Survey Research Methods (2014)
Vol. 8, No. 2, pp. 67-80
ISSN 1864-3361
http://www.surveymethods.org

Evaluating the Impact of Response Enhancement Methods on the Risk of Nonresponse Bias and Survey costs

Caroline Roberts
Institute of Social Sciences
University of Lausanne, Switzerland

Caroline Vandenplas
Centre for Sociological Research
KU Leuven, Belgium

Michèle Ernst Stähli
FORS – Swiss Centre of Expertise in the Social Sciences
Lausanne, Switzerland

The pursuit of high response rates to minimise the threat of nonresponse bias continues to dominate decisions about resource allocation in survey research. Yet a growing body of research has begun to question this practice because the cost of efforts to increase response rates is not always justified in terms of their impact on nonresponse bias. In this study, we assess the costs and benefits of different methods designed to increase response rates on the European Social Survey in Switzerland, by analysing data from a new sampling frame based on population registers to examine the changing composition of the respondent pool and the risk of bias, alongside the financial costs associated with additional fieldwork efforts. We compute an R-indicator to assess representativity with respect to the sampling register variables, and find little improvement in the sample composition as response rates increase. We then examine the impact of response rate increases on the risk of nonresponse bias based on Maximal Absolute Bias (MAB), and coefficients of variation between subgroup response rates, alongside the associated costs of different types of fieldwork effort. The results show that increases in response rate help to reduce MAB, while only small but important improvements to sample representativity are gained by varying the type of effort. The findings lend further support to research that has called into question the value of extensive investment in procedures aimed at reaching response rate targets and the need for more tailored fieldwork strategies aimed both at reducing survey costs and minimising the risk of bias.
Keywords: R-indicator; European Social Survey; response rates; coefficient of variation

1 Introduction

Response rates play a critical role in survey design and implementation. Expectations about likely response rates guide decisions about sampling and fieldwork procedures, and the setting of target response rates often determines the level of investment in methods aimed at encouraging sample members to respond (Wagner & Raghunathan, 2010). Meanwhile, final response rates remain the most widely-used indicators of survey quality (Biemer & Lyberg, 2003) and the risk of nonresponse bias (Wagner, 2012), and commonly form the basis of comparisons across different studies, both within and between countries. In the face of declining response rates in many countries over the past few decades (e.g. Brick & Williams, 2013; De Leeuw & De Heer, 2002), there have been growing concerns about the quality of the data collected, because of the threat of bias from nonresponse. This has led to the use of expensive response enhancement methods designed to help achieve response rates targets, such as additional contact attempts, more valuable incentives, and refusal conversion interviews. While increased efforts to reduce noncontacts and refusals can be shown to have a positive impact on response rates, it is not always clear whether they are worth it financially, in terms of achieving their goal of reducing or preventing nonresponse bias

In this study, we investigate the effect of efforts to increase participation and their relation to costs in the European Social Survey (ESS), a cross-national study, which since its launch has explicitly imposed a response rate target for participating countries of 70%, and encouraged countries to use a variety of response enhancement procedures in order to reach this target. We use data from Switzerland, where relatively low response rates in the earliest rounds of the ESS led to substantial investment in methods to increase levels of participation, and a consequent significant improvement in response rates in later rounds. However, it is not clear whether this investment has paid off in terms of its impact on the composition of the responding sample, its representativity, and the potential for nonresponse bias in key survey variables. In the following, we investigate these issues using data from round 5 of the ESS (European Social Survey Data Archive, 2012), and new data from the Swiss population register, which for the first time in this round provided

Caroline Roberts, Institute of Social Sciences, University of Lausanne, Géopolis – Room 5548, CH-1015 Lausanne, Switzerland (caroline.roberts@unil.ch)

the sampling frame for the survey[1]. Before describing our methodological approach, we first review current developments in the debate surrounding the value of response rates as indicators of survey quality and the search for alternatives, to serve both as quality indicators, and as guiding principles in data collection design.

Response rates as indicators of nonresponse bias

Response rates are popular for good reasons – they are generally easy to calculate and they provide an intuitive summary of the result of a survey, facilitating comparisons across multiple studies (assuming the same methods are used to compute response rates – see American Association for Public Opinion Research, 2011), and across different data collection protocols (Wagner, 2012). Continued reliance on response rates for the assessment of survey quality, and investment in methods aimed at limiting the rate of nonparticipation both rest on the assumption that the accuracy of a survey estimate improves with higher response rates, while the potential for bias declines.[2] A growing body of research evidence now calls this assumption into question (Keeter, Miller, Kohut, Groves, & Presser, 2000; Curtin, Presser, & Singer, 2000; Groves, 2006). Notably, a meta-analysis of studies investigating the relation between nonresponse bias and response rates by Groves and Peytcheva (2008) found only a weak correlation between bias and nonresponse, suggesting that high response rates do not always guarantee an absence of bias, while low response rates do not necessarily result in an increase in bias. This seemingly counter-intuitive finding makes sense given that nonresponse bias is a function not only of the presence of nonresponse, but also the extent and nature of differences between the respondents and nonrespondents. The extent and nature of observed differences between respondents and nonrespondents depends on the relation between variables influencing the decision to participate in a survey and responses to the survey questions (Bethlehem, 2002). An important implication of this is that bias is not something affecting the whole survey uniformly, but an item-level attribute (Groves, 2006; Groves & Peytcheva, 2008), with items, which are highly correlated with variables influencing the decision to participate being particularly at risk.

Research comparing the characteristics and responses of the most cooperative, early respondents to a survey with those who participate only after additional recruitment efforts, has found that late, or 'reluctant' respondents more closely resemble the final nonrespondents (Lin & Schaeffer, 1995). Thus, increasing efforts to contact the hard-to-contact, and persuade more reluctant sample members to participate can help to lower nonresponse rates and to reduce bias. However, not all strategies for reducing nonresponse rates are equally effective at reducing bias, as their effectiveness depends on their impact on different types of nonresponse (e.g. Peytchev, Baxter, & Carley-Baxter, 2009). Sample members who refuse to take part in a survey tend to be more likely to resemble eventual nonrespondents than sample members who are simply hard to contact, so addressing nonresponse from refusals is likely to have a bigger impact on bias reduction than simply trying to reduce noncontacts (Groves et al., 2006). The main implications of this are that response rates alone will mostly be poor predictors of bias, and that efforts to improve response rates may fail to reduce bias if the survey protocol continues to attract more respondents with the same characteristics as those who have already responded, or worse still, actually exacerbate bias if the mechanism underlying the decision to participate under a given protocol is correlated with important variables of interest.

Alternatives to response rates

In light of this understanding of the nature of nonresponse bias, survey methodologists have begun to question the value of response rates as indicators of survey quality and of setting response rate targets in survey design. This has prompted the development of alternative indicators of the risk of nonresponse bias (Groves & Peytcheva, 2008; Skalland, 2011; Wagner, 2012). A recent review of progress in this field by Wagner (2012) introduces a typology of such indicators based on the different data sources used in their estimation. Specifically, they include different combinations of the response indicator (i.e. whether a sample case responded or not) available frame or paradata (Couper, 1998), and the survey data themselves. Given the need for a pragmatic solution that can offer some of the advantages of response rates (being easy to calculate, and easily understood), the most promising alternatives are indicators that draw on a range of available data (unlike response rates, that only consider the response indicator) to provide summary information about the representativity of a responding sample and the likely risk of bias, without the added complexity associated with calculating and reporting estimate-specific indicators. Three examples of such approaches that are gaining popularity include sub-group response rates (Groves et al., 2008); the coefficient of variation of subgroup response rates, which is calculated as the ratio of the standard deviation of the subgroup response rates to the overall response rate (Groves, 2006; Wagner, 2012), and so-called Representativity Indicators, or 'R-indicators' (Schouten & Cobben, 2007), which measure the dispersion of estimated response propensities (the probability of taking part in the survey given certain observed attributes) based on available auxiliary data, to assess the extent to which the responding sample of a survey resembles the complete sample of respondents and nonrespondents. Each of these approaches involves the use of

[1] The ESS 2010 was the first survey conducted by FORS (The Swiss Centre of Expertise in the Social Sciences) to be based on the new 'SRPH' (Stichprobenrahmen fr Personen- und Haushaltserhebungen) sampling frame of the Swiss Federal Statistical Office. Previous rounds of the ESS were based on other sample frames, all household or address based.

