

Paradata in the European Social Survey: Studying Nonresponse and Adjusting for Bias

Ineke Stoop¹, Hideko Matsuo², Achim Koch³, Jaak Billiet⁴

¹SCP, P.O. Box 16164, 2500 BD The Netherlands, The Netherlands

²Katholieke Universiteit Leuven, CeSO, Parkstraat 45, 3601, 3000 Leuven, Belgium

³GESIS- Leibniz Institute for the Social Sciences, P.O. Box 12 21 55, 68072, Mannheim, Germany

Abstract

The ESS is a biennial face-to-face survey of attitudes, opinions and beliefs in around 30 European countries. The target response rate is 70%, but in practice response rates are often lower and vary across countries. This paper presents an overview of the paradata that are collected in the ESS and of the purposes for which they are used. It also describes the different auxiliary variables that have been used to assess and adjust for nonresponse bias: population statistics, interviewer observations, data from doorstep questionnaires and follow-up studies, and paradata derived from the ‘contact forms’. It also devotes attention to the balance between high national quality and optimal comparability across countries.

Key Words: Nonresponse, paradata, nonresponse bias, cross-national survey, optimal comparability

1. Introduction

Paradata (Couper, 2000) are collected for various purposes. They can be used to monitor the progress of fieldwork, to check whether fieldwork has been carried out according to specifications, to calculate response rates according to accepted standards (see American Association for Public Opinion Research, 2009), to identify reasons for nonresponse, to implement responsive designs (Groves & Heeringa, 2006), to explore whether non-response bias is probable, and to adjust for bias. Detailed paradata used to be scarce (see Stoop, 2005, pp. 60-63), especially in face-to-face studies, but nowadays many statistical offices and survey organizations keep detailed records of what goes on in the field.

This paper describes the collection and use of paradata in the European Social Survey (ESS). Optimal comparability of countries is pursued in this survey, which means that – ideally – all differences in substantive survey outcomes between countries should be ascribed to differences between populations and political, social and cultural institutions and systems, and not to differences in methodology and fieldwork efforts. The preparation of the ESS began at the end of the twentieth century. The collection of detailed paradata was planned right from the start. One of the main reasons for this was that the fieldwork for the ESS is carried out by individual survey agencies in around 30 countries; without standardized paradata collection, calculating harmonized response rates, studying the effects of field efforts, and assessing the impact of nonresponse bias would have been impossible. This paper aims to show that collecting paradata in a cross-national study like the ESS is important, necessary but far from simple. When trying to adjust for nonresponse, however, the best set of auxiliary variables – including paradata – should be used in each country.

2. European Social Survey: collecting paradata with contact forms

Collecting paradata in a cross-national survey, where many different organizations are conducting fieldwork, is more complicated than in a national survey conducted by a single organization. This applies in the case of the ESS (www.europeansocialsurvey.org), a biennial face-to-face survey of attitudes, beliefs and behavioral patterns held in more than 30 European countries, in which approximately 2,000 interviews are conducted in each country (Jowell et al., 2007). The ESS has an explicit methodological objective, namely to develop and demonstrate an approach to the conducting of rigorous quantitative multinational social surveys in Europe that matches that of the best national surveys. It aims to ensure a high level of consistency and standardization by applying a number of rigorous procedures and protocols in relation to sampling design, questionnaire and translation issues, and fieldwork specifications. One way to pursue optimal comparability in the ESS is by specifying a minimum target response rate of 70% and a maximum target noncontact rate of 3%.

Fieldwork in the ESS is closely monitored, and paradata are collected using detailed contact forms (see Stoop et al., 2003; Blom et al., 2008; Stoop et al., 2010) to record information on respondent selection within households, the timing and outcomes of each visit, and interviewer observations with regard to each contact as well as of characteristics of the household and the dwelling. The contact forms from each round of the survey are available at www.europeansocialsurvey.org (fieldwork documentation). They allow interviewers to record:

- interviewer number, target person name and telephone number;
- selection of households for address samples
- selection of persons for household samples
- call records: for each call
 - date, exact time, mode
 - result (complete/partial interview, different types of noncontact, ineligible)
- for every contact
 - outcome (interview, appointment, refusal (by target person or someone else)
 - different types of not being able to participate, language barrier)
- if ineligible
 - why invalid address
- for every refusal
 - reasons for refusal and judgments of possible cooperation at future calls
 - sex and age of doorstep contact
- for every sample unit:
 - neighborhood characteristic form (type of dwelling, physical state of building and neighborhood, presence of litter/rubbish, vandalism/graffiti in neighborhood).

Stoop et al. (2003) describe the history of the ESS contact forms; Blom et al. (2010) present a typology of the processes for generating contact data and discuss the implications for their utility for methodological research; and Stoop et al. (2010) illustrate in detail the difficulties of obtaining comparable and complete contact data in the ESS and the many uses that can be made of them. Bates et al. (2008) describe a similar instrument, developed at the US Census Bureau. One special feature of the ESS contact forms is that the resultant anonymized file is publicly available for secondary analysis (at <http://ess.nsd.uib.no/>), as are the ESS survey data. This means that researchers all over

the world can – and do – use this information to analyze nonresponse (for a particularly innovative analysis of contact sequences see Kreuter & Kohler (2009)).

The Central Coordinating Team (CCT) of the ESS uses information from the call records to provide feedback to fieldwork organizations for the next round, to check whether fieldwork has been carried out according to the specifications (which also serves as input for the design of fieldwork in the next round), and to analyze nonresponse. Using the contact forms, harmonized response rates can be calculated across countries, noncontacts can be distinguished from refusals, field efforts can be compared across countries, optimal visiting times can be predicted for each country, deviations from fieldwork requirements can be spotted, recommendations for fieldwork in future rounds can be drafted, and respondents can be classified according to field efforts in an attempt to analyze nonresponse bias. To date it has not been possible to use the information from the contact forms for responsive design purposes (Groves & Heeringa, 2006), but this may be possible in future rounds.

3. Response rates and nonresponse composition in the ESS

Figure 1 presents the ESS response rates in those countries that participated in each of the first four rounds¹. It will be clear that a 70% response rate is an ambitious target for most countries (though some countries that did not participate in every round exceeded the 70% target). A number of countries improve over time (e.g. Switzerland (CH), France (FR), Spain (ES)), whereas other countries (e.g. The Netherlands (NL), Germany (DE)) do worse.

