

# A Validation of R-Indicators as a Measure of the Risk of Bias using Data from a Nonresponse Follow-Up Survey

*Caroline Roberts<sup>1</sup>, Caroline Vandenplas<sup>2</sup>, and Jessica M.E. Herzing<sup>1</sup>*

*R*-indicators are increasingly used as nonresponse bias indicators. However, their effectiveness depends on the auxiliary data used in their estimation. Because of this, it is not always clear for practitioners what the magnitude of the R-indicator implies for bias in other survey variables, or how adjustment on auxiliary variables will affect it. In this article, we investigate these potential limitations of R-indicators in a case study using data from the Swiss European Social Survey (ESS5), which included a nonresponse follow-up (NRFU) survey. First, we analyse correlations between estimated response propensities based on auxiliary data from the register-based sampling frame, and responses to survey questions also included in the NRFU. We then examine how these relate to bias detected by the NRFU, before and after adjustment, and to predictions of the risk of bias provided by the R-indicator. While the results lend support for the utility of R-indicators as summary statistics of bias risk, they suggest a need for caution in their interpretation. Even where auxiliary variables are correlated with target variables, more bias in the former (resulting in a larger R-indicator) does not automatically imply more bias in the latter, nor does adjustment on the former necessarily reduce bias in the latter.

*Key words:* Nonresponse; R-indicator; propensity score weighting; nonresponse survey.

## 1. Introduction

High response rates have traditionally been regarded as a guarantee of survey data quality. Over the past two decades, however, obtaining high response rates in social surveys has become increasingly challenging (De Leeuw and De Heer 2002; Brick and Williams 2013; Kreuter 2013; Williams and Brick 2017; Beullens et al. 2018), and questions have been raised regarding the extent to which they can protect survey estimates from nonresponse bias (Groves 2006; Groves and Peytcheva 2008, Brick and Tourangeau 2017). Indeed, bias depends not only on the rate of nonresponse, but also on the difference in characteristics between respondents and nonrespondents (Groves and Couper 1998), and according to the stochastic view of nonresponse, the covariance between variables influencing the probability of responding and a given survey variable

<sup>1</sup> Institute of Social Sciences, University of Lausanne, Bâtiment Géopolis, Quartier Mouline, CH-1015 Lausanne, Switzerland. Emails: [caroline.roberts@unil.ch](mailto:caroline.roberts@unil.ch), and [Jessica.herzing@unil.ch](mailto:Jessica.herzing@unil.ch)

<sup>2</sup> Consulting, Chemin du Cyclotron 6, 1348 Ottignies-Louvain-la-Neuve, Belgium. Email: [caroline.vandenplas@b12-consulting.com](mailto:caroline.vandenplas@b12-consulting.com)

**Acknowledgments:** We are grateful to Michèle Ernst Stähli at FORS, the Swiss Centre of Expertise in the Social Sciences, for providing access to the sampling data analysed in this study. We would also like to thank the anonymous reviewers for their constructive feedback on earlier drafts of our manuscript, which enabled us to make a number of important improvements.

(Bethlehem 2002; Little and Rubin 2014; Brick and Tourangeau 2017). In other words, even in a survey with a high response rate, a variable that correlates highly with the probability of responding may have a larger nonresponse bias than a variable that does not, or a variable that only weakly correlates with the probability of responding in a survey with a lower response rate. Because nonresponse bias is, thus, variable-dependent, finding simple and intuitive methods for detecting its presence and assessing its impact poses an on-going challenge for survey methodologists and statisticians (Groves et al. 2008; Schouten 2018).

In response to this challenge, the last decade has seen the development of a number of new indicators for assessing the risk of nonresponse bias (Wagner 2012; Nishimura, et al. 2016). Of these, one that has rapidly gained popularity is the ‘*Representativity Indicator*’, or *R-indicator* (Schouten et al. 2009), together with its related ‘*partial R-indicators*’ (Schouten et al. 2011; Beullens and Loosveldt 2012). R-indicators offer an intuitive way of summarising the extent to which the respondents in a probability-based sample survey represent all the sample units that were selected from the sampling frame, and the risk of nonresponse bias in survey variables. Because of this, R-indicators have quickly attracted interest as a way to evaluate fieldwork outcomes and compare the effectiveness of different survey designs (e.g. Schouten et al. 2012; Luiten and Schouten 2013; Moore et al. 2018; Schouten and Shlomo 2017). Meanwhile, partial R-indicators are being used to plan adaptive survey designs or identify specific subgroups during fieldwork monitoring for targeted interventions, as in responsive designs (Groves and Heeringa 2006; Schouten et al. 2011a; Beullens and Loosveldt 2012; Schouten et al. 2016).

The utility of R-indicators depends in part, however, on the availability of suitable auxiliary data for their estimation (i.e. variables that correlate both with the probability of responding and key survey variables), which (in cross-sectional surveys at least) is often limited (e.g. Sakshaug and Antoni 2018). Given that R-indicators essentially summarise nonresponse bias in the auxiliary variables, the question is raised as to how effective they (and the auxiliary variables) are at identifying the risk of nonresponse bias on other survey variables – especially in the context of large-scale surveys covering a wide variety of topics. Relatedly, given that the same auxiliary data can also be used to adjust for nonresponse bias, a further question is raised as to what can be learned from the R-indicator about bias in other survey variables *after* adjustment on the auxiliary variables.

In this article, we investigate these potential limits of R-indicators in a case study using data from the Swiss European Social Survey (ESS), which in Round 5 (2010), included a nonresponse follow-up (NRFU) survey. We first evaluate the suitability of available auxiliary data (from the sampling frame based on population registers) for estimating R-indicators by examining how well they correlate with a selection of target variables. Then we assess how well the R-indicator predicts the presence of actual nonresponse bias on these variables, before and after adjustment, using data from the NRFU to estimate the difference between respondents and nonrespondents in the main survey. Before describing our research questions and analytic approach in more detail, we present an overview of R-indicators, their possible limitations, and the role of auxiliary variables, then review recent studies that have investigated their performance and interpretation.

## 2. Background

### 2.1. Using R-Indicators to Detect Nonresponse Bias – The Role of Auxiliary Variables

R-indicators describe the variance of the sample members' probability of responding to a given survey (for detailed accounts, including the statistical notation and formulae for estimating R-indicators, see [Schouten and Cobben 2007](#); [Schouten et al. 2009](#); [Schouten et al. 2011a](#); [Schouten and Shlomo 2017](#); [De Heij et al. 2015](#)). The higher the variance in the response probabilities, the more likely it is to have an unbalanced respondent sample (i.e., reduced 'representativity') and, theoretically, to have nonresponse bias on other variables that correlate with the response probability. R-indicators are normalised to range between zero and one, where one represents strong representativity, and zero the 'maximum deviation from representativity' ([Schouten et al. 2009](#), 104). As such, they provide an intuitive summary statistic for describing survey quality.

As we do not know the actual probability of responding of all members of the survey sample, in practice, the R-indicator is based on the standard deviation of the *estimated* response probabilities ([Schouten and Cobben 2007](#); [Schouten et al. 2009](#); [Schouten et al. 2011](#)), calculated based on auxiliary variables available both for respondents and nonrespondents (e.g. frame variables, linked contextual or administrative data, survey paradata – [Cornesse and Bosnjak 2018](#)), typically using a logistic regression model. The higher the standard deviation of the estimated response probabilities, the less representative is the sample across categories of the auxiliary variables. Thus, the notion of 'representativity' relates specifically to the extent to which the response sample represents the complete sample on the auxiliary variables included as covariates in the model ([Cornesse and Bosnjak 2018](#)).

The utility of R-indicators lies partly in their ability to translate nonresponse impact on the auxiliary variables used in the estimation to just one value. Partial R-indicators, which can be estimated at the variable or the category level (for details see [Schouten et al. 2011](#), 5–6), permit a more fine-grained investigation into which variables or subcategories of the auxiliary variables contribute most to a lack of representativeness ([Schouten et al. 2011](#); [Beullens and Loosveldt 2012](#)). This makes them useful for fieldwork management for example, as a basis for decisions to direct additional fieldwork effort to under-represented groups with the aim of achieving a balanced sample ([Schouten et al. 2016](#)). The intended use of the R-indicator may imply different considerations about which auxiliary variables are most suitable for their estimation (assuming such data are available to begin with). If the aim is to compare designs or monitor the evolution of fieldwork, then the indicators should ideally be estimated using the *same* auxiliary variables for each comparison, to allow an evaluation of the relative quality of responding samples. If intended to be interpreted in an absolute sense (i.e., for a single survey design at a single point in time), ideally as broad a range of variables as possible should be used to ensure the definition of representativeness is not too restricted ([Cornesse and Bosnjak 2018](#), 5). Irrespective of the intended use of R-indicators, the choice of auxiliary variables used in the estimation of response probabilities is key to their interpretation ([Nishimura et al. 2016](#)).

It is noteworthy that a number of other indirect nonresponse bias indicators have been proposed for similar purposes as the R-indicator (see [Wagner 2012](#), and more recently,

Nishimura et al. 2016, for reviews). These include the closely-related coefficient of variation of the response propensities (CV), which is more optimally suited to assessing the risk of bias in population means and totals (Schouten et al. 2009; Schouten and Shlomo 2017; Schouten et al. 2016; Schouten 2018); the coefficient of variation of subgroup response rates (Groves 2006; Wagner 2012); the coefficient of variation of nonresponse adjustments (Särndal and Lundström 2010); and the area under the curve or pseudo- $R^2$  (Nagelkerke 1991). Like the R-indicator, the interpretation – and utility – of such indicators similarly depends on which auxiliary variables are used in their estimation (Nishimura et al. 2016; Cornesse and Bosnjak 2018), and so the questions raised and addressed in the case study presented here have a broader relevance beyond R-indicators.

