UvA-DARE (Digital Academic Repository)

Who Gets Lost, and What Difference Does It Make? Mixed Modes, Nonresponse Follow-up Surveys and the Estimation of Turnout

Goldberg, A.C.; Sciarini, P.

DOI

[10.1093/jssam/smy025](https://doi.org/10.1093/jssam/smy025)

Publication date

2019

Document Version

Final published version

Published in

Journal of Survey Statistics and Methodology

[Link to publication](#)

Citation for published version (APA):

Goldberg, A. C., & Sciarini, P. (2019). Who Gets Lost, and What Difference Does It Make? Mixed Modes, Nonresponse Follow-up Surveys and the Estimation of Turnout. *Journal of Survey Statistics and Methodology*, 7(4), 520-544. <https://doi.org/10.1093/jssam/smy025>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

WHO GETS LOST, AND WHAT DIFFERENCE DOES IT MAKE? MIXED MODES, NONRESPONSE FOLLOW-UP SURVEYS AND THE ESTIMATION OF TURNOUT

ANDREAS C. GOLDBERG*
PASCAL SCIARINI

This article assesses whether—and to what extent—turnout bias in post-election surveys is reduced by adding a short nonresponse follow-up (NRFU) survey to a mixed-mode survey. Specifically, we examine how the NRFU survey influences response propensities across demographic groups and political factors and whether this affects data quality. We use a rich dataset on validated voter turnout data, collected across two different ballots. In addition to the main survey that comprises computer-assisted telephone interviews (CATI) and web respondents, both studies include a short follow-up mail survey for nonrespondents. The results demonstrate that collecting extra information from additional respondents on so-called “central” questions is worth the effort. In both studies, the NRFU survey substantially increases representativeness with respect to sociodemographic and participation variables. In particular, voters and politically active citizens are more accurately represented in the NRFU survey. This tends to result in better estimates of turnout

ANDREAS C. GOLDBERG is postdoctoral researcher at the Amsterdam School of Communication Research, University of Amsterdam, Postbus 15791, 1001 NG Amsterdam, the Netherlands. PASCAL SCIARINI is professor of Swiss and comparative politics with the Department of Political Science and International Relations, University of Geneva, Switzerland. Earlier versions of this paper were presented at the ECPR general conference in Oslo (2017/09/06–2017/09/09) and the conference of the European Survey Research Association in Lisbon (2017/07/17–2017/07/21). The authors thank the panel participants and the anonymous reviewers and editor of this journal for their constructive comments. We are grateful to Simon Hug for statistical advice and Nicolas Pekari and Fabio Wasserfallen for helpful remarks. We further thank the Service of Elections and Popular Votes (SEV) and the Statistical Cantonal Office (OCSTAT) of the Geneva state for providing us with the validated turnout data and SELECTS (Swiss electoral studies) for their support in linking the data to their survey. Part of the research time of A.C. Goldberg used for this project was funded by a Grant from the European Research Council (ERC), Grant No. 647316.

*Address correspondence to Andreas C. Goldberg, Amsterdam School of Communication Research, University of Amsterdam Postbus 15791, 1001 NG Amsterdam; E-mail: A.C. Goldberg@uva.nl.

determinants in the final (combined) sample than is seen from CATI/web respondents only. Moreover, the increase in response rate and the decrease in nonresponse bias comes at almost no price in terms of measurement errors. Vote overreporting is only slightly higher in the mail follow-up survey than in the main CATI/web survey.

KEYWORDS: Misreporting; Mixed mode; Nonresponse follow-up; Survey nonresponse; Turnout bias.

1. INTRODUCTION

Mixed-mode surveys have become increasingly popular. In a context of declining response rates and growing concerns about coverage, mixed-mode designs aim “to combine the best of all possible worlds by exploiting the advantages of different modes to compensate for their weaknesses” (de Leeuw 2010, p. 1). Yet the question of whether and to what extent mixed-mode designs keep such promises is far from settled. On the one hand, the meta-analysis by Medway and Fulton (2012) shows that offering concurrent mode choices results in lower (not higher) response rates. Moreover, using more than one mode does not necessarily decrease nonresponse errors, since the same groups may be overrepresented in each mode (Couper 2011). Even worse, combining different survey modes can lead to different kinds of measurement errors (Hox, de Leeuw, and Zijlmans 2015). On the other hand, however, results are more promising with respect to sequential mixed-mode designs (Couper 2011, but see Sakshaug and Eckman 2017). Adding a follow-up survey that uses a different mode to the main survey may lead to a higher response rate and lower nonresponse bias (Bandilla, Couper, and Kaczmarek 2014) and increase response quality (Voogt and Saris 2005).

In electoral research, studies that investigate the influence that combining modes has on response propensity and response equality are scarce (Atkeson, Adams, and Alvarez 2014). Electoral participation is a major variable of interest in postelection surveys, yet it is systematically overestimated because of voter overrepresentation (nonresponse bias) and vote overreporting (measurement error) (Selb and Munzert 2013). The aim of this article is to assess whether and to what extent turnout bias in postelection surveys can be reduced by adding a short mail nonresponse follow-up (NRFU) survey to a mixed-mode (CATI/web) survey. We make four contributions.

First, the article adds to the small extant number of studies that explore demographic differences between respondents of various modes. Our initial sample covers the whole target population and thus includes the true population characteristics. Therefore, we are able to assess the representativeness of CATI/web respondents for several sociodemographic variables and evaluate how the NRFU survey improves representativeness. Second, our dataset provides information on validated vote for all sampled citizens—respondents

and nonrespondents alike. This allows for a careful examination of voter overrepresentation and, more importantly, enables us to evaluate whether and to what extent the follow-up survey helps to reduce nonresponse bias. Third, we can additionally test whether the expected increase in response rate brought about by the follow-up survey leads to higher (or lower) vote overreporting and, hence, influences data quality. Finally, we assess data quality through the analysis of turnout determinants, i.e. through evaluating whether models including the NRFU survey lead to more accurate estimates of electoral participation.

2. MIXED-MODE DESIGNS, NONRESPONSE FOLLOW-UP, AND SURVEY ERRORS

Survey participation is a function of coverage, the accessibility of targeted persons (contact), and the ability and willingness of contacted persons to cooperate (Quinn, Gutek and Walsh 1980; Smith 1983; Groves and Mathiowetz 1984; Goyder 1987; Groves, Cialdini, and Couper 1992; Stoop, 2005). Survey modes influence participation in the survey and sample representativeness through a selection process. A selection process exists when the probability that certain subgroups will participate in the survey is not equal across different interview modes (Voogt and Saris 2005, p. 368).

The enduring decline of response rates has become a major concern in survey research (Singer 2006; Brick 2013). First, the rise of mobile phones and answering machines has reduced coverage for traditional telephone (CATI) surveys. Computer-assisted telephone-interview surveys miss those groups without a landline number, e.g. the so-called “mobile generation” of people between 25 and 40 years and people with changing social relationships such as separation and divorce (Keeter 2007; Carley-Baxter, Peytchev, and Black 2010). Coverage problems also affect internet (web) surveys, which still underrepresent older people and overrepresent younger people, males, and highly educated persons (Stern and Dillman 2006; Zickuhr and Smith 2012; Lipps and Pekari 2016). Second, internet surveys tend to suffer from high nonresponse rates (Messer, Edwards, and Dillman 2012; Stern, Bilgen, and Dillman 2014; Roberts, Joye, and Ernst Stähli 2016), whereas telephone surveys have faced increasing refusal rates (de Leeuw 2008).

