Received: 29 January 2021 | Revised: 13 June 2022 | Accepted: 20 June 2022

DOI: 10.1111/spsr.12539

ORIGINAL ARTICLE

SPSR Swiss Political Science Review
Schweizerische Zeitschrift für Politikwissenschaft
Revue Suisse de Science Politique
Rivista Svizzera di Scienza Politica

Conditional distributions of frame variables and voting behaviour in probability-based surveys and opt-in panels

Nicolas Pekari[1] | Oliver Lipps[1,2] | Caroline Roberts[3] | Georg Lutz[1]

[1]Swiss Centre of Expertise in the Social Sciences (FORS), Lausanne, Switzerland

[2]Universität Bern, Bern, Switzerland

[3]University of Lausanne, Lausanne, Switzerland

Correspondence
Oliver Lipps, c/o University of Lausanne
1015 Lausanne, Switzerland.
Email: oliver.lipps@fors.unil.ch

Abstract

Probability-based web surveys are increasingly challenged by decreasing response rates and high costs. A cheap and convenient solution is to use 'opt-in' online panels, which are based on non-probability samples. However, the quality of the data such panels produce is subject to debate. To improve our understanding in this regard, especially in the Swiss context, we compare conditional distributions of sociodemographic variables and voting behaviour of two probability-based web surveys and three opt-in panels. Indeed, point estimates in opt-in panels are well studied, but bivariate relationships between variables, arguably more important for researchers in political science research, have received less attention. Our analysis has the advantage of most variables of interest being included in the sampling frame and thus the true values are known for each conditional distribution. Our results show a lack of consistency and reproducibility in the results from opt-in panels, which leads us to recommend care when using this type of data.

Zusammenfassung

Wahrscheinlichkeitsbasierte Webumfragen sind zunehmend mit sinkenden Antwortraten und hohen Kosten konfrontiert. Eine billige und praktische Lösung ist die Verwendung von „Opt-in“ online-Panels, die auf Nicht-Wahrscheinlichkeitsstichproben beruhen. Allerdings ist die Qualität der Daten, die solche Panels produzieren, umstritten. Um unser Verständnis in dieser Hinsicht insbesondere im Schweizer Kontext zu verbessern, vergleichen wir die bedingten Verteilungen soziodemografischer Variablen und des Wahlverhaltens von zwei wahrscheinlichkeitsbasierten Webumfragen und drei Opt-in-Panels. Während viel über Punktschätzungen in Opt-in-Panels bekannt ist, haben Beziehungen zwischen Variablen, die für die wissenschaftliche Forschung wichtiger und für statistische Modelle entscheidend sind, weniger Aufmerksamkeit erhalten. Die meisten unserer interessierenden Variablen sind im Stichprobenrahmen enthalten, so dass die wahren Werte für jede bedingte Verteilung bekannt sind. Unsere Analysen zeigen mangelnde Konsistenz und Reproduzierbarkeit der Ergebnisse aus Opt-in-Panels, die Anlass zu Vorsicht bei der Verwendung dieser Art von Daten geben.

Résumé

Les enquêtes en ligne basées sur des probabilités sont de plus en plus confrontées à la baisse des taux de réponse et aux coûts élevés. Une alternative consiste à utiliser des panels en ligne « opt-in », qui sont basés sur des échantillons non probabilistes. Cependant, la qualité des données produites par ces panels est sujette à débat. Pour améliorer notre compréhension à cet égard, en particulier dans le contexte suisse, nous comparons les distributions conditionnelles des variables sociodémographiques et du comportement de vote de deux enquêtes en ligne basées sur des probabilités et de trois panels opt-in. On en sait beaucoup sur les estimations ponctuelles dans les panels opt-in, mais les relations entre les variables, sans doute plus importantes pour la recherche scientifique et cruciales pour les modèles statistiques, ont reçu moins d'attention. La plupart des variables qui nous intéressent sont incluses dans la base de sondage de sorte que les valeurs réelles sont connues pour chaque distribution conditionnelle. Le manque de cohérence et de reproductibilité des résultats des panels opt-in nous amène à recommander la prudence dans l'utilisation de ce type de données.