[2] In the ESS, response rate targets are also intended to help countries obtain comparable effective sample sizes, in the hope that bias from nonresponse will somehow be equivalent cross-nationally (Stoop, Billiet, Koch, & Fitzgerald, 2010).

complete auxiliary information typically relating to socio-demographic characteristics of a sample as the basis for comparisons between respondents and nonrespondents. Depending on the extent to which these characteristics co-vary with other survey variables and individual response propensities, the indicators may provide information about the potential for bias on key estimates (Wagner, 2012).

As well as providing an intuitive assessment of overall survey quality (and a basis for comparisons across studies), indicators of the risk of nonresponse bias are also being developed to serve as guides in fieldwork planning, to enable a better informed, more tailored approach to the reduction of nonresponse. Though still a long way from becoming common practice (at least in the European context), the aim of such developments – referred to as adaptive and responsive survey designs (Groves & Heeringa, 2006; Wagner, 2008; Schouten, Shlomo, & Skinner, 2011) – is to channel resources into nonresponse follow-ups that limit the risk of bias, rather than the blind pursuit of response rate targets, to achieve more cost- and error-efficient data collection designs. Key to such designs is the need to identify which sample cases should be prioritised in subsequent phases (Peytchev, Riley, Rosen, Murphy, & Lindblad, 2010) and rules about the optimal moment to stop a given phase of fieldwork altogether (Rao, Glickman, & Glynn, 2008; Wagner & Raghunathan, 2010).

R-indicators, in particular, have been developed with the twin aims of assessing survey quality and enabling tailored fieldwork strategies in mind. Schouten and Cobben (2007) define quality in terms of the representativity of the responding sample – i.e. the extent to which it is similar to the complete sample. In turn, they define representativity formally in terms of the distribution of response propensities across different categories of a covariate. The sample responding to the survey is said to be representative on a given covariate, if the average response propensity over the categories of the covariate is constant (Schouten, Cobben, & Bethlehem, 2009, p. 103). The R-indicator provides a measure of the dispersion of response propensities. The same authors illustrate how R-indicators can be used to assess the effect of response enhancement methods on sample representativity with an example from the 1998 Dutch Household Living Conditions survey (POLS). Telephone and face-to-face follow-up attempts in the second month of survey fieldwork succeeded in increasing the response rate from 47% to 60%. However, the R-indicator showed a drop in representativity, because the fieldwork procedures only succeeded in bringing into the respondent pool cases that were easier to contact (Schouten & Cobben, 2007, p. 6).

More recently, similar results were obtained in a simulation study by Beullens and Loosveldt (2012), designed to compare different strategies for following-up nonrespondents. These authors show that a strategy of implementing additional fieldwork efforts simply to maximise response rates is incompatible with strategies to minimise the variance across response propensities and to reduce bias because the former involves prioritising high propensity cases, while the latter implies the opposite. Efforts to increase response rates succeeded by bringing additional cases into the respondent pool that shared characteristics with other respondents with higher response propensities, just as was the case in Shouten and Cobben's (2007) study. They conclude that targeting fieldwork strategies may result in a lower response rate, but the risk of nonresponse bias is reduced considerably, and the statistical power of the sample (despite the reduced number of cases) can be substantially improved by minimising the need for large post-stratification weights.

The possibility of using adaptive and/or responsive survey designs requires an expertise and infrastructure that is unlikely to be readily available in smaller survey organisations, and is likely to limit the wholesale take-up of methods of tailoring fieldwork efforts in many countries for the present time. Such methods also depend upon the availability of certain types of auxiliary data (rich sampling frame data from registers, or paradata that correlate with response propensity and responses to key survey variables – Kreuter et al., 2010), making adaptive designs, unrealistic in many survey contexts. However, where such data are available, we believe an important first step in the direction of developing more cost-efficient fieldwork strategies for the future is to improve understanding of the relation between different types of fieldwork effort, response rates, sample representativity, the risk of nonresponse bias, and survey costs. The present study was motivated by this concern. The availability in the Swiss ESS 2010 of new sampling register data for both respondents and non-respondents presented for the first time the opportunity to evaluate the effectiveness of the various response enhancement techniques in use on the survey, and to explore the potential for developing more cost-effective nonresponse follow-up efforts in future waves. Using a variety of indicators of survey quality calculated on the basis of the register data, we address the following research questions:

1. How effective are different methods of response enhancement at improving response rates?
2. What impact do different response enhancement methods have on the representativity of the responding sample and the risk of nonresponse bias? In other words, how successful are different methods at diversifying the composition of the respondent pool by bringing in sample cases with lower response propensities?
3. Are the different methods in use worth the financial investment, given their impact on these different aspects of survey quality?

In the next section we describe the data and our analytic approach, before presenting the results of our analysis.

2 Method

Data

We use data from round 5 (2010) of the Swiss European Social Survey. As well as using questionnaire data from the main face-to-face survey interview, we additionally analyse data from interviewer contact forms designed to document the outcome of all visits to the addresses of sampled individuals and all centralised telephone contact attempts, and data from the sampling frame, which in 2010, was the Swiss Federal Statistical Office's (SFSO) individual sampling base, which is based on population registers maintained by municipalities. The population for the ESS is all resident adults (aged 15 and over) within private households, 'regardless of their nationality, citizenship, language or legal status' (European Social Survey Data Archive, 2012, p. 7) (though sample members not sufficiently competent in (Swiss) German, French and Italian may be unable to take part). The possibility to use the SFSO's new sampling frame meant that it was possible to obtain a single-stage equal probability systematic sample of individuals, with no clustering. The sample was proportionally stratified by the 7 NUTS regions of Switzerland[3]. The total number of issued sample units was 2850, and the final number of valid interviews was 1506 – a total response rate[4] of 53.3%.

Fieldwork protocol and response enhancement methods

The ESS specifications for participating countries encourage national teams to budget their surveys with a minimum target response rate of 70% and a maximum non-contact rate of 3% of all sample units in mind (European Social Survey, 2009). Response enhancement methods on the survey are, therefore, designed with a view to achieving these targets. In the Swiss ESS 2010, the methods included a combination of monetary incentives, extra contact attempts to minimise noncontacts, and 'refusal conversion' interviews. In addition, a nonresponse follow-up survey was carried out to gather additional data about the nonresponding sample. None of these methods were deliberately targeted at specific subgroups (apart from on the basis of non-contact and refusal status following previous contacts). The choice and the timing of the different phases of fieldwork were informed by the survey organisation's experiences in previous rounds of the survey, and by the targets. Furthermore, the sample management system used by the fieldwork agency did not offer detailed real-time information about fieldwork progress, so opportunities for targeting follow-up strategies would have been limited. The fieldwork procedures used are described in more detail in the following, and illustrated in Figure 1.

All respondents were sent an advance letter along with a leaflet about the survey. For a randomly selected 80% of the sample, the letter contained the offer of a 30.- Swiss Franc incentive conditional on completion of the interview, and the respondent could choose between cash, a voucher for Interflora or the cinema, or making a donation to their choice from a selection of charities. The remaining 20% of the sample were given an unconditional cash incentive of 30.- Swiss Francs (about 25 Euros) in the envelope with the advance letter.[5]

The face-to-face interviewers received one day's training on the fieldwork specifications, refusal avoidance techniques and the questionnaire. They were required to initially make up to five personal visits to the sampled individual's address on/at different days/times over the course of a two week period, at least one of which was required to be on a weekend and another in the evening to establish contact with the target respondent (the ESS specifications request a minimum of four face-to-face contact attempts). If a successful initial contact with the target respondent was made in person, further contacts to arrange an appointment for the face-to-face interview were permitted by telephone, if the interviewer was provided with one. Addresses were distributed to interviewers on a geographical basis. The precise timing of sending the contact letters was left up to each interviewer, who decided when to send them based on the geographical distribution of the addresses and personal availability.

All sample members who refused to participate *on the doorstep*[6] were reallocated to another, more experienced interviewer, chosen among those obtaining the best response rates, who had been specially trained in refusal conversion. Before re-contact attempts for refusal conversion began the cases were sent a personalised letter, and interviewers were given a special brochure containing findings from the 2008 ESS, which they could use to help persuade them to participate. This face-to-face refusal conversion procedure began 6 weeks after the start of fieldwork. During the same period, attempts were made to contact people who could not be contacted during the first five in-person visits, by sending a card to ask the best way to make contact and by additional face-to-face visits.