By using multiple methods to contact citizens and collect information, mixed-mode surveys offer a possible solution to these problems (Atkeson et al. 2014; de Leeuw 2005; Dillman, Smyth, and Christian 2014). The idea is that combining multiple modes decreases the specific drawbacks of each particular mode, compensating for them with the strengths of another mode (Couper 2011). This means better and more complete coverage but also the hope of an increased response rate and a reduction of nonresponse errors by reaching different subgroups of sample members (Dillman et al. 2009; de Leeuw 2010; Couper 2011; Hox et al. 2015).

The various possible combinations of modes alongside their related timings means it is not useful to talk about *one* mixed-mode design. De Leeuw (2005) distinguishes concurrent from sequential mixed modes of data collection. A concurrent design uses different modes in parallel, giving respondents the choice of which mode they prefer. A sequential design, meanwhile, uses different modes one after the other, typically one mode for the main data collection phase and a second mode for the nonresponse follow-up. A specific type of nonresponse follow-up offers nonrespondents the possibility of answering only a few questions central to the study while still altering the mode (e.g., Voogt and Saris 2005; Sakshaug and Eckman 2017). Overall, mixed-mode designs—independent of the specific mix at stake—are expected to increase the representativeness of the realized sample.

Among the different possible designs, sequential mixed-mode surveys especially are said to help to maximize the potentialities of mixing modes (Voogt and Saris 2005; Holmberg, Lorenc, and Werner 2010; Couper 2011; Messer and Dillman 2011; Bandilla et al. 2014; Stern et al. 2014; Roberts et al. 2016). Comparing the representativeness of an internet survey with an internet survey that added a follow-up mail survey, Bandilla et al. (2014) find that the sequential mixed-mode design increases response rates and reduces the selection bias arising from noncoverage and nonresponse. As a consequence, the authors' final (combined) sample is more in line with the population in terms of sociodemographic and attitudinal variables. However, other studies find that mixing modes in a sequential way increases demographic biases, since the same groups overrepresented in a given mode are also overrepresented in the other mode (e.g., Link and Mokdad 2006).

Generally, though, only a few studies can test representativeness in a strict sense using register data as we do in this article. For those few studies that also focus on the effects of NRFU surveys, the results are sometimes in line with the encouraging findings of sequential mixed modes in general, but not always. Two studies from the Netherlands report an increase in representativeness thanks to mixed modes, including NRFU surveys, one regarding demographic and socioeconomic characteristics (Schouten 2007) and the other regarding turnout (Voogt and Saris 2005). Yet a more recent study finds increased representativeness for some characteristics, i.e. employment measures, but not for others like demographic items (Sakshaug and Eckman 2017). The contradictory findings regarding the effect of NRFU surveys on demographic representativeness call for more research.

A common feature of sequential mixed modes—particularly NRFU surveys—is a higher overall response rate after adding a second or third mode. However, one should not equate higher response rates with lower nonresponse bias. Nonresponse bias is not a function of the response rate but a function of the interaction between response and respondents' characteristics. Empirical findings from survey research in various fields show that a low response rate does not necessarily lead to high nonresponse bias (e.g., Curtin, Presser, and Singer 2000; Groves, Presser, and Dipko 2004; Groves 2006; Groves and Peytcheva 2008).

The relationship between response rates and nonresponse bias is also disputed in the literature on turnout bias. Some argue that increasing the response rate helps to increase the share of nonvoters and, therefore, reduces voter overrepresentation in postelection surveys (e.g., [Burden 2000](#)). Others, however, object that unless the response rate approaches 100 percent, an increase in response rate may in fact bring an even more disproportionate share of active voters into the sample ([Lin and Schaeffer 1995](#); [Selb and Munzert 2013](#)).

In addition to influencing nonresponse errors, mixing modes may also affect the quality of the data through a change in the accuracy of responses. Nonresponse follow-up surveys are no exception. The quality of the data is affected whenever “the probability that a respondent will report the true value of a certain item in the questionnaire is different for different modes of interview” ([Voogt and Saris 2005](#), p. 368). If survey mode influences respondents’ answers, that is, if there are “measurement mode effects” ([Hox et al. 2015](#)), this may threaten the data’s validity. This would mean that differences in the data do not reflect actual differences in the population but rather only differences resulting from different survey modes.

Considering potential complications in terms of measurement errors ([Couper 2011](#)), the question is whether mixing different modes of data collection will lead to less (or more) biased data due to measurement error. In this context, it has been argued that mode effects are strongest on sensitive issues eliciting social desirability ([de Leeuw 1992](#); [Tourangeau et al. 2013](#)). Social desirability leads respondents to overreport good behaviors and underreport bad ones. This is most obvious and well-documented with regard to voter turnout (e.g., [Deufel and Kedar 2010](#); [Holbrook and Krosnick 2010](#); [Tourangeau et al. 2010](#); [Sciarini and Goldberg 2016](#)). Extant research shows that social desirability effects are stronger in telephone surveys and even more so in face-to-face surveys when an interviewer is present compared with self-administered (online or mail) surveys ([Atkeson et al. 2014](#); [Chang and Krosnick 2009](#)).

[Voogt and Saris \(2005\)](#) come to a similar conclusion in their study on turnout bias. Based on validated vote data from the 1998 Dutch national elections in one municipality, they compare three mode collections (telephone, mail and face-to-face) and additionally include a NRFU survey focusing on “central questions.” In addition to the reduction in nonresponse bias in estimates of voter turnout previously mentioned, their results also suggest that the increase in response rate does not come at the price of response quality. Quite the contrary, results suggest that offering nonrespondents the short questionnaire leads to a reported voter turnout estimate closer to the actual turnout in the population ([Voogt and Saris 2005](#)). However, in a different context (that of economic survey), a similar NRFU study by [Sakshaug and Eckman \(2017\)](#) leads to another paradoxical finding: adding the NRFU respondents to the main survey respondents results in greater bias in specific survey estimates despite reducing both nonresponse and measurement error bias separately. This may occur in the case of unequal and large counteracting biases.

2.1 The Present Study

The bulk of the literature on turnout bias in postelection surveys has focused on vote overreporting, i.e. the measurement error raised by respondents claiming they voted when they did not (Deufel and Kedar 2010; Ansolabehere and Hersh 2012). Yet recent work shows that voter overrepresentation among survey respondents has been on the rise and now contributes to an even greater extent to the overestimation of turnout than overreporting (Burden 2000; Berent, Krosnick, and Lupia 2011; Sciarini and Goldberg 2016, 2017). This means that the nonresponse process is more relevant for the correct estimation of turnout in surveys. Our goal in this study is to examine whether a sequential mixed-mode design helps to decrease turnout bias. More specifically, we first assess whether and to what extent the short mail NRFU survey helps to reduce nonresponse bias and increases the representativeness of the survey in terms of sociodemographic characteristics and voter (over)representation. Following the previously discussed virtues of sequential mixed-mode designs, we formulate two hypotheses:

H1: Adding an NRFU survey among initial nonrespondents decreases voter overrepresentation in the final sample.

H2: Adding an NRFU survey among initial nonrespondents improves the general representativeness of the final sample.