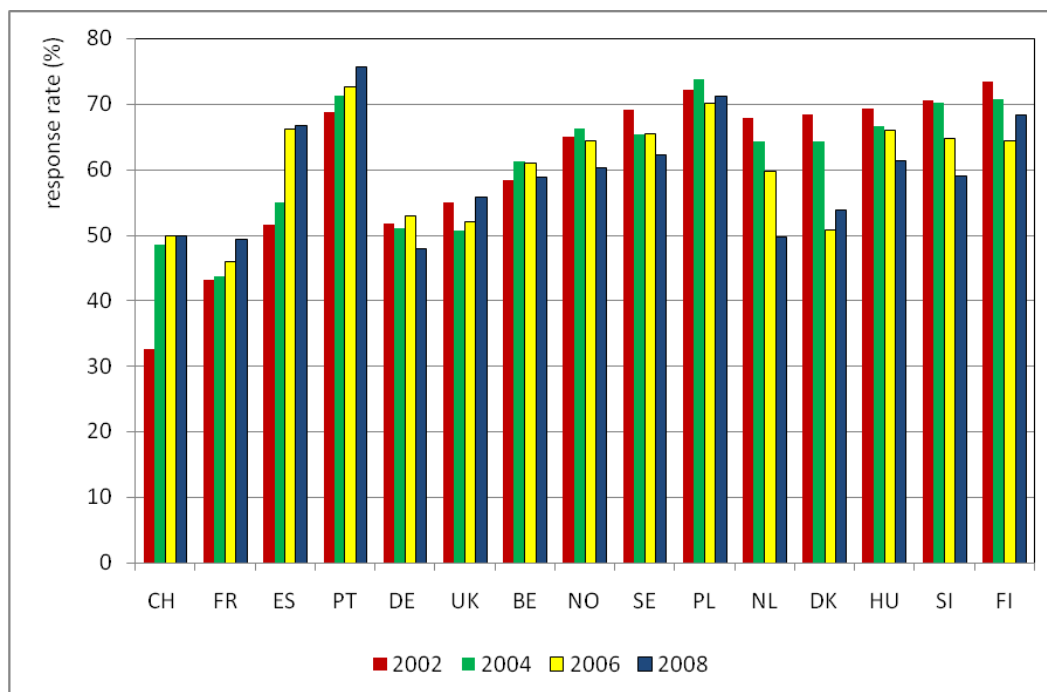


Figure 1: ESS response rates in countries participating in Rounds 1-4 (European Social Survey, 2010)

¹ This information is partly based on country reports as not all countries provided complete call records for the first four rounds of the ESS.

Figure 2 shows for Round 4 that the composition of the nonresponse (noncontact/refusal/other) differs across countries. However, in nearly all countries refusal constitutes the major part of nonresponse. The data for this figure are taken from Matsuo et al. (2010), who provide detailed information on all aspects of nonresponse in Round 4 for 24 countries, and also show how difficult it is to collect comparable data across countries.

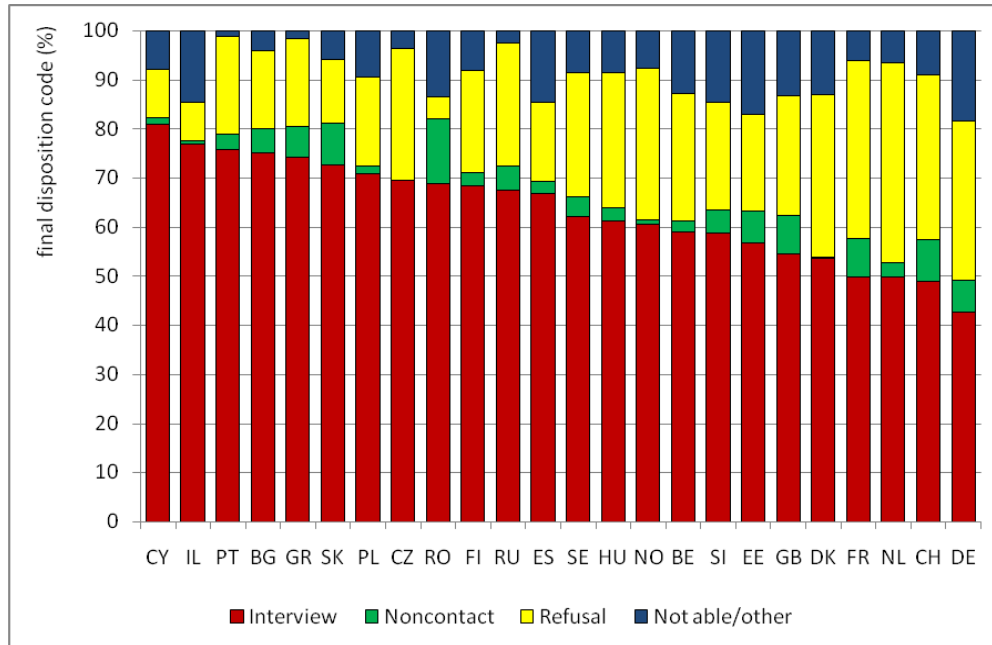


Figure 2: Final disposition codes in ESS Round 4 (Matsuo et al., 2010)

Further illustrations of the information to be obtained from the contact forms are given in figures 3 and 4, highlighting different aspects of ease of response and fieldwork efforts. One way to increase response rates is to increase the number of contact attempts; another is to re-approach initially reluctant respondents and try to convert them. These efforts should of course be seen in the context of general at-home behavior (few contact attempts are required when someone is at home in most households in a country) and willingness to participate (few refusal conversion attempts are required when most people are immediately willing to participate). A general indicator of survey difficulty (and of survey efforts) is the number of the call at which the interview is conducted. Figure 3 gives an overview of this indicator for 24 countries in Round 4 of the survey, based on Matsuo et al. (2010). It should be noted that in six countries a substantial proportion of first contact attempts were made by telephone (Sweden (SE) 91%; Finland (FI) 84%; Norway (NO) 77%; Germany (DE) 40%; Denmark (DK) 32%; and Estonia (EE) 28%). As the interview always had to be conducted face-to-face, response in these countries is much less often achieved at the first call.

Again, vast differences can be observed between countries. In The Netherlands (NL), for instance, the response rate at the first call was 3%, and a substantial percentage of respondents were only interviewed at the fifth or a later call. In Greece (GR), by contrast, a response rate of 71% was achieved at the first call, and all interviews were conducted during the first three calls. This could mean that no more calls were made after three unsuccessful visits in Greece, that additional calls were made without establishing contact, or that additional contacts did not result in an interview. The ESS contact forms contain the necessary information to be able to distinguish between these three situations.

