Substitutions made by interviewers can be affected not only by sampling methods but also by field procedures related to interviewer motivation to produce accurate survey data. Therefore, substitutions may vary based on the data collector (Hoffmeyer-Zlotnik 2006; Sodeur 1997; 2007) or controls used in a survey (Kohler 2007). Controls imply that a sample element is re-contacted to confirm the outcome produced by an interviewer. In addition to controls, methods of providing payment to interviewers can impact their performance. If interviewers are paid per completed interview they bear the risk of high costs due to long distances between selected addresses, numerous contact attempts or long interview times (Sodeur 2007). Consequently, a change of data collector, controls and payment should be considered when analysing the interviewer impact associated with substitutions. Apart from these factors it is interesting to see how such interviewer impact varies across time. For example, considering time in a cross-cultural context helps to indicate whether this interviewer impact is country-specific. A country-specific interviewer impact should be stable in a country across different survey rounds, even if the sampling method is changed.

3 Research hypotheses

If interviewer impact in terms of substitutions is operationalized using Sodeur’s method, it is expected that it can be observed in survey statistics as deviations from the 50/50 gender ratio in subsamples with respondents as representatives of heterosexual couples. This interviewer impact is expected to differ in terms of varying contactability or cooperation on the part of the partners. Partners differ in terms of contactability in households comprised of couples with young children in which men are the breadwinners (Hoffmeyer-Zlotnik 2006; Sodeur 2007; Stoop 2004). If interviewers apply substitutions the proportion of men should be significantly lower than the true value (50%) in such households, since men are more difficult to contact than women. This pattern changes when taking retired partners into consideration. Here, as previously discussed by Hoffmeyer-Zlotnik (2006), partners have comparable levels of contactability but they may differ in terms of their cooperation. For subsamples of retired couples, the proportion of males is expected to be significantly higher than 50% in the case of interviewer impact due to substitutions. The hypothesis describing this impact in different types of households is:

Hypothesis 1 (H1): Deviations from the true gender ratio (50/50) vary depending on the type of household. In households comprised of couples with (young) children the proportion of males is lower than 50%, while in households comprised of retired partners this proportion is higher than 50%.

As was shown in section 2, different sampling methods can be associated with varying degrees of freedom on the part of interviewers to substitute (figure 2.2). Therefore, the following differences between the sampling methods are expected:

Hypothesis 2 (H2): Deviations from the true gender ratio (50/50) vary depending on the sampling method used in a survey. They are lowest with PRS and highest with NRS samples.

If deviations from the population parameter are caused by interviewers deviating from prescribed standards, then they should vary with the sampling method used or with the type of household, which in turn is associated with varying levels of contactability or cooperation on the part of the partners. The deviations should be stable across time when keeping sampling methods constant. However the deviations can correlate with interviewer payment and control procedures, or with the data collectors, who are expected to differ in terms of practices related to interviewers' work motivation.

Hypothesis 3 (H3): Except for changes in sampling method in a country, deviations from the 50/50 gender ratio are independent of the influence of other changes over time. Thus, they do not vary across different survey rounds. However, interviewer payment, control procedures and change of data collector are expected to correlate with deviations from the 50/50 gender ratio.

4 Methods

4.1 Data

To isolate any effects due to the sampling method from other survey-specific effects, one can use data from a multi-country survey in which the various countries applied different sampling methods. Many rounds of a survey should be available in order to be able to consider the time effect. Therefore, data from rounds one to four of the ESS were used (European Social Survey Round 1-4 Data 2011). The ESS was conducted with between 20 and more than 30 countries, which differ in terms of their sampling methods. In addition, the ESS sets high standards for survey organisations, such as strict random sampling and extended contact procedures, or regarding field control procedures (Koch, Blom, Stoop and

Kappelhof 2009; Philippens and Billiet 2004). The effectiveness of the standards used in the ESS has been demonstrated by Kohler (2007), who showed that round 1 of the ESS had the fewest deviations from the 50/50 gender ratio compared to other surveys. In addition, the ESS has consistently improved data collection methods (Koch *et al.* 2009). Furthermore, the ESS provides detailed documentation of sampling procedures as well as data collection (*cf.* ESS Documentation Reports), which allows for operationalization of variables of interest.

4.2 Method for evaluating interviewer impact

The method developed by Sodeur (1997) was used for the analysis. This method helps to evaluate the net sample quality in probability samples. The quality of the random sample selection has often been examined by means of other statistics available in a country (external criteria). However, these external statistics are often unknown, leading Sodeur to suggest the use of internal criteria - that is to say use of information from the net sample only. Sodeur (1997) describes the method as consisting of the following steps: (1) separating a subsample from the entire sample to focus on respondents as representatives of heterosexual couples: the partners should live together within one household and both partners should belong to the target population of the survey; (2) defining units which should be dropped from the subsample: singles, partners not living together within a household and households with other relatives who belong to the target population. Then, step three entails (3) defining a survey statistic – *e.g.*, the percentage of males – as the dependent variable which should be compared with the population parameter.