Expanding existing research by Voogt and Saris (2005), we then analyze whether the additional NRFU survey results in a better estimation of the *determinants* of turnout, one of the ultimate goals of postelection surveys. As we are able to use validated vote data, our analysis does not face the usual problem of separating measurement errors from survey design effects. The corresponding hypothesis is as follows:

H3: Combining respondents from both the main survey and the NRFU survey results in more accurate estimates of the determinants of turnout.

In a complementary way, we also examine whether the attempt to reduce voter overrepresentation goes along with an increase in vote overreporting. Given the lack of clear theoretical expectations, we formulate the following research question:

RQ1: Does the expected reduction of voter overrepresentation through the NRFU survey come at the price of an increasing vote overreporting?

3. DATA AND METHODS

We rely on data from two similar postelection surveys conducted in the canton of Geneva, Switzerland. The first took place in 2012 after a compulsory referendum on the new cantonal constitution; the second took place in 2015 after the national elections. Both surveys use the official vote register of the Geneva

canton (more than 220,000 citizens) as the same sampling frame. Since the official vote register equals our target population (all eligible voters are automatically registered there), we achieve a 100 percent coverage and do not have to consider this source of error. Both surveys were announced with a letter sent to all sampled citizens one week before the election day. While the letter was signed by university professors, it mentioned that the Geneva state government sponsored the survey.

For the 2012 survey, we drew a random sample of 2,500 people. We relied on a CATI survey (87 percent of all respondents), complemented with a web survey (13 percent) for citizens without any telephone number and for citizens with whom no phone contact could be established during the first ten days of fieldwork.¹ This main phase of data collection achieved a response rate of 49 percent (AAPOR RR1). In 2015, we drew a random sample of 1,250 people. We relied on a web survey (83 percent of all respondents), complemented with CATI (17 percent). While the CATI survey officially started two weeks after the election day, some persons ($n = 13$) used the hotline offered in the invitation letter to fill in the questionnaire by phone before the official CATI start. The response rate amounted to 45 percent (AAPOR RR1).

Thus in both years, the main part of data collection was a mix between a sequential and a concurrent mixed-mode design. Data collection started with one main mode in the first ten to fourteen days of the fieldwork; we subsequently offered the second mode in parallel. We then complemented the main survey part with a short mail nonresponse follow-up survey. Nonrespondents received the mail questionnaire and a prestamped return envelope in the fifth week after election day. We excluded nonrespondents with an invalid postal address and nonrespondents who explicitly requested not to be contacted again or who were unable to answer (e.g., due to health issues). The mail questionnaire focused on a few central questions: level of education, professional status, state of health, interest in politics, trust in political institutions, and party preference. The 2015 survey additionally included a question about participation in the national elections. In both years, the response rate of the NRFU survey amounted to about 36 percent.

Given both the nonpure sequential character of the design in the main survey and the small number of questionnaires filled with the second alternative mode (2012: $n = 164$; 2015: $n = 97$), we refrain from analyzing the differences between the CATI and web surveys. We concentrate instead on the differences between the main survey and the NRFU survey.

1. Sampled citizens without a phone number received a letter asking for their phone number three weeks before the start of the field period. People who provided a phone number received the standard CATI invitation letter one week prior to election day. Persons who did not provide their phone number received a second letter offering them the possibility of calling a hotline to set up an appointment (or do the interview straight away) or to fill the questionnaire online through the internet. The possibility to fill the questionnaire online was also offered (via mail) to CATI non-contacts after ten days.

Both studies use highly reliable information on validated vote. Validation was achieved through an electronic reading of the voting card that each voter had to sign and send back by postal mail or put in the ballot box. The cantonal administration added this information to our two survey datasets. We thus have information about both self-reported vote as taken from the survey and validated vote from the official vote register. Moreover, given that validated vote data is available not only for survey respondents but also for nonrespondents, we can separate the two sources of turnout bias, namely nonresponse bias and vote measurement error.

Next, the survey data was merged with a larger dataset of official turnout data covering the whole cantonal electorate and including information on a handful of sociodemographic variables (sex, age, civil status, residence duration, and citizenship status).² Thus, we also have information about citizens' actual participation (or abstention) in all previous ballots held in the canton of Geneva since 1996—up to eighty referendums or elections on the national or cantonal level. We calculate citizens' past participation record by dividing the number of votes in which a respondent has participated by the total number of votes in which this person could have participated. Importantly, the sociodemographic information and vote history in the official dataset are available for the whole electorate, which obviously includes survey respondents and nonrespondents.

To address the issue of “who gets lost” and to assess the representativeness of the realized sample, we first compare the sociodemographic characteristics of the population to those of the two subsamples of CATI/web and NRFU respondents separately and combined. In addition to sociodemographic characteristics, we also compare information on (past) participation of sample members. For those descriptive analyses, we calculate one sample *t* tests. Yet as several of our variables may covary, e.g. young people are more often single and also vote less often than older respondents, we further examine the effects of sociodemographic and political variables on survey participation with multinomial logistic regression models using the final survey status, i.e. the type of (non)respondent, as dependent variable.

We then turn to “what difference does it make,” analyzing the effects of nonresponse bias. Using the initial sample as a baseline, we analyze the size and correlates of voter overrepresentation among the main survey respondents and the respondents to the follow-up questionnaire. We use logistic regression

2. We used this merging for a cross-check between the survey data and the official data. When checking the validity of sociodemographic variables, we detected some irregularities in the 2015 dataset where the gender information did not match (i.e., sometimes the husband/wife may have answered for the actual sampled respondent). These cases ($n = 7$) were deleted from our dataset, also reducing the number of the initial sample in the following analyses for 2015 to $N = 1,243$. Citizenship status refers to a cantonal citizenship, which in Switzerland exists next to the national one. Geneva residents with citizenship of another canton, mostly due to being born there, can still vote in Geneva.

models with validated turnout as the dependent variable and log likelihood ratio tests for possible differences between the model coefficients. Finally, we take a cursory look at the differences in vote overreporting in the main CATI/web survey and the NRFU survey. This is only possible for the 2015 survey since only in that year did the NRFU survey include the turnout question. Note that all following analyses are unweighted unless explicitly mentioned for the robustness checks.

4. EVALUATION OF REPRESENTATIVENESS

To evaluate the representativeness of the realized sample, we first present descriptive statistics of the different (sub)samples. Columns four, five, and six in [tables 1](#) and [2](#) show the distribution of all available variables across our two types of respondent (CATI/web in the main phase of data collection; mail questionnaire in the NRFU phase) separately and combined. The tables further include values for the whole cantonal population. These are not estimates as in many other studies but rather are taken directly from the vote register. We also show values for the initial sample and values for nonrespondents. The figures regarding the initial sample slightly differ from those of the population purely due to random sampling. We compare the population characteristics to the characteristics of the main study, NRFU, and combined sample respondents using one sample *t* test (see the last three columns).

Let us start with the comparisons between the population and our two separate groups of respondents. Respondents to the main survey show a much more biased composition compared with NRFU respondents. This pattern is present in both years but particularly in 2012. For instance, the main study respondents are older, more often married, and less often single and have resided for a longer time in Geneva. The biggest differences in the population, though, is seen in the two participation variables. Not only have the main study respondents turned out significantly more often in the vote under consideration but their past participation record is also significantly higher. Two variables that display no significant differences in the population are gender and Genevan citizenship. In contrast to this biased pattern of main study respondents, the NRFU respondents are more representative of the population, and this is seen especially with respect to the two main variables of interest (i.e., past participation and vote). This high representativeness was not necessarily expected because the sample of initial nonrespondents is strongly biased—in the opposite direction to the bias of the main study respondents.

