ISER Working Paper Series

Nonresponse Bias Adjustments: What Can Process Data Contribute?

Annelies G. Blom
GESIS - Leibniz Institute for the Social Sciences
ISER, University of Essex

No. 2009-21
July 2009

INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

www.iser.essex.ac.uk

Nonresponse Bias Adjustments: What Can Process Data Contribute?

Non-technical summary

Any kind of survey data is susceptible bias due to sampled units not being contacted or refusing to respond (so-called unit nonresponse bias). To minimise possible nonresponse bias survey researchers have two main strategies at their disposal. First, they can increase fieldwork efforts to increase the response rate. This decreases the *potential* for nonresponse bias; however, unless the efforts are specifically directed at the underrepresented groups, high response rates do not guarantee low nonresponse bias. Second, researchers can adjust for nonresponse bias *post hoc*, for example by means of nonresponse weighting, i.e. by giving underrepresented groups a higher weight than overrepresented groups. Such nonresponse adjustment is successful if the variables used to create the weight are correlated with both the nonresponse process and the survey estimate.

This paper investigates nonresponse weighting in the European Social Survey (ESS). The ESS is a biennial cross-national face-to-face survey of social and political attitudes across more than twenty countries in Europe. The analyses focus on nonresponse bias in Finland rounds 1-3, Poland rounds 1-3 and Slovakia round 2. Nonresponse weighting in cross-national surveys is hindered by a lack of comparative data to design weights. The analyses examine the suitability of nonresponse weights based on the ESS contact data, i.e. data on the fieldwork process, to adjust for nonresponse bias. These process weights are compared to other nonresponse weighting procedures that use demographic information about the sample units from the sampling frame data (frame weights) or about the target population from official population distributions (post-stratifications). Both population distributions and contact data are available for most ESS countries, while sampling frame data are not.

The analyses show that process weights in combination with demographic weights were most successful at reducing relative nonresponse bias. Furthermore, in the absence of sampling frame data, weights estimated from population distributions and contact data succeeded in reducing nonresponse bias in various estimates. An effective universal nonresponse bias adjustment strategy based on contact data and population distributions might therefore be possible across ESS countries.

Nonresponse Bias Adjustments: What Can Process Data Contribute?

Annelies G. Blom

Abstract

To minimise nonresponse bias most large-scale social surveys undertake nonresponse weighting. Traditional nonresponse weights adjust for demographic information only. This paper assesses the effect and added value of weights based on fieldwork process data in the European Social Survey (ESS). The reduction of relative nonresponse bias in estimates of political activism, trust, happiness and human values was examined. The effects of process, frame and post-stratification weights, as well as of weights combining several data sources, were examined. The findings demonstrate that process weights add explanatory power to nonresponse bias adjustments. Combined demographic and process weights were most successful at removing nonresponse bias.

Keywords: nonresponse weighting, propensity scores, post-stratification, paradata, contact data, European Social Survey

JEL codes: C81, C83

Contact: Annelies Blom, GESIS - Leibniz Institute for the Social Sciences,
P.O. Box 122155, 68072 Mannheim, Germany. Email: annelies.blom@gesis.org

Acknowledgements
Special thanks are due to Kari Djerf (Finland), Pawel Sztabinski and Zbyszek Sawiński (Poland) and Denisa Fedáková (Slovakia) for enabling access to ESS sampling frame data. I am grateful to Peter Lynn for guidance and encouragement and to the attendants of the session on "Non-response bias in cross-national surveys: an evaluation of designs for detection and adjustment" at the 2009 conference of the European Survey Research Association for questions and comments.

1 INTRODUCTION

Any kind of survey data is susceptible to unit nonresponse bias. To minimise possible nonresponse bias survey researchers have two main strategies at their disposal. First, they can increase fieldwork efforts to increase the response rate. This decreases the *potential* for nonresponse bias; however, unless the efforts are specifically directed at the underrepresented groups, high response rates do not guarantee low nonresponse bias (Groves 2006; Schouten, Cobben, and Bethlehem 2009). Second, researchers can adjust for nonresponse bias *post hoc*, for example by means of nonresponse weighting. Such nonresponse adjustment can render nonresponse ignorable, if the auxiliary variables used in the adjustment are correlated with both the nonresponse process and the survey estimate (Little and Vartivarian 2005; Groves 2006; Kreuter, Lemay, and Casas-Cordero 2007). In multi-purpose surveys tailoring nonresponse weights to a key survey estimate is impossible. Usually nonresponse weights thus aim for a more universal applicability.

In most large-scale social surveys researchers use demographic information about the sample units or the population in nonresponse weights. More recently nonresponse bias research using information on the nonresponse process itself has drawn attention (e.g. Olson 2006; Billiet et al. 2007). This paper discusses the suitability of process-based nonresponse weights at the example of the European Social Survey (ESS). In particular, the added value of process-based weights over and above demographic post-stratification and frame-data weights is examined.

If nonresponse bias is primarily associated with available demographic characteristics, demography-based nonresponse adjustment is optimal. However, if nonresponse bias is independent of standard demographics, such nonresponse adjustment is ineffective. Nonresponse weights that were derived from models predicting contact and cooperation by means of process variables (like the number of contact attempts until contact was achieved or whether any refusal conversion took place) offer an alternative to demographic nonresponse weights. Since such process weights adjust for the very process that generated nonresponse (and possibly nonresponse bias) in the first instance, they should be well-suited for nonresponse bias adjustment. Moreover, if the effect of process weights is (partially) independent of the effect of demographic weights, process weights can add value to nonresponse adjustments.

To test the effect and added-value of process-based nonresponse weighting in the ESS different kinds of nonresponse weights were generated: process, frame-data and post-stratification weights, as well as combinations of these three types of weights. The analyses compare the effects of these nonresponse weights on selected ESS survey estimates in the areas of political activism, trust, happiness and human values. Weights combining process and demographic data sources were found to remove more of the relative nonresponse bias than less complex weights. Comparing nonresponse weights that accounted for the process in addition to demography to nonresponse weights accounting for demography only, the analyses found added value in processes-based nonresponse weights.

2 Theoretical background

The magnitude of nonresponse bias is variable- and estimate-specific and defined by the association between the response propensity of the sample units and the measure examined. If one assumes that the nonresponse process is not static, i.e. that sample units do or do not respond to a survey with a certain probability, then the nonresponse bias in the variable mean is described by

$$B(\bar{y}_r) \approx \frac{\sigma_{y\rho}}{\bar{\rho}} . \qquad (1)$$

Nonresponse bias is thus a function of the correlation σ of the survey outcome y with the response propensity ρ and the mean response propensity $\bar{\rho}$ measured in the target population (Bethlehem 2002). For estimates of differences between two countries this means that, if there is nonresponse bias in the estimate in one of the countries, or if there is bias in both countries but of different magnitude or direction, then the cross-national comparison will be biased. Expanding on (1) the nonresponse bias in a difference in means between two countries A and B then is

$$B(\bar{y}) \approx \frac{\sigma_{counrtyA\,y\rho}}{\bar{\rho}_{countryA}} - \frac{\sigma_{countryB\,y\rho}}{\bar{\rho}_{countryB}} . \qquad (2)$$

With auxiliary information x available, nonresponse is ignorable (given x) if response is independent of the survey estimate y given x (Zhang 1999, pp.331/2). Furthermore, Rosenbaum and Rubin (1983) show that *ρ(x)* – the vector of response propensities – is the coarsest vector upon which response is independent of *x*. "Thus, if nonresponse is ignorable for *y* given *x*, then the partition of the data set induced by *ρ(x)* is a fine enough set of adjustment cells to avoid nonresponse bias" (Göksel, Judkins, and Mosher 1992, p.419). In other words, one can adjust a survey estimate *y* for nonresponse bias, if adjusting for sample units' response propensities *ρ(x)* renders the relationship between *y* and response independent.

Nonresponse weights adjust for nonresponse bias by weighting by the inverse of a sample unit's response propensity. There is a great variety of nonresponse weighting techniques (see Kalton and Flores-Cervantes 2003 for an overview). One can distinguish techniques using population distributions of key survey characteristics to adjust for nonresponse and non-coverage bias (e.g. post-stratification and raking) from techniques using auxiliary case-level data for respondents and nonrespondents to adjust for nonresponse bias only (e.g. logistic regression weighting). In either case, to be effective the nonresponse weights (and the auxiliary data they are derived from) need to be related to response and the survey outcome *y*. As a rule of thumb, weights based on variables related to response reduce nonresponse bias, while weights based on variables related to the survey outcome *y* make the sample more efficient, i.e. reduce the variance (see Kessler, Little, and Groves 1995; Little and Vartivarian 2005).

Due to differences in the magnitude and composition of nonresponse across ESS countries, there is a need to design nonresponse weights for the ESS in order to achieve better comparability of survey estimates. However, nonresponse adjustment in the ESS faces two important hurdles. First, like many social surveys, the ESS serves multiple purposes and no central estimate (or groups of estimates) can be identified. Since nonresponse bias is estimate-specific, ESS nonresponse adjustment needs to be optimal across a large variety of estimates. One way of dealing with this is to focus nonresponse adjustment on the nonresponse process instead of the survey outcome. If this adjustment rendered the survey outcomes independent of the nonresponse process, nonresponse would be ignorable; however, variances might be increased where these weights are insufficiently related to the survey estimates. Second, comparative auxiliary variables for cross-national surveys are scarce (due to differences in survey implementation and traditions across countries and data confidentiality) (see Blom, Jäckle, and Lynn forthcoming). However, with the ESS contact data one can model the probability of response for each sample unit comparatively across countries. If fieldwork processes are predictive of a sample unit's probability to respond and if the ESS contact data validly describe these fieldwork processes, then the so-derived response propensities will be valuable in nonresponse adjustments. Furthermore, nonresponse weights based on these contact data are then easily replicable and implementable across ESS countries.

While demographic nonresponse weights are generally accepted among data users, the rationale for basing nonresponse weights on process data might require further explanation. The underlying assumption of process weights is that respondents and nonrespondents who share the same process profile would have responded similarly during the interview. The theory assumes that the process indicators used to model response propensities are related to unobserved sample unit characteristics. For example, those difficult to reach are likely to be busy people who spent a lot of their time outside the household (e.g. because they are in employment, participate in leisure activities etc); those contacted but who (initially) do not participate in the survey are likely to be more socially excluded and less active in society (Groves and Couper 1998, ch.4-5; Groves, Singer, and Corning 2000). The sample units' process characteristics thus proxy other unobserved sample unit characteristics which are associated with substantive survey outcomes.

Post-stratification weights based on population distributions of age, gender and education have been tested in the ESS context (Vehovar n.d.; Vehovar and Zupanic n.d.[1]). These weights increased the variance of weighted means of key survey outcomes. Furthermore, "in most countries there have not been radical differences between the national [ESS] samples and the population structure regarding the gender and age structure" (Vehovar and Zupanic n.d., p. 42). Vehovar and Zupanic (n.d.) found differences in the structure of the educational level in the population and the ESS samples of rounds 1 and 2. However, the overall effect of weighting for nonresponse on the magnitude and direction of survey estimates was limited. Possible

[1] n.d. = no date

reasons for this are: (1) the auxiliary demographic variables used in the post-stratification might not have been the crucial drivers of nonresponse, while other demographics (e.g. household size or income) would have shown stronger effects of the post-stratification weights; (2) nonresponse bias was not a major problem in the ESS; or (3) nonresponse bias was not associated with sample unit demographics, though it might have been associated with other sample unit characteristics such as those described by the fieldwork processes.

Nonresponse weights based on the ESS contact data might be able to adjust for aspects of nonresponse bias that demographic weights cannot account for. Such process-based nonresponse weights appeal for three reasons: (1) They are based characteristics of the fieldwork process and thus, by their very nature, related to nonresponse. The propensity models in Tables 2 and 3 in a later section show that the ESS contact data were well-suited for predicting contact and cooperation. (2) Process characteristics are likely proxy sample unit characteristics that are related to various different types of survey outcomes including the social and political attitudes and behaviour measured in the ESS. (3) Finally, process weights appeal, because the data that these weights were derived from can be collected comparatively across countries and are already available for three rounds of the ESS.

If, given the auxiliary variables measured in the ESS contact data, being a respondent is independent of the answers given in the questionnaire, then nonresponse can be rendered ignorable. If these process-based weights show an effect that is independent of the effect of demographic weights, the process-based weights have an added value for nonresponse adjustments.