An analysis to determine the causes of deviations from the population parameter – for example, interviewer behaviour – requires additional specifications in steps 1 and 2 to ensure that interviewer behaviour (conceptually) varies with the contactability or cooperation of target persons. Such specifications have been made in this article in terms of definitions of different type of households (see hypothesis H1), whose selection is described in section 4.3.

The true 50/50 gender ratio in heterosexual couples is not related to any other gender ratios, such as that of the total population of residents in a country. Therefore, as Kohler (2007) argues, this gender ratio cannot be affected by any sort of measurement errors and it is unaffected by the household size since the analysis is restricted to two persons in the household and both persons belong to the target population.

Sodeur's method has advantages over other methods since no additional external information or data are required. However, Sodeur's method requires that the characteristics defined for selecting subsamples are known not only regarding the respondents but also regarding their partners (*e.g.*, gender of the partner). In addition, there should not be systematic gender differences in terms of refusal behaviour (differential refusal), which may occur even if interviewers work honestly. In practice, females have been found to be more reluctant than males (Pickery and Loosveldt 2002; Schnauber and Daschmann 2008; Stoop 2004; Williams *et al.* 2007). That also seems to be the case in the ESS, in which females were found to refuse more often than males. The author's own analysis of ESS1-ESS4 data from contact forms shows that 30.3% males and 37.9% females refused cooperation in the ESS1 (in some countries no data regarding the gender variable was provided; therefore the missing data was 32.4%). In the ESS2 there were 30.8% males and 37.9% females who refused cooperation (31.3% data missing); in the ESS3 33.8% males and 39.0% females (27.2% of data missing) refused cooperation and in the ESS4 there were 38.4% males and 45.8% females who refused (with a reduced 15.8% of data missing). Therefore, males being present in a subsample of ESS data less than 50% of the time can be plausibly explained by substitutions, while a frequency of males higher than 50% can be alternatively explained by differential refusal. However, if the percentage of males varied with a sampling method – as expected in hypothesis H2 – it would be hard to explain such a result only by differential refusal, which seems to be a quite stable feature.

4.3 Procedure

The following section provides a description of the procedures used for testing hypotheses H1 to H3. First, separation of subsamples from the entire ESS sample is described. Deviations d from the true 50/50 gender ratio in a subsample represent the dependent variable for all subsequent analyses. The values of d are compared between different households to test hypothesis H1. Second, operationalization of the "sampling method" variable (to test H2) is described. Third, hypothesis H3 is related to the variables time, change of data collector, payment and interviewer controls, whose operationalization is described in the last section. H2 and H3 were tested with the help of a Multivariate Analysis of Covariance (MANCOVA) with subsequent individual Analyses of Covariance (ANCOVA) in which the sampling method was used as the independent variable and the ESS round, change of data collector, payment bonus and interviewer controls were used as covariates.

Separation of subsamples

The ESS target population "contains in each country persons 15 years or older who are resident within private households, regardless of nationality and citizenship, language or legal status" (*e.g.*, ESS-1 2002 Documentation Report, page 2). Respondents ($n = 88{,}375$) who live together with a partner of the opposite sex who is 15 years or older were selected from the total ESS1-ESS4 sample ($n = 184{,}988$). This reduced the data base of the analysis to about half of the entire sample. However, such a selection is required to ensure the expected percentage of males of 50%.

Three household types were distinguished among the selected subsample: couples with children aged between 0-6 ($< 7; n = 18{,}791$), couples with children between 7-14 ($n = 53{,}651$), and couples in which both partners are of retirement age (retirees, $n = 15{,}933$). To determine retirement age the current state pension age in each country was used (see appendix). The first two groups with children were formed since it was assumed that differences in contactability between partners are particularly high in these households. For the third group it was assumed that gender differences in contactability are fairly modest, while men and women might differ in terms of cooperation.

The fact that men are breadwinners within the two subsamples containing households with children is supported by the author's own analysis using data from the ESS. Upon looking at respondents' activities within the last seven days in households with children younger than seven, it was shown that 58% of males and 42% of females were in paid work. In terms of respondents' partners, 64% of males and 36% of females were in paid work. Similar results were found for respondents in households with children aged 7 to 14 (for respondents 54% males and 46% females were in paid work and in terms of partners there were 60.5% males and 39.5% females). For households with retired partners it was found that 80.6% of respondents were retired, 11.5% did housework and 1.3% was sick or disabled on a long term basis. With respect to respondents' partners, 84.4% were retired, 17% did housework and 2.1% were sick or disabled on a long-term basis.

Categorisation of sampling methods

Documentation Reports for each ESS round (European Social Survey (2011): ESS 1-4 Documentation Reports) were used to classify the sampling methods. Table 4.1 summarises the main characteristics of the sampling methods used in the ESS. Table 4.2 shows which sampling methods were used in each country in each of the rounds. For more details on selection procedures in the ESS see the Documentation Reports or Lynn *et al.* (2007).