Besides their capacity to summarise representativeness with respect to the auxiliary variables, the utility of R-indicators (and other nonresponse bias indicators) also lies in their ability to detect bias in other survey variables. R-indicators provide an estimate of the upper bound of the nonresponse bias of a *hypothetical* survey variable under ‘worst case scenarios’ (Schouten et al. 2009, 107) – referred to as the Maximal Absolute Bias (MAB), which is equivalent to the coefficient of variation (CV) of the response propensities (Schouten et al. 2009; Beullens and Loosveldt 2012). As such, the magnitude of the R-indicator and MAB should be informative about the extent of actual bias in other survey variables. However, they cannot identify which survey variables are affected or by how much (Nishimura et al. 2016), nor whether bias will remain after adjustment on the auxiliary variables (Groves et al. 2008; Brick and Jones 2008; Kreuter and Olson 2011; Sakshaug and Antoni 2018). To be optimally informative, the choice of auxiliary variables used in the estimation of R-indicators is, once again, key (Schouten et al. 2016; Nishimura et al. 2016; Cornesse and Bosnjak 2018; Schouten 2018). The chosen auxiliary variables should not only be strongly related to the ‘real’ response propensities, but also to the variables of interest. Understanding that relationship is essential for interpreting (and evaluating the utility of) the nonresponse bias indicator, as well as the potential for reducing bias through adjustment.

In practice, the availability of auxiliary data for both respondents and nonrespondents is typically limited, leaving researchers little choice over which variables to use to build indicators of nonresponse bias (or nonresponse adjustment weights) – especially in the context of cross-sectional surveys (e.g. Sakshaug and Antoni 2018). If auxiliary data do exist, they typically consist of socio-demographic variables (e.g. on sampling frames), which may correlate only weakly with response probabilities and the variables of most interest to data users (Peytcheva and Groves 2009; Schouten 2018; Cornesse and Bosnjak 2018). In the case of general purpose (cross-sectional) social surveys (e.g. the International Social Survey Programme, the European Social Survey, the General Social Survey, the European and World Values Studies), where users may be interested in nonresponse impact on a diverse range of subjective variables covering many different topics, this limitation may be especially frustrating, and may, in turn, limit the value of R-indicators for nonresponse bias assessments in such studies.

## 2.2. Assessments of the Performance of R-Indicators

Because the utility of R-indicators (and other related indicators) is so dependent on the availability and power of the auxiliary variables used in their estimation, assessments of

their performance should ideally be focused on the latter. To date, however, relatively few studies have investigated how R-indicators perform under different conditions or how the choice of auxiliary variables used (and other factors) influence the utility of the information the indicators provide. This is partly because nonresponse biases on survey variables are usually unknown, rendering the validation of indirect indicators of bias risk challenging. This section provides a short review of available studies and their conclusions.

[Cornesse and Bosnjak \(2018\)](#) conducted a meta-analysis investigating the effect of different survey design variables on the representativeness of survey samples, including the number (though not type) of auxiliary variables used in the estimation and its relation to the magnitude of the R-indicator and the MAB. They hypothesised that the more auxiliary variables included, the more likely it is to detect bias (i.e., the smaller the value of the R-indicator and the larger the MAB ([Cornesse and Bosnjak 2018](#), 5). However, over 104 R-indicator studies, they did not find the anticipated relationship.

A theoretical contribution by [Schouten \(2018\)](#) considered the type of auxiliary variables used to detect bias (using the CV) and their degree of association with survey variables affects their capacity to detect bias on other variables. He presents a framework in which the socio-demographic auxiliary variables that are typically available (and used for bias detection and adjustment) are viewed as just one possible selection from the universe of potential variables on a population. Using simulations and an application to the problem of attrition bias in the Dutch online Longitudinal Internet Studies for the Social Sciences (LISS) panel, he attempts to show how the level of association between select survey and auxiliary variables (comparing standard socio-demographic variables with random draws of 20 alternative variables taken from prior waves of the panel) may influence the potential to detect bias. He concludes that auxiliary variables selected at random appear able to detect a (predictable) amount of the total bias (more bias in the auxiliary variables implying more expected bias in the survey variables). However, the standard socio-demographic covariates generally outperform any random selection.

[Schouten's \(2018\)](#) conclusions raise the question of whether larger bias detected by available (sociodemographic) auxiliary variables (which are not randomly selected) is also a sign of larger bias in other variables (i.e., above what would be predicted by a random selection of covariates). This question was addressed by [Schouten et al. \(2016\)](#) in a study investigating the usefulness of R-indicators in the context of adaptive survey design (the original motivation also for [Schouten's 2018](#) article). As the aim of such designs is to reduce bias by targeting fieldwork strategies to particular subgroups to optimise the balance of the response sample on auxiliary variables, it is of interest to know whether the (logistically more costly) targeted approach is more effective at reducing bias in survey estimates than simply adjusting on the same variables [Schouten et al. \(2016, 728\)](#), and hence, how informative the magnitude of the R-indicators are about the extent of bias after adjustment. Across 14 data sets, the authors found that achieving a balanced sample through adaptive design guided by such indicators was generally beneficial, resulting, on average, in less nonresponse bias in target survey variables even after adjustment (though the need for adjustment was not eliminated completely; [Schouten et al. 2016, 745](#)). However, due to some inconsistencies they observed, they conclude that more research is needed to provide further guidance as to the conditions under which a higher R-indicator

or lower CV (i.e., a more balanced response sample with respect to the auxiliary variables) implies less bias in the survey variables (Schouten et al. 2016, 745).

Finally, Nishimura et al. (2016) used simulation studies to compare the R-indicator to a number of alternative nonresponse bias indicators, under different scenarios varying response rates, missing data mechanisms (i.e., whether data are missing at random (MAR), missing completely at random (MCAR) or not missing at random, NMAR), and at varying levels of correlation between the auxiliary data and the survey data. They found that R-indicators did not perform well at indicating the magnitude of the bias on survey variables, though their effectiveness in this regard depended on the missing data mechanism (Nishimura et al. 2016, 54). Especially at low values, R-indicators give some indication of whether the data are MAR rather than MCAR. However, it is not possible to distinguish between MCAR and NMAR mechanisms, especially when the value of the indicator is large. On the assumption that available auxiliary data could be used to adjust bias on survey variables (as in Schouten et al.'s 2016 research), the same authors extended their analysis to investigate the circumstances in which nonresponse weight-adjusted means showed less bias than the unadjusted means, and obtained mixed findings across different survey variables. Though this is to be expected theoretically, it warrants further investigation to inform our understanding of what the magnitude of the R-indicator implies for bias in other survey variables after adjustment on the auxiliary variables (Nishimura et al. 2016, 59).

### 2.3. *The Present Study*

We address some of the issues raised above in the present case study, in which we investigate the effectiveness of R-indicators (and the related CV) as a measure of the risk of nonresponse bias in the context of the European Social Survey (ESS). Specifically, the study addresses the following research questions:

- RQ1: To what extent are available auxiliary data suitable for estimating response propensities and the risk of nonresponse bias using R-indicators? Or, specifically, how well do response propensities estimated on the basis of available auxiliary variables correlate with target survey variables?
- RQ2: To what extent are R-indicators based on the available auxiliary variables good predictors of nonresponse biases on target variables?
- RQ3: To what extent is the magnitude of the R-indicator informative about bias in the target variables once bias in the auxiliary variables has been adjusted? In other words, does more nonresponse bias on auxiliary variables (i.e., a lower R-indicator) imply more bias on other variables, even after adjustment for the auxiliary variables?

To tackle these questions, we use auxiliary data from a sampling frame based on population registers to estimate sample members' response propensities and the R-indicator/CV. We then assess the correlation between the response propensities with a selection of target survey variables. Finally, we examine the extent of the 'actual' nonresponse biases in the target variables, estimated on the basis of a nonresponse follow-up survey, and compare these to the predicted risk of bias provided by the R-indicator, before and after adjustment on the auxiliary variables.

### 3. Data and Methods

#### 3.1. Data

To address our research questions, we use data from Round 5 (2010) of the Swiss European Social Survey (ESS). The ESS is a biennial cross-national face-to-face survey of social values and attitudes. The questionnaire consists of a repeated core of items aimed to measure changing social attitudes and values, and two ‘rotating’ modules focused on specific topics, which change in each round. The ESS target population is defined as all resident adults (aged 15 and over) within private households, ‘regardless of their nationality, citizenship, language or legal status’ (ESS5 – 2010 Documentation Report 2012). The Swiss Federal Statistical Office (SFSO) supplied the Swiss ESS National Coordinator with a single-stage equal probability systematic sample of individuals from this population, with no clustering, proportionally stratified by the seven NUTS-2 regions of Switzerland (CH01 – Région lémanique; CH02 – Espace Mittelland; CH03 – Nordwestschweiz; CH04 – Zürich; CH05 – Ostschweiz; CH06 – Zentralschweiz; CH07 – Ticino). The total number of issued sample units was 2,850, and the final number of valid interviews was 1,506 – a total response rate of 53.2% (equivalent to AAPOR Response Rate 1). For more details on the fieldwork protocol and response enhancement methods used for the main survey, see Roberts et al. (2014a).

As well as using questionnaire data from the main survey interview, we analyse data from the Swiss Federal Statistical Office (SFSO)’s sampling frame of residents in Switzerland, based on population registers maintained by municipalities. In addition to individual names and addresses, the frame contains a number of socio-demographic variables, including the individual’s sex, date of birth, marital status, and nationality. On the basis of address information, additional contextual variables are derived, including the linguistic region of Switzerland (French, German, Romansch or Italian), and the degree of urbanicity. Telephone numbers were obtained for 61% of the sample via an automatic search by the fieldwork agency in the commercial database (‘AZ Direct’), so the auxiliary variables additionally include an indicator of whether or not a telephone number was available. This variable is known to be an important correlate of response propensity, in part because telephone contacts are used in refusal conversion and as a means to reduce the noncontact rate.