Thirteen weeks after the start of fieldwork, a final centralised contact procedure conducted by telephone was launched, with the principal aim of including people with whom there had so far been no successful contact attempt, or who had not been reached during the refusal conversion phase, but also people stating they were absent, temporarily ill or having other problems (selected for inclusion based on comments provided by interviewers). In this procedure, unlimited additional contact attempts were carried by interviewers at the fieldwork agency's centralised call centre (again, at different times and on different days), but the procedure was restricted to sample members for whom a telephone number was available. Telephone numbers were available for just over 61% of the total sample following an automated matching procedure with the commercial database AZ Direct.

Two months after the end of the main survey fieldwork,

[3] CH01: Région lémanique; CH02: Espace Mittelland; CH03: Nordwestschweiz; CH04: Zürich; CH05: Ostschweiz; CH06: Zentralschweiz; CH07: Ticino

[4] ESS response rates are calculated as the total number of completed interviews divided by the total number of interviews selected minus ineligibles. They are equivalent to the AAPOR Response Rate 1.

[5] Response rates in the unconditional incentive group were significantly higher (57.9%) than in the conditional incentive group (51.6%) (X^2=7.58, p<.001).

[6] In Switzerland, people who take the initiative to contact the survey organization to clearly announce their refusal to participate are considered as hard refusals and are not included in the refusal conversion procedure. Moreover, due to time constraints, some refusals are not followed up properly. This is regrettable, especially as cases that are more difficult to contact seem to be the ones concerned.

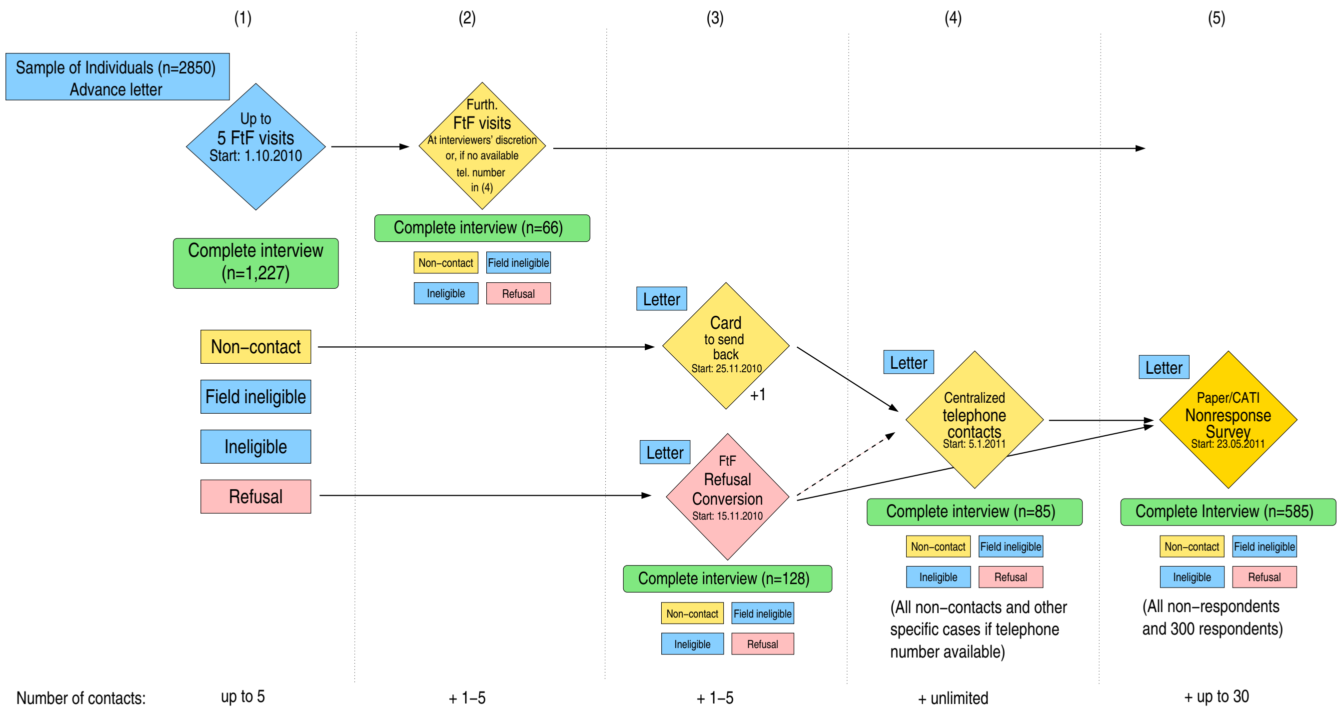

Figure 1. Schematized fieldwork procedures of the ESS 2010 in Switzerland

The figure aims at schematizing the fieldwork procedures as specified in the fieldwork contract. In principle, all nonrespondents after 5 face-to-face visits should be reconsidered. However, due to time constraints, certain refusals or noncontacts are not re-approached. There is no systematic selection of cases to be prioritized but it cannot be excluded that interviewers choose to re-approach cases based on perceptions about their likely cooperativeness.

a nonresponse follow-up survey was carried out, consisting of a single sheet (double-sided) paper questionnaire with around 20 questions sent with a personalised covering letter and a 10.- Swiss Franc (about 8 Euros) cash incentive to sample members who remained un-contacted or who had refused to participate in the main survey. Those who had not completed and returned the questionnaire within 4 weeks were then re-contacted by telephone (if a number was available) or by mail (if no number was available).

Survey fieldwork was conducted by M.I.S. Trend SA. Data collection began on 1st October 2010, and continued until 23rd March 2011 (the nonresponse follow-up was fielded between end of May and mid July 2011). A summary of response rates by fieldwork effort is provided in Table 1. Note that while the different procedures described were carried out in a broadly chronological order (as described above), the treatment of refusals occurred simultaneously alongside the face-to-face contact attempts, while the telephone contact attempts were undertaken towards the end of the main fieldwork period. Following the distinction drawn by Peytchev et al. (2009), we argue that additional contact attempts both face-to-face and by telephone represent an increase in the *level* of effort to solicit a survey response, while the refusal conversion protocol (and later, the nonresponse follow-up survey) represent changes in the *type* of effort used to recruit respondents. This distinction forms the basis of our analytic approach; however, we acknowledge that it is not a straightforward one given the differential treatment of sample members with and without telephone numbers. We return to this issue in the discussion.

Analytic approach

To recap, the aim of this study was to assess the value of the different response enhancement techniques used on the ESS5 2010, by evaluating their impact on response rates, representativity and the risk of nonresponse bias in relation to their associated costs. In the following we describe the analytic approach we use to assess the impact of different response enhancement techniques and to evaluate fieldwork costs.

Assessing the impact of different response enhancement techniques. Using contact form data, we differentiate respondents and nonrespondents at each stage of the fieldwork, identifying what procedures were used to achieve each completed interview. We then use sampling frame data to assess the extent to which the responding sample is representative of the sample as a whole, and the extent to which improvements in response rates, as a result of increasing the amount or changing the type of fieldwork effort either improve or worsen representativity and the risk of nonresponse

Table 1 ESS5 2010 Response rates by response enhancement efforts

	N	%
Break down of final response and nonresponse (n=2850)		
Refusal by respondent	713	25.0
Refusal by proxy (or household or address refusal)	76	2.7
No contact	278	9.8
Language barrier	67	2.4
Respondent mentally or physically unable to participate	64	2.3
Respondent unavailable throughout fieldwork period	109	3.8
Address ineligible[b]	20	0.7
Respondent moved abroad	10	0.4
Respondent deceased	7	0.3
Number of valid interviews	1506	52.8
Completed interviews resulting from (n=1506)		
First five in-person visits	1227	81.5
Additional in-person visits	66	4.4
FtF refusal conversions	128	8.5
Centralised telephone contacts	85	5.6
Completed interviews by incentive group		
SFR 30.- conditional	1176/2280	51.6
SFR 30.- unconditional	330/570	57.9
Total nonrespondents eligible for follow-up (n=1047)[b]		
Non-contacts	278	26.6
Refusals and refusals by proxy (excluding office refusals)	769	73.5
Completed follow-up questionnaires by nonrespondents		
On paper	530	50.6
By telephone	55	5.3

[a]Not residential, not occupied, not traceable or other ineligible.
[b]Does not include respondents who were sent the nonresponse follow-up questionnaire.

bias. To assess the representativeness of the responding sample, we use three different approaches. First, we compute a R-indicator (Schouten & Cobben, 2007; Schouten et al., 2009). Then, we examine actual variations in subgroup response rates across fieldwork stages, using the coefficient of variation proposed by Wagner (2012). Finally, we use socio-demographic data from the sampling register to analyse the characteristics of respondents and nonrespondents at each fieldwork stage to get a clearer idea of how changes in response enhancement methods impact on the representation of different subgroups.