3 DATA

The analyses used data from rounds 1 to 3 of the ESS. The ESS is a biennial cross-national face-to-face survey of social and political attitudes across more than twenty countries in Europe. It was first fielded in the winter of 2002/03. In addition to the main interview data the analyses draw upon three auxiliary data sources to derive the nonresponse weights: population distributions of age, gender and education, frame data on sample units' demographic characteristics and the ESS contact data. Only the data of Finland rounds 1 to 3, Poland rounds 1 to 3 and Slovakia round 2 were considered. In these countries and rounds auxiliary frame data were available.

3.1 The ESS main interview data

In the ESS (a translation of) the same questionnaire is implemented across the more than 20 participating countries. The ESS questionnaire includes two main sections: a 'core' module which remains relatively constant across rounds and two or more 'rotating' modules repeated at longer intervals. The core module aims to monitor change and continuity across a wide range of social variables. The rotating modules provide an in-depth focus on a series of particular academic or policy concerns.

The analyses focussed on measures that touch upon key sociological and political research questions. Most of these variables and scales stem from the core module, with the exception of one measure from a rotating module in round 3. They include variables related to (1) citizenship norms and political participation, (2) social trust (Rosenberg Trust Scale) and political trust, (3) happiness and depression (8-item CED depression measure) and (4) value orientations (the Schwartz human values scale). The variables and scales were selected to cover a wide range of subject areas. In addition, the selected variables may well be correlated with sample unit characteristics that are typically associated with either contactability (e.g. available at-home patterns) or cooperation (e.g. psychological predispositions or correlates thereof) (Groves and Couper 1998, ch.4-5).

3.2 Population distributions

As part of the ESS data documentation each participating country deposits population distributions on key demographic variables (see Appendix 1 to the ESS documentation reports (European Social Survey 2003; European Social Survey 2005; European Social Survey 2007)[2]). The population distributions provided vary across countries, but most countries provided some population distributions on age, gender, education and region. Vehovar (n.d.) and Vehovar and Zupanic (n.d.) found that the (cross-classifications of) the age, gender and education distributions were best suited

[2] Round 1: http://ess.nsd.uib.no/index.jsp?year=2003&country=&module=documentation
Round 2: http://ess.nsd.uib.no/index.jsp?year=2005&module=documentation&country=
Round 3: http://ess.nsd.uib.no/index.jsp?year=2007&module=documentation&country=

for post-stratifying the ESS samples. Following their research, my post-stratifications use the age, gender and education distributions that Finland, Poland and Slovakia provided.

3.3 The ESS contact data

In addition to the data collected during the interview, the ESS interviewers use standardised contact forms to collect information on the contacting and cooperation process and on the neighbourhood of all sample units (Stoop et al. 2003). Each country's contact form and contact data are available from the ESS data archive website (http://ess.nsd.uib.no/). Fieldwork process indicators used to estimate contact and cooperation propensities were derived from these contact data. The process weights were derived from the contact and cooperation propensity scores.

3.4 Frame data

Each country in the ESS drew their sample from the general population aged 15 and older by strict probability methods without substitution. Within these limitations, the countries used different sample frames and designs, depending on the access restrictions that the research teams faced. As a result, the auxiliary information available from the sampling frames differed across countries. Effectively, only countries that drew their samples from population registers had access to auxiliary case-level frame data. The ESS national coordinators of three countries provided their frame data for the nonresponse analyses in this paper. Finland and Poland provided frame data for rounds 1 through 3 and Slovakia for round 2. The type of information available varied across countries, but all three countries covered information on the sample unit’s age and gender, on region and/or urbanicity. Finland further provided information on household size and the language of the sample unit.

4 METHOD

This paper examines the effect of various different kinds of nonresponse weights on the relative nonresponse bias in ESS survey estimates. Most of the measures in the ESS are attitudinal or behavioural measures of social and political concepts. For this type of data there is little possibility for validation by means of external data. In fact, many of the variables are only ever measured in surveys. As a consequence it is impossible to examine *absolute* nonresponse bias in estimates of these measures.

Instead the paper examines *relative* nonresponse bias in ESS estimates by comparing weights with different types and combinations of auxiliary data for nonresponse adjustment. It is assumed that the more information was adjusted for with the nonresponse weights, the smaller the relative residual nonresponse bias was after weighting. If, in the worst case, the propensity models included variables that were not related to the survey outcome, only random variation would have been added. However, variables related to the nonresponse process reduce the relative nonresponse bias.

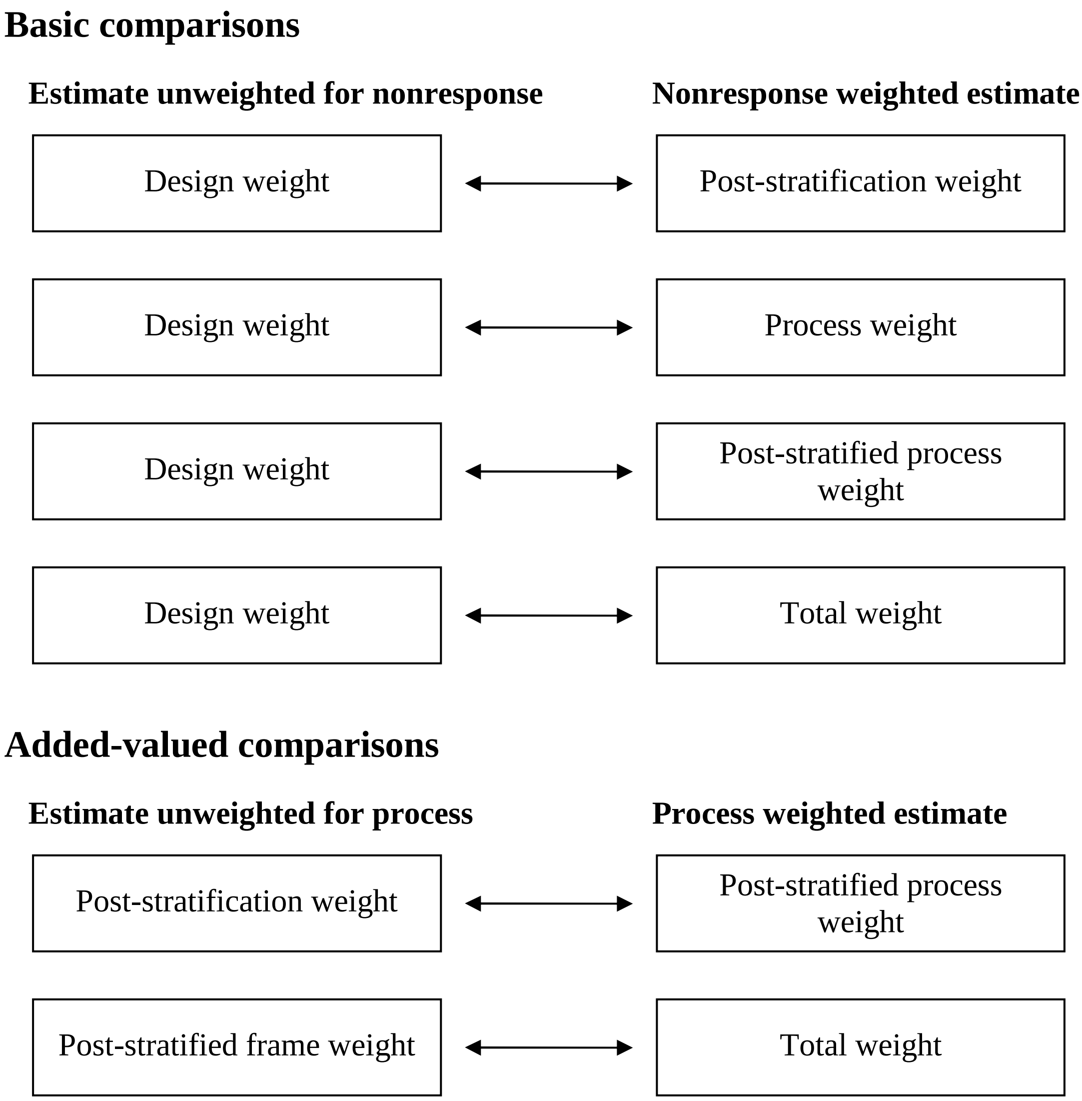
To test the effect and added-value of process-based nonresponse weighting process, frame and post-stratification weights were generated. In addition, to these basic weights combination weights were derived, i.e. a post-stratified frame weight, a post-stratified process weight and a post-stratified frame-and-process weight (the total weight). The next section describes the estimation of these weights. Table 1 provides a summary of the nonresponse weights. The premise of the analyses was that a more complex weight removed more of the relative nonresponse bias than a less complex weight.

The analyses considered a set of key political and sociological variables in the areas of political activism, trust, happiness and human values.[3] For political activism the analyses looked at the proportion of people who reported having taken various political actions: having voted in the last national election (compared to reporting not having voted); having contacted a politician or government official in the last 12 months; having taken part in a lawful demonstration in the last 12 months; being a party member. The examined happiness estimates were the mean of a general happiness scale and, for round 3, the proportion of people depressed according to the CED Depression Scale. The CED Depression Scale was derived by summing the answers to eight questions on 4-point scales. People with scores of 16 and higher were classified depressed. Furthermore, the analyses considered mean trust levels on the

[3] Please see Appendix B for the exact question wording of the measures considered.

Rosenberg Trust Scale and a political trust scale. The Rosenberg Trust Scale was derived by summing respondents' answers to three questions on interpersonal trust and dividing this sum by the number of valid responses. The derivation of the political trust scale followed the same procedure and contained four variables on trust in political institutions. Finally, mean estimates on Schwartz's human values scales were investigated. The values questions in the ESS described third-person actions and attitudes. Respondents were then asked how much they were like the vignette person. The scales distinguish ten basic motivational value orientations: security, conformity, tradition, benevolence, universalism, self-direction, stimulation, hedonism, achievement and power (see Schwartz 2003).

Figure 1: Nonresponse weight comparisons



Notes: All nonresponse weights also include the design weight; The total weight is a post-stratified frame-and-process weight.

The analyses made several comparisons of the effects of nonresponse weights on these mean and proportion estimates. First, the basic effects of the basic post-stratification weight (which is the ESS standard nonresponse weight), the process weight, the post-stratified process weight and the total weight compared to design weighted estimates were examined. Subsequently, comparisons of (1) the effect of the post-stratified process weight with the effect of the basic post-stratification weight and (2) the effect of the total weight with the effect of the post-stratified frame weight assessed the added value of the process component in the combination weights. (Figure 1). If a nonresponse weight with process data was better at reducing relative nonresponse bias in an estimate than a nonresponse weight without process data, then process data added value to nonresponse weighting.

The findings show various instances where more complex weights that included process data further reduced nonresponse bias in the ESS. Furthermore, in all but one instance is the relative nonresponse bias that was removed with the nonresponse weights was of the expected direction. For example, nonresponse weighting reduced the estimated proportion of people who reported that they voted in the last national election. Various studies show that non-voters are also less likely to participate in surveys (for example Jackman 1998; Keeter 2006). In addition, the analyses found an added value effect of the process within the total weight for a number of estimates.

5 DERIVING NONRESPONSE WEIGHTS

Having outlined the method of estimating the contribution of process weights to nonresponse weighting, this section describes how the various weights were derived. Table 1 provides an overview of all nonresponse weights used in the analyses. The basic nonresponse weights that used only one source of auxiliary data are described first. These are the post-stratification weight, the process weight and the frame weight. The latter two are logistic regression weights. They are obtained by (1) modelling response by means of logit models on unweighted sample data, (2) predicting response propensities for each sample unit based on these logit models, (3) taking the inverse of the response propensities for respondents to obtain the weights, and (4) dividing these weights by the mean weight to centre them on a mean of one. In the case of the process weight, contact and cooperation were modelled separately.

To derive the process weight for response the predicted contact and cooperation propensities were multiplied. For the frame weight my logit models predicted response. Following Little and Vartivarian (2003) the logit models were estimated on unweighted data, i.e. no design weights were applied. However, the final logistic regression weights were all multiplied by the design weight. Weights were derived separately for each country and round, although the propensities were modelled for each country across rounds. Having described the basic nonresponse weights the section turns to the combination weights which were estimated using two or more sources of auxiliary data.