For ARS it is important to see how multi-dwelling units at one address are dealt with since in this case interviewers also manage the situation. The survey documentation described this for only a few countries (Ireland, Israel, the Netherlands, and the United Kingdom). In Ireland, for example, interviewers listed the households and selected one of them using the Kish Grid (Kish 1965).

In Austria a NRS method was applied to only 50% of the sample, while the other 50% was selected based on an ARS. Since using NRS can lead to more substitutions compared to using only ARS it can be expected that the results in Austria are more similar to the results in countries with NRS than with ARS. Therefore, the author assigned Austria to NRS.

Table 4.1
Sampling methods in the countries of the EES (rounds 1-4)

	individual person register sample	**address/household register sample**	**non-register sample**
sampling frame	reliable lists of residents	reliable lists of addresses/households	regional areas (no lists of residents, addresses or households)
stage 1: Selection of PSUs			
definition of a unit	regional clusters, areas, municipalities	electoral sections, postal code areas	regional clusters, areas, municipalities
process of selection	systematic random sampling	systematic random sampling	systematic random sampling
result	community, municipality	electoral section, postal code section	geographical areas, municipalities
stage 2: Selection of households	not applicable		
definition of a unit		a household, an address	a household/dwelling unit
process of selection		simple or systematic random sampling	random route/ALS simple random sampling
result		addresses of households	a household/address/dwelling unit
stage 3: Selection of individuals			
definition of the unit	target person	target person	target person
process of selection	simple or systematic random sampling	random selection by interviewer by Kish Grid or last birth day method	random selection by interviewer by Kish Grid or last birth day method
result	name and address of sampled individuals	sampled individuals	sampled individuals

Table 4.2
Classification of ESS countries with respect to sampling methods

ESS round	individual person register sample (PRS)	address/household register sample (ARS)	non-register sample (NRS)
ESS 1	BE, DE, HU, NO, PL, SI, DK, FI, SE	Address: IE, IT, NL, GB, CH Household: CZ, LU, ES	FR, GR, PT, AT
ESS 2	BE, DE, HU, NO, PL, SI, DK, FI, SE, ES, EE, IS, SK	Address: IE, NL, GB, CH Household: LU, TR	FR, GR, PT, AT, CZ, UA
ESS 3	BE, DE, NO, PL, SI, DK, FI, SE, ES, EE, SK	Address: IE, NL, GB, CH, LV Household: CY, BG, HU	FR, PT, AT, UA, RU, RO
ESS 4	BE, DE, HU, NO, PL, SI, DK, FI, SE, ES, EE	Address: IE, NL, GB, CH, IL, LV Household: CZ, CY, LT, GR, KRO, TR, BG	FR, PT, SK, UA, RU, RO

Note Romania is not included in the ESS integrated data file; no information on sampling method has been provided for Italy in Documentation Reports ESS2-ESS4. Countries are labelled according to ISO 3166-1, see the appendix.

The kind of NRS method used in a country has rarely been described in the documentation. In the ESS1 it is evident that an ALS was used only in Greece. Usage of an ALS is described for the Czech Republic and Slovakia in later rounds. In the ESS4 Ukraine, Russia and Portugal report a procedure comparable to the ALS. However, in these countries interviewers (and not the offices) selected a fixed number of units from the lists generated by other interviewers.

Explanatory variables

Information related to the particular ESS round was used as a variable to control for the time effect. The Documentation Reports were used to obtain information related to other explanatory variables; change of data collector, payment and interviewer controls. Whether a country changed the data collector between rounds is shown in the appendix. Concerning payment, it was found that the ESS mainly employed a method involving payment per completed interview. An hourly rate of payment was used only in a few countries that also used PRS (in ESS1-2 in Norway and Sweden, as well as in ESS3-4 in Norway and Finland). Therefore, there was only a small variation in payment methods, and a corresponding data analysis was not possible. However, payment of bonuses varied across countries and rounds. Therefore, this information was used to generate a dichotomous control variable (bonus payment: yes/no).

Two variables are used to describe control procedures: the number of eligible sample elements selected for controls divided by the number of eligible sample elements (ratio selected), and the number of confirmed outcomes divided by the number of sample elements selected for controls (ratio confirmed). The first variable describes the number of controls in a country, while the second describes the effectiveness of these controls. The "ratio selected" varies between 10% for PRS, 13% for NRS and 16% for ARS. The "ratio confirmed" is somewhat higher for NRS ($M = 75.21$, $SD = 24.81$) than for the other two sampling methods (PRS: $M = 61.89$, $SD = 31.95$; ARS: $M = 66.49$; $SD = 32.56$).