Given the minor differences between the NRFU respondents and the population and the large differences between the main study respondents and the population, we should witness an improvement and smaller differences for the combined sample (last column). Indeed, for almost all variables, the combined sample is less biased in the sense that the differences to the true values in the

Table 1. CATI/Web Main Survey and NRFU Survey: Representativeness Evaluations (2012)

Variable in %	Means/shares in population and different (survey) samples					Comparisons (t-tests)			
	Population N = 221880	Initial sample N = 2500	Main study respondents N = 1231	NRFU respondents N = 362	Combined sample N = 1593	Nonrespondents N = 907	Population – Main study respondents	Population – NRFU respondents	Population – Combined sample
Female	54.9	55.6	57.0	55.8	56.8	53.5	-2.1	-0.9	-1.9
Age (avg. in years)	50.4	50.8	51.6	51.6	51.6	49.4	-1.2 ^a	-1.2	-1.2 ^a
Marital status									
Single	32.6	32.1	28.1	28.7	28.3	38.9	4.5 ^a	3.9	4.3 ^a
Married	44.9	45.0	51.7	45.0	50.2	35.8	-6.8 ^a	-0.1	-5.3 ^a
Divorced	15.6	15.9	14.4	17.7	15.1	17.2	1.2	-2.1	0.5
Widowed	6.9	7.0	5.9	8.6	6.5	8.1	1.0	-1.7	0.4
Residence duration									
<10 years	9.3	8.0	6.8	7.7	7.0	9.7	2.5 ^a	1.6	2.3 ^a
10-20 years	18.1	16.9	15.4	15.8	15.5	19.3	2.7 ^a	2.3	2.6 ^a
>20 years	72.6	75.1	77.7	76.5	77.5	71.0	-5.1 ^a	-3.9	-4.9 ^a
GE citizen	59.4	59.4	59.1	62.7	59.9	58.7	0.3	-3.3	0.5
Past participation	44.8	45.2	55.0	44.6	52.6	32.3	-10.2 ^a	0.2	-7.8 ^a
Vote	32.6	33.2	44.6	30.1	41.3	19.1	-12.0 ^a	2.5	-8.7 ^a

^aStands for significant differences at least at the 0.05 level.

Table 2. CATI/Web Main Survey and NRFU Survey: Representativeness Evaluations (2015)

Variable in %	Means/shares in population and different (survey) samples						Comparisons (t-tests)		
	Population N = 227130	Initial sample N = 1243	Main study respondents N = 563	NRFU respondents N = 229	Combined sample N = 792	Nonrespondents N = 451	Population – Main study respondents	Population – NRFU respondents	Population – Combined sample
Female	54.8	55.4	57.5	51.5	55.8	54.8	–2.7	3.3	–1.0
Age (avg. in years)	50.7	50.2	49.6	52.5	50.4	49.7	1.1	–1.8	0.3
Marital status									
Single	34.2	32.6	29.8	28.4	29.4	38.1	4.4 ^a	5.8	4.8 ^a
Married	43.4	44.9	47.1	49.4	47.7	39.9	–3.7	–6.0	–4.3 ^a
Divorced	15.8	15.9	17.6	13.5	16.4	15.1	–1.8	2.3	–0.6
Widowed	6.6	6.6	5.5	8.7	6.4	6.9	1.1	–2.1	0.2
Residence duration									
< 10 years	10.1	9.1	8.2	9.2	8.5	10.2	1.9	0.9	1.6
10-20 years	16.5	17.6	16.7	17.5	16.9	18.8	–0.2	–1.0	–0.4
>20 years	73.4	73.3	75.1	73.4	74.6	71.0	–1.7	0.0	–1.2
GE citizen	60.5	58.4	57.4	60.7	58.3	58.5	3.1	–0.2	2.2
Past participation	44.3	43.7	53.2	43.5	50.4	32.0	–8.9 ^a	0.8	–6.1 ^a
Vote	44.7	44.3	59.5	41.5	54.3	26.8	–14.8 ^a	3.2	–9.6 ^a

^aStands for significant differences at least at the 0.05 level.

population are in most cases smaller compared with the main study sample. In a strict statistical sense, though, all differences that are significant in the main study sample are still significant in the combined sample. For marital status in 2015, the combination of both samples even results in a slightly larger bias, i.e. singles and married people are now more strongly under- and overrepresented, respectively. This exception notwithstanding, adding NRFU respondents to CATI/web respondents increases the representativeness of the sample. The correction is most obvious for the two vote variables. The counterbalancing effect brought about by NRFU respondents helps to reduce the overrepresentation of both ballot voters and respondents with a high past participation record by about 25 percent (in 2012) to about 30 percent (in 2015).

5. MULTIVARIATE TEST OF SURVEY PARTICIPATION

Table 3 puts the analysis of survey participation and of the related contribution of the NRFU survey to a stricter, multivariate test. For each year, the first two columns stem from a multinomial logistic regression contrasting main study respondents and NRFU respondents to nonrespondents. The third column comes from a binary logistic regression contrasting the combined sample of respondents (main study and NRFU) to nonrespondents. To account for the common nonlinear relationship between age and survey participation (e.g., low participation among the “mobile generation” and also among very old people due to health reasons), we also include age in squared and cubic form.

Vote at the ballot of interest and past participation record again stand out as strong predictors of survey participation. This confirms that political participation and survey participation are “two sides of the same coin” (Brehm 1993). The NRFU respondents also participate politically more than nonrespondents, but this effect is either not significant (vote) or smaller in size compared with the main study respondents (past participation). Follow-up respondents thus appear as an in-between group in that respect. We further see similar curvilinear age effects in both years; however, this is not significant for the NRFU survey. We also note similar effects for marital status, although not all reach statistical significance in both years and across our three (sub)samples. In 2012, the main study and combined sample also show an overrepresentation of women and long-term residents in comparison with nonrespondents, and the main study additionally shows an overrepresentation of non-Genevan citizens. None of these latter three effects reaches statistical significance in 2015, though all point in the same direction as in 2012.

In sum, both descriptive statistics and regression analyses demonstrate that main study respondents have a very specific profile, which strongly differs from that of the population and that of nonrespondents. This holds especially—but not only—for political activism. By contrast, NRFU respondents

Table 3. Estimation of Survey Participation in 2012 and 2015 (Multinomial and Binomial Regression Models)

	2012			2015		
	Main study respondents (ref. nonrespondents)	NRFU respondents	Combined sample (ref. nonrespond.)	Main study respondents (ref. nonrespondents)	NRFU respondents	Combined sample (ref. nonrespond.)
Woman	0.311*** (0.097)	0.146 (0.130)	0.270*** (0.091)	0.171 (0.141)	-0.148 (0.170)	0.070 (0.129)
Age	-0.180*** (0.051)	-0.031 (0.066)	-0.133*** (0.047)	-0.207** (0.080)	-0.116 (0.087)	-0.143** (0.069)
Age ²	0.004*** (0.001)	0.0008 (0.001)	0.003*** (0.0009)	0.005*** (0.002)	0.002 (0.002)	0.003** (0.001)
Age ³	-0.00003*** (0.000006)	-0.000007 (0.000007)	-0.00002*** (0.000006)	-0.00003*** (0.00001)	-0.00002 (0.00001)	-0.00002*** (0.000008)
Marital status (ref. single)						
Married	0.505*** (0.142)	0.389** (0.194)	0.475*** (0.133)	0.257 (0.205)	0.403 (0.252)	0.314* (0.187)
Divorced	0.159 (0.174)	0.333 (0.232)	0.209 (0.162)	0.673*** (0.257)	0.291 (0.325)	0.558** (0.237)
Widowed	0.587** (0.263)	0.683*** (0.334)	0.616** (0.246)	0.956** (0.398)	0.836* (0.437)	0.909** (0.359)
Residence duration (ref. > 20 years)						
10 – 20 years	-0.250* (0.134)	-0.207 (0.182)	-0.236* (0.125)	-0.250 (0.191)	-0.028 (0.233)	-0.171 (0.175)
<10 years	-0.375** (0.177)	-0.233 (0.239)	-0.341** (0.165)	-0.302 (0.249)	0.018 (0.295)	-0.196 (0.225)
Vote	0.481*** (0.129)	0.115 (0.175)	0.400*** (0.124)	0.730*** (0.179)	0.201 (0.222)	0.566*** (0.166)