5.1 Basic post-stratification weight

Vehovar (n.d.) and Vehovar and Zupanic (n.d.) showed that the cross-classifications of gender (male and female), three age groups (15-34, 35-54 and 55+) and three education groups (up to lower secondary (ISCED2 or less), higher secondary (ISCED3) and post secondary (ISCED4-6))[4] were optimal for post-stratifications in the ESS. Building on their analyses I used the same variables and groups for my post-stratifications. (See Table A1 in Appendix A for the population distributions for age, gender and education in Finland, Poland and Slovakia.)

The post-stratification weight was estimated by (1) calculating the proportion of the population in each weighting cell, (2) calculating the proportion of the (design-weighted) sample in each weighting cell, and (3) assigning each sample member in the respective weighting cell the fraction of the population proportion and the sample proportion. In Finland and Slovakia the data of the age, gender and education population distributions were fully cross-classified, so that post- stratifications to each cross-classified weighting cell were possible. Therefore, each sample member could be assigned to exactly one weighting cell. In Poland round 1 age and gender distributions were cross-classified, while for education only the population frequencies were available. In Poland rounds 2 and 3 the age distribution was cross-classified with the gender distribution, and the education distribution was cross-classified with the age distribution. Consequently, for Poland raking (or iterative proportional fitting) according to the marginal distributions was applied. "The basic

[4] ISCED refers to the qualification groups of the International Standard Classification of Education.

idea of the technique is to make the marginal distributions of the various characteristics conform with the population distributions while making the least possible distortion to the pattern of the multi-way sample distribution." (Elliot 1991,

Table 1: Overview of nonresponse weights

Weight	**Description**
Basic	
Post-stratification weight	Post-stratification (in Finland and Slovakia) and raking (in Poland) of design-weighted sample data to known population distributions of age, gender and education. The derived post-stratification weight was multiplied with the design weight.
Process weight	The process weights are logistic regression weights using variables from the ESS contact data to predict contact and cooperation. The process weight for response is derived from the product of the predicted contact and cooperation propensities. The estimated weight was multiplied with the design weight.
Frame weight	The frame weight is a logistic regression weight using demographic information from the countries' sampling frames to predict response. The estimated frame weight was multiplied with the design weight.
Combination	
Post-stratified frame weight	Post-stratifying the design- and frame-weighted sample data and multiplying the resulting post-stratification weight with the design weight and the frame weight yielded this combined post-stratified frame weight.
Post-stratified process weight for response	Post-stratifying the design- and process-weighted sample data and multiplying the resulting post-stratification weight with the design weight and the process weight yielded this combined post-stratified process weight.
Total weight (post-stratified process and frame weight)	First, a combined frame and process weight was derived by modelling contact and cooperation logistic regressions using both frame and contact data. The frame-and-process weight for response is derived from the product of the predicted contact and cooperation propensities. Post-stratifying the design-, frame- and process-weighted sample data and multiplying the resulting post-stratification weight with the design weight and the combined frame-and-process weight yielded this total weight.

Note: The frame weight and the post-stratified frame weight only indirectly appeared in the analyses. Nevertheless, it was deemed important to describe their estimation. The post-stratified frame weighted estimates served as comparison group for the added value analysis of the process in the total weight. The frame weight was used to estimate the post-stratified frame weight.

p.27) First weights that align the (design-weighted) age-gender sample distribution with the population distribution were calculated. These weights were then applied to the sample and a new marginal distribution was formed for education (round 1) or education and age (rounds 2 and 3). The whole process was then repeated for this/these variable(s). The process was iterated for a second cycle.

Although post-stratification and raking are slightly different processes, unless otherwise stated, this paper uses the term post-stratification to refer to both post-stratification and raking.

5.2 Process weights

For the process weights the propensities of contact and cooperation (conditional on contact) were modelled separately and then multiplied to obtain the response propensities. This separate modelling had two reasons: First, it enabled observing the separate contribution of the contact and the cooperation propensities to the overall response propensities across countries. The analyses showed that while in Poland and Slovakia the sample units' overall response propensities were primarily determined by their cooperation propensities, in Finland they were primarily determined by their contact propensities. Second, the cooperation model included variables that could not have been included in an overall response model. The variables 'mode of first contact', 'time of first contact', 'no refusal during the cooperation stage' and 'change of interviewer during the cooperation stage' refer exclusively to the cooperation stage of the data collection process. In a model of response these variables would have been missing for all non-contacts resulting in non-contacted sample units not being included in the model.

I banded the top quintile of the process weight (that is the quintile with the lowest contact, cooperation and response propensities). Each sample unit with a top weight was assigned the average weight of the top quintile. This procedure was chosen, because it prevents extreme weights and because propensity score quintiles are often used for nonresponse weighting classes (see for example Olson 2006)[5].

The variables included in the models were chosen to optimally predict contact and cooperation. In addition, the analyses assume that all variables were also related to

[5] "Five propensity score subclasses are often found to be adequate for removing up to 90 percent of the bias in estimating causal effects" (Olson 2006, p.747 referring to Rosenbaum and Rubin 1984).

sample unit characteristics related to nonresponse bias in the substantive survey outcome. Interviewer process characteristics were included, because they were expected to be indirectly related to sample unit characteristics as household or regional characteristics (e.g. one might find lower interviewer contact and cooperation rates in urban areas). Variables unrelated to the survey outcome do not introduce bias; instead they add random variation making the weights less efficient. In addition to variables describing the contact and cooperation processes the models also included process variables that stemmed from interviewer observations such as the type and state of the building. These variables were also collected in the ESS contact data.

Tables 2 and 3 show the logit models for contact and cooperation, respectively. The contact and cooperation models and the predictors used therein are discussed in the following.

5.2.1 Contact propensity

For the contact propensities the models considered many of the variables that other researchers and my own previous research found relevant in predicting contact. The ESS is a face-to-face survey and the project specifications prescribe a minimum of four in-person contact attempts to non-contacted sample units (e.g. European Social Survey 2006). Successful contact was thus defined as in-person contact with the household. Some interviewers also attempted contact by phone. However, these contact attempts have been less well documented across countries and interviewers, since some interviewers failed to record unsuccessful phone calls (e.g. when the phone was ‘busy’ and the interviewer tried again a few minutes later).

The significance levels of the predictors in the models showed that primarily measures of the fieldwork process were significantly associated with contact propensity. The model fit of the contact models was moderately high, with the pseudo R^2 ranging from .242 in Slovakia to .481 in Finland.

Number of contact attempts. Traditionally a major predictor of contact is the number of contact attempts made to a sample unit (for example Goyder 1985; Groves and Couper 1998; Purdon, Campanelli, and Sturgis 1999; Olson 2006). The indicator of the number of contact attempts was primarily based on the number of in-person contact attempts that an interviewer made until contact was established. However, an additional contact attempt was added to this indicator, if an interviewer made at least one phone contact attempt to the sample unit. The number and mode of contact attempts differed quite substantially across ESS countries. For example among the countries included in the analysis, the Finnish fieldwork relied much more heavily on attempting contact by phone before visiting an address than fieldwork in Poland or Slovakia (see also Blom 2009). Since the aim is to derive efficient and relevant propensity weights with the same model specification across countries, the number of contact attempts was modelled as a dummy variable.

Table 2: Contact propensity models for process weights using ESS contact data

Contact	**Finland**		**Poland**		**Slovakia**		**combined model**	
	b		b		b		b	
Number of contact attempts								
2	4.22	***	0.56		-1.12	***	3.54	***
3	2.54	***	-2.38	***	-1.56	***	1.55	***
4	1.26	***	-3.47	***	-2.25	***	-0.12	
5 or more	0.32		-5.02	***	-3.57	***	-1.69	***
Ever f2f call in the evening	0.46	***	1.01	***	0.37		0.62	***
Ever f2f call on a Saturday	-0.72	**	0.97	***	0.39		-0.03	
Ever f2f call on a Sunday	1.56	**	0.72	**	0.01		0.01	
Physical state of building:								
Satisfactory	-0.24	**	0.14		0.23		-0.07	
Bad	-0.06		0.05		0.39		0.03	
Farm or single-unit housing	0.39	***	0.32		0.48	*	0.38	***
Interviewer cooperation rate	0.03	***	0.02	***	0.01	***	0.02	***
Interviewer f2f contact rate	0.07	***	0.04	**	0.08	***	0.07	***
Interviewer phone contacting	-0.07	***	-0.04	***	-0.02	**	-0.05	***
Interviewer f2f evening calling	0.00		-0.02	**	0.02		-0.01	*
Round 1 dummy	-0.02		-0.43				0.13	
Round 2 dummy	0.05		0.69	**			0.19	*
Poland dummy							0.29	
Slovakia dummy							-0.22	
Constant	-3.73	***	-0.89		-5.17	***	-4.38	***
Chi^2	4538		468		276		6452	
Pseudo R^2	0.481		0.254		0.242		0.450	
AIC	4925		1405		893		7924	
N	8522		7658		2359		18539	

Legend: *p<0.05; **p<0.01; ***p<0.001

Notes: Coding of dependent variable 'contact': 1 in-person contact with the household, 0 no in-person contact with the household; Table shows logit model coefficients

Table 2 shows that in Poland and Slovakia making more contact attempts was related to a lower contact probability. As Blom (2009) remarks a negative association between number of contact attempts and probability of contact can be due to the modelling of contact in logit models, where only the marginal effects of the total number of calls is examined. “Since contact attempts in the ESS are not randomly assigned, interviewers choose to call at times and days that they feel might be most productive and that suit them.” (Blom 2009, p.24) In Finland a similar pattern was found. However, since many interviewers pre-contacted sample units by phone, and since these contact attempts – by definition – cannot lead to a successful contact, many of the cases with one contact attempt were only attempted by phone. Consequently, in Finland cases with one contact attempt had a lower probability of contact than cases with two or more calls. From the second call onwards, however, additional calls led to decreasing probabilities of contact.

Timing of face to face contact attempts. In the literature calls on a weekday evening or at the weekend have been found most effective for making contact (see Groves and Couper 1998; Purdon, Campanelli, and Sturgis 1999; Stoop 2005). The contact model included variables indicating whether sample units were ever attempted in person on a weekday evening, on a Saturday and on a Sunday. In person evening and Sunday calls were associated with higher contact probabilities in Finland and Poland. In person Saturday calls were also positively associated with contact in Poland, although they were negatively associated with contact in Finland. However, sample units that were called by interviewers who made a large proportion of their contact attempts in the evening were less likely to be contacted in Poland.

Quality of housing. The quality of a sample unit’s housing can be associated with contact probability (for example Lipps and Benson 2005). In the contact models a satisfactory state of the building (compared to a good state) was associated with lower contact probabilities in Finland.

Type of housing. In accordance with the literature (Campanelli, Sturgis, and Purdon 1997; Groves and Couper 1998; Stoop 2005) the analyses found that in Finland and Poland those living in single-unit housing or farms were more likely to be contacted.

Interviewer contact and cooperation rates. The models controlled for interviewer contact and cooperation rates, both of which were positively associated with contact propensity. This also confirmed the notion that interviewers who are good at gaining cooperation are also good at making contact (O'Muircheartaigh and Campanelli 1999).

Interviewer phone pre-contacting. Blohm, Hox, and Koch (2007) found that interviewers' habit to pre-contact sample units by phone can have a detrimental effect on the response rate. In the three ESS samples, the proportion of cases an interviewer attempted by phone was negatively associated with contact propensity.

5.2.2 Cooperation propensity

Cooperation was modelled conditional on contact (Table 3). This means that only cases that were successfully contacted in person were considered in these models. Cooperation was defined as successful if there was an interview for the sample unit in the survey data. Since processes leading to cooperation differ from those leading to contact (Lynn and Clarke 2002), the cooperation models accounted for different variables than the contact models, although there was some overlap. Variables specific to the cooperation model included the mode and timing of the first successful contact with the household, whether during the cooperation process there was any (respondent or household) refusal outcome, and whether there was any change of interviewer after in-person contact with the household had been established.

The significance levels of the predictors in the models show that – similar to the contact models – primarily measures of the fieldwork process were significantly associated with cooperation propensity, although measures for the state of the building and type of building (farm or single unit versus multi unit housing) were also important. The model fit of the cooperation models was high, with the pseudo R^2 ranging from .458 in Slovakia to .567 in Finland.