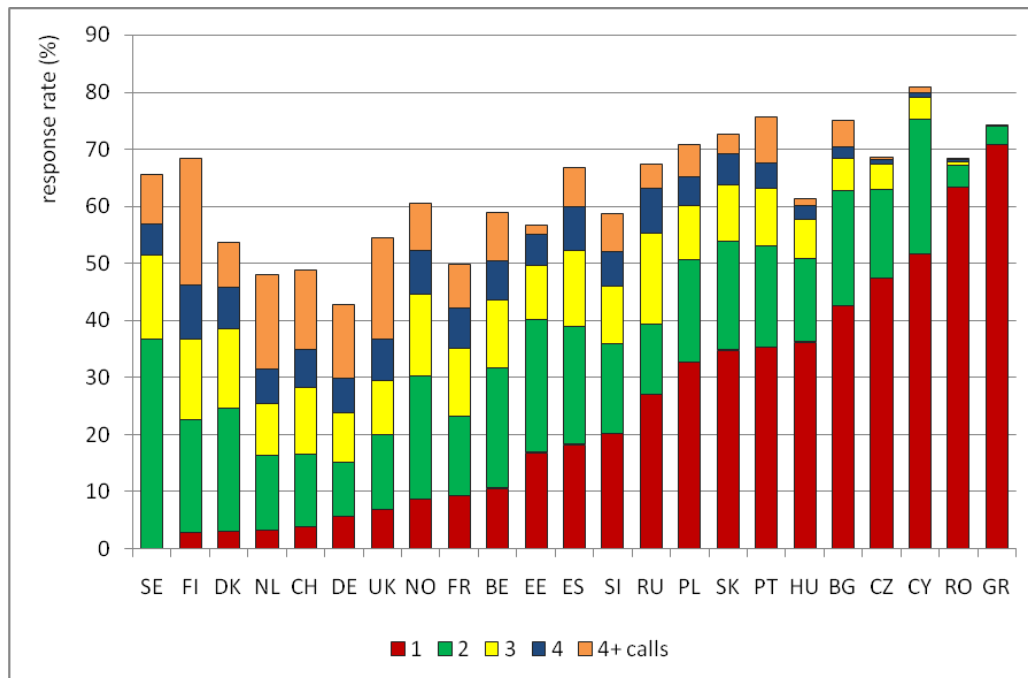


Figure 3: Response rate after one, two, three, four and more calls in ESS Round 4 (Matsuo et al., 2010)

The ESS project specifications prescribe at least four calls at different times of the day and on different days of the week. Matsuo et al. (2010) and Stoop et al. (2010) present detailed overviews of the effects of the number and timing of calls. These results make it possible to check whether the required number of calls was made in each country at the appropriate time, but also which time slot is most effective in each country. As expected, when calls are made in the evening or during the weekend, the probability of finding target persons at home is highest.

Figure 4 shows what happens in those cases where the first contact with a target person resulted in (initial) refusal. In Germany (DE), for instance, almost 2,800 persons said ‘no’ at the first contact. Almost half of them (1,283) were left alone; the others received a second visit. No contact was established with 53 of these 1,513 persons; 849 were re-contacted but did not participate, and 611 (22% of the initial refusals) ultimately cooperated. The pattern is quite different in The Netherlands (NL) where 1,882 target persons initially refused. In this country less than 10% were left alone, and 435 (23% of the initial refusals) ultimately cooperated. The efforts in The Netherlands were greater, but the net result was about the same. The costs of refusal conversion were therefore probably much higher in The Netherlands, but on the other hand, the efforts were directed at all refusals and not just at a selection. Evidence from previous rounds showed that most refusal conversion efforts in Germany were directed at the most promising cases (Stoop et al., 2010). In some other countries (Cyprus (CY) and Romania (RO)) hardly any refusal conversion attempts were made, though these countries had very or fairly high response rates and low refusal rates to begin with (see figure 2). In the remainder of this paper we will largely ignore response composition, because in the ESS noncontact rates can be kept low in most countries and the main cause of nonresponse is generally refusal.

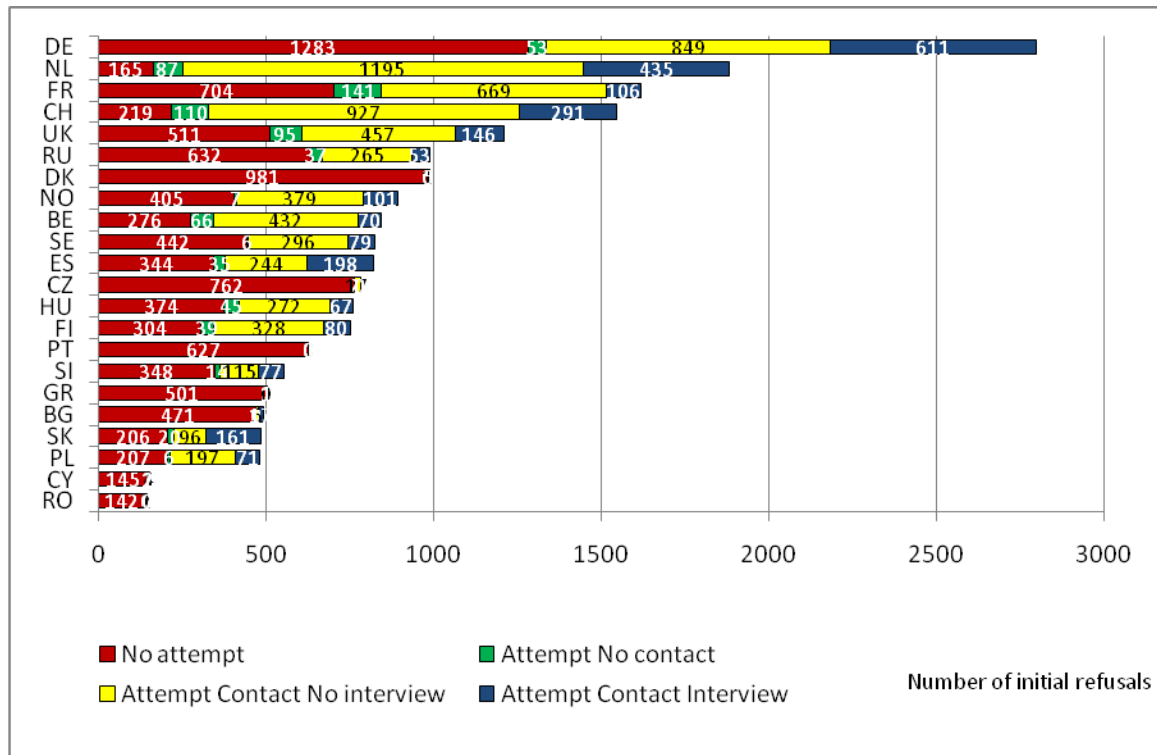


Figure 4: Effect of refusal conversion efforts in 24 ESS countries, Round 4 (Matsuo et al., 2010)