KEYWORDS
Conditional distribution, Opt-in online panels, Probability samples, Voting behaviour, Web survey

INTRODUCTION

Opt-in Panels Versus Probability-Based Surveys

In addition to high costs, probability-based surveys are faced with increasing difficulties in maintaining response rates at acceptable levels (Luiten et al., 2020). One alternative is to use opt-in online panels, which provide respondents at low prices in a quick and convenient way. While opt-in panels have been extensively compared with probability-based surveys, there are still questions as to whether opt-in panels can replace probability-based surveys and, crucially, under which conditions. For example, opt-in panels appear to be less suited to estimate accurate population values (Baker et al., 2010). Weighting has been shown to improve estimates up to a point, but it is not a sufficient solution (Dutwin & Buskirk, 2017). There is even evidence that weighting can reduce the accuracy of estimates (Yeager et al., 2011) and typically used demographic variables alone are not likely to correct for sampling differences (Pasek, 2016). One of the main problems is the lack of transparency and consistency in how surveys are run within a certain company or across different companies with considerable variations in accuracy among the findings (Cornesse et al., 2020; Yeager et al., 2011). In addition, the results are simply not reproducible in the way they would be with a probability-based sample. Some authors thus talk about "rolling the dice in terms of how accurate any one particular estimate may be" (Dutwin & Buskirk, 2017) when running these types of surveys. They might produce accurate results most of the time, but suddenly deviate from previous results without it being clear why this happened (Pennay et al., 2018). For example, results depend on variable types such as register variables or non-register variables (Brüggen et al., 2016). One reason for the highly varying quality may be that opt-in panels are a quickly evolving field and companies are constantly adapting their methods (Cornesse et al., 2020). Companies face increasing difficulties to recruit respondents since people are increasingly solicited by emails, messages, notifications and other requests for attention and the incentives offered by the companies are usually not attractive enough. The logical consequence is that those who enrol voluntarily into these types of panels constitute a subgroup that differs from the population that this subgroup is supposed to represent across a number of variables such as age, gender, education, marital status, homeownership (Yeager et al., 2011), health and life satisfaction (Brüggen et al., 2016), and election outcomes (Sturgis et al., 2018). Those who enrol voluntarily are also more interested in voicing their opinion (Callegaro et al., 2014). Companies running these panels, of course, try to actively attract groups that are underrepresented in their samples and many adjust the sample by means of calibration weighting or matching procedures (Cornesse et al., 2020).

Relationships Between Variables

One instance where opt-in online panels are often suggested as a viable alternative is for the estimation of relationships between variables (Pasek, 2016). As much of the research in political or social sciences focuses on relationships between variables, it could be argued that expensive probability-based surveys lose some of their advantage over opt-in panels especially with strongly falling response rates (Ansolabehere & Rivers, 2013). For example, if one is interested in explaining why certain persons turn out to vote or to know more about the composition of voters and non-voters, opt-in panels could offer a viable alternative.

Accuracy of the relationships between variables has been less studied in opt-in panels and results are somewhat mixed (Callegaro et al., 2014). Usually, one of the two following methods has been used: First, comparing whether statistical models run on data generated from probability-based samples and opt-in panels produce significantly different regression coefficients. This can provide interesting insights, but the results are very sensitive to the dependent and independent variables used, and the results are hard to generalise (Brüggen et al., 2016). The other way is to use the source of the data as an interaction term between one independent variable and the dependent variable and determining whether it is significant. This has the advantage of considering only one independent variable at a time and avoiding issues that could arise from a more complex model. However, interaction terms are not always easy to interpret unequivocally. The results from this type of research have been mixed: while a majority finds relatively small differences in relationships, others have shown that coefficients can change widely and thus even lead to different conclusions (Brüggen et al., 2016; Cornesse et al., 2020; Pasek & Krosnick, 2020).