5 Results

5.1 Differences between household types

Firstly, the results for testing hypothesis H1 are presented. This hypothesis expects deviations from the 50/50 gender ratio to vary according to the type of household. Figure 5.1 shows the differences (d) between the actual percentage of males and the expected true value of 50% in three subsamples. A 95% confidence interval (CI) was used to control for random fluctuation. As the expected proportion of men is $p = 0.5$, the variance averages $0.25/n$, whereby n is the number of cases in the subsample in a country. The 95% CI was calculated as follows (*cf.* Kohler 2007, page 59):

$$CI = 0.5 \pm 1.96 \times \sqrt{0.25/2}.$$

Figure 5.1 shows that for both subsamples covering households with children, significant values of d are negative in the majority of cases, meaning that the proportion of males in these subsamples is less than 50% (as expected by H1). Most of these d-values were approx. 10% or higher. Lower (approx. 5%) significant positive (unexpected) d-values are seen for three countries in which PRS was used (in the ESS1 in Belgium and Norway, in the ESS2 in Finland). However, these differences were not discernible in other rounds.

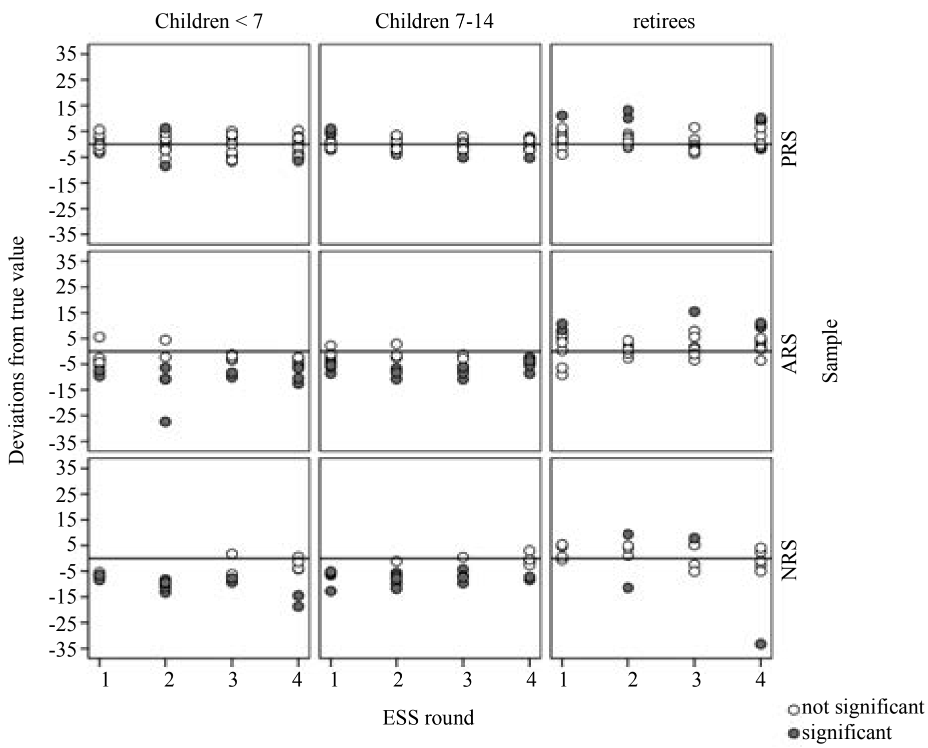

Figure 5.1 Deviations from the true value of 50% (d) in the percentages of males in different types of households of ESS1-ESS4

Regarding the results for the subsamples covering households with partners of retirement age (retirees) it is possible to see significantly high d-values (approx. 10% or higher) with the expected direction (positive, or that is to say the percentages of males are higher than 50%) for some countries across all sampling methods (in the ESS1 in Norway, the Czech Republic and the Netherlands; in the ESS 2 in Norway, Poland and France; in the ESS3 in Cyprus and Russia and in the ESS4 in Germany, Hungary, Cyprus and the United Kingdom). Interestingly, the proportion of men is markedly lower than 50% in Slovakia in the ESS4 (as low as approx. 33%) and in Portugal in the ESS2 (as low as approx. 11%). This result can be explained by specific patterns of role division between the partners. Here the woman appears to represent the household, even if the man is at home.

To summarise, significant deviations from true value in different types of households were mainly in line with the expectations of hypothesis H1.

5.2 Differences between sampling methods

The effect of sampling methods (as expected by H2) was tested by means of MANCOVA. The d-values for the three types of households (three isolated subsamples) were considered as values of three dependent variables, which were simultaneously analysed in the MANCOVA. Since the MANCOVA is based on an analysis of means the absolute values of d were considered. Otherwise it would not have been possible to take into account differences with an unexpected direction, which would also be associated with the effect of sampling methods. Since most of the differences were negative in the subsamples with children, the absolute d-values represent a proportion of men that is lower than 50%. With respect to the subsamples with partners of retirement age, it should be taken into account that the proportion of men was not only higher than 50% but also lower than 50% in Portugal (ESS2) and in Slovakia (ESS4). In addition, significant and non-significant differences are considered in order to enable a comparison between countries with low and high d-values.