Based on the R-indicator, we are able to estimate Maximal Absolute Bias (Schouten et al., 2009), which provides an indication of the risk of bias in a hypothetical survey variable in the worse case scenario by assessing the difference between the respondents and the full sample (described in more detail below). We also estimate the Maximal Absolute Contrast (Schouten, Shlomo, & Skinner, 2010), which measures the worse case scenario difference between the respondents and the non-respondents.

R-indicators and Maximal Absolute Bias. As the theory and method underlying the computation of R-indicators, their properties, and the implications of using alternative approaches in their estimation have all been discussed extensively elsewhere by the original developers (see, in particular, Schouten & Cobben, 2007; Schouten et al., 2009, 2011), we provide only a brief overview here. R-indicators are based on response-based estimates of individual response propensities and the average response propensity, derived from a model predicting the probability of participation in the survey from a selection of (categorical) covariates available for both respondents and nonrespondents alike.

To estimate the response propensities in this study, we estimated a logistic regression equation, using auxiliary data available for all sample members, including variables from the register-based sampling frame, and one variable specific to the survey design. The final list of covariates from the frame included respondent sex (coded 1 if male); dummy variables for the age categories (<30, 31-44, 45-64, leaving the group aged 65 and over as the reference category); marital status (coded 1 if married or in a legal partnership, 0 if single, divorced or widowed); dummy variables for nationality (being from a country bordering Switzerland, or from another non-bordering country compared with the reference category 'Swiss'); linguistic region variables (being from the French or Italian-speaking regions compared with the reference category 'German or Romansch-speaking regions'[7]); and living in an urban area (compared to an isolated

[7] People in the Romansch-speaking areas were interviewed by

town or rural community). The design-specific covariate was whether or not a telephone number was available from the AZ Direct database, which determined the possibility of including a case in the centralised telephone contact attempts. This list represents all the variables that were available, and we decided to include all of them in the final model irrespective of the strength of their relation to the decision to participate in the survey.[8] We discuss the implications of this later.

With the exception of sex, and linguistic region (those from the Italian-speaking region were slightly less likely to participate than those in the German-speaking region, but the effect only approached significance), all covariates were significantly associated with the probability of participating in the survey (Table not shown). It is noteworthy that, even when controlling for other variables, sample members with telephone numbers were almost one and a half times more likely to participate than those without (odds ratio = 1.47). Overall, Nagelkerke's R^2 was just 0.07, indicating only a weak relation between the available predictors and the likelihood of participating in the survey. However, Hosmer-Lemeshow's test (p=0.71) shows that actual response rates were not significantly different from those predicted by the model.

Computing a R-indicator based on the estimated response propensities involves an additional step to estimate the variation in the response propensities, essentially an assessment of how much they deviate from a situation where all response propensities are equal (Schouten et al., 2009, p. 106). In this respect, a R-indicator is based on the standard deviation of the estimated response probabilities; the more variation there is in the response probabilities, the less representative is the sample (across categories of the covariates included). R-indicators are normalised to range between 0 and 1, where 1 represents strong representativeness, and 0 the 'maximum deviation from representativeness' (p. 104). As such, they do not provide information about the presence of bias in the target survey variables. However, they can be used to estimate the likely impact of unrepresentative response on a hypothetical survey variable under 'worst case scenarios' (p.107) – i.e. the largest *possible* difference between the responding sample and the total sample on an estimate of a population mean in a survey with a response rate of less than 100% (Schouten et al., 2009). Following these authors, we estimate this so-called Maximal Absolute Bias (MAB) based on the covariance between the values of the target variable and the response probabilities, and the variance of the target variable, using the sample data and the response propensities to set the upper bound of the estimate. We additionally estimate the Maximal Absolute Contrast (MAC) by dividing the maximal absolute bias by the nonresponse rate.

Coefficient of variation in subgroup response rates. While R-indicators provide an indication of the overall representativity of a responding sample, they provide no information about the over- or under-representation of specific subgroups (i.e. about how response propensities vary across categories of the covariates). As such, on their own they are unsuitable for targeting and prioritising cases in adaptive or responsive survey designs. To discover which subgroups have the lowest response propensities, Schouten et al. (2011) have proposed the estimation of 'partial R-indicators', which decompose the variation in response propensities, making it possible to evaluate the contribution of a (...) specified auxiliary variable to a lack of representative response' (p. 3), either singly, as in unconditional partial R-indicators, or when controlling for other variables, as in conditional partial R-indicators. In this paper, we use an alternative approach to get an initial idea of which subgroups are most affected, proposed by Wagner (2012), which involves estimating the coefficient of variation of subgroup response rates. The indicator is computed as $\frac{\sigma_x}{\acute{x}}$, where $\acute{x}$ is the overall response rate (the weighted mean of the subgroup response rates) and σ_x is the standard deviation of the subgroup response rates (Wagner, 2012, p. 563)'. In practice, this is equivalent to presenting the MAB for each auxiliary variable used to compute the R-indicator individually, or to estimating unconditional partial R-indicators, the difference being that the latter do not take ratios over the response rate. To identify which subgroups contribute most to reduced representativity, we compute the coefficient of variation for all the variables used to estimate the response propensities, at each stage of the fieldwork. We then complete the picture by looking in closer detail at the distribution of respondents across the same variables.

Evaluating the costs of different types of fieldwork effort. Assessing the impact of the response enhancement methods used on response rates, sample representativity, and the risk of nonresponse bias addresses just one side of the cost-error trade-off that is integral to a Total Survey Error approach to survey design. If such analyses are to facilitate decisions about how to target fieldwork efforts more efficiently in future survey waves, then they must also take into consideration the financial costs each procedure entails. We calculated the fixed costs associated with each stage of fieldwork, the unit cost of a completed interview, and the mean cost of the number of visits made to a sampled unit required at each stage. Taken together these provide an indication of the overall cost of the first five face-to-face visits (including all fixed set-up and management costs), and the additional variable costs each subsequent effort entails. We present these data in the form of a percentage of the overall survey budget. Note, however, that these are rough estimations based on estimates provided by the survey agency, and should not be taken as indicative of the costs of different types of survey fieldwork generally.

All analyses were carried out in SAS (version 9.2). R-indicators were estimated using the R-Cockpit tools in SAS developed and supplied by the EC-funded 7th Framework

the German-speaking interviewers and were so few in number in this sample (n=10) that we decided it was acceptable to combine them with the German-speaking cases.

[8] Note that we could have additionally included dummy variables representing NUTS regions in which sample members were located, but in order to reduce complexity, we retained the linguistic region variable to provide a proxy for region.

Programme (FP7) RISQ (Representativity Indicators for Survey Quality) project.

3 Results

Taken together, the response enhancement methods used on round 5 of the ESS in Switzerland were successful in increasing the response rate by nearly 10 percentage points, from 43.1% after the first 5 face-to-face visits, to 52.8%[9] after all extra contact attempts and refusal conversion interviews (see Table 2). Extra contact attempts by face-to-face and telephone succeeded in reducing the non-contact rate from 14.6% to 9.8%, while the refusal conversion interviews helped to reduce the refusal rate from 32.5% to 26.6%.[10] This means that the majority of respondents were recruited in the five first face-to-face visits. Telephone contact attempts added a total of just 85 interviews, or 2.9% to the response rate, while additional face-to-face visits, added only 66 interviews or an extra 2.3% to the overall response rate. By contrast, refusal conversion added 128 interviews (or 4.5% to the response rate). The nonresponse follow-up survey, which was sent to a total of 769 refusal cases and 278 non-contacts, increased the number of participants from 1,506 to 2,089, for a total completion rate of 73.3% (but with a considerably reduced-length questionnaire).