The third source of data used in this study comes from a nonresponse follow-up (NRFU) survey (Ernst Stähli et al. 2018), which was a postal survey carried out two months after the end of the main survey fieldwork, and consisted of a single sheet (double-sided) paper questionnaire with around 20 questions. After removing ineligible sample units, the nonresponse questionnaire was sent to 1,047 non-respondents (186 refusals were not recontacted for reasons not known). Efforts to improve response rates if the questionnaire had not been completed and returned within four weeks included re-contacts by telephone (if a number was available) or by mail (if no number was available). The response rate for the NRFU was 55.7%, yielding a total of 583 cases for analysis. In total, therefore, 2,089 cases (73.3% of the total sample) responded to either the main survey or the NRFU, leaving 26.7% in a group we refer to here as ‘persistent nonrespondents’. Further details of final outcome rates are available in Table 1. From now on, we refer to the group of respondents to the main survey together with the respondents of the NRFU as the ‘reduced’ sample, in contrast to the original ‘complete’ sample, which includes the persistent non-respondents.

Table 1. ESS5 2010 final outcome rates (Switzerland).

<b>Break-down of final response and nonresponse:</b>	N	%
Total number of issued sample units	2,850	100.0
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 <sup>1</sup>	20	0.7
Respondent moved abroad	10	0.4
Respondent deceased	7	0.3
Number of valid interviews	1,506	52.8
<b>Total non-respondents eligible for follow-up<sup>2</sup>:</b>	1,047	100.0
Non-contacts	278	26.6
Refusals and refusals by proxy (excluding office refusals)	769	73.5
<b>Completed NRFU questionnaires by non-respondents:</b>		
On paper	530	50.6
By telephone	53	5.1

Notes. <sup>1</sup>Not residential, not occupied, not traceable or other ineligible. <sup>2</sup>Does not include respondents who were not sent the nonresponse follow-up questionnaire.

The selection of the 23 items included in the NRFU was based partly on decisions taken in collaboration with the Core Scientific Team of the ESS (see [Stoop et al. 2010](#); [Matsuo et al. 2010](#)). Items were selected on the assumption that they might be particularly likely to correlate with variables influencing the decision to participate in the survey and thus be at risk of nonresponse bias. The complete NRFU questionnaires are available in online Supplemental material. Details of the variables analysed here are shown in [Figure 1](#).

Data from nonresponse surveys can suffer from timing of the fieldwork, context, and mode of data collection effects, depending on the design of the shorter questionnaire and how it is administered ([Voogt and Saris 2005](#)). These artefacts may hinder comparisons with the answers given by respondents to the main survey questionnaire and hence, the overall assessment of nonresponse bias. To remedy this issue, the ESS NRFU questionnaire was additionally sent to a random subsample of 300 respondents to the main survey to enable an assessment of measurement differences resulting from the delayed timing of fieldwork, the change in mode and possible context effects from shortening the questionnaire. Based on an analysis of this sample, six out of the 23 items in the questionnaire were found to suffer from low reliability and were, therefore, excluded from the analysis of nonresponse bias reported here (see [Vandenplas et al. 2015](#) for further details). This resulted in a total of target variables measured in the main survey and the NRFU, for the assessment of nonresponse bias. Moreover, to further minimise the potential impact of differences between the two sources in the distribution of responses across ordinal response categories, we recoded them into binary variables.

The success of the nonresponse survey approach also depends on the extent to which respondents to NRFUs are representative of all non-respondents to the main survey ([Cobben 2009](#)). Following the continuum of resistance theory ([Lin and Schaeffer 1995](#); [Stoop 2004](#)),



1. **Living with a partner** (1 'Living with a partner' 0 'Not living with a partner')
2. **In paid work** (1 'In paid work' 0 'Other main activity')
3. **High school education only** (1 'Primary, secondary school or vocational/training school' 0 'Higher levels of education')
4. **Fixed line telephone** (1 'Fixed line telephone in accommodation' 0 'No fixed line telephone')
5. **Registered fixed line number** (1 'Fixed line number registered' 0 'No fixed line or fixed line number not registered')
6. **Mobile telephone** (1 'Respondent has a mobile phone' 0 'Respondent does not have a mobile phone')
7. **Registered mobile number** (1 'Registered' 0 'Not registered')
8. **Good health** (1 'Very good or good' 0 'Fair, bad or very bad')
9. **Extremely happy** (1 'Extremely happy (7, 8, 9, or 10 on 11-point scale)' 0 'Not extremely happy')
10. **Takes part in social activities** (1 'More or much more than most' 0 'About the same, less or much less than most')
11. **Meets people socially frequently** (1 'Several times a week or everyday' 0 'Once a week or less often')
12. **Very or quite interested in politics** (1 'Very or quite interested' 0 'Hardly or not at all interested')
13. **Satisfied with democracy** (1 'Extremely satisfied (7, 8, 9, 10 on 11-point scale)' 0 'Not extremely satisfied')
14. **Immigrants make country better** (1 'Immigrants make Switzerland a better place to live (7, 8, 9, or 10 on 11-point scale)' 0 'Immigrants do not make Switzerland a better place to live')
15. **Has complete trust in justice** (1 'Almost complete trust (7, 8, 9, 10 on 11-point scale)' 0 'Less than complete trust')
16. **Number of children** (0, 1, 2+)
17. **Number of people in the household** (1,2,3+)

*Fig. 1. Coding of variables in the nonresponse survey.*

which places respondents at the first contact attempt at one end of the continuum and nonrespondents at the other end, respondents to the NRFU can be considered to be situated somewhere between the respondents to the main survey and the persistent nonrespondents in terms of the characteristics measured by the survey (Lin and Schaeffer 1995; Stoop 2004). In this article, our analysis rests on the assumption that the nonrespondents participating in the NRFU are representative of all the nonrespondents to the main survey, including the persistent nonrespondents (and where possible, we validate this assumption with the sampling frame data). Additionally, we assume the answers to the NRFU survey are a good measure of the answers the respondents would have given had they participated in the main survey. We come back to these assumptions in the Discussion.

### 3.2. Analytic Approach

#### 3.2.1. Estimating Response Propensities

To estimate the response propensities, we estimated the parameters of logistic regression equations predicting each sample member's probability of participating in the survey

using covariates from the sampling frame data (Roberts et al. 2014a). These were respondent sex (coded 1 if male); age categories (<30 years, 31–44, 45–64, leaving the group aged 65 and over as the reference); marital status (coded 1 if married or in a legal partnership, 0 if single, divorced or widowed); nationality (coded 1 for those without Swiss citizenship, 0 if Swiss); linguistic region (coded 1 if from the French or Italian-speaking regions), 0 if German or Romansch-speaking (the ten Romansch-speaking respondents were interviewed by German speaking interviewers); urbanicity (coded 1 if living in an urban area and 0 if an isolated town or rural community); and availability of a telephone number (coded 1 if available 0 if not).

For our main analyses, we estimate the response propensities twice: first, for the complete sample (i.e., predicting response to the main survey among all sample members) and second, for the reduced sample (i.e., predicting response to the main survey among respondents to either the main survey or the NRFU). To assess the implications of only focusing on the responding nonrespondents and not all nonrespondents, we also estimate the response propensities for respondents to the NRFU compared with all the persistent nonrespondents.

### 3.2.2. Assessing the Relation Between Auxiliary Variables and Target Variables

To evaluate the effectiveness of the auxiliary variables included in the propensity model as indicators of bias in the target variables (RQ2), we first examine the correlations between the predicted response propensities estimated from the logistic regression model for the reduced sample and responses given in the main survey to the questions that were also included in the NRFU (Gummer and Blumenstiel 2018; Sakshaug and Antoni 2018). The items included 13 questions from the core questionnaire covering a variety of topics, five of which came from the socio-demographic module. The remaining four were country-specific items about having a fixed line or mobile telephone and whether the fixed line/mobile numbers were registered. All the items were recoded into dichotomous variables, where 1 represented a positive or affirmative response to the question. Household size and number of children in the household were kept as continuous measures. However, we also recoded them into categorical indicators, and then created dummy variables for each category where the first category was the reference (single person household/ household with no children, see Figure 1).

The list of coefficients includes a mix of Pearson's  $r$  (for continuous variables), biserial and point-biserial correlations (depending on whether the dichotomy reflects a discrete or continuous relation between the response options). For this reason, we convert the coefficients to  $z$ -scores (standard normal distribution) to facilitate comparisons between them. We focus our interpretation on whether or not the correlation was statistically significant at the 95% level.

In general, if there is a strong correlation between the estimated response propensities and the survey variable, the auxiliary variables should be considered suitable as predictors of bias for this variable (Little and Vartivarian 2005). Where the correlation between the estimated response propensities and the survey variable is low, we can have less confidence in the ability of the auxiliary variables to predict bias. Low correlations could also occur if the considered variable does not suffer from nonresponse bias. As a result, it is important to conjointly look at the correlations with the response propensities alongside the nonresponse biases with the help of the NRFU.

### 3.2.3. Assessing Nonresponse Bias

To investigate the extent to which nonresponse resulted in bias on the 17 target variables (RQ2), we compare estimates based on respondents to the main survey, and respondents to the NRFU, before and after adjustment for nonresponse bias on the auxiliary variables. To assess the size of the bias we compute the difference (contrast) in the proportion of respondents to the main survey and respondents to the NRFU selecting the categories coded 1, and use Chi-square tests of association to test whether the difference in proportions are statistically significant before and after adjustment on the auxiliary variables. To adjust for nonresponse bias in the auxiliary variables, we computed weights on the basis of the propensity scores from the logistic regression models (Little 1986), once for the complete sample, and once for the reduced sample. The weight for the respondents was estimated as the inverse of the propensity score, while that for the nonrespondents was calculated as one minus the inverse of the propensity score.

### 3.2.4. Assessing the R-Indicator As a Predictor of the Risk of Bias

The R-indicators were also estimated on the basis of the predicted response propensities from the logistic regression models using the R tool developed by De Heij et al. (2015). In addition, we also compute the adjusted coefficient of variation (CV) in the response propensities (following the formula provided by De Heij et al. (2015, 18 (14))), which is relevant when considering population means and totals (De Heij et al. 2015) and is equivalent to the Maximum Absolute Bias (MAB). As mentioned, the MAB is defined as the largest possible nonresponse bias on an estimate of a population mean in a survey with a response rate of less than 100%. To assess whether the R-indicator is a good predictor of the risk of bias (RQ2), we additionally estimate the Maximal Absolute Contrast (MAC) following the formula provided by Schouten et al. (2010). The MAC is defined as the largest possible difference between the respondents and non-respondents on an estimate of a population mean in a survey with a response rate of less than 100% (Schouten et al. 2010). If the estimated R-indicator is a good predictor of bias, the MAC should give a realistic upper limit for any actual difference observed between main survey and NRFU. Note that while the MAC gives the maximum possible difference between respondents and nonrespondents, the CV/MAB represents the maximum bias, in other words, the maximum difference between the respondents to the main survey and the total complete and reduced samples.