4. Assessing and adjusting for nonresponse bias

4.1 Studying nonresponse bias in a cross-national survey

As Groves (2006) has shown, there is no linear relationship between response rates and nonresponse bias *across* surveys; nor has it been shown that within a particular survey efforts to enhance response rates will automatically reduce nonresponse bias (see Keeter et al., 2000; Stoop, 2004). Until recently there were no empirical studies exploring whether nonresponse bias in a cross-national study will be greater in low-response countries than in high-response countries. Could it be that those who are not interested in politics cooperate in a high response rate country such as Cyprus (CY) but not in a low response rate country such as Switzerland (CH)? Could differences in nonresponse composition (noncontact and refusal) result in different types of nonresponse bias across countries despite the response rates being similar, for instance when refusal to participate in the survey is related to opinions on Europe and where countries with a similar response rate show quite different refusal rates (see the Czech Republic (CZ) and Romania (RO) in figure 2)? Using information on nonresponse from the European Social Survey we can now start to address these questions. It soon becomes clear, however, that assessing nonresponse bias in a study of values, attitudes and opinions is far from simple, as there is no gold standard, no true value, no external information with which to gauge the survey outcomes. We could compare ESS results with results from national surveys, but differences in question formulation, response rates in those national surveys and other survey quality issues will stand in the way, and we will not know which outcome is the best or most plausible.

We will therefore be more modest in our ambitions, and use four standard methods to estimate nonresponse bias and to suggest ways of adjusting for it (Billiet et al., 2009; Stoop et al., 2010). Firstly, the final sample composition in terms of background variables was weighted with population statistics (post-stratification). Weighted results could then be compared with unweighted results, helping to identify those variables that might be particularly prone to bias. Methods two and three used paradata, namely dwelling and neighborhood information as observed and recorded by the interviewers, and ease of cooperation derived from the call records. Observational data could be used as a proxy for individual household data to identify which factors might hinder interviewing, and using the third method we tried to ascertain whether reluctant respondents differ from cooperative respondents (and could thus possibly be used as a proxy for final refusals). Finally, in an effort to create a gold standard, in a few countries information was collected from refusals on the doorstep and from refusals, noncontacts and participants in a follow-up survey. This information could also shed light on the burning question of whether initial refusals are similar to final refusals.

4.2 Population-level information and post-stratification

Population-level information drawn from national statistics on age, sex and education was used by Vehovar (2007) for post-stratification weighting. He found small differences between population and ESS data for sex and age in most countries, and larger differences with respect to education. However, different groups were underrepresented in different countries. After weighting, the median relative change in the estimates of a large number of relevant ESS variables was about 1.4% of their values, generally due to the impact of education. The largest effects occurred in the areas of media use (especially newspaper reading), political interest, attitudes towards immigrants, and religion. Vehovar found that the higher the response rate, the lower the average absolute standardized bias² of the items he selected for his tests (figure 5), although the correlation is very modest. The size of the bubbles in the figure indicates the number of items with a significant bias. These results underline the importance of aiming for high response rates. One negative effect of weighting is inflation of variance. As Vehovar (2007) has shown, this effect should be taken into account when presenting the results of weighting.

Weighting with population-level background data has certain limitations. Firstly, the quality of the available population-level data differed across countries; in addition, the quality of educational variables in the survey itself differed. Secondly, in the ESS – as in many studies – socio-demographic data will not be the key variables that show high covariance with attitudes. Survey outcomes, whether univariate results or results derived from complex analyses, therefore hardly differ when they are weighted using post-stratification weights from the unweighted results. It is therefore questionable whether this procedure, based on this information, can correct for nonresponse bias. On the other hand, it can provide an indication as to which variables might be particularly susceptible to nonresponse bias. Questions identified by Vehovar as most likely prone to nonresponse bias and questions identified in previous studies (Billiet et al., 2007) (media use, political interest, attitudes towards immigrants, and voluntary work) are used in the methods for studying nonresponse bias presented below.

² Bias here is the difference between estimates in the unweighted and weighted sample; standardized bias divides this difference by the standard error; absolute standardized error ignores the sign and in the figure absolute standardized bias is averaged across items.

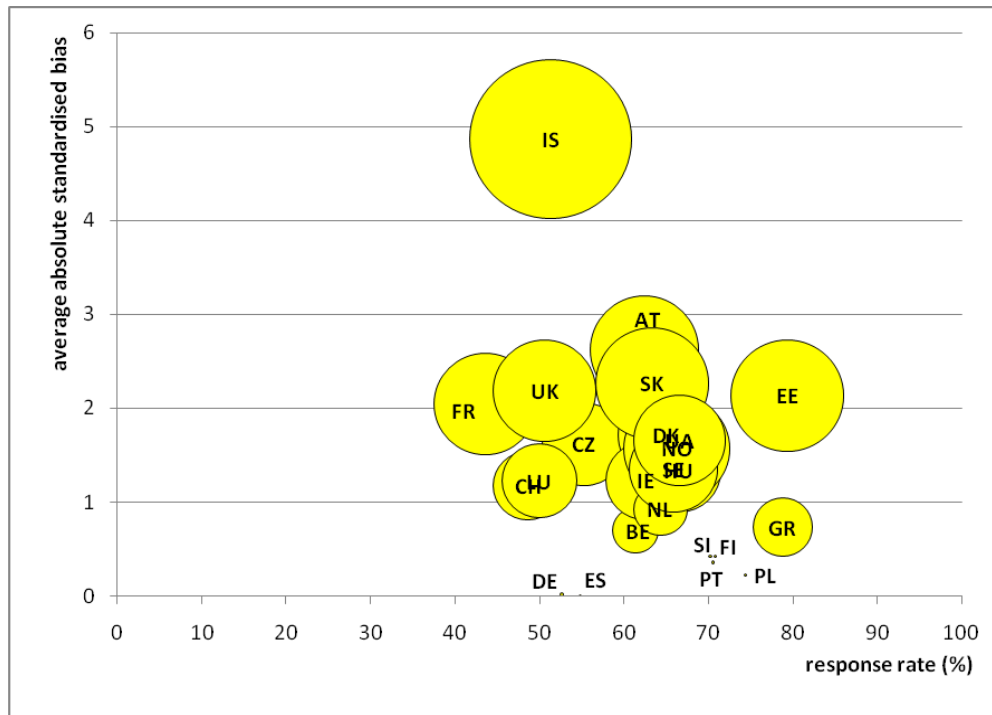


Figure 5: Average ASBias and response rate in ESS 2 (based on Vehovar, 2007)

4.3 Interviewer observations

Possibly a more useful source of information than national population statistics might be neighborhood data that are used for marketing, policymaking and administrative purposes. Neighborhood-level data (socioeconomic status, house prices, population density, crime rates - , see also Groves & Couper, 1998 and Johnson et al., 2007) are probably more closely related to household characteristics and core variables than population-level data, and thus more useful in adjusting for nonresponse. However, this type of information will not be available in every country in a cross-national study, and definitely not at the same level of detail or based on the same classifications.