Alongside the increasing response rate, table 2 shows our other quality indicators, the evolution of the composition of the respondent group measured by the estimated R-indicator, the risk of nonresponse bias shown by the MAB indicator, and the maximal absolute contrast (MAC) indicator. The R-indicator following the first 5 face-to-face visits is 0.80, and reduces slightly (though not significantly) to 0.78 as a result of extra contact attempts and refusal conversion interviews. In other words, even though the response rate rises, there is no improvement in the R-indicator, suggesting that the additional fieldwork efforts do little to change the overall composition of the sample. MAB reduces with the addition of extra cases resulting from each type of fieldwork effort; from 0.23 to 0.22 for the extra face-to-face visits, 0.22 to 0.21 for the refusal conversion. By contrast, MAC first decreases slightly after extra face-to-face attempts from 0.41 to 0.40, but then increases from 0.40 to 0.42 after refusal conversion and from 0.42 to 0.44 after the telephone contacts, indicating a small increase in the worst-case scenario difference between the respondents and nonrespondents. The increase in response rate resulting from the nonresponse follow-up survey is also associated with a small improvement in the R-indicator from 0.78 to 0.81, a drop in MAB from 0.21 to 0.13, but an increase of the MAC from 0.44 to 0.49. Given that the R-indicator stays stable after each extra fieldwork effort (including the nonresponse survey), the decrease in MAB can mainly be attributed to the increased response rate. The increase in MAC indicates that the potential difference between respondents and nonrespondents increases with more efforts. The MAB, which is the product of the nonresponse rate and MAC, is, however, lower after all fieldwork efforts, showing that the increase in MAC is counter-balanced by the decrease in the nonresponse rate. Overall, the R-indicator and hence

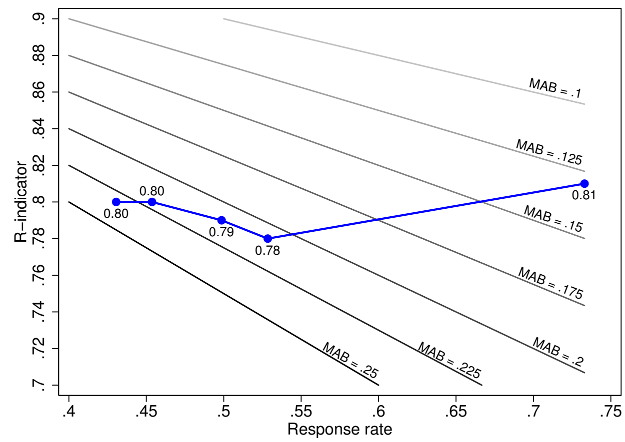

Figure 2. Comparison of the calculated R-indicator with R-indicators that correspond to a constant MAB

the representativity of the respondent sample does not change significantly. Even if an improvement in the R-indicator would have been desirable, these findings suggest that the extra fieldwork efforts manage to increase the response rate without only bringing 'low hanging fruit' into the respondent sample and increasing the bias.

Figure 2 illustrates these results. The purpose of the graph is to identify whether changes in the value of the R-indicator permit an overall decrease in MAB. The *x*-axis represents the response rates, and the *y*-axis the R-indicator, while the diagonal straight lines represent constant maximal absolute bias ranging from high to low, from the bottom to the top of the graph. With the addition of each response enhancement method, the R-indicator should ideally evolve from higher to lower MAB. In the present case, the R-indicator calculated for the different types of fieldwork effort evolves from close to the 0.25 MAB line after five face-to-face visits to between the 0.2 and 0.225 MAB lines after all the main survey fieldwork has been completed. Finally, the nonresponse follow-up survey brings the R-indicator close to the 0.125 MAB line, indicating a substantial decrease in the risk of nonresponse bias.

The interpretation of these results also depends on a consideration of the relative costs of the different methods used to increase response rates and reduce noncontacts following the first five face-to-face visits. In this context, additional face-to-face visits incur additional travel expenses, but are not associated with an increase in fixed costs. By contrast, setting up centralised telephone contact attempts involves

[9] Note that the percentages shown are not equivalent to the final ESS response rates which exclude ineligibles from the gross sample. We include ineligibles in the response rate calculation because the final number and rate of ineligibility is unknown until the end of the fieldwork. Moreover, after the main fieldwork only 1% of cases were coded as ineligible.

[10] Note that 17 cases recruited in the CATI phase were actually refusals that had not been re-contacted before the CATI phase.

Table 2 Quality and cost indicators following increased fieldwork effort

	Up to 5 face-to-face visits	Additional face-to-face visits	Telephone Refusal Conversion	Nonresponse contact attempts	Survey
Response rate (in %)[a]	43.1	45.4	49.9	52.8	73.3
Non-contact rate (in %)	14.6	13.1	13.1	9.8	5.9
Refusal rate (in %)[b]	31.6	32.5	26.6	27.6	12.7
R-indicator	0.80	0.80	0.79	0.78	0.81
Confidence interval	0.77-0.84	0.77-0.84	0.75-0.82	0.75-0.82	0.77-0.85
Max absolute bias	0.23	0.22	0.21	0.21	0.13
Max abs. contrast	0.41	0.40	0.42	0.44	0.49
Cost per effort type	0.73	0.02	0.16	0.09	0.07[c]
Cost per interview[d]	1.00	0.53	1.85	2.13	
Total Sample (N)	1227	1293	1421	1506	2089

[a]Response rate calculated here as total number of interviews divided by the sample size (i.e. it does not take account of ineligibles).
[b]Refusal rates include refusals by the target respondent and by other household members.
[c]Costs of the nonresponse follow-up are presented as a percentage of the total costs of the main survey.
[d]Costs per interview compared to the cost of an interview following the standard protocol (up to 5 face-to-face visits).

training and payment for CATI interviewers, programming a short CATI instrument, and the associated overheads of the call centre. Similarly, refusal conversions involve additional training for face-to-face interviewers, higher rates of payment for more experienced interviewers, and bonuses once a specified completion rate is achieved. Although telephone contacts have a positive impact on noncontact rates, helping to reduce them by 3.3% (compared to the 1.5% decrease offered by the extra face-to-face contacts), our cost analysis showed that the 85 interviews achieved following the telephone contact phase contribute 10% to the overall cost of the survey, compared to the 2% contributed by the 66 interviews achieved following extra face-to-face visits. This means that the per interview cost of an interview resulting from a centralised telephone contact is over double that of an interview obtained in the first phase of fieldwork, or through additional face-to-face contacts. Refusal conversions, though more expensive still in terms of their contribution to the overall cost of the survey, appear to be better value for money in terms of their success rate, as well as their impact on bias. These results raise doubts about the presumed benefits of using centralised telephone contact attempts in the Swiss ESS, particularly given that there is little evidence of a marked positive impact on bias (though they do help reduce the noncontact rate, bringing it closer to the ESS target).

To identify which groups contribute most to a lack of representative response, Table 3 shows the coefficients of variation in subgroup response rates for all the auxiliary variables that were used to calculate the R-indicator. After five face-to-face visits, the coefficients range between 0.04 for the variable 'marital status' and 0.13 for the variable 'nationality', meaning that the standard deviations of the subgroup response rates range between 4% of the mean response rate for marital status, and 13% of the mean for nationality (note that the contribution to the variance of each covariate category has been calculated proportional to the number of people in the category). The variables with the lowest variance between the subgroups are marital status (0.04) and gender (0.05), while those with the highest variance are linguistic region (0.08), availability of a telephone number (0.10), urbanisation (0.11), age group (0.11), and nationality (0.13). The variance reduces following each additional fieldwork effort for all variables with the exception of the telephone number indicator, which after refusal conversion and the extra telephone contact increases by 0.01; linguistic region, which after extra face-to-face interviews and refusal conversion increases by 0.01 and 0.02 respectively; and finally, marital status, which after refusal conversion increases by 0.01.

Additional face-to face visits had little effect on the coefficient of variation, though they helped reduce it for gender, marital status, urbanisation and the availability of a telephone number by one percentage point. The refusal conversion helped to reduce the discrepancy in response rates between the sexes from 0.04 to 0.03; between age categories from 0.11 to 0.08; and between Swiss sample members and those of other nationalities from 0.13 to 0.12. Telephone contact attempts helped reduce it for gender, marital status, linguistic region, and urbanisation by one or two percentage points. Most of these changes are not, however, statistically significant. Finally, the nonresponse follow-up survey was the most efficient method of reducing the variation in subgroup response rates, especially for linguistic region (which reduces from 0.09 to 0.02); for the telephone number indicator (from 0.11 to 0. 06); for urbanisation (from 0.09 to 0.05); for age (from 0.08 to 0.04); and finally, for nationality (from 0.12 to 0.10).