Past participation	1.931*** (0.190)	1.178*** (0.253)	1.740*** (0.179)	2.033*** (0.306)	0.990*** (0.365)	1.674*** (0.280)
Non-Geneva	0.198** (0.098)	-0.057 (0.133)	0.133 (0.093)	0.093 (0.143)	-0.098 (0.173)	0.027 (0.131)
Constant	1.568** (0.779)	-1.181 (1.043)	1.327* (0.720)	2.062* (1.214)	0.514 (1.368)	1.773* (1.055)
Observations	2500		2500		1243	1243
Pseudo R ²	0.075		0.097		0.084	0.095

NOTE.—For each year, the first two columns represent the coefficients from a multinomial logistic regression using three respondent categories as dependent variable (main study respondents, NRFU respondents, and nonrespondents as reference category) and the third column (“combined”) the coefficients from a binary logistic regression contrasting the combined sample respondents to nonrespondents. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

differ far less from nonrespondents, and they are fairly representative of the population. Therefore, our results confirm our first two hypotheses. Adding an NRFU mail survey among initial nonrespondents not only decreases voter overrepresentation but also brings the realized sample further in line with the electorate more generally.

6. ANALYSIS OF TURNOUT

We now consider to what extent the systematic differences for some variables between types of (non)respondents matter for the analysis of turnout. Tables 4 and 5 show the results of logistic regression models with participation in the vote of interest as the binary dependent variable (validated vote, not self-reported). We again include age in its squared and cubic form to account for the nonlinear relationship between age and vote. The first model is calculated for the initial sample, which thus represents the “true” estimates and serves as the baseline model. The second model is calculated for main study (CATI/web) respondents, the third for NRFU respondents, and the fourth for the combined sample. The closer the coefficients of the latter three models get to the model of the initial sample, the better the representativeness of the survey samples and the more accurate the estimates in comparison with the “true” population values.

For some variables, the coefficients of the baseline model differ substantially between both years. While these differences may deserve further investigation, they are not the focus of this article. More interesting to our present purpose are the results regarding the two subgroups of respondents and their combination in comparison with the baseline model.

In 2012, the model for main study respondents shows some of the same significant effects as in the model regarding the initial sample, i.e. the nonlinear age effect and a negative effect for non-Genevan citizenship. By contrast, the effects of gender, marital status, and residence duration do not show up among the main study respondents. Since these nonsignificant effects may partly be due to the small number of respondents in this model, we ran a robustness check. We weighted the data to have the same number of observations as in the combined or initial sample. This robustness check does not modify the picture. Turning to the model of NRFU respondents, we see no significant effects except for women. Again, correcting for the difference in the number of observations does not increase the number of significant coefficients.

Finally, the results of the fourth combined model are very similar to those of the initial sample (with the exception of residence duration). Therefore, the combination of our initial main study and NRFU respondents results in better estimates of participation than when looking at estimates among the two groups separately. Particularly relevant from an empirical perspective is the improvement compared with the main study sample, since usually surveys would stop after this response phase. A more formal check of differences between models using log likelihood ratio tests confirms that the combined

Table 4. Estimation of Turnout in 2012

	DV: turnout			
	Initial sample	Main study respondents	NRFU respondents	Combined sample
Woman	−0.130** (0.055)	−0.103 (0.076)	−0.346** (0.152)	−0.146** (0.067)
Age	−0.135*** (0.029)	−0.097** (0.044)	−0.055 (0.077)	−0.107*** (0.036)
Age ²	0.003*** (0.001)	0.002*** (0.001)	0.002 (0.001)	0.003*** (0.001)
Age ³	−0.00002*** (0.000003)	−0.00001*** (0.000005)	−0.00001 (0.000008)	−0.00002*** (0.000004)
Marital status (ref. single)				
Married	0.075 (0.084)	−0.039 (0.114)	0.179 (0.244)	0.010 (0.103)
Divorced	−0.238** (0.102)	−0.216 (0.145)	−0.135 (0.285)	−0.219* (0.127)
Widowed	−0.277* (0.143)	−0.298 (0.206)	−0.526 (0.394)	−0.369** (0.179)
Residence duration (ref. > 20 years)				
10 – 20 years	−0.174** (0.082)	−0.122 (0.115)	0.081 (0.232)	−0.094 (0.102)
<10 years	0.061 (0.106)	0.242 (0.152)	−0.165 (0.318)	0.156 (0.134)
Non-Geneva	−0.170*** (0.056)	−0.223*** (0.077)	−0.136 (0.157)	−0.196*** (0.068)
Constant	0.781* (0.458)	0.703 (0.666)	−0.642 (1.259)	0.694 (0.565)
Observations	2,500	1,231	362	1,593
Pseudo R-squared	0.089	0.072	0.128	0.078

Standard errors in parentheses; ***p < 0.01, **p < 0.05, *p < 0.1.

sample leads to more accurate turnout estimations. That is, we find fewer significant differences between the coefficients from the combined sample and from the initial sample than between the coefficients from the main study sample and from the initial sample.³

3. For this we compared the coefficients of the initial sample with the coefficients from models that also include all persons from the initial sample but further comprise the response mode (*not* being in the respective sample, main study, or combined sample) and related interactions with the variables of interest. To overcome multicollinearity problems, we ran separate models for each variable of interest, and we calculated log likelihood ratio tests to compare the overall effect of our categorical variables. We did the same for the simple, squared, and cubed terms of the age variable.