Therefore, basic information on neighborhoods and dwellings is collected by ESS interviewers themselves. On the contact forms they record information on the type and physical state of the dwelling and the presence of vandalism and litter in the neighborhood (see the *Neighborhood Characteristics Form* in table 1). An analysis of these data is presented in Cincinnato et al. (2008), Billiet et al. (2009) and Stoop et al. (2010). The physical state of the buildings showed a negative relationship with the education level of target persons. Refusals and noncontacts were more likely in areas characterized by poor physical state of the buildings and the presence of litter and/or vandalism. This could reflect characteristics of the residents, but we cannot rule out the possibility that the differences also to some degree reflect interviewers acting differently in different areas.

A great advantage of these observational data is that they can in theory be collected for every sample unit in every country (although in practice privacy rules in a number of countries forbid this). A disadvantage of the ESS observational data is that there are

missing data, and presumably also differences in the way different interviewers within countries and different survey agencies interpreted fairly abstract concepts such as the commonness of litter lying around. In fact, it is not simple to develop universal classifications for neighborhoods and dwellings that can be applied in the Ukrainian countryside, Alpine mountains, the City of London, the suburbs of Naples and Norwegian islands. Consequently, these classifications should always be seen within a national context.

Table 1: Neighborhood Characteristics Form used in the ESS

NEIGHBORHOOD CHARACTERISTICS FORM	
N1. In what type of house does the respondent live? (<i>dwelling type</i>)	
1 Farm	
Single-unit:	
2 Detached house	
3 Semi-detached house	
4 Terraced house	
5 The only housing unit in a building with another purpose (commercial property)	
Multi-unit:	
6 Multi-unit house, flat	
7 Student apartment, rooms	
8 Retirement home	
Other:	
9 House-trailer or boat	
10 Other:.....	
88 Don't know	
N2. In what physical state are the buildings or dwellings in this area? (<i>physical state</i>)	
1 In a very good state	
2 In a good state	
3 In a satisfactory state	
4 Poor state	
5 Very poor state	
N3. In the immediate area, how common is litter or rubbish lying around? (<i>litter</i>)	
1 Very common	
2 Fairly common	
3 Not very common	
4 Not at all common	
N4. How common is vandalism, graffiti or deliberate damage to property? (<i>vandalism</i>)	
1 Very common	
2 Fairly common	
3 Not very common	
4 Not at all common	

We have seen that there is a relationship between this kind of paradata (interviewer observations on neighborhoods and dwellings) and the likelihood of obtaining a response. To make these data really useful in nonresponse adjustment, however, they would have to be related to substantive variables of the survey. We have no conclusive evidence for this at present. Extra guidelines and training have been developed for ESS Round 5 to improve the collection of these observational data.

In a number of countries individual background data are available for every sample unit. Statistics Netherlands, for instance, has information on all sample units drawn from the population register which can be linked to a wide range of public records and administrative files containing information on country of birth, employment, social security benefits, etc. Successful examples of how this approach can adjust for bias come inter alia from the work carried out by Schouten and Cobben (2006) and Cobben and Schouten (2008) on indicators of representativeness. Unfortunately, the number of

countries where this is possible is limited, and there is generally no access to these auxiliary data outside National Statistical Institutes.

In order to have some information on each sample unit, both respondents and nonrespondents, ESS has again developed an alternative. Interviewers were instructed to record the sex and approximate age of the target respondent (refusers, refusal proxy) on the contact form ('dwelling' can also be seen as an individual characteristic; this was discussed above). Two problems arise here. Firstly, there has to be contact with the target person in order to know whether it is a man or a woman and to estimate his or her age. Secondly, it will not always be clear whether a contact person who refuses to cooperate is the target person or simply another household member. In practice, especially when the sample was not a sample of individuals, there were too many item-missings in these auxiliary data for them to be of much use. Moreover, even if age and sex information had been available for every sample unit, the experiences from the post-stratification study cast doubt on the usefulness of weighting by sex and age.

4.4 Paradata from refusal conversion

Earlier analyses (Stoop et al., 2010) of the ESS paradata showed that those respondents who required many calls (mainly because they were not at home) did differ substantially and systematically from those who were easy to contact with respect to age, labor market participation and education, but differed relatively little, and unsystematically, as regards core variables of the survey such as ethnocentrism, political interest and social trust. Final noncontact can be minimized, however, and the main and persistent cause of nonresponse in the ESS is refusal (see figure 2). Therefore, we will focus our overview of the uses of paradata on the process and analysis of refusal conversion.

The differential efforts and outcomes of refusal conversion can be illustrated by the first results from ESS Round 4 (Matsuo et al., 2010), presented in figure 4. The number of initial refusals, the percentage of initial refusals who are re-approached, the percentage of re-approached refusals who are found at home, and the percentage of re-approached, contacted initial refusals who finally participate, vary substantially across countries. Due to vast differences in the number of converted refusals, using paradata on refusal conversion as a means of studying nonresponse bias is feasible in only a small number of countries.

Billiet et al. (2009) (see also Stoop et al., 2010) compare the refusal conversion process in two ESS countries (Germany and The Netherlands) in which large enough numbers of initial refusals were converted. In their explanatory models of refusal conversion success, they factored in whether or not the refusal was by the target respondent or a household member (proxy refusal), the interviewer's assessment of future cooperation after a refusal,³ and whether the refusal conversion attempt was made by the same or a different interviewer⁴. However, the reissuing process differs in the two countries; in Germany the decision to re-approach initial refusers seems to be based partly on the interviewer's assessment of future success (higher reissue rates for sample units with a higher chance of future cooperation), whereas in The Netherlands almost all initial refusals were re-

³ Dutch and German interviewers estimate that the target respondent will participate after an initial refusal when the reason for refusal was 'no time' or 'not interested', and that they will not participate when they state that surveys are 'a waste of time' or that they 'never do surveys'.