Finally, to address the question of whether more nonresponse bias on the auxiliary variables (i.e., a lower R or higher CV) implies more nonresponse bias on the survey variables after adjustment for the auxiliary variables (RQ3), we estimate the difference in the contrast between respondents and nonrespondents in the reduced sample, after adjustment on the auxiliary variables, and compare the effects of the two different propensity score weights.

## 4. Results

The presentation of the results is organised around our research questions. We start by presenting the results of the logistic regression models for estimating the response

propensities for the complete and the reduced sample, and the correlations between response propensities and the target variables measured in the main survey (RQ1). Then, we present the unadjusted biases on the target variables, and the R-indicators and related risk-of-bias indicators (RQ2). Finally, we present the nonresponse adjusted biases on the target variables for both samples, together with their contrasts (RQ3).

#### 4.1. Predicted Response Propensities

Coefficients for the parameters of the logistic regression models estimated for the complete and reduced samples are given in Table 2. With the exception of sex, all variables included in the model for the complete sample were significantly associated with the propensity to respond to the main survey. Living in an urban area, residing in the French or Italian regions of Switzerland compared with the German region, and being a foreigner were all negatively associated with responding to the survey, while being aged 15–30 or 45–65 (compared to being over 65), being married, and having a registered telephone number were positively associated with responding. When we replace the complete sample with the reduced sample, living in an urban area, being married, being a foreigner and being aged 45–65 (compared to older) are no longer significantly associated with responding to the survey. This could be an indication that the NRFU fails to increase the level of participation for certain subgroups. Overall, however, it implies that there is somewhat less nonresponse bias on the auxiliary variables in the reduced sample compared with the complete sample.

Table 2. Parameter coefficients for logistic regression equations estimating response propensities for the reduced and complete samples.

Parameter	Complete sample			Reduced sample		
	$\hat{\beta}$	<i>p</i>	SE	$\hat{\beta}$	<i>p</i>	SE
Male	0.08		0.08	0.18 <sup>+</sup>		0.10
Urban	−0.31***		0.09	−0.24		0.11
Linguistic region (ref. German)						
French	−0.34***		0.09	−0.48***		0.11
Italian	−0.55**		0.19	−0.70**		0.23
Nationality (ref. Swiss)						
Bordering countries	−0.41**		0.14	−0.07		0.19
Other countries	−0.68***		0.13	−0.12		0.17
Age category (ref. 65+)						
15–30 years	0.54***		0.13	0.37*		0.17
31–44 years	0.09		0.12	−0.09		0.15
45–65 years	0.30**		0.11	0.12		0.14
Married	0.25**		0.09	0.13		0.11
Telephone number available	0.38***		0.08	0.34**		0.11
Constant	−0.07		0.14	0.82***		0.18
N		2,850			2,089	
Nagelkerke R <sup>2</sup>		0.07			0.04	
Hosmer-Lemeshow's Test		0.73			0.31	

Notes:  $\hat{\beta}$  = unstandardized beta coefficient; SE = standard error; ref. = reference category; <sup>+</sup>p < .1, \*p < .05, \*\*p < .01, \*\*\*p < .001; Data source: ESS 2010.

Note that both models have a poor fit (p-value for the Hosmer-Lemeshow's Test of 0.31 for the reduced sample and 0.73 for the complete sample and 60% versus 62% correspondence between expected and observed outcomes), indicating that taken together, the available auxiliary data explain little of the variance in the probability of responding to the survey. Nagelkerke's  $R^2$  for both models is also low, which could be taken as a positive indication of the overall magnitude of bias on these variables.

#### 4.2. Correlations Between the Response Propensities and the Target Variables

Of the 17 target variables analysed, 11 were significantly correlated with the predicted response propensities (right-hand side of Table 3, columns 4 and 5). Nine of these variables were positively correlated with the response propensities, while the other two were negatively correlated (number of children reported to be living in the household and believing that immigrants do not make Switzerland a better place to live). The six variables not significantly correlated with the response propensities were: having a registered mobile telephone number, being in paid work, being in good health, frequently meeting people socially, taking part more often in social activities, and having complete trust in the justice system. Of these, the first five were selected for the NRFU because they were presumed to relate to a sample member's contactability/time availability to participate in the survey and hence be at risk of bias if noncontact rates indeed vary as a function of these characteristics. If this is the case and these target variables are consequently affected by bias, then the socio-demographic variables used to estimate the response propensities would be ineffective for predicting the risk of bias on these measures. However, as mentioned, low correlations with the response propensities may also result from a lack of bias in the target variables, so for this reason, we need to assess the correlations alongside the nonresponse biases in the target variables, which we do in more detail in the following.

#### 4.3. Actual Nonresponse Bias

Shown in the left-hand side of Table 3 are the unadjusted estimates for the target variables based on the main survey respondents (column 1) and the NRFU respondents (column 2), together with the contrast (column 3 – ordered according to size). In total, nine out of 17 estimates were affected by nonresponse bias (statistically significant differences between the respondents and nonrespondents). Seven of these were among the variables that were correlated significantly with the response propensities (shaded in grey). These included five factual variables: number of people and number of children in the household, education, having a fixed line telephone and a registered fixed line telephone number; and two subjective variables: extremely happy, and satisfied with democracy. Note that the correlation with the 'telephone' variables and to a certain extent the number of people in the household (from the main survey and nonresponse follow-up) was to be expected given that variables about having a registered phone number and marital status (from the frame data) were included in the response propensity model. For all these variables, the differences between respondents and nonrespondents were in the expected direction, depending on the sign of the correlation coefficient: nonrespondents scoring lower if the correlation was positive and higher if it was negative (which, in fact, was only the case of

Table 3. Differences in unadjusted estimates based on the main survey and NRFU, with correlation coefficients between predicted response propensities and target variables.

Target variables	(1)		(2)		(3)		(4)		(5)	
	<i>n</i>	Main survey ( <i>n</i> = 1506) %(SE)	NRFU ( <i>n</i> = 583) %(SE)	Contrast %(SE)	<i>p</i>	<i>r</i>	<i>z</i>	<i>p</i>		
1. Takes part in social activities more	2,063	18.0 (1.0)	18.0 (1.6)	0.0 (1.8)		rb = -0.06	-0.43			
2. In paid work	2,057	57.7 (1.3)	57.5 (2.1)	-0.0 (2.5)		rpb = -0.02	-0.66			
3. Very or quite interested in politics	2,083	58.9 (1.3)	58.3 (2.1)	-0.6 (2.5)		rb = 0.16	5.62***			
4. Mobile telephone	2,062	87.7 (0.9)	88.5 (1.3)	0.8 (1.6)		rpb = 0.04	2.24*			
5. Living with a partner	2,079	62.3 (1.3)	63.2 (2.0)	0.9 (2.4)		rpb = 0.05	1.90*			
6. Good health	2,084	81.4 (1.0)	79.2 (1.7)	2.2 (1.9)		rb = -0.01	-0.21			
7. Immigrants make country better	2,044	29.2 (1.2)	26.1 (1.8)	-3.2 (2.1)		rb = -0.12	-3.95***			
8. Registered mobile number <sup>1</sup>	1,798	7.3 (0.7)	9.8 (1.3)	2.5 (1.5)		rpb = 0.08	0.01			
9. Number of people in household (mean) Single household	2,052	2.8 (0.0)	2.6 (0.06)	-0.2 (0.1)*		r = 0.10	3.20***			
	378	17.8 (1.0)	20.0 (1.7)	-2.1 (1.9)		rpb = -0.16	-6.31***			
Two household members	750	35.5 (1.2)	39.6 (2.1)	-4.1 (2.4) <sup>+</sup>		rpb = 0.05	1.85 <sup>+</sup>			
Three or more household members	924	46.7 (2.1)	40.5 (1.3)	6.2 (2.4)*		rpb = 0.08	3.03***			
10. High school education only	2,076	59.7 (1.2)	54.8 (2.1)	-4.9 (2.4)*		rpb = 0.10	3.83***			
11. Number of children in household (mean) Household without child	2,071	0.5 (0.0)	0.58 (0.04)	0.1 (0.0)**		r = -0.06	2.40**			
	1,513	74.1 (1.1)	70.3 (1.9)	3.8 (2.1) <sup>+</sup>		rpb = 0.12	4.55**			
Household with one child	256	11.8 (0.8)	14.0 (1.5)	-2.2 (1.6)		rpb = -0.07	-2.52**			
Household with two or more children	302	14.1 (0.9)	15.8 (1.5)	-1.6 (1.7)		rpb = -0.09	-3.37***			
12. Has complete trust in justice	2,046	53.8 (1.3)	46.8 (2.1)	-7.0 (2.2)**		rb = 0.03	0.47			
13. Fixed line telephone	2,082	90.6 (0.8)	83.7 (1.5)	-6.9 (1.7)**		rpb = 0.18	6.88***			
14. Meets people socially frequently	2,082	52.3 (1.3)	43.1 (2.1)	-9.2 (1.4)**		rb = 0.01	0.19			
15. Extremely happy	2,081	89.2 (0.8)	80.0 (1.7)	-9.2 (1.9)**		rb = 0.10	2.44*			
16. Satisfied with democracy	2,045	69.7 (1.2)	57.5 (2.1)	-12.1 (2.4)**		rb = 0.08	2.30*			
17. Registered fixed line number <sup>1</sup>	1,798	94.9 (0.6)	79.5 (1.9)	-15.4 (2.0)**		rpb = 0.15	8.75***			

Notes. <sup>1</sup>Questions only asked of respondents who reported having a fixed-line or mobile telephone. *r* denotes Pearson's *r*, *r*<sub>pb</sub> the point-biserial correlation and *r*<sub>b</sub> the biserial correlation. \*\*\**p* < .001, \*\**p* < .01, \**p* < .05, <sup>+</sup> *p* < .1. Shaded cells show variables with both significant contrasts and correlations.

number of children in the household). For this set of target variables, therefore, a post-survey nonresponse adjustment or a targeted fieldwork based on the available auxiliary variables should substantially reduce the nonresponse bias (and this is indeed the case – see below).