Mode of first successful household contact. Blohm, Hox, and Koch (2007) found that "interviewers who report that they normally show up unannounced [without making a prior appointment by phone] to conduct an interview achieve higher cooperation rates." (p.105) The ESS models found the opposite: where the first successful contact was by phone the probability of cooperation was significantly higher in all three countries. However, this may well be due to a difference in operationalisation; most importantly, in the ESS the mode of contact attempts was not randomised over interviewers and sample units. In addition, Blohm, Hox, and Koch (2007) used interviewers' reported habits of pre-contacting by phone, while the analyses at hand found an association between the observed contacting mode and cooperation.

Table 3: Cooperation propensity models for process weights using ESS contact data

Cooperation	**Finland**		**Poland**		**Slovakia**		**combined model**	
	b		b		b		b	
Mode of first contact: phone	2.57	***	2.79	***	1.75	***	2.42	***
Time of first f2f contact: evening	0.10		-0.29	*	-0.08		-0.16	
No refusal	5.21	***	4.49	***	3.38	***	4.37	***
Change of interviewer	-1.35	**	1.16	***	1.61	*	0.88	***
Number of contact attempts								
2	0.09		-0.37	*	-0.52	*	-0.36	***
3	-0.11		-1.00	***	-0.81	*	-0.88	***
4 or more	-1.10	*	-0.98	***	-0.24		-1.04	***
Physical state of building:								
Satisfactory	-0.51	**	-0.04		-0.44	**	-0.21	**
Bad	-1.46	***	-0.31	*	-0.53		-0.52	***
Farm or single-unit housing	0.33	*	0.42	***	0.20		0.33	***
Interviewer cooperation rate	0.05	***	0.04	***	0.05	***	0.04	***
Interviewer f2f contact rate	-0.07	**	-0.01		0.00		-0.01	
Interviewer f2f evening calling	0.00		0.01	**	0.01		0.01	**
Round 1 dummy	0.20		0.06				0.03	
Round 2 dummy	0.06		0.26	**			0.13	
Poland dummy							0.37	**
Slovakia dummy							0.25	
Constant	-0.22		-4.35	***	-4.72	***	-4.71	***
Chi^2	2114		4067		1260		8234	
Pseudo R^2	0.567		0.479		0.458		0.517	
AIC	1646		4459		1519		7739	
N	6461		7460		2205		16126	

Legend: *p<0.05; **p<0.01; ***p<0.001
Notes: Coding of dependent variable 'cooperation': 1 interviewed, 0 not interviewed but in-person contact with the household achieved; Table shows logit model coefficients

Timing of first successful face-to-face household contact. While evening calls were found positively associated with making contact (see also Groves and Couper 1998; Purdon, Campanelli, and Sturgis 1999; Stoop 2005), regarding cooperation sample units that were first contacted in the evening were less likely to cooperate in Poland.

(Initial) refusal. Unsurprisingly, across all countries cases that never refused participation were more likely to finally cooperate than cases that (initially) refused. Since the aim of this model is to subsequently derive cooperation propensities for nonresponse weights, including this indicator in the model should result in lower cooperation propensities for initially refusing sample units. In this way the process

weights account for possible differences between directly cooperative and initially refusing respondents (see also Billiet et al. 2007).

Change of interviewer. The model accounted for whether there was any change of interviewer once in-person contact with the household had been achieved. There were significant effects in all three countries. However, while in Poland and Slovakia a change of interviewer was positively associated with gaining cooperation, changing the interviewer in Finland was associated with lower cooperation propensities. One should again note that no experimental setting was used to analyse the effect of a change in interviewer. Therefore, the differences in effects might well be due to differences across countries in fieldwork strategies regarding when an interviewer change took place.

Number of contact attempts. The relationship between the number of contact attempts until contact was made (see contact model for operationalisation) and cooperation was significant and of the same direction as the relationship between the number of contact attempts and contact; however for cooperation this relationship was considerably weaker. This shows that cases that were difficult to contact were also difficult to gain cooperation from. Interestingly, this finding was in contrast with much of the literature according to which there is no association between difficulty of making contact and cooperation propensity (for example Lepowski and Couper 2002; Stoop 2005).

Quality of housing. There was a significant association between the state of a sample unit's housing and their likelihood of cooperation; the better the state of the housing the more likely was cooperation. Those living in satisfactory housing (compared to good housing) were less likely to cooperate in Finland and Slovakia, and those living in bad housing (compared to good housing) were less likely to cooperate in Finland and Poland.

Type of housing. The contact model showed that sample units living in single-unit housing were more contactable in Finland and Poland. In addition, the cooperation model showed sample units living in single-unit housing were also more likely to cooperate in Finland and Slovakia.

Interviewer strategies. The interviewers' cooperation rates were positively associated with achieving an interview in all three countries. In Finland the interviewer contact rate was negatively associated with cooperation propensity. In Poland interviewers who carried out a large proportion of their contact attempts in the evening were more likely gain cooperation.

5.2.3 Response propensity

Response propensities were estimated as the product of the estimated contact and cooperation propensities. The process weights were then derived by inverting the response propensities for respondents, centring them on a mean of one and banding the top quintile of the resulting process weights for response.

5.2.4 The relationship between contact, cooperation and response propensities

Table 4 shows that the composition of the overall nonresponse rate differed between the countries considered. While the nonresponse rate ranged from 63.9 percent in Slovakia round 2 to 73.7 percent in Poland round 2, the in-person contact and cooperation rates varied much more. In Poland and Slovakia the in-person contact rates were above 90 percent in all rounds and the cooperation rates were around 70 percent. In Finland however, the in-person contact rates ranged from 70.4 percent in round 3 to 80.2 percent in round 1. The cooperation rates in Finland were always above 90 percent. As mentioned previously, the Finnish fieldwork relied heavily on phone pre-contacting. The lower in-person contact rates (and the resulting high cooperation rates that are conditional on contact) are likely to have resulted from this.

Table 4: Outcome rates

Country and round	Overall response rate %	In-person contact rate %	Cooperation rate %
Finland round 1	73.3	80.2	91.5
Finland round 2	70.8	77.2	91.7
Finland round 3	64.5	70.4	91.6
Poland round 1	72.2	95.4	75.7
Poland round 2	73.7	98.8	74.6
Poland round 3	70.6	97.3	72.5
Slovakia round 2	63.9	93.4	68.4

Notes: Overall response rate: number of interviewed sample units / number of eligible sample units
In-person contact rate: number of sample units, where in-person contact with the household was made / number of eligible sample units
Cooperation rate: number of interviewed sample units / number of sample units, where in-person contact with the household was made

Further analyses looked at the relationship between the probability of contact and cooperation with final response probabilities by arranging the sample units' contact, cooperation and response propensities into strata of quintiles. The discovered patterns differed across countries (see Tables A2 and A3 in Appendix A). In Finland the highest response propensity strata consisted exclusively of cases with the highest contactability. Some cases with high contactability were also found in low response propensity quintiles, but not vice versa. Regarding the relationship between cooperation and response propensity the cases were more evenly spread, i.e. most cases that had a high (low) cooperation propensity also had a high (low) response propensity, but there were also cases with a high (low) cooperation propensity and a low (high) response propensity.

Poland and Slovakia many cases with high (low) contact propensity also had a high (low) response propensity. However, there were also cases with low contact propensities, but a high response propensity; although in Slovakia cases with the lowest contact propensity were not found in the highest response propensity stratum. Regarding cooperation propensities the picture was a different one. In Poland and Slovakia almost all cases with a high (low) cooperation propensity also had a high (low) response propensity. None of the cases in the bottom cooperation propensity stratum were found in the top three response propensity strata. Similarly, hardly any cases with high cooperation propensities had low response propensities.

This shows that while in Finland the distribution of cases over the final response propensity strata was driven by the cases' contactability, in Poland and Slovakia it was driven by their likelihood to cooperate. This means that in Finland a low/high response propensity was largely associated with a low/high contact propensity, while in Poland and Slovakia a low/high response propensity was largely associated with a low/high cooperation propensity. The differential contribution of contact and cooperation propensities to the overall response across countries emphasises the importance of including both contact and cooperation processes in process-based nonresponse weights. If nonresponse weights were only based on cooperation propensities, this would result in effective weights for Poland and Slovakia, but ineffective weights for Finland (effective in terms of reducing relative nonresponse bias). Similarly, nonresponse weights based exclusively on contact propensities would lead to effective nonresponse weights in Finland but ineffective nonresponse weights in Poland and Slovakia.

5.3 Frame weight

The frame weight was solely used in estimating the post-stratified frame weight and is thus only briefly described here. The frame weight was estimated by means of logistic regressions. The models used all available frame variables that were significantly associated with response. The regressions in Table A4 in Appendix A show that the available significant variables varied across countries. However, in all three countries the models included some indicator of region or urbanicity. In all countries living in more rural areas was associated with higher response propensities. While in Finland being male was associated with lower response propensities, in Poland and Slovakia no such association was found. Furthermore, a negative correlation between age and response propensity found in Poland and Slovakia was not observed in Finland.

While many of the frame variables included in these models showed high levels of significance, the measures of the model fit were quite low. This indicated a weak association between the frame variables and the probability of response.

5.4 Combination weights

The combination weights (1) used the derived logistic regression nonresponse weights (i.e. the frame weight and the process weight), (2) multiplied these with the design weight and (3) post-stratified the so-weighted samples. The resulting post-stratification weight was multiplied with the design and nonresponse weights. Three combination weights were thus derived: (1) the post-stratified frame weight, (2) the post-stratified process weight and (3) the total weight (i.e. the post-stratified frame-and-process weight). The post-stratifications for the combination weights applied the same population distributions as the basic post-stratifications (see Table A5 in Appendix A).

For the total weight, first a nonresponse weight based on both frame and process data needed to be generated. For this contact and cooperation were modelled separately in logistic regressions using all variables that were also included in the frame model and the contact or cooperation model. The logit coefficients of these combined models are shown in Tables A5 and A6 in Appendix A. The significant variables in the models primarily stem from the contact data, however, both Finland and Poland contribute significant frame variables to the contact and cooperation models. The predicted contact and cooperation propensities were multiplied and inverted for the frame-and-process weight and the top weight quintile was banded.

6 DESCRIPTIVES OF THE WEIGHTS

When weighting survey data for nonresponse researchers are often concerned about an increase in variance, which can lead to a loss of statistical power for testing hypotheses. The following descriptive statistics investigate the variance of the nonresponse weights and the increase in variance due to nonresponse weighting. Table 5 displays the minimum, maximum, standard deviation, the coefficient of variation (CV) and the variance inflation factor (VIF)[6] for the design weight and all nonresponse weights used in the subsequent analyses.

Across most countries and rounds the VIFs of the basic post-stratification weight were considerably smaller than the VIFs of the process weight. With 1.54 Finland round 3 had the highest process weight VIF. In Finland and Poland the VIF of the process weight increased from round 1 to round 3, pointing at a greater variation in process-related response propensities in later rounds. For the combined weights, i.e. the post-stratified process weight and the total weight, the VIFs are marginally higher than for the process weights.

Regarding the variation in the design weights one should note that in Finland and Slovakia gross samples were drawn as simple random samples. Therefore, no design weighting was necessary and the design weights equalled one for each sample unit (Gabler et al. 2008; Gabler and Ganninger forthcoming). For the Polish sample design the country was divided into two parts. The first part contained the larger towns (with 100,000 or more inhabitants in round 1 and with 50,000 and more inhabitants in

---

[6] Based on Kish (1965) the coefficient of variation for the weight variable is $CV = \frac{\sqrt{s_w^2}}{\bar{w}}$. The variance inflation factor "expresses the increase of the sampling variance of a weighted sample in comparison with the sample variance (of the same sample size) where there would be no need for weights" (Vehovar and Zupanic n.d.): $VIF = 1 + CV^2$

rounds 2 and 3), where the sample was drawn by simple random sample. The second part of the sample was stratified and clustered. In addition, the Polish design accounted for lower expected response rates in the larger towns by oversampling these (see Gabler, Häder, and Lynn 2006 for a more detailed description of the Polish design and its impact on design effects; see European Social Survey 2003; European Social Survey 2005; European Social Survey 2007 for descriptions of sampling frames and designs across all ESS countries).