⁴ An additional indicator is the number of times the target respondent refused before finally cooperating. The Netherlands was however the only country where a sizeable number of refusals was re-approached more than once.

approached. Surprisingly, the success rate of a conversion attempt was related to the interviewer's judgment in Germany, but not in The Netherlands. In both countries new interviewers were much more likely to convert initial refusals than the original interviewers.

Using the contact forms in the European Social Survey, sample units who cooperate immediately can be distinguished from those who initially refused (the so-called reluctant respondents or converted refusals). In a few countries it was also possible to subdivide converted refusals into those who were easy to convert and those who were hard to convert. Billiet et al. (2007) present the results of a comparison between those who cooperated immediately in ESS Round 1, the easy-to-convert refusals and the hard-to-convert refusals. Easy-to-convert refusals are those who were easy to persuade and who agreed to cooperate at the second contact with the interviewer. Hard-to-convert refusals participated after several attempts and sometimes only when special incentives were brought in to persuade them. One of the major findings of that study was that converted refusals did differ on some attitudinal scales and background variables from cooperative respondents, and easy-to-convert from hard-to-convert initial refusals, but the pattern of differences varied across countries.

Earlier studies (Curtin et al. 2000; Keeter et al., 2002; Stoop, 2005) found that reluctant respondents (or converted refusals) were similar to cooperative respondents. In addition, in some cases the process of refusal conversion actually makes the composition of the final sample worse: subgroups that were overrepresented before refusal conversion are more likely to be converted than underrepresented subgroups (Stoop, 2005). A similar result was found by Billiet et al. (2009) and Stoop et al. (2010) for the ESS. This suggests that refusal conversion may not be the best strategy either to increase response rates or to minimize bias.

Paradata can provide information on ease of cooperation, but also on the underlying reasons for cooperation and refusal. A number of studies have attempted to distinguish between different types of refusals and related these, as well as doorstep comments from cooperative respondents, initial refusals and ultimate nonparticipants, to response outcomes (Smith, 1984; Hox et al. 1995; Couper, 1997; Campanelli et al., 1997; Rogelberg et al., 2001, 2003; Voogt, 2004; Bates et al., 2008). These studies conclude that target persons supply meaningful comments and that different reasons for refusal can have a different impact on nonresponse bias. In the ESS, no information on the doorstep interaction between immediate respondents and interviewers is available, but those who initially refuse can be classified by their reason or reasons for refusal. Stoop (2007) compared “no time” and “too busy” as reasons for refusal across ESS countries and found that pressure of time is mentioned (or at least recorded) much more often in some countries than in others, but that there is little evidence that busy individuals cooperate less in surveys.

Given the differences in need, procedure, and effect of refusal conversion between countries, the study of converted refusals as a proxy for final refusals is limited to a very small number of countries. Refusal conversion does not necessarily improve the composition of the final sample in terms of socio-demographics. The remaining question is whether converted refusals are (more) similar to final refusals. This question will be addressed in the next section.

4.5 Core information on nonrespondents

Direct information on nonrespondents can be obtained by asking nonrespondents about the topics of the survey. Two options are available here, the doorstep approach and the follow-up survey. In a follow-up survey refusals and noncontacts (and also respondents, in order to control for context, mode and timing effects) are surveyed subsequent to the data collection in the original study. Stoop (2004, 2005) achieved a very high response rate in a follow-up study and found that converted refusals were not similar to final refusals. A study by Prybysz et al. (2007) in Poland compared refusals in a follow-up of the pilot-ESS in 2004, and again of the main study. The types of bias in the two studies, based on an identical questionnaire, were different. In the main study, the response rate in the follow-up survey was lower; probably because the reminder was sent on the day the Pope died.

In ESS Round 3 a follow-up survey was carried out in three countries, namely Poland (PL), Norway (NO) and Switzerland (CH). The first two usually have a high response rate in the ESS, while the third is below average (see figure 1). An abbreviated questionnaire was compiled in two versions, one of four pages with 17 questions, and an even shorter one with seven questions. The results from this semi-experiment, as reported in Stoop et al. (2010) seem promising. As might be expected, response rates among the nonrespondents in the main study were modest: 23% in Poland, 30% in Norway, and 52% in Switzerland. In the follow-up survey, respondents in the main survey were also re-approached: their response rate was much higher (around 60% in Poland and Norway, and 84% in Switzerland). It is clear that response rates in the follow-up survey, from both participants and nonrespondents in the main survey, were lower in those countries that had a high response in the main survey.

The substantive results of the study differed from country to country. In Norway, for instance, nonrespondents and respondents who participated in the follow-up survey differed from each other. Respondents who participated in the follow-up survey ('double cooperators') also differed from all respondents in the main survey. This could be due to selection, timing, mode, and/or context effects. Billiet and Matsuo (as reported in Stoop et al., 2010) developed propensity weights that could adjust for nonresponse bias. One slightly disappointing result was that there was virtually no difference between the weighted and unweighted distributions of the Norwegian respondents in the main survey. This could indicate that there was no bias, or that the weighting procedure was not able to adjust for bias. A more promising result was that differences between respondents in the main survey and nonrespondents who participated in the follow-up survey disappeared after weighting.

In the doorstep approach, refusals are asked to answer at least a small number of questions; this approach is also called the Basic Question Approach of Kersten and Bethlehem (1984) (see also Elliot, 1991) or the Pre-Emptive Doorstep Administration of Key Survey Items (PEDAKSI) approach by Lynn (2003). Lynn found that nonrespondents who provided Key Survey Items were in some respects similar to survey respondents (entryphone, ethnic minority neighborhood) and in other respects similar to nonrespondents (security devices, house condition, social status of neighborhood). They appeared to have experienced fewer crimes than the respondents. Van Ingen et al. (2009) conducted a similar doorstep study as part of a Time Use Survey. They found that the nonrespondents who provided key information read fewer newspapers, were less interested in politics, did less voluntary work, were less involved in sports and spent less time traveling and working. They were not busier than the respondents.

In the ESS the doorstep approach was used in Belgium during the Round 3 fieldwork. 45% of the refusals provided information for the doorstep questionnaire. This approach can in principle be used in every country to obtain information about refusals. No information on noncontacts can be obtained in this way, but – as mentioned earlier – the noncontact rate is usually low in the ESS. A potential problem is that when individual sample frames are used it is more likely that the person who answers the basic questions is the target person than is the case with address samples. Consequently, this approach is less feasible for use in combination with household or address samples. A second problem is that this approach usually has a low response rate. Finally, it is important to avoid the Basic Question Survey being used as an easy way out for the interviewer, which will make refusal more acceptable.