The two remaining variables that were affected by nonresponse bias but not significantly correlated with the response propensities were frequently meeting people socially and having complete trust in justice. This finding suggests that these two variables are strongly related to the nonresponse mechanism (nonrespondents having less trust in justice and meeting people less frequently), but that the estimated response propensities used to build the R-indicator fail to predict bias because the variables are not related to the auxiliary variables. For these variables, therefore, using the available auxiliary data for predicting the risk of bias, or for the purposes of post-survey adjustment or targeted fieldwork strategies would fail to correct the nonresponse bias (and this is also confirmed below).

For the eight variables where no nonresponse bias was observed (i.e., where there were no significant differences between the respondents and nonrespondents), four (living with a partner, having a mobile phone, being very or quite interested in politics, and believing immigrants make the country better) were among those that were significantly correlated with the estimated response propensities. The correlations suggest that bias could potentially arise as a result of nonresponse, but with the reduced sample of nonrespondents observed here, no bias is detected. For the remaining four variables where no bias was detected, the correlation with the estimated response propensities was not significantly different from zero. These included taking part in social activities more often than other people, being in paid work, having a registered mobile phone and being in good health.

#### 4.4. R-Indicators As Predictors of the Risk of Bias

The response rate for the complete sample (i.e., the actual survey outcome, without taking the ineligible into account) was 52.8% (see column 1 of Table 4), and the adjusted R-indicator was 0.79. The response rate increases to 72.1% and the value of the R-indicator increases to 0.86 (column 2 of Table 4), when they are calculated on the basis of the reduced sample (i.e., when the ‘persistent nonrespondents’ are removed from the sample).

Table 4. Response rates, adjusted R-indicators, coefficients of variation (CV) and maximum absolute contrast (MAC) for the complete and reduced samples and for the non-respondents.

	(1) Complete sample (n = 2,850)	(2) Reduced sample (n = 2,089)	(3) NRFU compared to all non-respondents (n = 1,344)
Response sample size	1,506	1,506	583
Response rate <sup>1</sup>	52.8%	72.1%	43.4%
Adjusted R-indicator	0.79	0.86	0.83
Confidence interval	(0.75–0.82)	(0.82–0.90)	(0.78–0.88)
Adjusted CV	0.20	0.10	0.20
MAC	0.44	0.35	0.35

Notes. <sup>1</sup>Response rate calculated here as total number of interviews divided by the sample size (i.e., it does not take account of ineligible). Data source: ESS 2010.

The coefficient of variation (MAB) for the complete sample is 0.21 (21%), while for the reduced sample it is 0.10 (10%), so using the reduced sample instead of the complete sample underestimates the risk of bias by almost 11% (according to R-indicator estimated on the basis of the available auxiliary variables). This suggests the respondents to the NRFU are indeed situated somewhere ‘between’ the respondents to the main survey and the extreme non-respondents, and shows how the NRFU may fail to detect bias on certain variables.

To fully assess predictions of the risk of bias in target variables provided by the R-indicator and CV, we also consider their implications for the bias that remains after adjusting on the auxiliary variables (RQ3). We compare the contrasts (i.e., the differences between estimates based on the respondents to the main survey and respondents to the NRFU) before and after applying the propensity score weighting adjustment a) based on the logistic regression predicting participation for the complete sample (columns 1–3 of [Table 5](#)) and b) based on the logistic regression predicting participation for the reduced sample (columns 4–6 of [Table 5](#)). With both the complete and reduced sample adjustment, bias is ‘removed’ from three of the variables correlated with the response propensities: mean number of people (though for two person households specifically, it persists) and number of children in the household and having only completed high school education. However, bias remains in six of the target variables, of which four were significantly correlated with the response propensities – having a fixed line telephone and registered telephone number, being extremely happy and being satisfied with democracy. In all but the latter, the size of the contrast is reduced by the adjustment, but not removed. For the two variables not correlated with the response propensities – trust in justice and meets people socially – bias remains after adjustment (the size of the contrast increases slightly for the former and reduces slightly for the latter).

The pattern of results when the two sets of weights are applied is very similar (the absolute size of the contrasts when adjusting to the reduced sample propensity scores is actually slightly larger for six of the variables). There is one exception, however. In one other variable – believing immigrants makes the country better – adjustment to the complete sample increases the contrast such that it becomes statistically significant. This variable was one of the four variables correlated with the response propensities for which no bias was observed initially. With the reduced sample adjustment, the contrast on this variable also increases, but the difference between the respondents and nonrespondents is not significant. Thus, adjustment on the auxiliary variables (irrespective of whether the complete or reduced sample is considered) has – as anticipated – a mixed, and not altogether positive, impact on the estimates. In this respect, the lower R-indicator associated with the complete sample (where there is more nonresponse bias in the auxiliary variables) does imply more bias on the other variables, even after adjustment – but also due to the adjustment. However, overall, the gain in bias from the complete sample adjustment is greater than with the reduced sample adjustment.

A final observation can be made about the size of the bias on each of the variables (see column 3 of [Table 4](#)). We hypothesised that a ‘good’ R-indicator estimation based on the ‘reduced’ sample should predict bias ‘correctly’ and that this can be verified by examining whether the value for the MAC – the maximum difference between respondents and nonrespondents – exceeds the observed differences between respondents and



Table 5. Differences in estimates based on the main survey and NRFU respondents, adjusting for bias on the sociodemographic variables using a) response propensity scores predicted for the complete sample; and b) response propensity scores predicted for the restricted sample.

Target variables	Complete vample weights				Restricted sample weights				
	(1) Main survey (n=1,506) %(SE)	(2) NRFU (n=583) %(SE)	(3) Contrast %(SE) p	(4) Gain in bias <sup>2</sup> %	(5) Main survey (n=1,506) %(SE)	(6) NRFU (n=583) %(SE)	(7) Contrast %(SE) p	(8) Gain in bias <sup>2</sup>	(9) Diff. (7)-(8)
1. Takes part in social activities more	2,063	17.9 (1.0)	18.6 (1.7)	0.7	18.1 (1.0)	18.7 (1.7)	0.6 (2.0)	0.6	-0.1
2. In paid work	2,057	57.5 (1.3)	57.3 (2.1)	-0.2 (2.5)	57.9 (1.3)	57.2 (2.1)	-0.8 (2.4)	-0.8	-0.6
3. Very or quite interested in politics	2,083	57.1 (1.3)	59.1 (2.1)	2.0 (2.5)	58.0 (1.3)	59.6 (2.1)	1.5 (2.5)	2.1	-0.1
4. Mobile telephone	2,062	87.2 (0.9)	88.8 (1.4)	1.6 (1.6)	87.6 (0.9)	88.7 (1.4)	1.1 (1.6)	0.3	-0.5
5. Living with a partner	2,079	60.9 (1.3)	64.0 (2.1)	3.2 (2.4)	62.0 (1.3)	63.4 (2.1)	1.4 (2.4)	0.5	-1.8
6. Good health	2,084	81.2 (1.0)	79.5 (1.7)	-1.7 (2.0)	81.5 (1.0)	79.1 (1.8)	-2.4 (2.0)	-4.6	-0.1
7. Immigrants make country better	2,044	30.6 (1.3)	24.1 (1.8)	-6.4 (2.2)**	29.8 (1.2)	24.3 (1.8)	-5.5 (2.2)	-2.3	-1.0
8. Registered mobile number	1,798	7.2 (0.7)	10.3 (1.4)	3.1 (1.6)	7.2 (0.7)	10.3 (1.5)	3.1 (1.6)	0.6	-0.0
9. Number of people in household (mean)	2,052	2.7 (0.0)	2.6 (0.1)	-0.1 (0.1)	2.7 (0.0)	2.6 (0.1)	-0.1 (0.1)	0.1	-0.0
Single household	378	19.8 (1.1)	18.5 (1.7)	-1.3 (1.9)	18.5 (1.0)	18.6 (1.7)	0.1 (1.9)	2.2	-1.4
Two household members	750	34.6 (1.3)	39.8 (2.2)	5.2 (2.5)*	35.2 (1.3)	39.9 (2.2)	4.7 (2.5) +	8.8	0.5
Three or more household members	924	45.7 (1.3)	41.7 (2.2)	-3.9 (2.6)	46.3 (1.3)	41.5 (2.2)	-4.8 (2.6) +	-11.0	0.9
10. High school education only	2,076	59.3 (1.3)	56.6 (2.1)	-2.8 (2.5)	59.1 (1.3)	56.7 (2.1)	-2.4 (2.5)	2.5	-0.4
11. Number of children in household (mean)	2,071	0.5 (0.0)	0.5 (0.0)	0.1 (0.1)	0.5 (0.0)	0.5 (0.0)	0.0 (0.0)	-0.1	-0.0
Household without child	1,513	73.5 (1.2)	69.4 (2.0)	-4.0 (2.3) +	73.5 (1.2)	69.9 (2.0)	-3.7 (2.3)	-7.5	-0.3