The MANCOVA revealed a high significant multivariate effect of the factor "sampling method" (Wilks Lambda (WL) $F_{(6,174)} = 6.87, p < 0.001,$ effect size $\eta^2 = 0.21$). In contrast, no significant results for explanatory variables were found ($p > 0.10,$ max $\eta^2 = 0.04$). In order to consider d-values in different household types univariate analyses of covariance (ANCOVAs) were employed. Variance homogeneity – as a presupposition for an ANCOVA – is given according to the Levene test in the subsample with retirees, and also according to the $F_{\max}$ test in the subsamples with children. Significant mean differences of d-values between sampling methods were found using the ANCOVAs in both subsamples with children (table 5.1). The variances explained in the ANCOVAs for these subsamples are quite high (see R^2 in table 5.1). On average the lowest d-value can be seen for the PRS, while the highest d-value is seen for the NRS (table 5.1 and figure 5.2). However, post-*hoc* single comparisons using subsamples with children show significant differences only between PRS and the other two sampling methods (table 5.2). Also, no remarkable differences in d-values were found between the countries with ALS and with Random Route samples.

Overall, the results show that hypothesis H2 can be partially supported if households with children are considered.

Table 5.1
Descriptive statistics $(M(SD))$ and results of the ANCOVAs for comparison of d in the three types of household

	types of household			
	children <7	**children 7-14**	**retirees**	***n* (countries)**
Sampling method (treatment)				
PRS	3.28(2.07)	2.21(1.37)	3.34 (3.35)	43
ARS	6.61(4.98)	4.87 (2.74)	4.94(3.83)	31
NRS	7.85 (4.4)	5.92 (3.55)	5.78(6.87)	21
$F(df_1 = 2, df_2 = 88)$	14.52***	20.9***	1.93	
Time: ESS round				
1	4.49(2.67)	4.08(2.94)	4.75(3.22)	22
2	6.92(5.73)	4.33(3.3)	3.63(3.71)	24
3	4.78(3.04)	4.02(3.18)	3.74(3.44)	23
4	5.23(4.41)	3.24(2.22)	5.39(6.66)	26
$F(df_1 = 1, df_2 = 88)$	0.00	1.18	0.02	
Payment bonus				
no	5.83(4.37)	4.41(3.10)	4.10(3.73)	54
yes	4.78(3.99)	3.23(2.52)	4.81(5.49)	41
$F(df_1 = 1, df_2 = 88)$	0.57	3.21^{+}	0.49	
Ratio controlled				
$F(df_1 = 1, df_2 = 88)$	0.11	0.51	1.09	
Ratio confirmed				
$F(df_1 = 1, df_2 = 88)$	3.11^{+}	0.11	0.00	
R^2	0.22	0.31	0.01	

Notes $***p < 0.001$, $^{+}p < 0.10$.

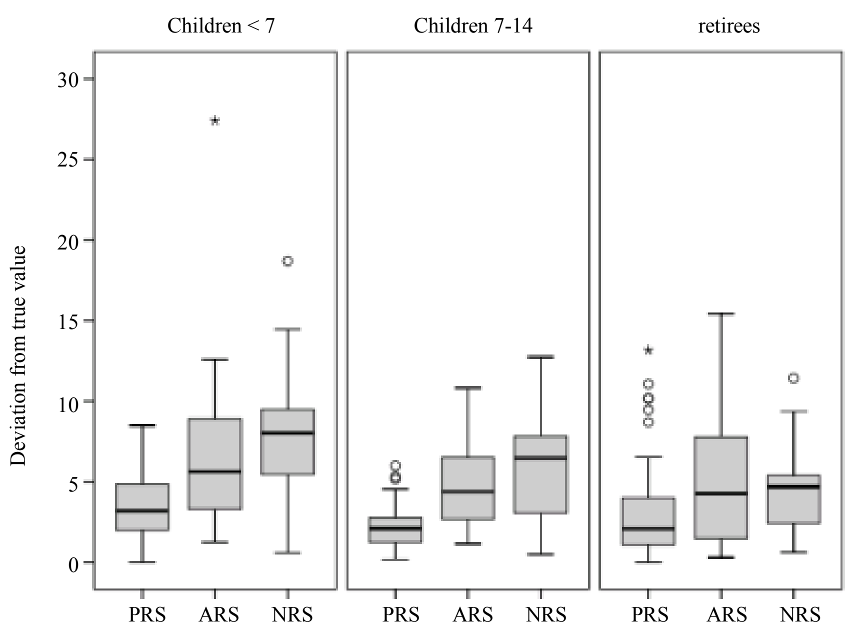

Figure 5.2 Box plots for absolute d-values are shown for different sampling methods in the three types of households

Table 5.2
Mean differences of $(d\,(SD))$ between sampling methods in subsamples with children

	children <7	children 7-14
differences between		
PRS and ARS	-3.34 (0.89)**	-2.66 (0.58)**
PRS and NRS	-4.58 (1.0)**	-3.71 (0.65)**
ARS and NRS	-1.24 (1.07)	-1.05 (0.7)

Note $**\ p < 0.01$. Single post *hoc* tests with Bonferroni correction.