Contribution of This Paper

Clearly more research on the relationships between variables is needed. While existing studies mostly rely on "gold-standard" benchmark surveys for all variables (e.g., Dutwin & Buskirk, 2017; Erens et al., 2014; MacInnis et al., 2018; Malhotra & Krosnick, 2007; Pasek & Krosnick, 2020), we use sociodemographic variables from the population register and administrative data as the main benchmark. The population register[1] of the Swiss Federal Statistical Office is updated on a quarterly basis using information gathered at the municipal level and covers all legal residents. The data is first aggregated and checked at the regional (cantonal) level and then again at the country level. Switzerland also has strict laws in terms of informing of any changes to one's situation. Although these data cannot be considered totally error free, it is likely much less biased than sociodemographic variables from survey data, which suffer from non-observation and measurement issues (Groves et al., 2011). Census data is close to register data in overall quality, but still relies on a survey process and as such suffers from the same sources of error, despite very high response rates and usually mandatory participation. However, the biggest drawback of census data is obviously related to temporality, as censuses are usually conducted every five or ten years and will thus most often not be up to date to conduct comparisons. Measurement issues may also occur in register-based data. However, for the sociodemographic variables we focus on, measurement problems should be very minimal in the register data. To complement sociodemographic data, we add information regarding voting behaviour, for which no conditional distribution is available from registers. Instead, we use the gold standard concerning voting behaviour in Switzerland as the benchmark: The Selects mixed-mode post-electoral survey (Lutz, 2016), financed by the Swiss National Science Foundation and overseen by faculty members of seven Swiss universities, as well as the Swiss Federal Statistical Office (FSO).

To our knowledge, this is the first study that compares non-probability and probability-based surveys in Switzerland. In addition, register-based data have only been used by Brüggen et al. (2016) in this context. The authors conducted a similar study, including both register and non-register variables, and also used probability surveys as benchmarks for

[1]See https://www.bfs.admin.ch/bfs/en/home/basics/census/natonal-census-integrated-system/sampling-frame.html (accessed Nov 29, 2021).

the variables which were not included in the register: the Dutch Labor Force Survey for employment and the Statistics on Income and Labor Conditions (SILC) for education, health, and life satisfaction.

Substantively, our analysis compares the conditional distribution of sociodemographic register and voting variables in probability-based surveys and in opt-in online panels. While the basic sociodemographic variables are directly available from the sampling frame owned by the FSO, income is matched from additional administrative data. The second part will focus on the two main variables in political science research: participation and vote choice. The true values here are available from official voting results at the aggregate level. Finally, we conduct several robustness checks.

DATA AND METHODS

Data

Data for this study was collected in the context of the Swiss Electoral Study (Selects) 2015 (Selects Post-Electoral Study, 2015). The main Selects post-electoral survey is the benchmark in Switzerland for voting behaviour in national elections. For this survey, the FSO drew a random sample of 10,391 individuals (excluding the oversampling of individuals from Geneva, and excluding Ticino completely) with the right to vote in Switzerland, with Zurich and some smaller cantons oversampled to have enough observations based on the political structure of each (Steenbergen, 2014).[2] The survey was conducted by the survey institute DemoSCOPE as a sequential mixed mode starting with Web on the day after the elections, on October 19, and adding the telephone[3] component after two weeks. Around ten days before the fieldwork started, sample members received a pre-notification letter including a flyer presenting the study. The invitation letter that followed included a 10 Sfr. postal check as an unconditional incentive. Two reminders were then sent on October 28 and November 6. On December 1, a short paper nonresponse survey was sent to the remaining sample members. The AAPOR RR1 (*AAPOR*, 2016) for the web survey was 36%. This increased to 45% when including the telephone and to 53% when adding the paper mode (Lipps & Pekari, 2021). The authors showed that adding the telephone mode improves sample representation in terms of register variables. While adding the paper nonresponse questionnaire does not show further improvements in register variables, adding the telephone, and in particular, the paper questionnaire, reduce bias in terms of voting behaviour.