Table 5: Descriptives of weights by country and round

Country and round					
Design weight	**Min.**	**Max.**	**Std. Dev.**	**CV**	**VIF**
Finland round 1	1.00	1.00	0.00	0.00	1.00
Finland round 2	1.00	1.00	0.00	0.00	1.00
Finland round 3	1.00	1.00	0.00	0.00	1.00
Poland round 1	0.73	1.16	0.14	0.14	1.02
Poland round 2	0.77	1.12	0.12	0.12	1.02
Poland round 3	0.46	1.38	0.22	0.22	1.05
Slovakia round 2	1.00	1.00	0.00	0.00	1.00
Post-stratification weight					
Finland round 1	0.84	1.29	0.14	0.14	1.02
Finland round 2	0.69	1.41	0.16	0.16	1.03
Finland round 3	0.67	1.94	0.24	0.24	1.06
Poland round 1	0.47	2.02	0.16	0.16	1.03
Poland round 2	0.61	1.18	0.14	0.14	1.02
Poland round 3	0.32	1.84	0.27	0.27	1.07
Slovakia round 2	0.62	1.77	0.26	0.26	1.07
Process weight					
Finland round 1	0.77	1.77	0.39	0.39	1.15
Finland round 2	0.70	2.05	0.52	0.52	1.28
Finland round 3	0.59	2.47	0.74	0.74	1.54
Poland round 1	0.55	2.02	0.31	0.32	1.10
Poland round 2	0.60	2.07	0.39	0.39	1.15
Poland round 3	0.31	2.95	0.54	0.56	1.31
Slovakia round 2	0.68	2.00	0.50	0.50	1.25
Post-stratified process weight					
Finland round 1	0.66	2.36	0.41	0.41	1.17
Finland round 2	0.49	2.87	0.55	0.55	1.30
Finland round 3	0.36	4.19	0.79	0.79	1.63
Poland round 1	0.38	3.37	0.33	0.33	1.11
Poland round 2	0.45	2.23	0.39	0.40	1.16
Poland round 3	0.19	3.50	0.57	0.59	1.34
Slovakia round 2	0.37	4.09	0.55	0.56	1.31
Total weight (post-stratified process and frame weight)					
Finland round 1	0.65	2.30	0.42	0.43	1.18
Finland round 2	0.52	2.89	0.56	0.56	1.31
Finland round 3	0.37	4.06	0.75	0.75	1.56
Poland round 1	0.37	3.31	0.33	0.34	1.12
Poland round 2	0.44	2.22	0.39	0.40	1.16
Poland round 3	0.18	3.62	0.58	0.60	1.36
Slovakia round 2	0.34	4.38	0.62	0.62	1.38

Notes: All weights include the design weight. The Finnish and Slovakian samples were simple random samples; consequently each Finnish and Slovakian sample unit has a design weight of one.

In Poland both the design weight and the frame weight were derived using information on the size of the town a sample unit lives in. The total weight therefore accounts for both the unequal selection probabilities in the sampling frame and the differential nonresponse regarding town size.

In addition to the variance analyses I looked at correlations between the various nonresponse weights (Table A7 in Appendix A) to find out more about similarities between different weights. Across all countries and rounds the highest correlations were found in weights containing process information. Correlations of the process weight and the post-stratified process weight were above 90 percent in Finland and Poland and 78 percent in Slovakia. However, the correlation between the post-stratification and the post-stratified process weight was only between 11 percent in Poland round 1 and 47 percent in Slovakia round 2. Apparently, the post-stratification of the process weight changed the overall structure of the weight only marginally; however adding a process component to the demographic post-stratifications changed the weight substantially. Finally, a low correlation between the basic post-stratification weight and the basic process weight indicates that these two types of nonresponse adjustments indeed accounted for different aspects of nonresponse. Whether the weights also had a different effect on survey estimates is examined in the following.

7 FINDINGS

Looking into the effects of the various nonresponse weights on survey estimates is at the heart of this research. The bias analyses considered a set of key political and sociological variables in the areas of political activism, trust, depression and human values. The examined estimates are described in detail in the methods section of this paper.

The analyses made several comparisons of the effects of nonresponse weights on these ESS estimates (Tables 6a-d). First, the basic effects of the post-stratification weight, the process weight, the post-stratified process weight and the total weight compared to design weighted estimates were examined[7]. Subsequently, the analyses

[7] All nonresponse weights always also included the design weight.

looked at the added value of the process component within the combination weights. This added value analysis compared the effect of the post-stratified process weight to the effect of the basic post-stratification and the effect of the total weight to the effect of the post-stratified frame weight.

For the basic effects, significance was evaluated by calculating five-percent and ten-percent confidence intervals for the nonresponse-weighted estimates and examining whether the design-weighted estimate fell within these confidence intervals. For added value effects, the five- and ten-percent confidence intervals of the estimate of the more complex weight were calculated. Subsequently, it was assessed whether the estimate based upon the less complex weight fell within these confidence intervals. Since the differently weighted estimates are all based on the same sample, standard significance tests are not applicable. Instead, this approach examined the relative change in size of an estimate due to nonresponse weighting. Therefore, where 'significant' differences were found, this referred to these pseudo significance tests.

Looking at the number of instances where the basic weights and the combined weights compared to the design weights brought about significant weighting effects gives an overview of the basic effects. The combined weights were more successful at reducing relative nonresponse bias than the basic weights. The post-stratification weight and the process weight only reduced relative nonresponse bias in three and six estimates, respectively. In contrast, the post-stratified process weight had a significant effect on 16 estimates and the total weight on 17 estimates. Therefore, combining demographic and process weighting was a valuable strategy for removing relative nonresponse bias.

Furthermore, there was some added value of the process within the combination weights. In three instances the post-stratified process weight added significantly to removing relative nonresponse bias from estimates compared to the post-stratification weight only. In one instance the total weight was of added value compared to the post-stratified frame weight. The limited added value effect in the total weight might be due to the frame characteristics adjusting for some of the same aspects of nonresponse bias as the process characteristics. Furthermore, this might indicate that the combination of post-stratification and frame weighting already considerably reduced nonresponse bias, so that adjusting for additional process characteristics was of limited added value.

The effects of the weights differed across variables, countries and rounds, for both demographic and process weights. However, within subject area there were some consistencies; for example, the only significant effects on the political activism estimates were found in Finland. Any relative nonresponse bias in the measures of trust and the CED depression scale could not be corrected with any of the nonresponse weights.

Three of the four political activism estimates showed nonresponse weighting effects. For indicators of political activism one would expect that people with low response propensities are also less likely to be politically active (see for example Groves and Couper 1998; Groves, Singer, and Corning 2000; Tourangeau 2004). In Finland the proportion of those eligible to vote who reported having voted in the last national election was negatively affected by the process weight and the combination weights. Previous studies have demonstrated a negative association between voting and survey participation (Jackman 1998; Keeter et al. 2006); the effects of the nonresponse weights thus support previous findings. Moreover, the analyses showed that basic post-stratification weights would have been unsuccessful at reducing the relative bias in this estimate.

In Finland round 2 the combination weights reduced the estimated the proportion of people who reported being a party member. The effects were of the expected direction since party membership is generally assumed to be related to survey participation (see for example Keeter et al. 2006). In contrast, in Finland round 2 the proportion of people that took part in a lawful demonstration in the last year was higher when the post-stratified process weight was applied. This was the only unexpected finding, since the politically activism is generally found to be positively related to survey participation.

The nonresponse weights also affected estimates of level of happiness. When weighting with the process or the post-stratified process weights the estimated mean happiness of people in Finland round 3 was significantly lower. According to Groves, Cialdini, and Couper's (1992) theory of survey participation happiness generally enhances compliance with the survey request (p.485). The findings are thus consistent with their theory.

The Schwartz human values scales measure people's general values structure. Literature on the relationship between nonresponse and the human values measured by Schwartz is scarce. The findings showed that the value structure can be affected by nonresponse bias which process weights adjust for. The effect of different weighting strategies on the Schwartz human values scales varied according to the scale examined. However, every value scale was affected by relative nonresponse bias that at least one of the nonresponse weights could reduce. For eight out of the ten human values scales the total weight corrected for relative nonresponse bias. In addition, the universalism and power scales showed added value effects of the process and are therefore described in more detail.

In Finland round 3, Poland round 3 and Slovakia round 2 the means of the universalism scale were decreased when weighting with process, post-stratified process or total weights. In addition, there were added value effects of process weighting in both combined weights. The universalism value orientation measures people's understanding, appreciation, tolerance and protection for the welfare of all people (Schwartz 2003). The findings demonstrate that such values were positively related to survey participation. At the same time, in Poland round 3 and Slovakia round 3 the means of the power scale were increased when weighting with process, post-stratified process or total weights. The power scale measures people's evaluation of social status and prestige. Since participating in social surveys does not 'pay off' in terms of social status or prestige, and since reduced norms of helping behaviour can be related to nonresponse (O'Muircheartaigh and Campanelli 1999), the nonresponse bias was of the expected direction. The intuitiveness of the direction of the effects on these values scales is further support for the usefulness of considering process indicators in nonresponse weights.

Table 6a: Weighting effects on estimates of political activism

Country and round	Design weight	Post-stratification weight		Process weight		Post-stratified process weight			Total weight		
Proportion 'yes'	%	%	sign. (dweight)	%	sign. (dweight)	%	sign. (dweight)	sig. (post-stratified dweight)	%	sig. (dweight)	sig. (post-stratified frame*dweight)
Voted											
Finland round 1	82	82		81		81			81		
Finland round 2	79	78		78		77	*		77	*	
Finland round 3	84	83		82	*	82	*		81	*	
Poland round 1	66	66		66		66			66		
Poland round 2	65	64		65		64			64		
Poland round 3	66	66		66		65			65		
Slovakia round 2	74	74		74		73			73		
Contacted politician											
Finland round 1	24	24		23		23			23		
Finland round 2	22	22		21		21			21		
Finland round 3	19	18		18		18			18		
Poland round 1	10	10		10		10			9		
Poland round 2	7	7		7		7			7		
Poland round 3	6	6		6		6			6		
Slovakia round 2	7	7		8		7			7		
Demonstration											
Finland round 1	2.0	1.9		2.1		2.0			2.0		
Finland round 2	2.0	2.2		2.4		2.6	*		2.6	*	
Finland round 3	2.2	2.2		2.5		2.5			2.4		
Poland round 1	1.3	1.3		1.3		1.3			1.3		
Poland round 2	1.6	1.6		1.7		1.7			1.7		
Poland round 3	1.4	1.3		1.3		1.2			1.3		
Slovakia round 2	3.7	3.6		4.2		4.1			4.1		
Party member											
Finland round 1	7.3	7.4		6.7		6.7			6.7		
Finland round 2	7.2	7.0		6.6		6.3	*		6.3	*	
Finland round 3	7.7	7.2		7.2		6.9			7.0		
Poland round 1	1.7	1.7		1.8		1.8			1.8		
Poland round 2	1.0	1.1		1.0		1.0			1.0		
Poland round 3	1.1	1.1		0.8		0.8			0.8		
Slovakia round 2	2.1	1.8		2.1		1.6			1.6		

Legend: Pseudo significance levels: * less complex estimate fell within 0.1 confidence interval of more complex estimate; ** less complex estimate fell within 0.05 confidence interval of more complex estimate
Note: All weights include the design weight

Table 6b: Weighting effects on estimates of social and political trust

Country and round	Design weight	Post-stratification weight		Process weight		Post-stratified process weight			Total weight		
	mean	mean	sign. (dweight)	mean	sign. (dweight)	mean	sign. (dweight)	sig. (post-stratified dweight)	mean	sig. (dweight)	sig. (post-stratified frame*dweight)
Social trust											
Finland round 1	6.34	6.34		6.31		6.32			6.32		
Finland round 2	6.35	6.33		6.34		6.31			6.31		
Finland round 3	6.44	6.43		6.41		6.41			6.41		
Poland round 1	3.78	3.77		3.80		3.79			3.79		
Poland round 2	3.77	3.76		3.76		3.75			3.77		
Poland round 3	4.13	4.14		4.18		4.18			4.19		
Slovakia round 2	4.07	4.04		4.04		4.03			4.06		
Political trust											
Finland round 1	6.32	6.31		6.31		6.31			6.31		
Finland round 2	6.44	6.43		6.44		6.43			6.43		
Finland round 3	6.51	6.52		6.49		6.49			6.51		
Poland round 1	3.73	3.72		3.73		3.73			3.72		
Poland round 2	2.99	2.99		2.99		2.98			3.00		
Poland round 3	3.40	3.39		3.36		3.35			3.35		
Slovakia round 2	3.38	3.37		3.45		3.42			3.42		

Legend: Pseudo significance levels: * less complex estimate fell within 0.1 confidence interval of more complex estimate; ** less complex estimate fell within 0.05 confidence interval of more complex estimate