5.3 The effect of explanatory variables

The effect of explanatory variables was analysed to test hypothesis H3, which expects deviations from the 50/50 gender ratio to be stable across time and to correlate with payment, interviewer controls and change of data collector.

Some countries in the ESS changed their sampling method procedures and/or data collector between the rounds (see appendix). The results showed that neither multivariate effects $(\text{WL}\ F_{(3,85)} = 0.81, p > 0.10)$ nor univariate effects are significant for the change of data collector. Thus, table 5.1 presents the ANCOVA results without this variable. If the "change of data collector" is included in the analyses, then the effect of the variable "ratio confirmed" is no longer significant, but this does not impact the effects of any of the other variables. This result shows that a change of data collectors may correlate with control procedures. The differences in d-values across the ESS rounds are not significant either, neither within multivariate analysis $(\text{WL}\ F_{(3,86)} = 0.51, p > 0.10)$ nor within the univariate analyses (for the latter see table 5.1).

Table 5.1 shows that in subsamples with children d-value means are lower if a payment bonus is used as compared to when it is not used. However, this difference is significant only on a 10% level $(p < 0.10)$ and only in households with older children. Hence, this result shows that payment methods may play a role, thereby reducing deviation from the true value in the case of higher payments.

Regarding control procedures, the number of controls ("ratio selected") is not related to the value of d (table 5.1). The success rate in controls ("ratio confirmed") is related to the value of d in the subsample with children younger than seven years old. This relationship is negative $(B = -0.06; SE = 0.04)$, meaning that the lower the confirmed control rates are, the higher the values of d are. However, this relationship is also significant only at a 10% level.

Concerning hypothesis H3, it has been shown that the effect of sampling methods is independent of the time effect. The results support the expectation of H3 concerning interviewer payment and controls. However, the results for these variables show that these effects are only weak and they can only be found in some household types.

6 Summary and conclusion

The results of the present study show that significant deviances from the population parameter (50% males) were seen in many ESS countries and that these were associated with the contactability or cooperation of partners in heterosexual couples (support for hypothesis H1). The magnitude of these deviances was found to differ between sampling methods when partners also differed in terms of their contactability (in subsamples with children). Thus, hypothesis H2 was partially supported. In subsamples with children, PRS was associated with the best data quality since the lowest deviances from the population parameter were found with this sampling method. However, the results for subsamples with retired partners show that highly pronounced deviances are also possible in the case of PRS.

The results for subsamples with children are in line with the explanation that interviewer behaviour related to substitutions is involved, since as expected deviations from the population parameter varied with the degree of interviewer freedom in influencing sample realization. Comparable results were reported by Sodeur (1997) and Kohler (2007). It is less plausible to explain interviewing males less than 50% of the time by differential refusal since in such cases the proportion of males is expected to be higher than 50%. Next, differential refusal is not expected to vary between different sampling methods. For retirees, interviewing males more than 50% of the time was found in several countries, but only in single rounds. This low stability of deviations from a 50% gender ratio can also be associated with interviewer impact instead of differential refusal, since the latter would be quite stable in a country over the period of time considered in the analysis. However, since the present study did not apply an experimental design it is important to address differential refusal and substitutions through further research in order to allow for better differentiation as well as for causal references.

Even though the deviations from a 50% population parameter varied in some countries across the rounds, overall their magnitude did not significantly change during the passage of time despite improved data collection procedures in the ESS (*cf.* Koch *et al.* 2009). Furthermore, deviations from the population parameter did not depend on the data collector nor were they country-specific.

The results also imply that interviewer payment and control procedures may reduce substitutions. However, it should be noted that only limited consideration of payment and control procedures was possible due to either their low variation in the data or limited information available in the survey documentation.

It should also be considered that the results presented here are based on specific subsamples and they cannot be used to generalise about the entire sample of the ESS. However, the “absence of a bias in the subsamples does not guarantee the absence of bias for the entire sample” (Kohler 2007, page 55). In addition, an analysis focused on special groups can often be of interest (*e.g.*, what are the opinions of parents with children or of employed people).

The results of the present study imply that PRS is associated with higher sample quality, meaning a lower non-representation bias in cross-cultural surveys than with other sampling methods. This is shown more clearly by the current study than by previous studies. Analyses using ALLBUS data by Sodeur (2007) and Hoffmeyer-Zlotnik (2006) only compared several rounds of one survey in a single country (Germany); in the analysis by Kohler (2007) a sampling method effect was confounded with the survey effect (see section 1). This has been avoided in the analysis presented here.