Table 5. Estimation of Turnout in 2015

	DV: turnout			
	Initial sample	Main study respondents	NRFU respondents	Combined sample
Woman	-0.054 (0.076)	-0.212* (0.116)	0.138 (0.183)	-0.080 (0.095)
Age	-0.090** (0.041)	-0.068 (0.072)	-0.003 (0.095)	-0.059 (0.053)
Age ²	0.003*** (0.0008)	0.002 (0.002)	0.001 (0.002)	0.002** (0.001)
Age ³	-0.00002*** (0.000005)	-0.00001 (0.00001)	-0.00001 (0.00001)	-0.00002** (0.000007)
Marital status (ref. single)				
Married	0.212* (0.111)	0.066 (0.166)	-0.216 (0.261)	-0.041 (0.138)
Divorced	-0.199 (0.138)	-0.210 (0.202)	-0.467 (0.331)	-0.277 (0.170)
Widowed	0.118 (0.197)	-0.054 (0.317)	-0.187 (0.429)	-0.086 (0.250)
Residence duration (ref. > 20 years)				
10 – 20 years	-0.020 (0.105)	0.160 (0.159)	0.010 (0.266)	0.081 (0.133)
<10 years	-0.027 (0.135)	0.370* (0.215)	-0.515 (0.349)	0.068 (0.173)
Non-Geneva	0.036 (0.077)	-0.167 (0.117)	-0.346* (0.187)	-0.171* (0.097)
Constant	0.165 (0.644)	0.355 (1.076)	-1.278 (1.530)	-0.002 (0.816)
Observations	1,243	563	229	792
Pseudo R-squared	0.078	0.094	0.113	0.081

Standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The results for the 2015 data are less conclusive. Both subgroups of main study respondents and NRFU respondents show partly other significant effects than the baseline model (see the coefficients for women and residence duration in the main study and for non-Geneva citizenship in the NRFU survey). However, only the citizenship variable remains also weakly significant in the combined sample and is thus different from the initial sample. Conversely, being married has a significant effect in the initial sample but not in the combined sample. In contrast, the found age effect again appears when combining the two samples. Log likelihood ratio tests confirm that the combined model is closer to the initial sample for some variables but not for others. The reasons for these mixed results may be the smaller number of respondents in each

sample, which hinders the finding of significant effects in the first place, and the overall less significant turnout determinants in the initial sample.

In sum, we can partly confirm our third hypothesis (H3) because the determinants of turnout in 2012 indeed get closer to those in the initial sample if we combine the main study and NRFU samples. While the findings for 2015 are less conclusive, the results overall highlight the possible virtue of the NRFU questionnaire focusing on a small number of questions, which by including additional respondents into the survey, may improve the analysis of turnout.

7. EVALUATING TURNOUT BIAS THROUGH RESPONSE ACCURACY

Finally, we still need to check whether the better representativeness of our combined sample comes at the price of worse response accuracy with respect to the turnout question. For this last test, we can only use the 2015 data since in the 2012 survey the NRFU questionnaire does not include the turnout question. [Table 6](#) (third row) first confirms that in the main CATI/web sample nonresponse bias is high. While official turnout amounts to 44.2 percent in the initial sample, it is 15.9 percentage points higher among the combined main study respondents. By contrast, voter overreporting is very small (2.5 percent). Overall, then, voter overrepresentation accounts for the bulk of total turnout bias (four-fifths), which is even higher than in previous work ([Sciarini and Goldberg 2016, 2017](#)).

For this specific analysis, we also examine the CATI and web surveys separately. The results overall show a similar pattern for both survey modes but less bias among CATI respondents, the second mode offered during the main response phase.⁴ Interestingly, although social desirability is usually more problematic in interviewer-administered surveys, in our study measurement bias is smaller among CATI than web respondents.

More importantly, the fourth row of [table 6](#) shows there is no voter overrepresentation at all in the NRFU questionnaire. In fact, citizens who participated in the 2015 elections are slightly *underrepresented* among NRFU respondents (by 4.0 percentage points). However, it is important to remember that there was no random assignment into our different samples, i.e. part of the shown nonresponse bias may be due to self-selection into a response mode. The presented validated turnout in the initial sample of 44.2 percent is thus the correct comparison for the combined sample. However, the value would probably be slightly higher for main study respondents and slightly lower for NRFU

4. Excluding the thirteen CATI respondents who answered already before the official start of the CATI survey (see data description), the turnout bias in the CATI subsample is even smaller (10.0) and is due to a significantly lower nonresponse bias (8.8) and a roughly similar measurement bias (1.2). Hence, the thirteen respondents who actively asked to conduct the CATI survey before its official start are politically active citizens who partly blurred even bigger differences in turnout bias between the two main survey modes.

Table 6. Size of Overall Turnout Bias, Nonresponse Bias and Measurement Bias in the CATI/Web and NRFU Questionnaire in Percentage Points (2015)

	Validated turnout in the initial sample	Validated turnout among respondents	Reported turnout among respondents	<i>Total turnout bias</i>	<i>Bias estimates</i>	
					<i>Non response</i>	<i>Measurement</i>
Web (n = 455)		60.9	63.7	19.5	16.7	2.8
CATI (n = 96)		56.3	57.3	13.1	12.1	1.0
CATI/web (n = 551)	44.2 (n = 1,221)	60.1	62.6	18.4	15.9	2.5
NRFU respondents (n = 219)		40.2	46.6	2.4	-4.0	6.4
All respondents combined (n = 770)		54.4	58.1	13.9	10.2	3.7

NOTE.— Respondents who did not answer the turnout question are excluded from the table (twelve in the main survey and ten in the mail questionnaire).

respondents since voters tend to participate earlier and more voluntarily in postselection surveys.

Results are a little less straightforward with respect to vote overreporting, which is around four percentage points higher among NRFU respondents than among main study respondents (6.4 percent against 2.5 percent). Yet the magnitude of overreporting is extremely low in the main survey, and it is still fairly low in the NRFU survey. Moreover, a closer look at the data suggests that the share of overreporting in the NRFU survey might be due to the late timing and long field period of that survey. The follow-up survey remained open for twelve weeks (over the Christmas break), which means that some people sent us back their questionnaire up to four months after the election day. As [table A1](#) in the appendix shows, among citizens who answered in the first five weeks, the measurement error is 2.5 (i.e., exactly as high as among the main study respondents). By contrast, the measurement error is far higher (8.5) among persons who answered later (i.e., in the last seven weeks), which accounts for the higher measurement error between the main study and the NRFU survey.⁵ Reasons for the higher measurement error among later NRFU

5. In addition to measurement error, we also checked the composition of the NRFU survey depending on when people sent back their questionnaire. [Table A2](#) in the appendix shows that there are also compositional differences between the subsamples having answered in the first five weeks than those having answered later on (e.g. for gender and marital status). However, given the now even smaller number of respondents in the two subsamples and particularly for some of the categorical variables, we refrain from discussing the results in more detail.

respondents may be memory failure but also social desirability, given the higher share of nonvoters among them.

Finally, the last row of [table 6](#) presents the results when combining the main study and NRFU survey. The results again highlight that, as far as the turnout question is concerned, adding the follow-up questionnaire to the main survey brings the realized sample more in line with the initial sample. While voters are still overrepresented, their overrepresentation strongly diminishes (by one-third, from 16 percent to 10 percent) once we add the NRFU respondents to the respondents to the main survey. Moreover, adding NRFU respondents only weakly increases measurement errors (vote overreporting) by 1.2 percentage points (2.5 to 3.7). Therefore, to answer our research question (RQ1), it is fair to conclude that the NRFU questionnaire increases response rate and decreases voter overrepresentation, without or almost without weakening the quality of survey responses with respect to the turnout question. In addition, starting the NRFU survey earlier than we did and making additional efforts to reduce its duration would, arguably, help to further reduce the risk of misreporting.

8. CONCLUSION

To curb the long-lasting decline in response rates and to decrease nonresponse bias, survey research has placed great hopes in mixed-mode surveys. However, whether and to what extent mixed-mode designs keep such promises remains an open question. In this article, we have examined how adding a short mail NRFU survey to the main CATI/web survey helps to reduce nonresponse bias and increase sample representativeness. Moreover, we have taken a cursory look at measurement errors by asking whether the gains in sample representativeness come at the price of responses quality. The focus of this article is the sensitive voter turnout question, which we analyzed based on a rich set of validated vote data gathered in the context of two ballots in the Geneva state.