Note: All weights include the design weight

Table 6c: Weighting effects on estimates of happiness and depression

Country and round	Design weight	Post-stratification weight		Process weight		Post-stratified process weight			Total weight		
			sign. (dweight)		sign. (dweight)		sign. (dweight)	sig. (post-stratified dweight)		sig. (dweight)	sig. (post-stratified frame*dweight)
Happiness	**mean**	**mean**		**mean**		**mean**			**mean**		
Finland round 1	8.03	8.03		8.01		8.02			8.02		
Finland round 2	8.06	8.06		8.02		8.02			8.02		
Finland round 3	8.00	7.99		7.95	*	7.94	*		7.95		
Poland round 1	6.43	6.42		6.40		6.39			6.39		
Poland round 2	6.72	6.72		6.73		6.72			6.72		
Poland round 3	6.96	6.99		6.97		6.96			6.96		
Slovakia round 2	6.24	6.19		6.28		6.22			6.24		
Depression	**%**	**%**		**%**		**%**			**%**		
Finland round 3	19	18		20		19			19		
Poland round 3	35	34		34		34			34		

Legend: Pseudo significance levels: * less complex estimate fell within 0.1 confidence interval of more complex estimate; ** less complex estimate fell within 0.05 confidence interval of more complex estimate
Note: All weights include the design weight

Table 6d: Weighting effects on estimates of Schwartz's human values

Country and round	Design weight	Post-stratification weight		Process weight		Post-stratified process weight			Total weight		
	mean	mean	sig. (dweight)	mean	sig. (dweight)	mean	sig. (dweight)	sig. (post-stratified dweight)	mean	sig. (dweight)	sig. (post-stratified frame*dweight)
Security											
Finland round 1	-0.52	-0.52		-0.52		-0.52			-0.52		
Finland round 2	-0.48	-0.47		-0.46		-0.46			-0.46		
Finland round 3	-0.42	-0.40		-0.41		-0.38	*		-0.40		
Poland round 1	-0.71	-0.72		-0.70		-0.70			-0.70		
Poland round 2	-0.65	-0.65		-0.65		-0.66			-0.65		
Poland round 3	-0.57	-0.56		-0.58		-0.57			-0.58		
Slovakia round 2	-0.59	-0.62		-0.59		-0.60			-0.61		
Conformity											
Finland round 1	-0.07	-0.07		-0.06		-0.06			-0.07		
Finland round 2	-0.06	-0.04		-0.04		-0.02			-0.02		
Finland round 3	-0.06	-0.03		-0.05		-0.02			-0.02	*	
Poland round 1	-0.38	-0.38		-0.36		-0.37			-0.37		
Poland round 2	-0.30	-0.30		-0.29		-0.30			-0.28		
Poland round 3	-0.31	-0.31		-0.31		-0.31			-0.30		
Slovakia round 2	-0.17	-0.20		-0.17		-0.20			-0.20		
Tradition											
Finland round 1	0.20	0.20		0.21		0.21			0.20		
Finland round 2	0.12	0.14		0.14		0.15			0.15		
Finland round 3	0.10	0.12		0.11		0.12			0.12		
Poland round 1	-0.27	-0.27		-0.25		-0.25			-0.25		
Poland round 2	-0.25	-0.25		-0.23		-0.24			-0.23		
Poland round 3	-0.23	-0.22		-0.21		-0.20			-0.20		
Slovakia round 2	-0.30	-0.36	*	-0.30		-0.35	*		-0.36	*	
Benevolence											
Finland round 1	-0.68	-0.67		-0.68		-0.67			-0.67		
Finland round 2	-0.71	-0.70		-0.70		-0.70			-0.70		
Finland round 3	-0.75	-0.75		-0.73		-0.73			-0.73		
Poland round 1	-0.48	-0.48		-0.49		-0.49			-0.49		
Poland round 2	-0.55	-0.55		-0.55		-0.55			-0.54		
Poland round 3	-0.47	-0.47		-0.49		-0.48			-0.49		
Slovakia round 2	-0.41	-0.44		-0.42		-0.44			-0.45	*	

Table 6d (continued): Weighting effects on estimates of Schwartz's human values

Country and round	Design weight	Post-stratification weight		Process weight		Post-stratified process weight			Total weight		
	mean	mean	sig. (dweight)	mean	sig. (dweight)	mean	sig. (dweight)	sig. (post-stratified dweight)	mean	sig. (dweight)	sig. (post-stratified frame*dweight)
Universalism											
Finland round 1	-0.76	-0.76		-0.76		-0.75			-0.76		
Finland round 2	-0.74	-0.73		-0.73		-0.73			-0.73		
Finland round 3	-0.79	-0.78		-0.78		-0.77			-0.76	*	
Poland round 1	-0.51	-0.51		-0.52		-0.52			-0.51		
Poland round 2	-0.48	-0.48		-0.48		-0.48			-0.48		
Poland round 3	-0.55	-0.54		-0.57	*	-0.57	*	*	-0.57	*	*
Slovakia round 2	-0.46	-0.47		-0.49	*	-0.49	*	*	-0.50	*	
Self-direction											
Finland round 1	-0.44	-0.45		-0.46		-0.46			-0.46		
Finland round 2	-0.48	-0.47		-0.49		-0.49			-0.49		
Finland round 3	-0.46	-0.46		-0.45		-0.45			-0.45		
Poland round 1	-0.17	-0.17		-0.20		-0.19			-0.20		
Poland round 2	-0.18	-0.17		-0.19		-0.18			-0.18		
Poland round 3	-0.15	-0.16		-0.17		-0.16			-0.16		
Slovakia round 2	-0.25	-0.21	*	-0.28		-0.23			-0.23		
Stimulation											
Finland round 1	0.46	0.46		0.45		0.45			0.45		
Finland round 2	0.50	0.48		0.48		0.46	*		0.47	*	
Finland round 3	0.52	0.48	*	0.49		0.45	**		0.46	*	
Poland round 1	0.73	0.74		0.71		0.72			0.72		
Poland round 2	0.76	0.76		0.74		0.74			0.72		
Poland round 3	0.71	0.69		0.71		0.69			0.69		
Slovakia round 2	0.68	0.70		0.69		0.69			0.70		

Table 6d (continued): Weighting effects on estimates of Schwartz's human values

Country and round	Design weight	Post-stratification weight		Process weight		Post-stratified process weight			Total weight		
	mean	mean	sig. (dweight)	mean	sig. (dweight)	mean	sig. (dweight)	sig. (post-stratified dweight)	mean	sig. (dweight)	sig. (post-stratified frame*dweight)
Hedonism											
Finland round 1	0.26	0.27		0.26		0.25			0.26		
Finland round 2	0.29	0.27		0.27		0.25	*		0.25	*	
Finland round 3	0.29	0.25		0.26		0.22	*		0.22	*	
Poland round 1	0.92	0.92		0.90		0.90			0.90		
Poland round 2	0.93	0.93		0.92		0.92			0.91		
Poland round 3	0.89	0.87		0.89		0.87			0.87		
Slovakia round 2	0.77	0.80		0.80		0.81			0.82		
Achievement											
Finland round 1	0.72	0.73		0.73		0.73			0.73		
Finland round 2	0.72	0.71		0.71		0.70			0.70		
Finland round 3	0.74	0.73		0.73		0.72			0.71		
Poland round 1	0.39	0.39		0.40		0.40			0.40		
Poland round 2	0.30	0.31		0.30		0.30			0.30		
Poland round 3	0.32	0.31		0.33		0.33			0.34		
Slovakia round 2	0.29	0.32		0.32	*	0.35	*		0.36	**	
Power											
Finland round 1	1.21	1.21		1.22		1.21			1.22		
Finland round 2	1.20	1.19		1.19		1.19			1.19		
Finland round 3	1.22	1.22		1.22		1.22			1.22		
Poland round 1	0.75	0.74		0.77		0.77			0.76		
Poland round 2	0.67	0.67		0.70		0.69			0.69		
Poland round 3	0.65	0.65		0.69	*	0.69	*	*	0.69	*	
Slovakia round 2	0.67	0.70		0.69		0.72	*		0.73	*	

Legend: Pseudo significance levels: * less complex estimate fell within 0.1 confidence interval of more complex estimate; ** less complex estimate fell within 0.05 confidence interval of more complex estimate

Note: All weights include the design weight

8 DISCUSSION AND CONCLUSIONS

This paper set out to examine the suitability of nonresponse weights based on the ESS contact data to adjust for nonresponse bias in Finland rounds 1 to 3, Poland rounds 1 to 3 and Slovakia round 2. The analyses showed that such process weights succeeded in reducing relative nonresponse bias in various ESS estimates. Combining contact data with frame data and population distributions when deriving nonresponse weights benefited the nonresponse adjustment. Moreover, in the absence of frame data, the analyses showed that the post-stratified process weight (estimated from contact data and population distributions) succeeded in reducing relative nonresponse bias in various estimates. Both population distributions and contact data are available for most ESS countries. An effective universal nonresponse bias adjustment strategy across ESS countries is therefore possible.

The analyses compared the effects of various nonresponse weights on ESS estimates of political activism, happiness, trust and human values. The aim was to investigate (1) the effect of weights based on process data, frame data and population distributions (basic effects) and (2) the added value of adjusting for fieldwork processes in addition to adjusting for demographic characteristics from frame data and population distributions (added value effects). Therefore, the basic effects of a post-stratification weight, a process weight, a post-stratified process weight and a total weight compared to design weighted estimates were studied. In addition, the added value analysis compared the effect of a post-stratified process weight to the effect of a basic post-stratification and the effect of a post-stratified frame-and-process weight (total weight) to the effect of the post-stratified frame weight.

The findings emphasise the estimate-specificity of nonresponse bias, as the effect of all nonresponse weights differed across variables, countries and rounds. At the same time, some consistency of effects was found. For example, significant effects in the political activism variables were only found in Finland and no significant effects were found in the trust scales.

Overall the weighting effects were of the expected direction. Weighting with combination weights (i.e. with the post-stratified process weight or the total weight) reduced estimates of the number of people who reported voting in the last national election and who reported being a party member. This was in line with previous research showing that the politically active are more likely to participate in surveys (Groves and Couper 1998; Groves, Singer, and Corning 2000; Tourangeau 2004). Similarly, the estimated mean happiness was reduced when weighting with process or post-stratified process weights. Since nonresponse theories link happiness to cooperation (Groves, Cialdini, and Couper 1992) this finding confirmed the effectiveness of the derived nonresponse weights. Finally, estimates of human values were also affected by nonresponse weights that included process information. In particular the power and universalism scales showed nonresponse bias that the combination weights reduced.

The analyses demonstrated that combining demographic and process data when designing nonresponse weights was the most valuable strategy. For some of the estimates the process component of the combination weights added value to the nonresponse adjustment. Accordingly, the process weight showed significant contributions to nonresponse weighting in the ESS. However, this effect was limited in the total weights, which already adjusted for frame information and population distributions. Apparently, the process and frame data are – at least to some extent – related to similar aspects of nonresponse bias. Nevertheless, some nonresponse bias in the ESS was partially independent of standard demographic characteristics; especially, when demographic nonresponse adjustments were limited to post-stratifications according to age, gender and education distributions. The alternative way of adjusting for nonresponse bias by accounting for fieldwork processes added value to ESS nonresponse adjustments. Since frame data are never available in all countries of a cross-national survey, using process information in nonresponse weights to proxy unobserved sample unit characteristics could improve cross-national nonresponse weights.

In addition to being an alternative way of addressing nonresponse bias in multi-purpose surveys, providing process weights to secondary data analysts could be worthwhile. Such weights provide information about nonresponse that non-methodologists do not ordinarily have familiarity with or access to. Most social scientists control for standard demographics in their models. However, adjusting for fieldwork processes would be a novel, and possibly fruitful, approach. Future research on the suitability of process-based nonresponse weights should therefore start exploring the effect of such process weights on more complex sociological, political or economic models.

REFERENCES

Bethlehem, Jelke G. 2002. "Weighting nonresponse adjustments based on auxiliary information." In *Survey nonresponse*, eds. Groves, Robert M., Don Dillman, John L. Eltinge, and Roderick J. A. Little, pp.275-288. New York: Wiley & Sons.