In conclusion, significant deviations from the population parameter, which appear to be associated with substitutions by interviewers, were observed in many countries of the ESS. In order to decrease this interviewer impact it is preferable to use sampling methods, such as PRS, with which the interviewers’ degree of freedom in selecting respondents and in influencing sample quality is reduced. In addition, survey procedures that increase interviewers’ motivation to produce accurate survey data are highly relevant and should be addressed by further research as well as by survey practices.

Appendix

Coding of ESS countries, change in sampling method and data collector, and state pension age for males and females in each country.

Country coding: ISO 3166-1	country	change sampling method (between rounds)	change data collector: between rounds	state pension age	
				males	females
BE	Belgium		1-2; 2-3	65	65
BG	Bulgaria		3-4	63	60
DE	Germany			65	65
DK	Denmark			65	65
EE	Estonia		2-3; 3-4	63	60
ES	Spain	ARS-NRS (1-2)	2-3	65	65
FI	Finland			65	65
HU	Hungary	PRS – ARS (2-3) and back (3-4)	2-3	62	62
NO	Norway			67	67
PL	Poland			65	60
SE	Sweden			65	65
SI	Slovenia			63	60
SK	Slovakia	PRS – NRS (3-4)	2-3; 3-4	62	59
CH	Switzerland			65	64
CZ	Czech Republic	ARS-NRS (1-2)	1-2	65	62
CY	Cyprus		3-4	65	65
GB	United Kingdom		1-2; 3-4	65	60
GR	Greece	NRS-ARS (2-4)		65	60
IE	Ireland		3-4	65	65
IL	Israel			67	64
IT	Italy			65	60
LU	Luxembourg			65	65
NL	Netherlands			65	65
TR	Turkey			47	44
AT	Austria			65	60
FR	France			60	60
PT	Portugal			65	65
RU	Russian Federation			60	55
UA	Ukraine			60	55

Notes Sources for state pension age:
1) http://www.oecd-ilibrary.org/finance-and-investment/pensions-at-a-glance-2011_pension_glance-2011-en
2) http://ec.europa.eu/employment_social/missoc/db/public/compareTables.do
3) Israel: http://www.btl.gov.il/

References

AAPOR (2003). *Interviewer Falsification in Survey Research: Current Best Methods for Prevention, Detection and Repair of Its Effects*. (http://www.aapor.org/pdfs/falsification.pdf; May 14 2009).

Arber, S. (2002). Design samples. In *Researching Social Life*, (Ed., N. Gilbert), Thousand Oaks: Sage, 58-84.

Billiet, J., Phillipsen, M., Fitzgerald, R. and Stoop, I. (2007). Estimation of nonresponse bias in the European Social Survey: Using information from reluctant respondents. *Journal of Official Statistics*, 23, 135-162.

Couper, M.P., and Groves, R.M. (1992). The role of the interviewer in survey participation. *Survey Methodology*, 18, 2, 263-277.

Couper, M.P., and De Leeuw, E.D. (2003). Nonresponse in cross-cultural and crossnational surveys. In *Cross-Cultural Survey Methods*, (Eds., J.A. Harkness, F.J.R. van de Vijver and P.Ph. Mohler), New York: John Wiley & Sons, Inc., 157-177.

De Heer, W. (1999). International response trends: Results of an international survey. *Journal of Official Statistics*, 15, 129-142.

De Heer, W., and Israis, A.Z. (1992). Response Trends in Europe. Paper presented to the American Statistical Association, August 1992.

De Leeuw, E., and De Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In *Survey Nonresponse*, (Eds., R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little), New York: John Wiley & Sons, Inc., 41-54.

De Leeuw, E.D., and Hox, J. (1996). The effect of the interviewer on the decision to cooperate in a survey of the elderly. In *International Perspectives on Nonresponse*, (Ed., S. Laaksonen). Helsinki: Statistics Finland, 46-52.

Dohrenwend, B.S., and Dohrenwend, B.P. (1968). Sources of refusals in surveys. *The Public Opinion Quarterly*, 32(1), 74-83.

Durrant, G.B., Groves, R.M., Staetsky, L. and Steele, F. (2010). Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly*, 74, 1-36.

ESS1-ESS4 data from Contact forms. Bergen, European Social Survey Data Archive, Norwegian Social Science Data Services. Retrieved June 8, 2013 from: http://ess.nsd.uib.no/ess/round1/download.html.

European Social Survey Round 1-4 Data (2011). Data file edition ESS1-4e01.0_F1. Norwegian Social Science Data Services, Norway – Data Archive and distributor of ESS data.

European Social Survey (2011). ESS-4 2008 Documentation Report. Edition 4.0. Bergen, European Social Survey Data Archive, Norwegian Social Science Data Services.