Table 5. Continued

Target variables	Complete sample weights			Restricted sample weights			(9) Diff. (7)-(3)			
	(1) Main survey (n=1,506) %(SE)	(2) NRFU (n=583) %(SE)	(3) Contrast %(SE) p	(4) Gain in bias <sup>2</sup> %	(5) Main survey (n=1,506) %(SE)	(6) NRFU (n=583) %(SE)		(7) Contrast %(SE) p	(8) Gain in bias <sup>2</sup>	
Household with one child	256	12.2 (0.9)	13.9 (1.5)	1.8 (1.7)	4.0	12.0 (0.9)	14.1 (1.5)	2.1 (1.8)	4.3	-0.3
Household with two or more children	302	14.4 (0.9)	16.6 (1.6)	2.2 (1.9)	3.8	14.5 (0.9)	16.1 (1.6)	1.6 (1.7)	3.2	0.6
12. Has complete trust in justice	2,046	54.0 (1.3)	44.9 (2.1)	-9.0 (2.5)***	-2.0	53.7 (1.3)	45.4 (2.1)	-8.3 (2.5)***	-1.3	-0.7
13. Fixed line telephone	2,082	89.0 (0.9)	85.5 (1.4)	-3.5 (1.7)*	3.4	89.9 (0.8)	85.4 (1.5)	-4.5 (1.6)**	2.4	-1.0
14. Meets people socially frequently	2,082	52.1 (1.3)	43.2 (2.1)	-8.9 (2.5)***	0.3	52.3 (1.3)	43.2 (2.1)	-9.1 (2.5)***	0.1	-0.2
15. Extremely happy	2,081	88.5 (0.9)	80.4 (1.7)	-8.1 (1.9)***	1.1	89.0 (0.8)	80.6 (1.7)	-8.5 (1.9)***	0.7	-0.4
16. Satisfied with democracy	2,045	69.4 (1.2)	57.4 (2.1)	-12.0 (2.5)***	0.1	69.4 (1.2)	57.8 (2.2)	-11.6 (2.5)***	0.5	-0.4
17. Registered fixed line number <sup>1</sup>	1,798	94.0 (0.7)	82.1 (1.8)	-11.9 (1.9)***	3.5	94.6 (0.6)	82.3 (1.8)	-12.2 (1.9)***	3.2	-0.3

Notes. Data source: ESS 2010. <sup>1</sup>Questions only asked of respondents who reported having a fixed-line or mobile telephone. <sup>2</sup>Gain in bias = adjusted contrast - unadjusted contrast shown in table 3, column 3 (positive values indicate reduction in the absolute size of the contrast). + p < .1, \* p < .05, \*\* p < .01, \*\*\* p < .001. Shaded cells show variables with both significant contrasts and correlations when unadjusted.

nonrespondents. This is indeed the case – none of the observed differences were greater than 0.35 (the value for MAC for the reduced sample).

## 5. Discussion and Conclusion

In the search for indicators of the risk of nonresponse bias to supplement response rates, indicators of the representativeness of the responding sample ('R-indicators' – and the related coefficient of variation (CV) of response propensities) offer considerable appeal. Yet the utility of such indicators depends on a) the availability of suitable auxiliary data for their estimation, b) how well they predict nonresponse bias on other variables in the survey, and c) whether their magnitude (i.e., what they tell us about the extent of bias in the auxiliary variables) is also informative about bias on other variables after adjustment for bias on the auxiliary variables. We investigated these issues in a case study using data from a nonresponse survey to assess the extent of actual bias in estimates of socio-demographic and attitudinal measures from the Swiss ESS, by treating respondents to an NRFU survey as though they were the complete sample of non-respondents (or at least, perfectly representative of them). Though not unproblematic (discussed further below), this set-up allowed us to address questions raised in previous research (e.g. [Schouten 2018](#); [Schouten et al. 2016](#); [Nishimura et al. 2016](#)) about whether the presence of more nonresponse bias in auxiliary variables necessarily translates into more bias in survey variables (the issue raised by one anonymous reviewer of whether 'where there's smoke there's fire'), and how adjusting for nonresponse on auxiliary variables affects this relationship (e.g. [Schouten 2018](#); [Schouten et al. 2016](#); [Nishimura et al. 2016](#)).

By examining the correlations between estimated response propensities used to build the R-indicator and variables included in the NRFU, we assessed the suitability of the available auxiliary variables (socio-demographic variables from a sample frame based on population registers) for detecting the observed bias, before and after adjustment. We then assessed the value of the R-indicator, the CV (maximum absolute bias) and the maximum absolute contrast as summary statistics of the risk of nonresponse bias, by comparing their predictions with the biases detected by the NRFU.

Our results with respect to the auxiliary variables (RQ1) were, on the one hand, reassuring. Of the nine variables that were affected by bias, seven were significantly correlated with the estimated response propensities used to calculate the R-indicator, and the observed bias was consistent with the sign of the correlation coefficients. Five of these variables were socio-demographic or other factual variables, of which three (fixed telephone number, registered telephone number and number of people in the household) were directly correlated with two auxiliary variables that were included in the response propensity model (telephone number available and marital status). On the other hand, the absence of significant correlations for the remaining variables included in the NRFU (including some affected by bias) suggests some limits to the socio-demographic variables used for detecting bias on subjective measures. The two variables affected by bias that were not correlated with the estimated response propensities were 'has complete trust in justice' (nonrespondents had lower levels of trust), and 'meets socially frequently' (nonrespondents were less likely to meet). Thus, the R-indicator and CV based on the auxiliary variables available in this study were only partially informative about the extent of bias in the survey variables (RQ2).

Whilst perhaps not surprising given that their primary purpose is to translate bias in the auxiliary variables to smaller dimensions, this finding highlights potential limitations of R-indicators for practitioners and analysts. In addition, it has implications for the possibility to adjust nonresponse bias using the same auxiliary variables in post-stratification weights. Indeed, we found that when adjusting for nonresponse using a propensity score weighting method based on the response propensities predicted for the complete sample, the contrast in estimates for respondents and non-respondents for ‘trust in justice’ and ‘meets people socially’ remained statistically significant (i.e., bias was, unsurprisingly, not reduced by the adjustment). We also found that the contrast for the variable ‘immigrants make the country better’ became significant only *after* adjustment on the propensity scores for the complete sample (which was not the case when using the reduced sample weights).

Thus, in this case study, the presence of *more* nonresponse bias in the auxiliary variables (resulting in a lower R-indicator), did imply slightly more bias in the target variables both before and after adjustment (RQ3) – more smoke indicating more fire. Nevertheless, there were relatively few differences in the effectiveness of the complete and reduced sample adjustment methods, so the slightly larger R-indicator obtained for the reduced sample did not imply *much less* bias in the target variables than was the case for the complete sample. Even after adjustment on the auxiliary variables, bias remained on four of the variables correlated with the response propensities for which the unadjusted contrasts were significant (as well as the two which were not). These included self-reported measures of having a fixed-line telephone and telephone number (which apparently do not concur with the auxiliary data on number availability, which came from a commercial database), and feeling extremely happy and being satisfied with democracy. The results, therefore, make it difficult to draw strong conclusions about whether a survey design with less nonresponse bias on auxiliary variables also has, on average, less bias on other variables. We recommend that future research investigates this question further.

The finding that variation in the magnitude of the R-indicator is only partially informative of the risk of bias on other variables (irrespective of the effects of weighting) concurs with the findings of other studies (e.g. [Nishimura et al. 2016](#); [Schouten et al. 2016](#)). This limitation may be particularly relevant where subjective variables are concerned, however, and may not be entirely due to a lack of, or only weak correlations with the auxiliary variables. In particular, subjective variables may additionally be affected by substantial measurement biases ([Roberts and Vandenplas 2017](#)), which could account for some of the results observed in the comparison between the main survey and the NRFU. Nevertheless, the results suggest the need for some caution when interpreting the magnitude of the R-indicator – as well as that of related bias indicators. For example, while the MAC represents an upper limit of the contrast detected by the NRFU, we found that the values of the CV and MAC were somewhat exaggerated; a maximum nonresponse bias of 10% (reduced sample) or 21% (complete sample) as given by the CV, or a maximum contrast of 35% (reduced) or of 44% (complete) as predicted by the MAC, would likely be unacceptably high. In this sense, the predictions of the R-indicator do not map directly onto the observed nonresponse bias, meaning the R-indicator provides only a broad-brushed measure of the likely impact of nonresponse error. While it may be unrealistic to expect one value to represent the impact of nonresponse on multiple

variables, it is important for practitioners using R-indicators to be aware of the need for a more narrow interpretation of their meaning.

This having been said, and as previously discussed, partial R-indicators (and partial CVs) can, and do, provide far richer insight into how different variables (and categories of variables) are affected by nonresponse and in turn, contribute to a reduction in sample representativeness. In recognition of this, we extended the analyses presented here (Tables A1 and A2 in the online Supplemental material.) by calculating partial indicators for the reduced sample, for the auxiliary variables, and again for the target variables (using the latter to predict the response propensities). The results illustrate the advantages of having variable-level information (on the same metric) about the extent of nonresponse bias on specific variables, and lend further support to the findings reported here relating to the impact of adjusting on the auxiliary variables (namely, that adjustment has a mixed and not altogether positive effect on bias in the survey variables, and hence the magnitude of R-indicators).

While our findings are informative about some of the potential drawbacks of using R-indicators, it is important to recognise the limitations of the case study presented. As already alluded to, a principal concern is that our conclusions are sensitive to the methodological limits of the NRFU survey used to estimate bias (namely, nonresponse and differential measurement errors in the reduced sample). We treated the latter as though it was the complete sample, but our results suggest that the assumption that the NRFU respondents are representative of all the non-respondents to the ESS may not hold. While, the R-indicator for the reduced sample was substantially higher than for the complete sample, and the value for the CV and the MAC was substantially lower, there is evidence to suggest that the respondents to the NRFU were more similar to the respondents to the main survey than they were to the persistent nonrespondents, and that consequently, bias would be underestimated by the nonresponse survey. These findings are in line with previous research using these data. [Roberts et al. \(2014a\)](#) analysis found that the NRFU survey in the ESS Round 5 was successful in bringing into the overall responding sample more people from urban areas, from the French-speaking region of Switzerland, and without an available telephone number, as well as in balancing the different age categories. However, they found that it failed to improve the representation of non-Swiss citizens and the unmarried population. In this respect, it is perhaps not surprising that the nonresponse survey underestimates nonresponse bias on certain variables. It implies, however, that our analysis of the utility of the available auxiliary variables for detecting bias was restricted to a somewhat peculiar target population (of main survey and NRFU respondents), which may arguably somewhat limit the validity of our conclusions.