Billiet, Jaak, Michel Philippens, Rory Fitzgerald, and Ineke Stoop. 2007. "Estimation of nonresponse bias in the European Social Survey: Using information from reluctant respondents." *Journal of Official Statistics*, 23:135-162.

Blohm, Michael, Joop Hox, and Achim Koch. 2007. "The influence of interviewers' contact behavior on the contact and cooperation rate in face-to-face household surveys." *International Journal of Public Opinion Research*, 19:97-111.

Blom, Annelies G. 2009. "Explaining Cross-Country Differences in Contact Rates." ISER Working Paper Series 2009-08. Colchester: Institute for Social and Economic Research, University of Essex.

Blom, Annelies G., Annette Jäckle, and Peter Lynn. forthcoming. "The use of contact data in understanding cross-national differences in unit non-response." In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, eds. Harkness, Janet, Michael Braun, Brad Edwards, Timothy Johnson, Lars Lyberg, and Peter Ph. Mohler. New York: Wiley & Sons.

Campanelli, Pamela, Patrick Sturgis, and Susan Purdon. 1997. *Can You Hear Me Knocking? An Investigation into the Impact of Interviewers on Survey Response Rates*. London: Social and Community Planning Research.

Elliot, Dave. 1991. *Weighting for Non-Response: A Survey Researcher's Guide*. London: Office for Population Censuses and Surveys, Social Survey Division.

European Social Survey. 2003. *ESS-1 2002 Documentation Report*. 6.1 ed. Bergen: European Social Survey Data Archive, Norwegian Social Science Data Services.

-----. 2005. *ESS-2 2004 Documentation Report*. 3.2 ed. Bergen: European Social Survey Data Archive, Norwegian Social Science Data Services.

-----. 2006. *Round 3 Specifications for Participating Countries*. London: Centre for Comparative Social Surveys.

-----. 2007. *ESS-3 2006 Documentation Report*. 3.2 ed. Bergen: European Social Survey Data Archive, Norwegian Social Science Data Services.

Gabler, Siegfried and Matthias Ganninger. 2009. "Gewichtung." In *Handbuch der sozalwissenschaftlichen Datenanalyse*, eds. Best, Henning and Christof Wolf. Wiesbaden: VS-Verlag.

Gabler, Siegfried, Matthias Ganninger, Sabine Häder, and Ralf Munnich. 2008. "Design effects." In *Encyclopedia of Survey Research Methods*, eds. Lavrakas, Paul J., pp.193-197. Los Angeles: Sage Publications.

Gabler, Siegfried, Sabine Häder, and Peter Lynn. 2006. "Design effects for multiple design samples." *Survey Methodology*, 32:115-120.

Göksel, Hüseyin, David R. Judkins, and William D. Mosher. 1992. "Nonresponse adjustments for a telephone follow-up to a national in-person survey." *Journal of Official Statistics*, 8:417-431.
Goyder, John. 1985. "Nonresponse on surveys: A Canada - United States comparison." *Canadian Journal of Sociology,* 10:231-251.
Groves, Robert M. 2006. "Nonresponse rates and nonresponse bias in household surveys." *Public Opinion Quarterly*, 70:646-675.
Groves, Robert M., Robert B. Cialdini, and Mick P. Couper. 1992. "Understanding the decision to participate in a survey." *Public Opinion Quarterly*, 56:475-495.
Groves, Robert M. and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons, Inc.
Groves, Robert M., Eleanor Singer, and Amy Corning. 2000. "Leverage-saliency theory of survey participation: Description and an illustration." *Public Opinion Quarterly,* 64:299-308.
Jackman, Simon. 1998. "Correcting surveys for non-response and measurement error using auxiliary information." *Electoral Studies*, 18:7-27.
Kalton, Graham and Ismael Flores-Cervantes. 2003. "Weighting methods." *Journal of Official Statistics*, 19:81-97.
Keeter, Scott, Courtney Kenedy, Michael Dimock, Jonathan Best, and Peyton Craighill. 2006. "Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey." *Public Opinion Quarterly*, 70:759-779.
Kessler, Ronald C., Roderick J. A. Little, and Robert M. Groves. 1995. "Advances in strategies for minimizing and adjusting for survey nonresponse." *Epidemiologic Reviews*, 17:192-204.
Kish, Leslie. 1965. *Survey Sampling*. New York: Wiley.
Kreuter, Frauke, Michael Lemay, and Carolina Casas-Cordero. 2007. "Using proxy measures of survey outcomes in post-survey adjustments: examples from the European Social Survey (ESS)." In *Proceedings of the Survey Research Methods Section*, pp. 3142-3149. Alexandria, VA: American Statistical Association.
Lepowski, James M. and Mick P. Couper. 2002. "Nonresponse in the second wave of longitudinal household surveys." In *Survey Nonresponse*, eds. Groves, Robert M., Don Dillman, John L. Eltinge, and Roderick J. A. Little. Chichester: Wiley.
Lipps, Oliver and Grant Benson. 2005. "Cross-national contact strategies." In *Proceedings of the Survey Research Methods Section*, pp. 3905-3914. Alexandria, VA: American Statistical Association.
Little, Rod and Sonya Vartivarian. 2003. "On weighting the rates in non-response weights." *Statistics in Medicine*, 22:1589-1599.
Little, Rod and Sonya L. Vartivarian. 2005. "Does weighting for nonresponse increase the variance of survey means?" *Survey Methodology*, 31:161-168.
Lynn, Peter and Paul Clarke. 2002. "Separating refusal bias and non-contact bias: Evidence from UK national surveys." *The Statistician*, 51:319-333.
O'Muircheartaigh, Colm and Pamela Campanelli. 1999. "A multilevel exploration of the role of interviewers in survey non-response." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162:437-446.
Olson, Kristen. 2006. "Survey participation, nonresponse bias, measurement error bias, and total bias." *Public Opinion Quarterly,* 70:737-758.
Purdon, Susan, Pamela Campanelli, and Patrick Sturgis. 1999. "Interviewers' calling strategies on face-to-face interview surveys." *Journal of Official Statistics*, 15:199-216.

Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika Trust*, 70:41-55.
-----. 1984. "Reducing bias in observational studies using subclassification on the propensity score." *Journal of the American Statistical Association*, 79:516-524.
Schouten, Barry, Fannie Cobben, and Jelke Bethlehem. 2009. "Indicators for the representativeness of survey response." *Survey Methodology*, 35:101-113.
Schwartz, Shalom H. 2003. "A proposal for measuring value orientations across nations." In *ESS Questionnaire Development Report*. London: Centre for Comparative Social Surveys. City University.
Stoop, Ineke. 2005. *The Hunt for the Last Respondent*. The Hague: SCP.
Stoop, Ineke, Silke Devacht, Jaak Billiet, Geert Loosveldt, and Michel Philippens. 2003. *The Development of a Uniform Contact Description form in ESS*. Paper presented at the 14th International Workshop on Household Survey Nonresponse, Leuven 22-24 September 2003.
Tourangeau, Roger. 2004. "Survey research and societal change." *Annual Review of Psychology*, 55:775-801.
Vehovar, Vasja. n.d.. *Weighting in the ESS - Round 1*. University of Ljubljana, Faculty of Social sciences.
Vehovar, Vasja and Tina Zupanic. n.d.. *Weighting in the ESS - Round 2*. University of Ljubljana, Faculty of Social sciences.
Zhang, Li-Chun. 1999. "A note on post-stratification when analyzing binary survey data subject to nonresponse." *Journal of Official Statistics*, 15:329-334.

APPENDIX A

Table A1: Population distributions used in the post-stratifications

Finland round 1

Male	Level of education		
Age	ISCED 2 or less	ISCED 3	ISCED 4 or more
15-34	242656	318804	104103
35-54	213190	351758	232169
55+	356698	123037	119672

Female	Level of education		
Age	ISCED 2 or less	ISCED 3	ISCED 4 or more
15-34	200769	273682	162066
35-54	171914	318497	286598
55+	512099	165060	110542

Finland round 2

Male	Level of education		
Age	ISCED 2 or less	ISCED 3	ISCED 4 or more
15-34	239949	317341	100897
35-54	190024	352510	235994
55+	361343	146316	138819

Female	Level of education		
Age	ISCED 2 or less	ISCED 3	ISCED 4 or more
15-34	197124	273971	157918
35-54	144921	313896	301419
55+	506377	190235	130581

Finland round 3

Male	Level of education		
Age	ISCED 2 or less	ISCED 3	ISCED 4 or more
15-34	240715	322293	99623
35-54	170131	348444	238216
55+	362848	171330	155933

Female	Level of education		
Age	ISCED 2 or less	ISCED 3	ISCED 4 or more
15-34	197605	278002	156804
35-54	123168	303761	313884
55+	498398	216750	150771

Slovakia round 2

Male	Level of education		
Age	ISCED 2 or less	ISCED 3	ISCED 4 or more
15-34	180544	571724	65059
35-54	91689	538228	115866
55+	147919	230413	59162

Female	Level of education		
Age	ISCED 2 or less	ISCED 3	ISCED 4 or more
15-34	172391	545738	73086
35-54	155011	504289	102437
55+	392412	206109	30062

Poland round 1

Age	Male	Female
15-34	6099477	5881868
35-54	5667999	5765815
55+	3390305	4807649

Level of education		
ISCED 2 or less	ISCED 3	ISCED 4 or more
12645245	15964622	2971633

Poland round 3

Age	Male	Female
15-34	6117750	5914357
35-54	5417528	5514733
55+	3704843	5206477

	Level of education		
Age	ISCED 2 or less	ISCED 3	ISCED 4 or more
15-34	6101000	3870800	1560300
35-54	5553000	3595300	1837400
55+	5571400	1717800	829400

Poland round 2

Age	Male	Female
15-34	6056762	5854838
35-54	5516669	5627507
55+	3492633	4955893

	Level of education		
Age	ISCED 2 or less	ISCED 3	ISCED 4 or more
15-34	6101000	3870800	1560300
35-54	5553000	3595300	1837400
55+	5571400	1717800	829400

Table A2: Contact versus final response propensity strata

Finland

		response propensity quintile					
		lowest	2	3	4	highest	Total
contact propensity quintile	lowest	980	204	0	0	0	1184
	2	80	900	205	0	0	1185
	3	46	41	883	213	0	1183
	4	45	18	72	891	166	1192
	highest	33	25	20	79	1017	1174
	Total	1184	1188	1180	1183	1183	5918

Poland

		response propensity quintile					
		lowest	2	3	4	highest	Total
contact propensity quintile	lowest	623	320	117	30	21	1111
	2	218	304	286	234	66	1108
	3	129	211	290	271	209	1110
	4	93	182	273	292	269	1109
	highest	48	97	155	269	540	1109
	Total	1111	1114	1121	1096	1105	5547

Slovakia

		response propensity quintile					
		lowest	2	3	4	highest	Total
contact propensity quintile	lowest	145	92	51	16	0	304
	2	64	113	82	34	7	300
	3	45	58	73	87	38	301
	4	24	28	66	95	94	307
	highest	24	11	30	69	162	296
	Total	302	302	302	301	301	1508

Note: In these propensity strata only responding sample units are considered

Table A3: Cooperation versus final response propensity strata

Finland

		response propensity quintile					
		lowest	2	3	4	highest	Total
cooperation propensity quintile	lowest	861	229	78	14	2	1184
	2	153	496	345	154	37	1185
	3	62	274	403	311	132	1182
	4	45	120	240	414	367	1186
	highest	63	69	114	290	645	1181
	Total	1184	1188	1180	1183	1183	5918

Poland

		response propensity quintile					
		lowest	2	3	4	highest	Total
cooperation propensity quintile	lowest	1003	110	0	0	0	1113
	2	55	890	166	0	0	1111
	3	8	68	876	153	0	1105
	4	5	5	44	848	207	1109
	highest	40	41	35	95	898	1109
	Total	1111	1114	1121	1096	1105	5547

Slovakia

		response propensity quintile					
		lowest	2	3	4	highest	Total
cooperation propensity quintile	lowest	258	44	0	0	0	302
	2	35	211	57	0	0	303
	3	8	35	185	75	0	303
	4	0	4	21	188	86	299
	highest	1	8	39	38	215	301
	Total	302	302	302	301	301	1508

Note: For reasons of simplicity the propensities were considered by country but across rounds. The final process weights were derived separately by country and round.