Our results show that sampled citizens who get lost, i.e. who do not participate in the main survey, have a specific sociodemographic profile and, even more so, a specific political profile. Those citizens are younger, less often married and more often single, and have recently arrived in Geneva. Moreover, they are likely to have abstained in the vote of interest and, more generally, to belong to the group of seldom voters. This makes a strong difference for the study of political participation. Yet we also find that the NRFU survey helps to cope with the nonresponse bias associated with voter overrepresentation without affecting the quality of survey responses. In that sense, our results confirm the positive effects of an NRFU found in earlier studies ([Voogt and Saris 2005](#); [Bandilla et al. 2014](#)) and disconfirm the negative effects found in others ([Link and Mokdad 2006](#); [Sakshaug and Eckman 2017](#)). They highlight the added value of a follow-up survey focusing on a handful of so-called central questions.

First, the NRFU sample is more representative of the initial sample than the main CATI/web survey. Granting initial nonrespondents the opportunity to fill in a short mail questionnaire increases the representativeness of the sample, and this is true especially regarding political participation variables. While citizens who participated in the ballot under study and frequent voters more generally are strongly overrepresented in the main survey, they are accurately represented among NRFU respondents. Therefore, adding the NRFU respondents to initial (CATI/web) respondents reduces voter overrepresentation and brings the realized sample closer to the initial sample.

Second, sociodemographic determinants of turnout tend to be more in line with true values (i.e., values in the initial sample) if we combine the main study and the NRFU survey. Yet in a strict statistical sense, we could confirm this improved accuracy of turnout determinants only in one of the two surveys under study. Therefore, more research is needed to confirm our findings.

Finally, bringing additional respondents in with a follow-up questionnaire does not worsen the quality of the data. True, vote overreporting is higher among NRFU respondents than among CATI or web respondents. However, the amount of overreporting remains fairly low. Thus, it is safe to conclude that the costs in terms of measurement errors do not offset the gains in terms of sample representativeness, at least with respect to the simple but sensitive turnout question.

Whether this conclusion also holds for more complex attitudinal scale-questions is obviously an open question (see the different results for economic variables in [Sakshaug and Eckman \[2017\]](#)). Another open question regards the influence of the specific mode and length of the NRFU survey. As our follow-up survey comprised both a mode switch and a substantial cut in the length of the questionnaire, we cannot separate the influence of these two aspects. Future research should thus separate them and examine in more detail which mode change leads to the best results and how long a shortened follow-up survey should be. Another limitation of our study is the nonstrict sequential character of the mixed-mode design in the main survey, which prevented us from systematically analyzing the differences between CATI and web respondents, i.e. between an interview- and a self-administered survey. As our main response phase already combined both types of surveys, we could not give any additional insight on the extent to which such a general switch from a purely interview-based survey to a self-administered mail survey in the NRFU section matters. Lastly, the differences between our two surveys regarding order and share of CATI and web mode in our main response phase may account for the partly stronger and more coherent effects found in the 2012 survey. However, we could not test the influence of these design differences in more detail.

While our results are drawn from a given state in a small country, they have broader implications. They speak to survey research in general and to electoral research in particular. Overall, they confirm that a sequential mixed-mode design offering initial nonrespondents the opportunity to fill in a short mail

questionnaire is worth the effort in terms of sample representativeness. Further, this design represents a simple and relatively low-cost solution to partly overcome the problem of turnout bias in postelection surveys. In the present study, as in other recent studies, turnout bias mainly stems from voter overrepresentation. Assuming this holds more generally, a significant decrease in turnout bias can be achieved by complementing a web or a telephone survey with a short mail questionnaire. Such a design is likely to decrease the main problem of voter overrepresentation, while hardly worsening the (more minor) vote over-reporting problem.

Appendix

Table A1. Measurement Error in the 2015 NRFU Survey Depending on Timing of Response

	Validated turnout	Reported turnout	Measurement bias
Weeks 1–5 (n = 78)	44.9%	47.4%	2.5
Weeks 6–12 (n = 141)	37.6%	46.1%	8.5

NOTE.— Respondents who did not answer the turnout question are excluded from the table.

Table A2. Representativeness in the 2015 NRFU Survey Depending on Timing of Response

Variable in %	Full NRFU questionnaire N = 229	NRFU questionnaire Week 1–5 N = 82	NRFU questionnaire Week 6–12 N = 147
Female	51.5	48.8	53.1
Age (avg. in years)	52.5	52.8	52.4
Marital status			
Single	28.4	34.2	25.2
Married	49.4	43.9	52.4
Divorced	13.5	15.9	12.2
Widowed	8.7	6.1	10.2
Residence duration			
< 10 years	9.2	7.3	10.2
10–20 years	17.5	19.5	16.3
>20 years	73.4	73.2	73.5
GE citizen	60.7	59.8	61.2
Past participation	43.5	44.7	42.9
Vote	41.5	47.6	38.1

REFERENCES

- Ansolabehere, S., and E. Hersh (2012), "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate," *Political Analysis*, 20, 437–459.
- Atkeson, L. R., A. N. Adams, and R. M. Alvarez (2014), "Nonresponse and Mode Effects in Self- and Interviewer-Administered Surveys," *Political Analysis*, 22, 304–320.
- Bandilla, W., M. P. Couper, and L. Kaczmirek (2014), "The Effectiveness of Mailed Invitations for Web Surveys and the Representativeness of Mixed-Mode versus Internet-Only Samples," *Survey Practice*, 7, 1–9.
- Berent, M. K., J. A. Krosnick, and A. Lupia (2011), "The Quality of Government Records and Over-estimation of Registration and Turnout in Surveys: Lessons from the 2008 ANES Panel Study's Registration and Turnout Validation Exercises." Working paper nes012554, American National Election Studies, Ann Arbor, MI; Palo Alto, CA, Available at: <http://www.electionstudies.org/resources/papers/nes012554.pdf> (retrieved at 2018/10/02).
- Brehm, J. (1993), *The Phantom Respondents: Opinion Surveys and Political Representation*, Ann Arbor: University of Michigan Press.
- Brick, J.M. (2013), "Unit nonresponse and weighting adjustments: A critical review", *Journal of Official Statistics*: 29, 329–353.
- Burden, B. C. (2000), "Voter Turnout and the National Election Studies," *Political Analysis*, 8, 389–398.
- Carley-Baxter, L. R., A. Peytchev, and M. C. Black (2010), "Comparison of Cell Phone and Landline Surveys: A Design Perspective," *Field Methods*, 22, 3–15.
- Chang, L., and J. A. Krosnick (2009), "National Surveys via RDD Telephone Interviewing versus the Internet Comparing Sample Representativeness and Response Quality," *Public Opinion Quarterly*, 73, 641–678.
- Couper, M. P. (2011), "The Future of Modes of Data Collection," *Public Opinion Quarterly*, 75, 889–908.
- Curtin, R., S. Presser, and E. Singer (2000), "The Effects of Response Rate Changes on the Index of Consumer Sentiment," *Public Opinion Quarterly*, 64, 413–428.
- de Leeuw, E. D. (1992), *Data Quality in Mail, Telephone and Face to Face Surveys.*, ERIC Amsterdam: T. T. Publikaties.
- . (2005), "To Mix or Not to Mix Data Collection Modes in Surveys," *Journal of Official Statistics*, 21, 233–255.
- . (2008), "Choosing the Method of Data Collection," in *International Handbook of Survey Methodology*, eds E. D. de Leeuw, J. J. Hox, and D. A. Dillman, pp. 113–135, London: Taylor and Francis.
- . (2010), "Mixed-Mode Surveys and the Internet," *Survey Practice*, 3, 1–5.
- Deufel, B. J., and O. Kedar (2010), "Race and Turnout in US Elections Exposing Hidden Effects," *Public Opinion Quarterly*, 74, 286–318.
- Dillman, D. A., J. D. Smyth, and L. M. Christian (2014), *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (4th ed.), New Jersey: Wiley.
- Dillman, D. A., G. Phelps, R. Tortora, K. Swift, J. Kohrell, J. Berck, and B. L. Messer (2009), "Response Rate and Measurement Differences in Mixed-Mode Surveys Using Mail, Telephone, Interactive Voice Response (IVR) and the Internet," *Social Science Research*, 38, 1–18.
- Goyder, J. (1987), *The Silent Minority: Nonrespondents on Sample Surveys*, Boulder, CO: Westview Press.
- Groves, R. M. (2006), "Nonresponse Rates and Nonresponse Bias in Household Surveys," *Public Opinion Quarterly*, 70, 646–675.
- Groves, R. M., R. B. Cialdini, and M. P. Couper (1992), "Understanding the Decision to Participate in a Survey," *Public Opinion Quarterly*, 56, 475–495.
- Groves, R. M., and N. A. Mathiowetz (1984), "Computer Assisted Telephone Interviewing: Effects on Interviewers and Respondents," *Public Opinion Quarterly*, 48, 356–369.
- Groves, R. M., and E. Peytcheva (2008), "The Impact of Nonresponse Rates on Nonresponse Bias a Meta-Analysis," *Public Opinion Quarterly*, 72, 167–189.