As a second probability-based survey, we include a panel survey that uses the Web as the only mode (Selects Panel / Rolling Cross-Section Study, 2015). For this survey, the FSO drew a simple random sample of 29,548 individuals representative of the Swiss voting population. The first wave out of a total of four was fielded in June/July 2015. Respondents were contacted first with a pre-notification letter and a flyer and then with an invitation letter including a link and a 10 Sfr. postal check as in the post-electoral survey. In addition, a raffle of five iPads was organised among all respondents to the first three waves to boost enrolment in the panel and reduce attrition. Two reminders were sent on June 26 and July 3. The AAPOR RR1 response rate of the first wave was 37%, and 11,009 individuals participated. All respondents were asked at the end of the first wave to provide a valid email address to be recontacted for further interviews, which 90% did. These people were invited to take part in the second and the third wave. A response to the second wave was not required for invitation to participate at the third wave. The second wave lasted during the last 60 days of the campaign, ending on the day before the vote of October 18. For this wave, a rolling

[2]This oversampling made design weights necessary.

[3]The sampling frame of the FSO includes landline numbers if applicable and whether these are listed in the telephone directory.

cross-section design (Johnston & Brady, 2002) was used, and invitations and reminders were sent via the email addresses obtained in the first wave, as well as a postal mail pre-notification. Due to the rolling cross-section design, which required daily communication, those without an email address were excluded from this wave. The total number of respondents was 7,295, and the AAPOR RR1 response rate among those invited among the first wave respondents was 74%. The third wave was conducted after the elections. Invitations were sent mostly via email, but, given that the rolling cross-section method was not used for this wave, also by letter for those who hadn't provided an email address in the first wave. A total of 7,587 respondents took part in the survey with an AAPOR RR1 response rate of 67% of those invited. Questions regarding socio-demographics were asked in the first wave, whereas participation and vote choice were asked in the second and third waves, depending on whether the person had already voted by mail when responding to the second wave. Due to selective attrition, which means that for instance politically interested individuals are more likely to remain in the survey, this survey's results regarding political behaviour differ from those of the main post-electoral survey.

Parallel to the probability-based post-electoral surveys, a slightly reduced version of the questionnaire was administered using opt-in online samples from three different companies. The criteria for choosing the companies were that 1) they had responded to the 28 ESOMAR questions, an industry standard list of questions to help researchers compare providers (ESOMAR, 2012); 2) all respondents could be redirected to the same survey software (Qualtrics); and 3) the companies were able to provide 1000 respondents from the Swiss population from the German- and French-speaking language regions meeting quotas for age, gender, and language. One of the selected companies was Swiss and the two others were international providers. The panel from Provider 1 differed from those of the other providers in some respects: it included education in the quotas, and its core consists in the customer bases of two large companies.

Fieldwork in all three online panels started on the same day as the post-electoral survey and lasted ten days on average.

Table 1 provides an overview of the surveys compared in this paper.

Variables

For the part focusing on register variables, we use the four variables that are available in all surveys as well as being part of the sampling frame: gender, age, language, and household size. We use the simple random sample of 29,548 individuals drawn for the Selects panel survey of which 28,324 with Ticino excluded as the definition of the target population for all surveys. In the variable combination that includes income, the benchmark consists of the 10,391 individuals drawn at random (with small cantons oversampled) for the Selects post-electoral survey. The data from the sample members and the members of their households of the Selects post-electoral survey were matched with data from the Central Compensation Office (CCO)[4] using a deterministic matching based on the Social Security Number. The variables used include the sociodemographic variables age group (18–30, 31–44, 45–58, 59–65),[5] gender, language (Swiss-German, French), household size (1, 2, 3, 4+

[4]As data for retired individuals can be incomplete in this register, we limited ourselves in theses analysis only on those under 66 years of age.

[5]We use the same age groups as in Lipps and Pekari (2021), but limit age by the smallest common maximum age in our samples (65 years, in the panels from Provider 2 and 3; see table 1). Limiting age to 65 years was also necessary because retirement pensions were not included in the income register (see conclusion).

TABLE 1 Surveys Compared

Name	Type	Age	Sample Size	Response rate	Quotas
Selects post-electoral survey	Probability	18+	5391	53%	None
Selects panel survey	Probability	18+	11009	37%	None
Panel Provider 1	Nonprobability	18–80	1000		Age, gender, language, (education)
Panel Provider 2	Nonprobability	18–65	1000		Age, gender, language
Panel Provider 3	Nonprobability	18–65	1000		Age, gender, language

persons), and monthly taxed household income (3 groups with cut-off values 5,776 Sfr. and 9,920 Sfr.).