Table A4: Response propensity models using maximum information from frame data

Response	**Finland**	**Response**	**Poland**	**Response**	**Slovakia**
	b		b		b
Male	-0.10*	Year of birth	0.01***	Year of birth	0.01**
Urbanicity[1]		Urbanicity[2]		Regions[3]	
urban	-0.13	Village	1.25***	Bratislava	-0.29
rural	0.17*	Town < 10k	0.45**	Trnava	-0.36*
Helsinki	-0.30***	Town 10k-19k	0.56***	Trencin	-0.16
Household size	0.11***	Town 20k-49k	0.70***	Nitra	-0.89***
Language non-Scandinavian	-0.80***	Town 50k-99k	0.34**	Zilina	0.75***
		Town 100k-199k	0.07	Banska Bystrica	-0.34*
		Town 200k-499k	0.23*	Presov	0.07
		Town 500k-999k	0.06		
Round 1 dummy	0.41***	Round 1 dummy	0.14*		
Round 2 dummy	0.28***	Round 2 dummy	0.17*		
Constant	0.52***	Constant	-9.28***	Constant	-12.14*
Chi^2	188	Chi^2	348	Chi^2	116
Pseudo R^2	0.018	Pseudo R^2	0.039	Pseudo R^2	0.038
AIC	10321	AIC	8701	AIC	2992
N	8522	N	7661	N	2363

Legend: *p<0.05; **p<0.01; ***p<0.001

Notes: Coding of dependent variable 'response': 1 interviewed, 0 not interviewed; Table shows logit model coefficients

[1] Semi-urban omitted; [2] Town size over 1m inhabitants omitted; [3] Region of Kosice omitted

Table A5: Contact propensity models using ESS contact and frame data

Contact	Finland		Poland		Slovakia
	b		b		b
Number of contact attempts					
2	4.22***		0.59*		-1.07***
3	2.54***		-2.36***		-1.61***
4	1.24***		-3.50***		-2.22***
5 or more	0.31		-5.15***		-3.64***
Ever f2f call in the evening	0.47***		1.01***		0.37
Ever f2f call on a Saturday	-0.71**		0.94***		0.33
Ever f2f call on a Sunday	1.58**		0.78**		-0.04
Physical state of building:					
Satisfactory	-0.25**		0.07		0.22
Bad	-0.07		0.03		0.36
Farm or single-unit housing	0.34***		0.18		0.47*
Interviewer cooperation rate	0.03***		0.02***		0.01**
Interviewer f2f contact rate	0.07***		0.05***		0.09***
Interviewer phone contacting	-0.07***		-0.05***		-0.01*
Interviewer f2f evening calling	0.00		-0.01*		0.00
Male	-0.09	Year of birth	0.00	Year of birth	-0.01
Urbanicity1		Urbanicity2		Regions3	
urban	0.06	Village	-0.04	Bratislava	0.23
rural	0.31*	Town < 10k	-0.55	Trnava	0.62
Helsinki	-0.01	Town 10k-19k	-0.26	Trencin	-0.32
Household size	-0.01	Town 20k-49k	-0.63	Nitra	-0.51
Language non-Scandinavian	-0.36	Town 50k-99k	-0.76*	Zilina	-0.11
		Town 100k-199k	0.07	Banska Bystrica	-0.64
		Town 200k-499k	-0.58*	Presov	-0.22
		Town 500k-999k	-0.76**		
Round 1	-0.02		-0.34		
Round 2	0.06		0.73**		
Constant	-3.64***		-3.55		5.96
Chi^2	4550		484		287
Pseudo R^2	0.48		0.26		0.25
AIC	4925		1407		897
N	8522		7658		2359

Legend: *p<0.05; **p<0.01; ***p<0.001

Notes: Coding of dependent variable 'contact': 1 in-person contact with the household, 0 no in-person contact with the household; Table shows logit model coefficients

[1] Semi-urban omitted; [2] Town size over 1m inhabitants omitted; [3] Region of Kosice omitted

Table A6: Cooperation propensity models using ESS contact and frame data

Cooperation	**Finland**		**Poland**		**Slovakia**
	b		b		b
Mode of first contact: phone	2.52 ***		2.86 ***		1.77 ***
Time of first f2f contact: evening	0.08		-0.29 *		-0.07
No refusal	5.27 ***		4.54 ***		3.43 ***
Change of interviewer	-1.42 **		1.19 ***		1.47 *
Number of contact attempts	0.09		-0.38 *		-0.54 *
2	-0.06		-0.99 ***		-0.86 *
3	-1.10 *		-0.96 ***		-0.16
4 or more	-0.48 **		-0.02		-0.40 **
Physical state of building:					
Satisfactory	-1.40 ***		-0.30 *		-0.52
Bad	0.30		0.44 ***		0.28 *
Farm or single-unit housing	0.05 ***		0.04 ***		0.05 ***
Interviewer cooperation rate	-0.06 **		-0.01		0.01
Interviewer f2f contact rate	0.00		0.01 **		0.01
Interviewer f2f evening calling	-0.24		2.86 ***		1.77 ***
Male	2.52 ***	Year of birth	0.00	Year of birth	0.00
Urbanicity1		Urbanicity2		Regions3	
urban	-0.18	Village	0.31	Bratislava	0.48
rural	-0.54 *	Town < 10k	-0.11	Trnava	-0.24
Helsinki	0.26	Town 10k-19k	0.18	Trencin	0.50
Household size	0.11	Town 20k-49k	0.66 **	Nitra	-0.29
Language non-Scandinavian	-1.72 ***	Town 50k-99k	0.28	Zilina	0.31
		Town 100k-199k	0.24	Banska Bystrica	0.16
		Town 200k-499k	0.42 *	Presov	0.08
		Town 500k-999k	0.43 *		
Round 1	0.18		0.05		
Round 2	0.01		0.27 **		
Constant	-0.66		0.13		-8.38
Chi^2	2144		4089		1275
Pseudo R^2	0.58		0.48		0.46
AIC	1629		4455		1521
N	6461		7460		2205

Legend: *p<0.05; **p<0.01; ***p<0.001

Notes: Coding of dependent variable 'cooperation': 1 interviewed, 0 not interviewed but in-person contact with the household achieved; Table shows logit model coefficients

[1] Semi-urban omitted; [2] Town size over 1m inhabitants omitted; [3] Region of Kosice omitted

Table A7: Correlations of the design and nonresponse weights

Finland round 1	**Design weight**	**Post-stratification weight**	**Process weight**	**Post-stratified process weight**	**Total weight**
Design weight	.				
Post-stratification weight	.	1.00			
Process weight	.	0.00	1.00		
Post-stratified process weight	.	0.34	0.93	1.00	
Total weight	.	0.33	0.87	0.94	1.00
Finland round 2					
Design weight	.				
Post-stratification weight	.	1.00			
Process weight	.	-0.01	1.00		
Post-stratified process weight	.	0.29	0.93	1.00	
Total weight	.	0.29	0.88	0.94	1.00
Finland round 3					
Design weight	.				
Post-stratification weight	.	1.00			
Process weight		0.03	1.00		
Post-stratified process weight	.	0.31	0.94	1.00	
Total weight	.	0.32	0.88	0.95	1.00
Poland round 1					
Design weight	1.00				
Post-stratification weight	0.89	1.00			
Process weight	-0.01	-0.01	1.00		
Post-stratified process weight	0.00	0.11	0.96	1.00	
Total weight	0.01	0.12	0.80	0.84	1.00
Poland round 2					
Design weight	1.00				
Post-stratification weight	0.89	1.00			
Process weight	0.01	-0.01	1.00		
Post-stratified process weight	0.05	0.13	0.96	1.00	
Total weight	0.03	0.11	0.80	0.84	1.00
Poland round 3					
Design weight	1.00				
Post-stratification weight	0.82	1.00			
Process weight	0.19	0.17	1.00		
Post-stratified process weight	0.20	0.32	0.95	1.00	
Total weight	0.18	0.29	0.85	0.91	1.00
Slovakia round 2					
Design weight	.				
Post-stratification weight	.	1.00			
Process weight	.	-0.11	1.00		
Post-stratified process weight	.	0.47	0.78	1.00	
Total weight	.	0.42	0.72	0.91	1.00

Note: All nonresponse weights include the design weight

APPENDIX B: Question wording

Voted in last national election
"Some people don't vote nowadays for one reason or another. Did you vote in the last [country] national election in [month/year]?"
Answer categories: Yes, No, Not eligible to vote

Contacted a politician / Took part in a lawful demonstration
"There are different ways of trying to improve things in [country] or help prevent things from going wrong. During the last 12 months, have you done any of the following? …
… Contacted a politician, government or local government official?
… Taken part in a lawful public demonstration?"
Answer categories: Yes, No

Party membership
"Are you a member of any political party?"
Answer categories: Yes, No

Happiness
"Taking all things together, how happy would you say you are?"
Extreme points on the 11-point scale: "Extremely unhappy" and "Extremely happy"

CED depression scale
"I will now read out a list of the ways you might have felt or behaved during the past week… [P]lease tell me how much of the time during the past week…
… you felt depressed?
… you felt that everything you did was an effort?
… your sleep was restless?
… you were happy?
… you felt lonely?
… you enjoyed life?
… you felt sad?
… you could not get going?"
Answer scale: 1 "None or almost none of the time" 2 "Some of the time" 3 "Most of the time" 4 "All or almost all of the time"

Rosenberg trust scale
"[G]enerally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?"
Extreme points on the 11-point scale: "You can't be too careful" and "Most people can be trusted"
[D]o you think that most people would try to take advantage4 of you if they got the chance, or would they try to be fair?"
Extreme points on the 11-point scale: "Most people would try to take advantage of me" and "Most people would try to be fair"
"Would you say that most of the time people try to be helpful or that they are mostly looking out for themselves?"
Extreme points on the 11-point scale: "People mostly look out for themselves" and "People mostly try to be helpful"

Political trust
"[P]lease tell me on a score of 0-10 how much you personally trust each of the institutions I read out. 0 means you do not trust an institution at all, and 10 means you have complete trust. …
… [country]'s parliament?
… the legal system?
… the police?
… politicians?"
Extreme points on the 11-point scale: "No trust at all" and "Complete trust"

Schwartz human values scale
These questions were part of the ESS supplementary questionnaire. In Finland the supplementary questionnaire was implemented as a self-completion paper questionnaire. In Poland and Slovakia the supplementary questionnaire was implemented face-to-face as a continuation of main interview.

Male and female respondents were asked how much they were like a described person. Male respondents received questions about male third persons and female respondents about female. Otherwise the questions were identical. The below questions were for female respondents.

For documentation on how to derive the scales from these measures see http://ess.nsd.uib.no/index.jsp?year=2003&country=&module=other.

"Here we briefly describe some people. Please read each description and tick the box on each line that shows how much each person is or is not like you.
- Thinking up new ideas and being creative is important to her. She likes to do things in her own original way.
- It is important to her to be rich. She wants to have a lot of money and expensive things.
- She thinks it is important that every person in the world should be treated equally. She believes everyone should have equal opportunities in life.
- It's important to her to show her abilities. She wants people to admire what she does.
- It is important to her to live in secure surroundings. She avoids anything that might endanger her safety.
- She likes surprises and is always looking for new things to do. She thinks it is important to do lots of different things in life.
- She believes that people should do what they're told. She thinks people should follow rules at all times, even when no-one is watching.
- It is important to her to listen to people who are different from her. Even when she disagrees with them, she still wants to understand them.
- It is important to her to be humble and modest. She tries not to draw attention to herself.
- Having a good time is important to her. She likes to "spoil" herself.
- It is important to her to make her own decisions about what she does. She likes to be free and not depend on others.
- It's very important to her to help the people around her. She wants to care for their well-being.
- Being very successful is important to her. She hopes people will recognise her achievements.

- It is important to her that the government ensures her safety against all threats. She wants the state to be strong so it can defend its citizens.
- She looks for adventures and likes to take risks. She wants to have an exciting life.
- It is important to her always to behave properly. She wants to avoid doing anything people would say is wrong.
- It is important to her to get respect from others. She wants people to do what she says.
- It is important to her to be loyal to her friends. She wants to devote herself to people close to her.
- She strongly believes that people should care for nature. Looking after the environment is important to her.
- Tradition is important to her. She tries to follow the customs handed down by her religion or her family.
- She seeks every chance she can to have fun. It is important to her to do things that give her pleasure."

Answer scales headed by "How much like you is this person?"
Answer scales: 1 " Very much like me" 2 " Like me" 3 "Some-what like me" 4 " A little like me" 5 " Not like me at all"