European Social Survey (2011). ESS-3 2006 Documentation Report. Edition 3.3. Bergen, European Social Survey Data Archive, Norwegian Social Science Data Services.

European Social Survey (2011). ESS-2 2004 Documentation Report. Edition 3.3. Bergen, European Social Survey Data Archive, Norwegian Social Science Data Services.

European Social Survey (2011). ESS-1 2002 Documentation Report. Edition 6.2. Bergen, European Social Survey Data Archive, Norwegian Social Science Data Services.

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2004). *Survey Methodology*. New Jersey: Wiley.

Haeder, S., and Lynn, P. (2007). How representative can a multi-nation survey be? In *Measuring Attitudes Cross-Nationally. Lessons from the European Social Survey*, (Eds., R. Jowell, C. Roberts, R. Fitzgerald and E. Gillian), London *et al.*: Sage, 33-52.

Hoffmeyer-Zlotnik, J.H.P. (2006). Stichprobenziehung in der Umfragepraxis. Die unterschiedlichen Ergebnisse von Zufallsstichproben in face-to-face umfragen. In *Stichprobenqualität in Bevölkerungsumfragen*, (Eds., F. Faulbaum and Ch. Wolf). Informationszentrum Sozialwissenschaften: Bonn, 19-36.

Hox, J.J., and De Leeuw, E.D. (2002). The Influence of interviewers' attitude and behavior on household survey nonresponse: An international comparison. In *Survey Nonresponse*, (Eds., R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little), New York: John Wiley & Sons, Inc., 103-118.

Johnson, T.P., O'Rourke, D., Burris, J. and Owens, L. (2002). Culture and survey nonresponse. In *Survey Nonresponse*, (Eds., R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little), New York: John Wiley & Sons, Inc., 55-69.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons, Inc.

Koch, A. (1995). Gefälschte Interviews: Ergebnisse der Interviewerkontrolle beim ALLBUS 1994. *ZUMA-Nachrichten*, 36, 89-105.

Koch, A., Blom, A.G., Stoop, I. and Kappelhof, J. (2009). Data collection quality assurance in cross-national surveys at the example of the ESS. *Methoden Daten Analysen*, 3, 219-247.

Kohler, U. (2007). Surveys from inside: An assessment of unit nonresponse bias with internal criteria. *Survey Research Methods*, 1, 2, 55-67.

Kreuter, F., and Kohler, U. (2009). Analyzing contact sequences in call record data. Potential and limitations of sequence indicators for nonresponse adjustments in the European Social Survey. *Journal of Official Statistics*, 25, 203-226.

Lynn, P., Haeder, S., Gabler, S. and Laaksonen, S. (2007). Methods for achiving equivalence of samples in cross-national surveys: The European Social Survey Experience. *Journal of Official Statistics*, 1, 107-124.

Pickery, J., and Loosveldt, G. (2002). A multilevel multinomial analysis of interviewer effects on various components of unit non response. *Quality and Quantity*, 36, 427-437.

Philippens, M., and Billiet, J. (2004). Monitoring and evaluating nonresponse issues and fieldwork efforts in the European Social Survey. Paper presented at the *European Conference on Quality and Methodology in Official Statistics*. Mainz, Germany.

Schnauber, A., and Daschmann, G. (2008). States oder traits? Was beeinflußt die Teilnahmebereitschaft an telefonischen Interviews? *Zeitschrift Für Empirische Sozialforschung*, 2, 97-123.

Schnell, R., Hill, P.B. and Esser, E. (2011). *Methoden der Empirischen Sozialforschung.* München: R. Oldenbourg Verlag.

Smith, T.W. (2007). Survey nonresponse procedures in cross-national perspective: The 2005 ISSP non-response survey. *Survey Research Methods*, 1, 45-54.

Sodeur, W. (1997). Interne kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen. *ZA-Information*, 41, 58-82.

Sodeur, W. (2007). Entscheidungsspielräume von Interviewern bei der Wahrscheinlichskeitsauswahl. *Methoden Daten Analysen*, 1, 2, 107-130.

Stoop, I.A.L. (2004). Surveying nonrespondents. *Field Methods*, 16, 23-54.

Stoop, I.A.L. (2005). *The Hunt for the Last Respondent*. The Hague: Social and Cultural Planning Office.

Symons, K., Matsuo, H., Beullens, K. and Billiet, J. (2008). *Response Based Quality Assessment in the ESS – Round 3: An Update for 19 countries*. London: Centre for Comparative Social Surveys, City University.

Vehovar, V. (1999). Field substitution and unit nonresponse. *Journal of Oficial Statistics*, 2, 335-350.

Williams, B., Irvine, L., McGinnis, A.R., McMurdo, M.E.T. and Crombie, I.K. (2007). When “no” might not quite mean “no”; the importance of informed and meaningful non-consent: results from a survey of individuals refusing participation in a health-related research project. *BMC Health Services Research*, 7, 59.