- Groves, R. M., S. Presser, and S. Dipko (2004), "The Role of Topic Interest in Survey Participation Decisions," *Public Opinion Quarterly*, 68, 2–31.
- Holbrook, A. L., and J. A. Krosnick (2010), "Social Desirability Bias in Voter Turnout Reports Tests Using the Item Count Technique," *Public Opinion Quarterly*, 74, 37–67.
- Holmberg, A., B. Lorenc, and P. Werner (2010), "Contact Strategies to Improve Participation via the Web in a Mixed-Mode Mail and Web Survey," *Journal of Official Statistics*, 26, 465–480.
- Hox, J. J., E. D. de Leeuw, and E. A. O. Zijlman (2015), "Measurement Equivalence in Mixed Mode Surveys," *Frontiers in Psychology*, 6, 87–111.
- Keeter, S., C. Kennedy, A. Clark, T. Tompson, and M. Mokrzycki (2007), "What's Missing from National Landline RDD Surveys?: The Impact of the Growing Cell-Only Population," *Public Opinion Quarterly*, 71, 772–792.
- Lin, I.-F., and N. C. Schaeffer (1995), "Using Survey Participants to Estimate the Impact of Nonparticipation," *Public Opinion Quarterly*, 59, 236–258.
- Link, M. W., and A. Mokdad (2006), "Can Web and Mail Survey Modes Improve Participation in an RDD-Based National Health Surveillance?" *Journal of Official Statistics*, 22, 293–312.
- Lipps, O., and N. Pekari (2016), "Sample Representation and Substantive Outcomes Using Web with and without Incentives Compared to Telephone in an Election Survey," *Journal of Statistics*, 32, 165–186.
- Medway, R. L., and J. Fulton (2012), "When More Gets You Less: A Meta-Analysis of the Effect of Concurrent Web Options on Mail Survey Response Rates," *Public Opinion Quarterly*, 76, 733–746.
- Messer, B. L., and D. A. Dillman (2011), "Surveying the General Public over the Internet Using Address-Based Sampling and Mail Contact Procedures," *Public Opinion Quarterly*, 75, 249–269.
- Messer, B. L., M. Edwards, and D. A. Dillman (2012), "Determinants of Item Nonresponse to Web and Mail Respondents in Three Address-Based Mixed-Mode Surveys of the General Public," *Survey Practice*, 5, 1–9.
- Quinn, R. P., B. A. Gutek, and J. T. Walsh (1980), "Telephone Interviewing: A Reappraisal and a Field Experiment," *Basic and Applied Social Psychology*, 1, 127–153.
- Roberts, C., D. Joye, and M. E. Stähli (2016), *Mixing Modes of Data Collection in Swiss Social Surveys: Methodological Report of the LIVES-FORS Mixed Mode Experiment*, (Vol. LIVES Working Papers 48). Lausanne: University of Lausanne.
- Sakshaug, J. W., and S. Eckman (2017), "Following up with Nonrespondents via Mode Switch and Shortened Questionnaire in an Economic Survey: Evaluating Nonresponse Bias, Measurement Error Bias, and Total Bias," *Journal of Survey Statistics and Methodology*, 5, 454–479.
- Schouten, B. (2007), "A Follow-up of Nonresponse in the Dutch Labour Force Survey," Discussion Paper 07004, Statistics Netherlands, Available at <https://www.cbs.nl/-/media/imported/documents/2007/18/2007-04-x10-pub.pdf> (retrieved at 2018/12/02).
- Sciarini, P., and A. C. Goldberg (2016), "Turnout Bias in Postelection Surveys: Political Involvement, Survey Participation, and Vote Overreporting," *Journal of Survey Statistics and Methodology*, 4, 110–137.
- . (2017), "Lost on the Way: Nonresponse and Its Influence on Turnout Bias in Postelection Surveys," *International Journal of Public Opinion Research*, 29, 291–315.
- Selb, P., and S. Munzert (2013), "Voter Overrepresentation, Vote Misreporting, and Turnout Bias in Postelection Surveys," *Electoral Studies*, 32, 186–196.
- Singer, E. (2006), "Introduction: Nonresponse Bias in Household Surveys," *International Journal of Public Opinion Quarterly*, 70, 637–645.
- Smith, T. W. (1983), "The Hidden 25 Percent: An Analysis of Nonresponse on the 1980 General Social Survey," *Public Opinion Quarterly*, 47, 386–404.
- Stern, M. J., and D. A. Dillman (2006), "Community Participation, Social Ties, and Use of the Internet," *City and Community*, 5, 409–424.
- Stern, M. J., I. Bilgen, and D. A. Dillman (2014), "The State of Survey Methodology: Challenges, Dilemmas, and New Frontiers in the Era of Tailored Design," *Field Methods*, 26, 284–301.

- Stoop, I. A. L. (2005), *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*, The Hague: SCP, Social and Cultural Planning Office of the Netherlands.
- Tourangeau, R., R. M. Groves, and C. D. Redline (2010), "Sensitive Topics and Reluctant Respondents Demonstrating a Link between Nonresponse Bias and Measurement Error," *Public Opinion Quarterly*, 74, 413–432.
- Tourangeau, R., F. G. Conrad, and M. P. Couper (2013), "Nonresponse in Web Surveys." in *The Science of Web Surveys*, eds. R. Tourangeau, F. G. Conrad, and M. P. Couper, pp. 35–56, Oxford: Oxford University Press.
- Voogt, R. J. J., and W. E. Saris (2005), "Mixed Mode Designs: Finding the Balance between Nonresponse Bias and Mode Effects," *Journal of Official Statistics*, 21, 367–387.
- Zickuhr, K., and A. Smith (2012), "Digital Differences," *Pew Internet & American Life Project*, 13, 1–41.