For the part on political variables, the reference is the Selects post-electoral survey, including the paper nonresponse survey. There are no official nationwide statistics on the sociodemographic characteristics of voters and non-voters and obviously none for party choice. However, these two variables are chosen as they represent the main dependent variables in electoral research. For vote choice, we use a dichotomous variable on the vote for the largest party in Switzerland, the right-wing populist Swiss People's Party (SVP).

ANALYTICAL METHOD

In this section, we describe the methods used to compare bias of conditional estimates computed from the cross-tabulation of pairs of register and political variables between the three opt-in panels, the mixed-mode post-electoral survey, and the panel survey. Following Dutwin and Buskirk (2017), we focus on the conditional distribution of a variable given a value of the other. We do not conduct multivariate regressions because predictors can be specified in many ways, and results may differ depending on the combination of predictors included (Malhotra & Krosnick, 2007). In addition, every variable added in a model is an additional potential source of error, e.g., due to measurement error or item missingness, but also due to overspecification. Results from conditional distributions are thus likely to lead to higher external validity compared to multivariate regressions.

Specifically, we might for instance look at the distribution of gender within the four different age categories, using the register as the benchmark. We compare these two distributions with those from the register data and calculate eight absolute percentage point differences. From these eight differences, we calculate the mean (main analysis) and the maximum. This is done twice for each pair of variables: e.g., age within gender (resulting in two distributions) and gender within age (resulting in four distributions). For political variables, we look for example at the distribution of age groups within those who voted for the SVP, using the Selects post-electoral survey as the benchmark.

We do these analyses in three steps: (1a,b) across all pairs of register variables; (2) across all pairs of register variables and participation, and (3) across all pairs of register variables and participation and SVP party choice. The reason to distinguish between (1a,b), (2), and (3) are the different weights used because of the changing benchmark, depending on whether only register variables are included. In 1a), the benchmark is the sample drawn for the Selects panel survey. In 1b), the benchmark is the sample drawn for the Selects post-electoral survey with income included, so design weights are necessary because the post-electoral sample uses oversampling. In 2), because participation is included, the benchmark is the combined mixed mode post-electoral survey. Voters and non-voters are given a single weight respectively in order to replicate the true participation figures from the total population.[6] In 3), the benchmark is the party choice weighted combined mixed mode post-electoral survey for SVP party choice. Note that for participation and SVP party choice, the weights adjust for the known mean values in the population[7] based on election results in all surveys included in this study. The weighted benchmark distribution of one variable within another is based on the frequency-weighted data.

Finally, to account for the surveys' different sample sizes and the risk of higher sampling errors with smaller samples, we limit the sample from the larger probability-based surveys to 1000 random respondents to coincide with the three opt-in panels. In the last step, we run our analyses 100 times using random sampling with replacement and calculate the mean across the simulations for each test sample and each bias combination. In our view, this results in a fairer comparison of surveys with different sample sizes (see also Brüggen et al., 2016).

We then run several robustness studies. First, we compare the full original samples (with no simulation). Second, still based on the full original samples, we slightly deviate from the procedure used by MacInnis et al. (2018): instead of the mean absolute difference of *all* categories of a variable within another variable (our main analysis), we compare the maximum absolute difference of the categories of a variable within another variable using the full samples. The idea is to acknowledge bigger errors which can be costly for a researcher (MacInnis et al., 2018). We do not present statistical significance tests (Amrhein et al., 2019), since these highly depend on the sample sizes of the data compared. In fact, the samples used in the paper have different sample sizes, which would confuse the interpretation of differences between the main studies and the robustness studies. Borderline significant results could be insignificant due to other errors, such as measurement or representation errors. This makes clear that dichotomisation as statistically significant or not may lead to too rigid statements. Last, but not least, such assessments hinge on the correctness the statistical assumptions used to compute the significance. Instead, we focus on the size or importance of an effect.