Even though we have been able to identify the most problematic variables using the coefficient of variation in subgroup response rates, we still miss more detailed information about which sub-categories of each variable have the lowest response propensities, and which might benefit most from a more targeted fieldwork strategy. Table 4 shows the charac-

Table 3 Coefficients of variation of the subgroup response rate (and standard errors) for all auxiliary variables used to calculate the R-indicator

	Up to 5 face-to-face visits		Additional face-to-face visits		Refusal conversion interviews		Telephone contact attempts		Nonresponse follow-up survey	
	b	S.E.	b	S.E.	b	S.E.	b	S.E.	b	S.E.
Response rate (in %)[a]	43.1		45.4		49.9		52.8		73.3	
Gender	0.05	0.03	0.04	0.02	0.03	0.02	0.02	0.02	0.01	0.01
Age	0.11	0.04	0.11	0.04	0.08	0.04	0.08	0.03	0.04	0.03
Marital Status	0.04	0.02	0.03	0.02	0.04	0.02	0.03	0.02	0.02	0.01
Nationality	0.13	0.02	0.13	0.02	0.12	0.02	0.12	0.02	0.10	0.01
Linguistic region	0.08	0.02	0.09	0.02	0.11	0.02	0.09	0.02	0.02	0.01
Urbanisation	0.11	0.03	0.10	0.02	0.09	0.02	0.09	0.02	0.05	0.01
Telephone	0.10	0.02	0.09	0.02	0.10	0.02	0.11	0.02	0.06	0.01

[a]Response rates are calculated here as total number of interviews divided by the sample size (i.e. it does not take account of ineligibles). Variances are weighted by the number of cases in each category of the covariate.

teristics of respondents and nonrespondents following each type of fieldwork effort[11], and compares the frequency distributions across the register variables at each stage with those following the first five face-to-face visits.

Compared with the earlier respondents, respondents recruited as a result of additional face-to-face visits were significantly less likely to be married and more likely to be divorced. They were also more likely to be aged between 30 and 44, more likely to live in a city or town centre, and less likely to have a telephone number (differences approaching significance). Respondents recruited as a result of refusal conversion methods were significantly less likely to be male, to be aged 15-29, and more likely to be aged 65 and over. They were also significantly more likely to be from the German-speaking region. Refusal conversion respondents were also more likely to have telephone numbers. Finally, the sample responding as a result of the telephone contacts included significantly fewer people below the age of 30, more people in the French-speaking region of Switzerland, and fewer respondents in the German-speaking regions. This group was also more likely to be divorced compared with the earlier respondents, and as already mentioned, to have a telephone number (differences approaching significance).

Looking at columns 5 and 7 shows the final bias on sociodemographic variables. People aged 30-44 are underrepresented. Swiss nationals are overrepresented and foreigners underrepresented, particularly if they are not from a bordering country. The German linguistic region is overrepresented while the French and Italian regions are underrepresented. A lower proportion of people living in city/town centres participated, and this is matched by a higher rate of participation among people living in rural areas. Finally, people with a registered phone number are overrepresented in the final sample.

4 Discussion and Conclusion

In light of recent research into the relation between survey nonresponse rates and nonresponse bias (Groves & Peytcheva, 2008), the appropriateness of blindly pursuing higher response rates as a fieldwork strategy is being increasingly brought into doubt (Beullens & Loosveldt, 2012; Wagner, 2012). We addressed this issue in the present study by exploiting new auxiliary data available for the Swiss European Social Survey to evaluate the costs and benefits of different response enhancement methods used to pursue the survey's target response rate of 70% and non-contact rate of 3% (European Social Survey, 2009). Given that a substantial portion of the Swiss fieldwork budget (estimated as 27% of the total) is invested in efforts to reduce nonresponse (including extra contact attempts, refusal conversion interviews, and a nonresponse follow-up survey), it makes sense to assess how this investment impacts on survey quality, using indicators other than the specified targets. We focused here on the representativeness of the responding sample at each stage of the survey fieldwork, and the risk of nonresponse bias, as measured by an R-indicator and its associated estimates of Maximal Absolute Bias (MAB) and Maximal Absolute Contrast (MAC) (Schouten et al., 2009, 2010). To gain a clearer picture of which sample subgroups contribute most to a lack of representative response, we additionally examined the coefficient of variation between subgroup response rates. The objective was to evaluate whether different types of fieldwork approach succeed in varying the respondent pool by bringing in sample members with lower response propensities, and if not, which subgroups might benefit from more targeted efforts to improve their representation in the survey in future rounds.

Overall, the findings were encouraging, suggesting that investments in response enhancement methods generally pay off in the Swiss context. The fieldwork enhancement techniques contributed overall to a 9.7% increase in the num-

[11] As noted, the number of respondents participating as a result of extra contact attempts both by telephone and in person was low, making it difficult to carry out a meaningful analysis of the differences between the samples. Nevertheless, some observations from Table 4 are informative.

Table 4 Socio-demographic characteristics of respondents and nonrespondents following extra contact attempts, refusal conversion and NRS

	(1) Up to 5 face-to-face visits		(2) Additional face-to-face visits		(3) Refusal conversion interviews		(4) Telephone contact attempts		(5) All standard interviews		(6) Nonresponse follow-up survey		(7) Total Sample	
	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.
Male	52.7	1.4	47.0	6.1	42.2	4.4*	47.1	5.4	51.3	1.3	46.1	2.1*	50.3	0.9
Age														
<30	24.7	1.2	24.2	5.3	14.1	3.1**	15.3	3.9*	23.2	1.1	18.7	1.6*	21.5	0.8
30-44	22.1	1.2	31.8	5.7	23.4	3.7	29.4	4.9	23.0	1.1	28.3	1.9*	25.6	0.8
45-64	35.4	1.4	33.3	5.8	33.6	4.2	41.2	5.3	35.5	1.2	33.6	2.0	33.3	0.9
65+	17.8	1.1	10.6	3.8	28.9	4.0**	14.1	3.8	18.3	1.0	19.4	1.6	19.5	0.7
Marital Status														
Single	29.9	1.3	37.9	6.0	24.2	3.8	29.4	4.9	29.7	1.2	29.7	1.9	30.6	0.9
Married	56.6	1.4	42.4	6.1*	60.2	4.3	54.1	5.4	56.2	1.3	54	2.1	54.3	0.9
Widowed	4.7	0.6	3.0	2.1	7.0	2.3	2.4	1.6	4.7	0.5	5.7	1.0	5.6	0.4
Divorced	8.7	0.8	16.7	4.6*	8.6	2.5	14.1	3.8	9.4	0.8	10.3	1.3	9.4	0.5
Legal partnership	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.2*	0.1	0.1
Nationality														
Swiss	85.1	1.0	78.8	5.0	82.0	3.4	84.7	3.9	84.5	0.9	80.4	1.6*	79.0	0.8***
Bordering country	6.5	0.7	10.6	3.8	7.8	2.4	9.4	3.2	7.0	0.7	8.9	1.2	8.8	0.5*
Other	8.4	0.8	10.6	3.8	10.2	2.7	5.9	2.6	8.5	0.7	10.6	1.3	12.1	0.6
Linguistic Region														
German	74.6	1.2	80.3	4.9	87.5	2.9**	61.2	5.3**	75.2	1.1	63.8	2.0***	71.0	0.8***
French	21.1	1.2	15.2	4.4	11.7	2.8*	34.1	5.1**	20.8	1.0	29.5	1.9***	24.0	0.8*
Italian	3.5	0.5	4.5	2.6	0.8	0.8	3.5	2.0	3.3	0.5	6.2	1.0*	4.4	0.4
Romansh	0.8	0.3	0.0	0.0	0.0	0.0	1.2	1.2	0.7	0.2	0.5	0.3	0.6	0.1
Urbanisation														
City/town centre	25.6	1.2	36.4	5.9	27.3	3.9	25.9	4.8	26.2	1.1	29.3	1.9	29.4	0.9*
City/town suburbs	43.8	1.4	40.9	6.1	44.5	4.4	44.7	5.4	43.8	1.3	46.1	2.1	44.4	0.9
Isolated town	0.6	0.2	0.0	0.0	2.3	1.3*	0.0	0.0	0.7	0.2	0.7	0.3	0.6	0.1
Rural community	30.0	1.3	22.7	5.2	25.8	3.9	29.4	4.9	29.3	1.2	23.8	1.8*	25.5	0.8**
Telephone number	66.1	1.4	54.5	6.1	75.0	3.8*	75.3	4.7	66.9	1.2	57.5	2.0***	61.2	0.9***
Observations	1227		66		128		85		1506		583		2850	

***p <.001,
**p<.01,
*p <.05, p<.10. Chi-square tests comparing columns (2) to (4) with column (1) and columns (6) and (7) with column (5).

ber of completed interviews (and a reduction in the noncontact rate), and enabled the survey to exceed response rates achieved in all earlier rounds. Unlike in Schouten and Cobben's (2007) study, and the simulation described by Beullens and Loosveldt (2012), there was no significant decrease in the R-indicator (or corresponding increase in the MAB) as response rates increased, suggesting that the additional efforts were reasonably successful at maintaining a varied respondent pool as well as at reducing the risk of nonresponse bias. However, this was brought into doubt by the increase observed in the MAC, which suggests that the worse case scenario difference between respondents and non-respondents is aggravated by additional fieldwork efforts (i.e. that late respondents share more characteristics with early respondents than they do with the final non-respondents).