The results of the analysis of the Belgian doorstep questions (Stoop et al., 2010, p. 260) indicate that refusal conversion and doorstep questions pick up different types of refusals (see figure 6). Converted refusals (called reluctant respondents here) come in between cooperative respondents and refusals in terms of political interest, but are quite similar to cooperative respondents with regard to social participation. As regards education level, the results are mixed. Converted refusals therefore cannot automatically be used as proxies for final refusals, a result also found in other studies. It should be borne in mind that these studies of final refusals have one serious shortcoming: they ignore possible nonresponse bias on the doorstep and follow-up studies.

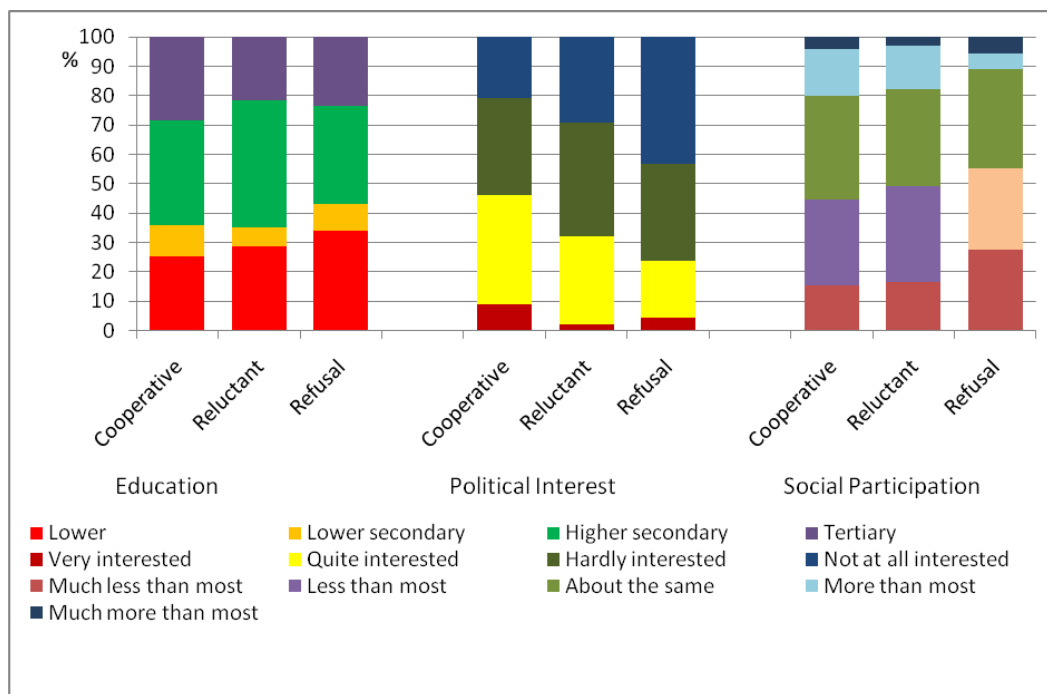


Figure 6: Responses to key questions in Belgian Doorstep Survey, ESS Round 3

5. Auxiliary data in the ESS

To adjust for nonresponse bias it is necessary to have auxiliary data – including paradata – that are related to response behavior and to target survey data (Bethlehem, 2002). Using paradata to adjust for nonresponse in a cross-national survey is complicated for three

reasons. Firstly, paradata should be related to target survey data; in a cross-national study this might not be the case in all countries, or the relationship might be different in different countries. Secondly, it is difficult to obtain standardized paradata when different survey organizations are involved, working in different survey cultures and with different national traditions, regulations and legislations that have to be taken into account. And finally, a trade-off has to be made between two potentially conflicting goals. One goal, which also applies in national surveys, is to optimally adjust for nonresponse bias. However, in a cross-national survey comparability across countries is also a major goal, and a core criterion of survey quality (Lyberg et al., 2001; Eurostat, 2005). In the ESS the aim of optimal comparability is pursued by following the ideal model of input harmonization (Körner & Meyer, 2005; see also Lynn, 2003a). This means that the instrument and procedures are kept as near the same as possible, as is shown by the standard specifications for the survey (see e.g. European Social Survey, 2007). The question is therefore whether relatively simple but universally implementable weighting procedures (such as post-stratification) are better suited for the ESS than innovative and probably more effective weighting procedures that can be implemented only in a small number of countries, due to differences in the availability and quality of paradata and other auxiliary variables. Put differently, excellent adjustment for nonresponse in one country might be less than excellent if there is no or poor adjustment in other countries.

From the present studies of paradata in the ESS, we can conclude that in a cross-national study paradata need to be collected to control and compare the data collection process, and to suggest ways to analyze and reduce nonresponse. In addition, the different types of paradata in the ESS need to be improved through standardization, interviewer training, clearer guidelines and other quality efforts. Collecting core information on nonrespondents is a promising approach, but the results from one country cannot be generalized to other countries, if only because the response rates vary so much. Given the restrictions, it seems unlikely that one single approach is the most appropriate way to correct for nonresponse. The most promising approach would seem to be to select the best auxiliary data for each country and to combine the auxiliary data described in the previous section to create one propensity score model. It will not be clear whether this strategy will make the nonresponse adjusted data more comparable across countries, and one important basic principle is therefore that the paradata from the contact forms should be as comparable as possible, and that additional information should be collected on national/organizational fieldwork procedures and on practical decisions that are taken in the field.

References

- American Association for Public Opinion Research (2009) *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 6th edition. Lenexa, Kansas, AAPOR.
- Bates, N., J. Dahlhamer and E. Singer (2008) Privacy Concerns, Too Busy, or Just Not Interested: Using Doorstep Concerns to Predict Survey Nonresponse. *Journal of Official Statistics*, 24, 4, 591-612.
- Bethlehem, J. G. (2002) Weighting Nonresponse Adjustments Based on Auxiliary Information. In: R. M. Groves, D.A. Dillman, J. L. Eltinge and R.J.A. Little (eds) *Survey Nonresponse*. New York: Wiley, 265-287.
- Billiet, J., H. Matsuo, K. Beullens and V. Vehovar (2009) Non-response Bias in Cross-national Surveys: Designs for Detection and Adjustment in the ESS. *ASK: Society, Research, Methods* (18/2009).
- Billiet, J., M. Philippens, R. Fitzgerald and I. Stoop (2007) Estimation of Nonresponse Bias in the European Social Survey: Using Information from Reluctant Respondents. *Journal of Official Statistics*, 23, 2, 135-162.
- Blom, A.G., P. Lynn and A. Jäckle (2008) *Understanding Cross-National Differences in Unit Non-Response: The Role of Contact Data*. ISER Working paper 2008-01.
- Campanelli, P., P. Sturgis and S. Purdon (1997) *Can You Hear Me Knocking: An Investigation into the Impact of Interviewers on Survey Response Rates*. London: The Survey Methods Centre at SCPR.