RESULTS

Univariate distributions

First, we look at the univariate distributions comparing the opt-in panels to our reference surveys (Table 2). Where possible, we provide the official statistics as a benchmark. As the opt-in panels had quotas for age, gender, and language, comparisons for these three variables are of less interest. Age quotas were done by age groups, which explains some of the differences. Provider 1's panel seems to best represent the mean age in this regard, whereas the other two show signs of over-representing younger people within the quotas. Differences in household size are small but the opt-in panels over-represent smaller households. Finally, the panels from Providers 2 and 3 over-represent low-income individuals.

Regarding the substantive variables, participation and vote choice, there is a substantial difference between the panel from Provider 1 and the two others. The panel from Provider 1 is

[6]This means that the voters are given the population mean divided by the sample mean and the non-voters are given (1-the population mean) divided by (1-the sample mean). We calculated the SVP party choice weight in the same way.

[7]Because our sample includes only people 65 and younger, we cannot use the results of all voters as a direct benchmark. To account for the specific age distribution, we assume, based on available data, a 2 percentage points lower participation compared to all Swiss adult citizens and a 1 percentage point lower SVP percentage in our sample of people 65- compared to all voters.

TABLE 2 (Unweighted) Univariate Statistics

	Variable Statistic	Age [years]	Income [1000 Sfr.]	Household size*	Sex [Male]	German-speaking	Turnout rate	SVP Vote
Official Mean		42.6	8,5	2.9	0.50	0.77	0.49	0.29
Survey								
Selects post-electoral survey	N (analytical sample)	3250	3250	3250	3250	3250	3250	2115
	Mean	43.5	8.8	2.7	0.50	0.75	0.67	0.25
	Median	46	8.5	3	0	1	1	0
	Standard Deviation	13.6	4.8	1.1	0.50	0.43	0.47	0.43
Selects panel survey	N (analytical sample)	8581	8206	8581	8568	8581	6007	4516
	Mean	43.7	8.5	2.8	0.50	0.75	0.76	0.23
	Median	46	7.5	3	0	1	1	0
	Standard Deviation	13.4	3.9	1.1	0.50	0.43	0.42	0.43
Panel Provider 1	N (analytical sample)	819	676	811	819	819	815	617
	Mean	42.3	7.8	2.4	0.48	0.74	0.77	0.21
	Median	42	7	2	0	1	1	0
	Standard Deviation	13.4	3.7	1.1	0.50	0.44	0.42	0.40
Panel Provider 2	N (analytical sample)	997	799	993	997	997	966	553
	Mean	41.3	6.2	2.5	0.50	0.69	0.60	0.38
	Median	42	6	2	0	1	1	0
	Standard Deviation	13.1	3.6	1.1	0.50	0.46	0.49	0.49
Panel Provider 3	N (analytical sample)	916	726	907	916	916	892	531
	Mean	40.7	6.5	2.5	0.48	0.73	0.61	0.36
	Median	41	6	2	0	1	1	0
	Standard Deviation	13.0	3.8	1.1	0.50	0.44	0.49	0.48

[a]topcoded at 4.

relatively close to the reference survey, which over-represents voters and underrepresents SVP voters. The panels from the two other providers better represent non-voters as well as SVP voters, the latter of which are actually overrepresented, which is rare in probability-based surveys (Lipps & Pekari, 2021).

Bivariate bias

We first consider the results from the mean absolute bias analyses for the register variables, while for the probability samples the means of the 100 simulation results using 1000 respondents each are reported (Figure 1).

As expected, the opt-in panel surveys have larger errors than the probability-based surveys. The panel from Provider 1 fares best out of the opt-in panels, with many estimates close to those of the probability-based surveys, which remain below the 4% error limit for almost all variables. Particularly striking in these results are the peaks for certain pairs of variables. The error is relatively evenly distributed in both probability-based samples, with a few exceptions involving income in the Selects panel. In turn, the differences in the opt-in panels are sometimes large, with mean absolute biases reaching over 10 percentage points. This means that for example when looking at gender within age groups in the panel from Provider 3 (the longest bar in the graph), the average error of each cell in the cross tabulation amounts to more than