The effect of extra contact attempts (whether by telephone or face-to-face) and of refusal conversions was comparable (though, of course, each made a unique contribution to the noncontact and refusal rates). Only by looking more closely at the coefficients of variation in subgroup response rates and the composition of the samples responding at each stage do we see small but important differences in the effectiveness of different types of fieldwork effort at recruiting different types of respondent.

Each step was successful in reducing the coefficient of variation, except on two variables: whether or not a telephone number was available, and linguistic region. Indeed, the difference in response rates between people with a registered phone number and those without increased after the refusal conversion, as well as after the telephone contacts. It is not clear, however, whether this is because people with a registered phone number have a higher propensity to respond overall, or whether they are more likely to be targeted for refusal conversion because they are easier to re-contact. The higher variation in response rates between the linguistic regions appeared after the refusal conversions; again, it is not clear whether this is due to genuine variation in the willingness to participate in surveys across the linguistic regions, or simply the result of different strategies used by the interviewers in the field and decisions made by fieldwork supervisors about how to re-allocate cases. Given the potential for bias introduced by differential nonresponse across these subgroups, adaptations to fieldwork planning in future waves should address how cases are selected for refusal conversion and how they are allocated to the available interviewers.

Additional face-to-face visits helped reduce the underrepresentation of several groups that are known to be 'hard-to-reach' in surveys (e.g. divorced people and people living in city or town centres, people aged between 30 and 44), as well as reduce the over-representation of married people, and of those with a registered telephone number. The refusal conversion interviews succeeded in eliminating bias from the overrepresentation of men, of people aged younger than 30, and the underrepresentation of people older than 65. By comparison, the telephone contacts reduced the overrepresentation of the younger than 30 and the German-speaking region whilst increasing the number responding from the French-speaking region, but this had little impact on the bias between these groups observed in the final responding sample.

The advantages of using different types of response enhancement method must be evaluated in relation to the additional costs each one entails over and above the standard fieldwork procedures. In this context, the extra face-to-face contact attempts seem to be the most cost-effective method, helping to increase the response rate and to reduce the underrepresentation of groups that are known to be 'hard-to-reach'. By rendering such additional visits compulsory for the interviewers, the reduction of bias could be even stronger, though the costs would, of course, be higher[12]. Although understandably more expensive than the extra visit attempts, the refusal conversions prove to be relatively good value for money in terms of their impact on the response rate, as well as their possible impact on nonresponse bias. By contrast, the telephone contact attempts appear to be relatively inefficient cost-wise, especially given their low impact on bias. Taking all the methods used together, the overall gain in response rate of 9.7% seems to justify the cost of the extra fieldwork, particularly considering that there is no evidence of a *negative* impact on nonresponse bias. However, though theoretically, a mix of different types of fieldwork effort seems to make sense as a way of reducing bias (Peytchev et al., 2009), the present findings would suggest that the appropriateness of telephone contact attempts in addition to face-to-face visits be reconsidered in future rounds of the Swiss ESS, given the high costs involved and their relatively low impact on response rates and bias.

It is also of interest to consider the impact of the nonresponse follow-up survey, which represents an important change in the type of fieldwork effort (notably, a mode switch from CAPI to self-administered paper, a substantial reduction in the length of the questionnaire, and the offer of an additional incentive). Setting aside the limitations of the method (including the potential for confounded mode, timing, and context effects), it is noteworthy that the follow-up succeeded in obtaining data from a further 20% of the sample, significantly increased the R-indicator, and decreased the maximal absolute bias from 0.21 to 0.13. Furthermore, it had a bigger effect on the coefficient of variation than any of the main survey fieldwork effort types, almost eliminating or substantially reducing bias on many of the variables analysed (despite leading to an increase in the MAC indicator). Though the nonresponse follow-up survey cannot be seen as part of the main survey, it has the potential to provide important information about the characteristics of nonrespondents and the presence of bias on key survey variables, and can serve as a tool to correct for nonresponse bias, at relatively minimal additional cost compared to other methods. The results of the nonresponse follow-up in this study highlight the benefits of varying contact attempts in the pursuit of nonrespondents, and the potential value of further research into how to improve the design of follow-up studies in order to maximise their utility for survey methodologists and data users.

After all fieldwork efforts for the main survey, bias remained on four of the variables analysed. Nationality was the variable with the highest coefficient of variation, with Swiss-citizens overrepresented compared to non-Swiss from bordering countries and to a higher extent, the non-Swiss from other countries. Possible explanations for this imbalance are the language barrier, as well as perhaps a lack of interest or sense of obligation. These potential causes of non-participation among non-Swiss residents could be taken into account in fieldwork planning, perhaps through targeted announcements that highlight the importance of the survey for these specific groups, or by providing translations of the questionnaire in minority languages (see Laganà, Elcheroth, Penic, Kleiner, & Fasel, 2011) for a more detailed consideration of these issues). The German linguistic region is overrepresented whilst the Italian and French regions are underrepresented. In this case, it is hard to disentangle possible interviewer effects from genuine differences in the propensity to answer between regions. But as the underrepresentation of linguistic regions can vary from one round to another (according to technical reports supplied by the fieldwork agency), the hypothesis of interviewer effects is stronger. There remains an imbalance between the city/town centre and the rural areas, most likely due to variation in contactability, people in urban areas being well known to be harder to contact than people in rural areas (Stoop et al., 2010, p. 120). Both regional variations in response propensity suggest potential benefits could be gained from targeted fieldwork efforts. Lastly, the final sample over-represented people with a registered telephone number. Other studies have found that people with registered telephone numbers more readily participate in surveys (e.g. Cobben & Schouten, 2007), but this may be emphasized in the present study by the decision to contact people by phone when possible. We plan

[12] It indeed seems likely that interviewers initiate a larger number of additional contact attempts with easier-to-contact cases (the so-called 'low hanging fruit'), or with cases requiring minimal additional effort for them.

to investigate this problem in more detail in an extension to the research presented here.

This study provides an illustration of how existing fieldwork practices can be evaluated retrospectively with a view to developing more targeted strategies in the future in an adaptive or responsive design. We used a combination of approaches to build up a detailed picture of nonresponse and the risk of bias in the Swiss ESS2010, based on socio-demographic variables available from a previously unavailable register-based sampling frame. One limitation of this is that we do not address the problem of whether under- or over-representation of population subgroups results in actual bias in key survey estimates other than socio-demographic variables (except through our assessment of MAB and MAC). Neither do we investigate the effect of the increasing and varying fieldwork efforts on measurement error, which in future research, would provide a more thorough picture of the relation between response enhancement methods and total survey error.

R-indicators are 'motivated by the potential for systematic differences on auxiliary variables between respondents and nonrespondents to be predictive of nonresponse bias' (Schouten et al., 2011, p. 232). To this extent, their effectiveness depends on the strength of the relation between the selected auxiliary variables and the key survey variables likely to be most vulnerable to bias. This poses two potential challenges. For one, as Peytcheva and Groves (2009) have shown, it is quite common for socio-demographic variables to be only weakly correlated to the variables most of interest to data users, meaning they may not be the most suitable candidates for building an R-indicator. For another, in a general social survey like the ESS, users are interested in a wide variety of variables, making it difficult to develop a 'one-size-fits-all' bias indicator.