- Cincinatto, S., K. Beullens & J. Billiet (2008) *Analysis of Observable Data in Call Records ESS – R2*. Deliverable n° 6 of ESSi-JRA2. CeSO, K.U. Leuven.
- Cobben, F., and B. Schouten (2008) An empirical validation of R-indicators. Voorburg, Statistics Netherlands.
- Couper, M.P. (1997) Survey Introductions and Data Quality. *Public Opinion Quarterly*, 61, 317-338.
- Couper, M. P. (2000) Usability Evaluation of computer-assisted survey instruments. *Social Science Computer Review*, 18, 384-396.
- Curtin, R., S. Presser and E. Singer (2000) The Effects of Response Rate Changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, 64, 413-428.
- De Heer, W. (1999) International Response Trends, Results of an International Survey. *Journal of Official Statistics*, 15, 2, 129-142.
- Elliot, Dave. 1991. *Weighting for non-response. A survey researcher's guide*. London: Office of Population Censuses and Surveys.
- European Social Survey (2007) *Round 4 Specification for participating countries*. London: Centre for Comparative Social Surveys, City University London. www.europeansocialsurvey.org.
- European Social Survey (2010) *ESS4-2008 Documentation Report*. Edition 3.01. Bergen, European Social Survey Data Archive, Norwegian Social Science Data Services.
- Eurostat (2005) *European Statistics Code of Practice*.
- Groves, R.M. (1989) *Survey Errors and Survey Costs*. New York: John Wiley & Sons, Inc.
- Groves, R.M. (2006) Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, Vol. 70, 646-675.
- Groves, R.M., and M.P. Couper (1998) *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, R.M., and S. Heeringa (2006) Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 169, Part 3, 439-457
- Hox, J., E. de Leeuw and H. Vorst (1995) Survey Participation as Reasoned Action; a Behavioral Paradigm for Survey Nonresponse? *Bulletin de Méthodologie Sociologique*, 48, September 1995, 52-67.
- Johnson, T.P., Y.I. Cho, R.T. Campbell and A.L. Holbrook (2006) Using Community-Level Correlates to Evaluate Nonresponse Effects in a Telephone Survey. *Public Opinion Quarterly*, 70, 704-719.
- Jowell, R., M. Kaase, R. Fitzgerald and G. Eva (2007) The European Social Survey as a measurement model. In: R. Jowell, C. Roberts, R. Fitzgerald and G. Eva (eds) *Measuring Attitudes Cross-Nationally. Lessons from the European Social Survey*. London, Sage, 1-31.
- Keeter, S., C. Miller, A. Kohut, R.M. Groves and S. Presser (2000) Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly*, 64, 125-148.
- Kersten, H.M.P., and J.G. Bethlehem (1984), Exploring and reducing the nonresponse bias by asking the basic question, *Statistical Journal of the U.N. Economic Commission for Europe*, 2, 369 – 380.
- Kreuter, F., and U. Kohler (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 (2), 203–226.
- Lyberg, L., et al. (2001) *Summary Report from the Leadership Group (LEG) on Quality, SPC*.
- Lynn, P. (2003) PEDAKSI: Methodology for Collecting Data about Survey Non-Respondents. *Quality and Quantity*, 37, 239-261.
- Matsuo, H., J. Billiet & G. Loosveldt (2010) *Response-based quality assessment of ESS Round 4: Results for 24 countries based on contact files*. Leuven: European Social Survey, University of Leuven.
- Przybysz, D., P. Sztabiński, F. Sztabiński (2007) Are Non-respondents similar to respondents? Findings from the ESS-2004 in Poland. *ASK: Society, Research, Methods* (16/2007), 25-54.
- Rogelberg, S.G., J.M. Conway, M.E. Sederburg, C. Spitzmüller, S. Aziz, W.E. Knight (2003) Profiling Active and Passive Nonrespondents to an Organizational Survey, *Journal of Applied Psychology*, 88, 6, 1104–1114.
- Rogelberg, S.G., G.G. Fisher, D.C. Maynard, M.D. Hakel and M. Horvath (2001) Attitudes Toward Surveys: Development of a Measure and Its Relationship to Respondent Behavior. *Organizational Research Methods*, 4 1, 3-25.
- Schouten, B., and F. Cobben (2006) R-Indices for the Comparison of Different Fieldwork Strategies and Data Collection Modes. Paper presented at the International Household Survey Nonresponse Workshop, Omaha, 2006. Voorburg, Statistics Netherlands.
- Smith, T.W. 1984. Estimating Nonresponse Bias with Temporary Refusals. *Sociological Perspectives*, Vol. 27, N° 4, 473-489.
- Stoop, I.A.L (2004) Surveying Nonrespondents. *Field Methods*, Vol. 16, nr. 1, 23-54.
- Stoop, I.A.L (2005) *The Hunt for the Last Respondent*. The Hague: SCP.
- Stoop, I. (2007) No time, too busy. Time strain and survey cooperation. In: G. Loosveldt, M. Swyngedouw and B. Cambré (eds) *Measuring Meaningful Data in Social Research*. Leuven: Acco, 301-314.
- Stoop, I., S. Devacht, J. Billiet, G. Loosveldt and M. Philippens (2003) *The development of a uniform contact description form in the ESS*. Paper presented at the 14th International Workshop on Household Survey Nonresponse, Leuven, September 2003.
- Van Ingen, E., I. Stoop, and K. Breedveld (2009) Nonresponse in the Dutch Time Use Survey: Strategies for Response Enhancement and Bias Reduction. *Field Methods*, Vol.21, No. 1, 69-90.
- Vehovar, V. (2007) Non-response Bias in the European Social Survey. In: G. Loosveldt, M. Swyngedouw and B. Cambré (eds) *Measuring Meaningful Data in Social Research*. Leuven: Acco, 335-356.
- Voogt, R. (2004) *"I'm Not Interested". Nonresponse bias, response bias and stimulus effects in election research*. Academisch Proefschrift. Amsterdam: Universiteit van Amsterdam.