Another factor alluded to that is likely to hinder comparisons between the main survey and the nonresponse survey concerns the possibility that survey participation propensity and measurement error are interrelated. A large literature exists on whether reluctant respondents (refusals or hard-to-contact) are more likely to give inaccurate answers than motivated respondents (e.g. [Roberts et al. 2014b](#); [Peytchev et al. 2010](#); [Olson 2006](#); [Tancreto and Bentley 2005](#); [Yan et al. 2004](#)). In [Olson's \(2013\)](#) review of the published literature, she found considerable support for the conclusion that data from respondents recruited after many follow-ups or refusal conversion procedures were of lower quality.

However, it is not clear what to expect in terms of data quality for the NRFU surveys conducted for the purpose of nonresponse bias detection. On the one hand, there is a risk that respondents to the NRFU may be less motivated to answer and more inclined to reduce the effort to give accurate answers, which would result in bias due to measurement, and not necessarily due to selection effects. At the same time, the NRFU questionnaire is a lot shorter than the full interview, which should decrease response burden, and consequently, improve measurement quality. Similarly, it is not always clear whether and how mode differences in measurement will affect estimates. Given the lack of clarity on this matter, we consider the threat of persistent selection biases in the nonresponse bias estimates to be a greater cause for concern, but measurement bias should not be ignored. However, in the absence of alternative sources of information about the actual nonresponse bias on target variables in the ESS, we believe that the NRFU survey analysed here is still able to offer valuable insights. Nevertheless, it should be noted that the specificities of the design of this case study may restrict the possibility to extrapolate to other surveys, and so future studies should consider replicating our approach on other types of survey, with access to more and different auxiliary data, to see whether different conclusions are drawn.

Other methodological decisions we made might also have affected our conclusions, such as the use of logistic regression to estimate the response propensities, which may not produce the best propensity scores (e.g. [Olmos and Govindsamy 2015](#)). Alternative approaches such as Classification and Regression Tree models or Generalised Boosted Regression may outperform logistic regression in this regard ([Olmos and Govindsamy 2015](#); [McCaffrey et al. 2004](#)), and deserve consideration, particularly where large numbers of auxiliary variables are available. Similarly, the method of propensity score weighting we used is not the only way to adjust for nonresponse bias in the auxiliary data. We opted for both methods for pragmatic reasons and consistency, but also because of the limited number of auxiliary variables available from the sampling frame.

These caveats aside, our analyses underline the need to carefully consider how to select auxiliary variables and survey variables for nonresponse bias assessments, and the implications of this for interpreting R-indicators. Both influence conclusions drawn about sample representativeness and the risk of bias, and sometimes unpredictably. Nevertheless, we believe the findings of this study also highlight the considerable value to be gained from using the R-indicator to summarise bias in the auxiliary variables, given the additional information it contains, compared to the response rate. Combined with partial indicators ([Schouten and Shlomo 2017](#)), the R-indicator as a summary statistic has proven useful for monitoring and managing fieldwork progress, as well as for comparing the representativeness of different survey designs equipped with the same auxiliary variables ([Luiten and Schouten 2013](#); [Schouten et al. 2012](#)). Together, such indicators can be used to ensure that population subgroups are adequately represented in surveys to allow meaningful comparisons between subgroups or to highlight where adjustments should be made to ensure greater balance in the response sample. Given the implications of greater balance for bias on other survey variables may be less obvious, however, practitioners should be encouraged to fully assess the implications of alternative model specifications when estimating response propensities for conclusions about representativeness and bias risk.

## 6. References

- AAPOR, American Association for Public Opinion Research. 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition.
- Bethlehem, J.G. 2002. "Weighting Non-response Adjustment Based on Auxiliary Information." In *Survey Non-response*, edited by R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little, 41–54. New York: Wiley.
- Beullens, K. and G. Loosveldt. 2012. "Should High response Rates Really be the Primary Objective?" *Survey Practice* 5(3): 1–5. DOI: <https://doi.org/10.29115/SP-2012-0019>.
- Beullens, K., G. Loosveldt, C. Vandenplas, and I. Stoop. 2018. "Response Rates in the European Social Survey: Increasing, Decreasing, or a Matter of Fieldwork Efforts?" *Survey Methods: Insights from the Field*. DOI: <https://doi.org/10.13094/SMIF-2018-00003>.
- Brick, J.M. and M.E. Jones. 2008. "Propensity to respond and nonresponse bias." *METRON – International Journal of Statistics*, LXVI(1), 51–73.
- Brick, J.M. and D. Williams. 2013. "Explaining rising Non-response Rates in Cross-sectional Surveys." *The ANNALS of the American Academy of Political and Social Science* 645: 36–59. DOI: <https://doi.org/10.1177/0002716212456834>.
- Brick, J.M. and R. Tourangeau. 2017. "Responsive Survey Designs for Reducing Non-response Bias." *Journal of Official Statistics* 33(3): 735–752. DOI: <http://dx.doi.org/10.1515/JOS-2017-0034>.
- Cobben, F. 2009. *Non-response in Sample Surveys. Methods for Analysis and Adjustment*. PhD thesis, The Hague: Statistics Netherlands. Available at: <https://hdl.handle.net/11245/1.312964> (accessed June 2020).
- Cornesse, C. and M. Bosnjak. 2018. "Is there an association between survey characteristics and representativeness? A meta-analysis." *Survey Research Methods* 12(1): 1–13. DOI: <https://doi.org/10.18148/srm/2018.v12i1.7205>.
- De Heij, V., B. Schouten, and N. Shlomo. 2015. *RISQ manual 2.1. Tools in SAS and R for the computation of R-indicators, partial R-indicators and partial coefficients of variation*. RISQ Project. Available at: [www.risq-project.eu](http://www.risq-project.eu).
- De Leeuw, E. and W. de Heer. 2002. "Trends in Household Survey Non-response: A Longitudinal and International Comparison." In *Survey Non-response*, edited by R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little, 41–54. New York: Wiley.
- Ernst Stähli, M., D. Joye, M. Sapin, A. Pollien, M. Ochsner, and A. van den Hende. 2018. "Non Response Surveys (NRS): ESS 2006, EVS 2008, ESS 2010, MOSAiCH 2011, ESS 2012, ESS 2014." [Dataset]. Distributed by FORS, Lausanne. <https://doi.org/10.23662/FORS-DS-697-1>.
- ESS Round 5: European Social Survey 2016: ESS-5 2010 Documentation Report. Edition 4.1. Bergen, European Social Survey Data Archive, NSD – Norwegian Centre for Research Data for ESS ERIC. Available at: [https://www.europeansocialsurvey.org/docs/round5/survey/ESS5\\_data\\_documentation\\_report\\_e04\\_2.pdf](https://www.europeansocialsurvey.org/docs/round5/survey/ESS5_data_documentation_report_e04_2.pdf) (accessed June 2020).
- ESS Round 5: European Social Survey Round 5 Data. 2010. *Data file edition 3.4*. NSD – Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. Available at: <https://www.europeansocialsurvey.org/download.html?file=ESS5CH&c=CH&y=2010>.

- Groves, R.M. 2006 . “Non-response Rates and Non-response Bias in Household surveys.” *Public Opinion Quarterly* 70, Special Issue: 646–675. DOI: <https://doi.org/10.1093/poq/nfi033>.
- Groves, R.M. and S. Heeringa. 2006. “Responsive design for household surveys: tools for actively controlling survey errors and costs.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(3): 439–457. DOI: <https://doi.org/10.1111/j.1467-985X.2006.00423.x>
- Groves, R.M. and E. Peytcheva. 2008. “The Impact of Non-response Rates on Non-response Bias – A Meta-Analysis.” *Public Opinion Quarterly* 72: 167–189. DOI: <https://doi.org/10.1093/poq/nfn011>.
- Groves, R.M. and M.P. Couper. 1998. *Non-Response in Household Interview Survey*. New York: John Wiley & Sons.
- Groves, R.M., J.M. Brick, M.P. Couper, W. Kalsbeek, B. Harris-Kojetin, F. Kreuter, B.-E. Pennell, T. Raghunathan, B. Schouten, T. Smith, R. Tourangeau, A. Bowers, M. Jans, C. Kennedy, R. Levenstein, K. Olson, E. Peytcheva, S. Ziniel, and J. Wagner. 2008. “Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools.” *Survey Practice* 1(3): 1–6. DOI: <https://doi.org/10.29115/SP-2008-0013>.
- Gummer, T. and J.E. Blummenstiel. 2018. “Experimental Evidence on Reducing Nonresponse Bias through Case Prioritization: The Allocation of Interviewers.” *Field Methods* 30(2): 124–139. DOI: <https://doi.org/10.1177/1525822X18757967>.
- Kreuter, F. 2013. “Facing the Nonresponse Challenge.” *The Annals of the American Academy of Political and Social Science* 645(1): 23–35. DOI: <https://doi.org/10.1177/0002716212456815>.
- Kreuter, F. and K. Olson. 2011. “Multiple Auxiliary Variables in Nonresponse Adjustment.” *Sociological Methods & Research* 40(2): 311–332. DOI: <https://doi.org/10.1177/0049124111400042>.
- Lin, I.-F. and N.C. Schaeffer. 1995. “Using Survey Participants to Estimate the Impact of Nonparticipation.” *Public Opinion Quarterly* 59(2): 236–258, DOI: <https://doi.org/10.1086/269471>.
- Little, R.J.A. and S. Vartivarian. 2005. “Does Weighting for Non-Response increase the Variance of Survey Means?” *Survey Methodology* 31(2): 161–68. DOI: <https://doi.org/10.2307/1403140>.
- Little, R.J.A. 1986. “Survey Nonresponse Adjustments for Estimates of Means.” *International Statistical Review / Revue Internationale de Statistique* 54(2), 139–157. JSTOR. DOI: <https://doi.org/10.2307/1403140>.
- Little, R.J.A. and D.B. Rubin. 2014. *Statistical Analysis with Missing Data*. London, UK: John Wiley & Sons.
- Luiten, A. and B. Schouten, 2013. “Tailored Fieldwork Design to increase Representative Household Survey Response: An Experiment in the Survey of Consumer Satisfaction.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176(1): 169–189. DOI: <https://doi.org/10.1111/j.1467-985X.2012.01080.x>.
- Matsuo, H., J. Billiet, G. Loosveldt, F. Berglund, and Ø. Kleven. 2010. “Measurement and Adjustment of Non-response Bias based on Non-response Survey: the Case of Belgium