This highlights one of the difficulties of working with R-indicators. When selecting variables for the response propensity model, we decided to restrict our analysis to auxiliary data from the sampling register (paradata in the form of interviewer observations about neighbourhood characteristics are also available on ESS), and to a survey-specific variable that seemed particularly likely to influence response propensities, opting to retain all the available covariates in the model rather than selecting the subset with the strongest relation with the probability of responding. This is a defensible strategy (e.g. Lee & Vaillant, 2008, p. 178), and makes sense given that the main appeal of R-indicators is that they can serve as a basis for comparisons across surveys; using only variables readily available on the sampling frame would greatly facilitate such comparisons in the Swiss context. However, the choice of covariates will depend on the motivation for evaluating representativity (e.g. for comparisons across countries participating in the ESS a different choice of covariates would likely be more appropriate than those used in within-country cross-wave comparisons). The danger of an inclusive approach to model selection, such as was used here, is that it runs the risk of reducing the overall ability of the R-indicator to highlight where the impact of non-representative response is likely to be most damaging to estimates of interest. One way to assess this risk is by drawing on data from the nonresponse follow-up survey to see how well correlations between response propensity scores and the key survey variables predict actual bias observed in the follow-up. This analysis forms the focus of a separate paper based on the ESS data analysed here (Vandenplas, Roberts, & Ernst Stähli, 2013).

Notwithstanding some of the potential pitfalls of using R-indicators, in combination with information about variation in response propensities across subgroups, they provide a valuable complementary indicator to the response rate for data quality, and rightly divert attention from potentially counterproductive fieldwork targets. We used the coefficient of variation in subgroup response rates, which has the advantage of being relatively simple to compute and intuitive, but the disadvantage that it is not possible to identify which categories of a given covariate contribute the most to a lack of representativity, nor whether the covariates are strongly collinear. For this reason, we also studied the distribution of respondents across covariate categories following each type of fieldwork effort. Though these strategies combined are equivalent to using unconditional partial R-indicators, future analyses might additionally benefit from using conditional partial R-indicators (Schouten et al., 2011), which provide information about the contribution of different subgroups to a lack of representativity, while controlling for the full vector of variables included in the R-indicator. Whatever the choice of indicator used, the present study illustrates the potential advantages of using alternatives to response rates to facilitate decisions about how to focus fieldwork efforts in a way that maximises the possibility of reducing the risk of bias.

Acknowledgements

The Swiss European Social Survey is financed by the Swiss Science Foundation. We are grateful to the Swiss Federal Statistical Office for supplying a sample from their frame based on population registers maintained by municipalities. We would particularly like to thank the reviewers for their helpful comments and suggestions for improving the paper.

References

American Association for Public Opinion Research. (2011). *Standard definitions: Final dispositions of case codes and outcome rates for surveys.* (7th ed.). Retrieved 20/5/2014, from http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156

Bethlehem, J. (2002). Weighting nonresponse adjustments based on auxiliary information. In R. Groves, D. Dillman, J. Eltinge, & R. Little (Eds.), *Survey nonresponse* (p. 275-288). New York: Wiley.

Beullens, K., & Loosveldt, G. (2012). Should high response rates really be a primary objective? *Survey Practice*, *5*(3), 1–5.

Biemer, P., & Lyberg, L. (2003). *Introduction to survey quality*. Hoboken: Wiley.

Brick, J., & Williams, D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. *The Annals of the American Academy of Political and Social Science*, *645*, 36-59.

Cobben, F., & Schouten, B. (2007). *A follow-up with basic questions of nonrespondents to the dutch labour force survey.* (Statistics Netherlands Discussion Paper 07011, CBC, Voorburg)

Couper, M. (1998). *Measuring survey quality in a casic environment.* (Pro-ceedings of the Section on Survey Research Methods of the American Statistical Association, p. 41–49)

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly*, *64*(4), 413-428.

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

European Social Survey. (2009). *Round 5 specification for participating countries*. London: Centre for Comparative Social Surveys, City University London.

European Social Survey Data Archive. (2012). *ESS5 – 2010 documentation report* (3.0 ed.). Bergen, Norway: NSD.

Groves, R. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*, 646–675.

Groves, R., Brick, J., Couper, M., Kalsbeek, W., Harris-Kojetin, B., Kreuter, F., … Wagner, J. (2008). Issues facing the field: Alternative practical measures of representativeness of survey respondent pools. *Survey Practice*, *1*(3), 1–6. Retrieved 20/05/2014, from http://surveypractice.org/2008/10/30/issues-facing-the-field/

Groves, R., Couper, M., Presser, S., Singer, E., Tourangeau, R., Piani Acosta, G., & Nelson, L. (2006). Experiments in producing nonresponse bias. *Public Opinion Quarterly*, *70*, 720–36.

Groves, R., & Heeringa, S. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society, Series A*, *169*, 439–457.

Groves, R., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. *Public Opinion Quarterly*, *72*, 1–23.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, *64*, 125–148.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T., Casas-Cordero, C., … Raghunathan, T. (2010). Using proxy measures and other correlates of survey outcomes to adjust for nonresponse: Examples from multiple surveys. *Journal of the Royal Statistical Society, Series A*, *173*, 389–407.

Laganà, F., Elcheroth, G., Penic, S., Kleiner, B., & Fasel, N. (2011). *National minorities and their representation in Swiss surveys (II): What practices make a difference?* (FORS Working Paper Series, 2011-03)

Lee, S., & Vaillant, R. (2008). Weighting telephone samples using propensity scores. In J. Lepkowski et al. (Eds.), *Advances in telephone survey methodology.* New York: John Wiley & Sons.

Lin, I.-F., & Schaeffer, N. (1995). Using survey participants to estimate the impact of nonparticipation. *Public Opinion Quarterly*, *59*, 236–258.

Peytchev, A., Baxter, R. K., & Carley-Baxter, L. R. (2009). Not all survey effort is equal: Reduction of nonresponse bias and nonresponse error. *Public Opinion Quarterly*, *73*(4), 785-806.

Peytchev, A., Riley, S., Rosen, J., Murphy, J., & Lindblad, M. (2010). Reduction of nonresponse bias in surveys through case prioritization. *Survey Research Methods*, *4*, 21–29.

Peytcheva, E., & Groves, R. (2009). Using variation in response rates of demographic subgroups as evidence of nonresponse bias in survey estimates. *Journal of Official Statistics*, *25*(2), 193–201.

Rao, R., Glickman, M., & Glynn, R. (2008). Stopping rules for surveys with multiple waves of nonrespondent follow-up. *Statistics in Medicine*, *27*(12), 2196-2213.

Schouten, B., & Cobben, F. (2007). *R-indexes for the comparison of different fieldwork strategies and data collection modes.* (Statistics Netherlands Discussion Paper, 07002, CBC, Voorburg)

Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, *35*(1), 101-113.

Schouten, B., Shlomo, N., & Skinner, C. (2010). *Indicators for representative response.* (Paper presented at Q2010, Helinski, Finland)

Schouten, B., Shlomo, N., & Skinner, C. (2011). Indicators for monitoring and improving representativeness of response. *Journal of Official Statistics*, *27*(2), 231–253.

Skalland, B. (2011). An alternative to the response rate for measuring a surveys realization of the target population. *Public Opinion Quarterly*, *75*(1), 89–98.

Stoop, I., Billiet, J., Koch, A., & Fitzgerald, R. (2010). *Improving survey response: lessons learned from the European Social Survey.* New York: John Wiley & Sons.

Vandenplas, C., Roberts, C., & Ernst Stähli, M. (2013). *Use of a non-response follow-up survey to assess the validity of R-indicators as a measure of the risk of bias.* (Paper presented at the 5th conference of the European Survey Research Association, Ljubljana, Slovenia, July 15-19, 2013)

Wagner, J. (2008). *Adaptive survey design to reduce nonresponse bias.* Retrieved 25/6/2014, from http://deepblue.lib.umich.edu/bitstream/handle/2027.42/60831/jameswag_1.pdf?sequence=1 (PhD Thesis, University of Michigan, U.S.A.)

Wagner, J. (2012). A comparison of alternative indicators for the risk of nonresponse bias. *Public Opinion Quarterly*, *76*(3), 555-575.

Wagner, J., & Raghunathan, T. (2010). A new stopping rule for surveys. *Statistics in Medicine*, *29*(9), 1014-1024.