- and Norway in the European Social Survey Round 3.” *Survey Research Methods* 4(3): 165–178. DOI: <http://dx.doi.org/10.18148/srm/2010.v4i3.3774>.
- McCaffrey, D.F., G. Ridgeway, and A.R. Morral. 2004. “Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies.” *Psychological Methods* 9: 403–425. DOI: <https://doi.org/10.1037/1082-989X.9.4.403>.
- Moore, J.C., G.B. Durrant, and P.W. Smith. 2018. “Data Set Representativeness during Data Collection in Three UK Social Surveys: Generalizability and the Effects of Auxiliary Covariate Choice.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181(1): 229–248. DOI: <https://doi.org/10.1111/rssa.12256>.
- Nagelkerke, N.J. 1991. “A Note on a General Definition of the Coefficient of Determination.” *Biometrika* 78(3): 691–692. DOI: <https://doi.org/10.1093/biomet/78.3.691>.
- Nishimura, R., J. Wagner, and M. Elliott. 2016. “Alternative Indicators for the Risk of Non-response Bias: A Simulation Study.” *International Statistical Review* 84(1): 43–62. DOI: <https://doi.org/10.1111/insr.12100>.
- Olmos, A. and P. Govindsamy. 2015. “A Practical Guide for using Propensity Score Weighting in R.” *Practical Assessment, Research & Evaluation*, 20(13). Available at: <http://pareonline.net/getvn.asp?v=20&n=13>. (accessed June 2020).
- Olson, K. 2006. “Survey participation, Non-Response Bias, Measurement Error Bias, and Total Bias.” *Public Opinion Quarterly* 70(5): 737–758. DOI: <https://doi.org/10.1093/poq/nfl038>.
- Olson, K. 2013. “Do Non-response Follow-ups Improve or Reduce Data Quality?: A Review of the Existing Literature.” *Journal of the Royal Statistical Society: Series A: Statistics in Society* 176(1): 129–145. DOI: <http://doi.org/10.1111/j.1467-985X.2012.01042.x>.
- Peytchev, A., E. Peytcheva, and R.M. Groves. 2010. “Measurement Error, Unit Nonresponse, and Self-Reports of Abortion Experiences.” *Public Opinion Quarterly* 74: 319–327. DOI: <https://doi.org/10.1093/poq/nfq002>.
- Peytcheva, E. and R.M. Groves. 2009. “Using Variation in Response Rates of Demographic Subgroups as Evidence of Non-response Bias in Survey Estimates.” *Journal of Official Statistics* 25: 193–201. Available at: [https://www.researchgate.net/publication/282119961\\_Using\\_Variation\\_in\\_Response\\_Rates\\_of\\_Demographic\\_Subgroups\\_as\\_Evidence\\_of\\_Nonresponse\\_Bias\\_in\\_Survey\\_Estimates](https://www.researchgate.net/publication/282119961_Using_Variation_in_Response_Rates_of_Demographic_Subgroups_as_Evidence_of_Nonresponse_Bias_in_Survey_Estimates) (accessed June 2020).
- Roberts, C. and C. Vandenplas. 2017. “Estimating Components of Mean Squared Error to Evaluate the Benefits of Mixing Data Collection Modes.” *Journal of Official Statistics* 33(2): 303–334. DOI: <https://doi.org/10.1515/jos-2017-0016>.
- Roberts, C., C. Vandenplas, and M. Ernst Stähli. 2014a. “Using Register Data to assess the Impact of Response Enhancement Methods on the Risk of Non-response Bias.” *Survey Research Methods* 8(2): 67–80. DOI: <http://dx.doi.org/10.18148/srm/2014.v8i2.5459>.
- Roberts, C., N. Allum, and P. Sturgis. 2014b. “Non-response and Measurement error in an online panel: Does additional Effort to Recruit Reluctant Respondents Result in Poorer Quality Data?” In *Online Panel Research: A Data Quality Perspective*, edited by M. Callegaro, R. Baker, J. Bethlehem, A.S. Göritz, J.A. Krosnick, and P.J. Lavrakas. Hoboken: Wiley, Survey Methodology Series.

- Sakshaug, J.F. and M. Antoni. 2018. "Evaluating the Utility of Indirectly Linked Federal Administrative Records for Nonresponse Bias Adjustment." *Journal of Survey Statistics and Methodology* 7(2): 227–249. DOI: <https://doi.org/10.1093/jssam/smy009>.
- Särndal, C.-E. and S. Lundström. 2010. "Design for Estimation: Identifying Auxiliary Vectors to Reduce Nonresponse Bias." *Survey Methodology* 36(2): 131–144.
- Schouten, B. 2018. "Statistical inference based on randomly generated auxiliary variables." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80(1): 33–56. DOI: <https://doi.org/10.1111/rssb.12242>.
- Schouten, B. and N. Shlomo. 2017. "Selecting Adaptive Survey Design Strata with Partial R-indicators." *International Statistical Review* 85(1): 143–163. DOI: <https://doi.org/10.1111/insr.12159>.
- Schouten, B. and F. Cobben. 2007. "R-Indexes for the Comparison of Different Fieldwork Strategies and Data Collection Modes." *Statistics Netherlands. Discussion Paper*, 07002, CBC, Voorburg. Available at: <http://hummedia.manchester.ac.uk/institutes/cmist/risq/schouten-cobben-2007-a.pdf> (accessed June 2020).
- Schouten, B., F. Cobben, and J. Bethlehem. 2009. "Indicators for the Representativeness of Survey Response." *Survey Methodology* 35(1): 101–113. Available at: <https://pdfs.semanticscholar.org/aa59/4bf03a7cc219ccc6da01d1e3cb14a125d67a.pdf> (accessed June 2020).
- Schouten, B., F. Cobben, P. Lundquist, and J. Wagner. 2016. "Does more Balanced Response imply less Non-response Bias?" *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 179(3): 727–748. DOI: <https://doi.org/10.1111/rssa.12152>.
- Schouten, B., J. Bethlehem, K. Beullens, Ø. Kleven, G. Loosveldt, A. Luiten, K. Rutar, N. Shlomo, and C. Skinner. 2012. "Evaluating, Comparing, Monitoring, and Improving Representativeness of Survey Response Through R-Indicators and Partial R-Indicators." *International Statistical Review* 80(3): 382–399. DOI: <https://doi.org/10.1111/j.1751-5823.2012.00189.x>.
- Schouten, B., N. Shlomo, and C. Skinner. 2010. "Indicators for Representative Response." Paper presented at Quality in Official Statistics Conference 2010, Helsinki, Finland. Available at: <https://q2010.stat.fi/> (accessed June 2020).
- Schouten, B., N. Shlomo, and C. Skinner. 2011. "Indicators for Monitoring and Improving Representativeness of Response." *Journal of Official Statistics* 27(2): 231–253. Available at: <https://www.scb.se/contentassets/ca21efb41fee47d293bbe55bf7be7fb3/indicators-for-monitoring-and-improving-representativeness-of-response.pdf> (accessed June 2020).
- Stoop, I.A. 2004. "Surveying Nonrespondents." *Field Methods* 16(1): 23–54. DOI: <https://doi.org/10.1177/1525822X03259479>.
- Stoop, I., J. Billiet, A. Koch, and R. Fitzgerald. 2010. *Improving Survey Response: Lessons Learned from the European Social Survey*. London, UK: John Wiley and Sons Ltd.
- Tancreto, J.G. and M. Bentley. 2005. "Determining the Effectiveness of Multiple Non-response Follow-up Contact Attempts on Response and Data Quality." In Proceedings of the Section on Survey Research Methods: American Statistical Association, 2005. 3626–3632. Minneapolis, MN: American Statistical Association. Available at: <http://www.asasrms.org/Proceedings/y2005f.html> (accessed June 2020).

- Vandenplas, C., D. Joye, M. Ernst Stähli, and A. Pollien. 2015. “Identifying Pertinent Variables for Nonresponse Follow-Up Surveys. Lessons Learned from 4 Cases in Switzerland.” *Survey Research Methods* 9(3): 141–158. DOI: <https://doi.org/10.18148/srm/2015.v9i3.6138>.
- Voogt, R.J.J. and W.E. Saris, 2005. “Mixed Mode Designs: Finding the Balance Between Non-response Bias and Mode Effects.” *Journal of Official Statistics*, 21(3): 367–387. Available at: <https://www.scb.se/contentassets/ca21efb41fee47d293bbec5bf7be7fb3/-mixed-mode-designs-finding-the-balance-between-nonresponse-bias-and-mode-effects.pdf> (accessed June 2020).
- Wagner, J. 2012. “A Comparison of Alternative Indicators for the Risk of Nonresponse Bias.” *Public Opinion Quarterly* 76(3): 555–575. DOI: <https://doi.org/10.1093/poq/nfs032>.
- Williams, D. and J.M. Brick. 2017. “Trends in U.S. Face-To-Face Household Survey Nonresponse and Level of Effort.” *Journal of Survey Statistics and Methodology* 6(2): 186–211. DOI: <https://doi.org/10.1093/jssam/smx019>.
- Yan, T., R. Tourangeau, and Z. Arens. 2004. “When less is more: Are reluctant respondents poor reporters?” In *Proceedings of the Section on Survey Research Methods: American Statistical Association*, 2004. 4632–4651. Toronto: American Statistical Association. Available at: <http://www.asasrms.org/Proceedings/y2004/files/Jsm2004-000169.pdf> (accessed July 2020).

Received July 2018

Revised March 2020

Accepted May 2020