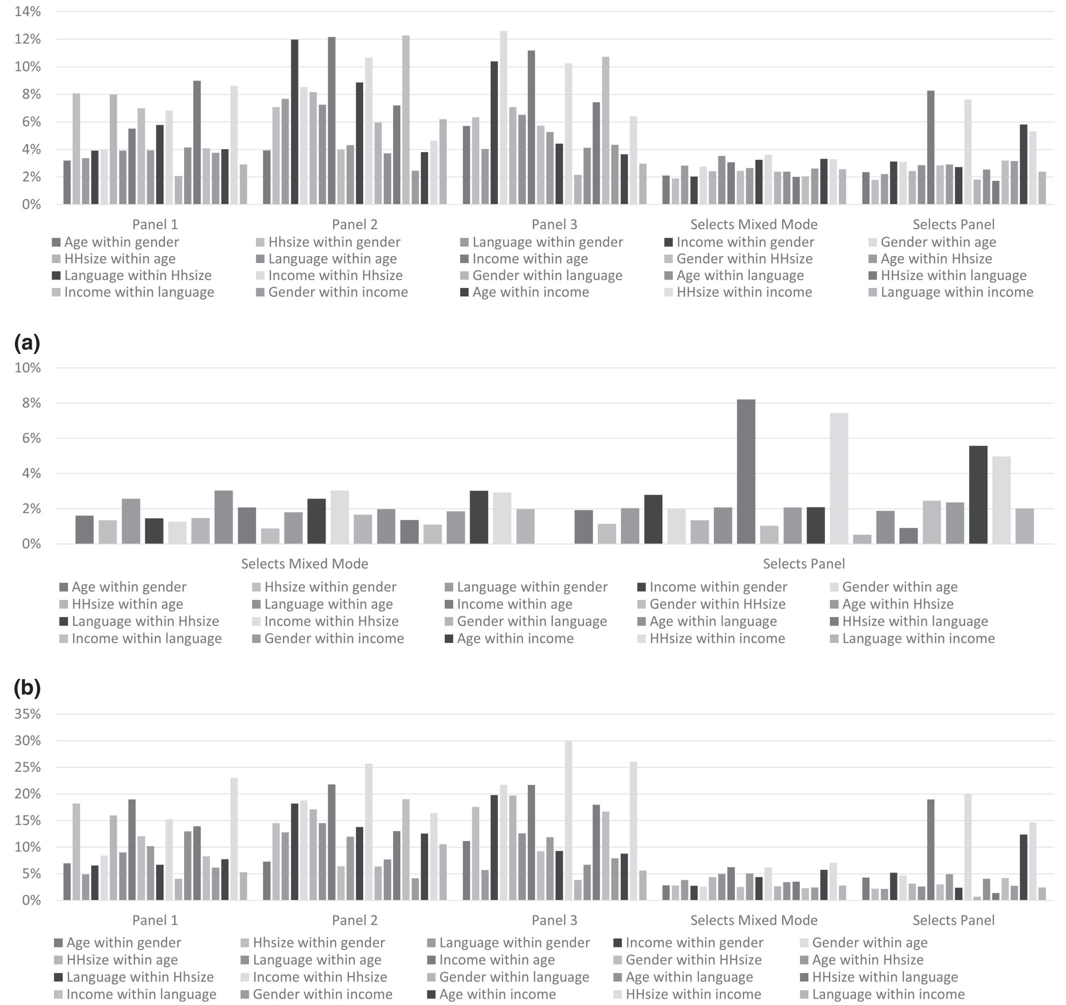

FIGURE 1 Mean Absolute Errors of the Conditional Distribution of the Sociodemographic Variables: Full panels and N = 1000 for the probability samples, mean of 100 simulations. (a) Mean Absolute Errors of the Conditional Distribution of the Register Variables: Full probability samples. (b) Maximum Absolute Errors of the Conditional Distribution of the Register Variables: Full panels and full probability samples.

12 percentage points. Despite the exceptions in the Selects Panel survey, the two probability-based surveys show much more similar patterns and better consistency than the three opt-in panels.

On average, the mean errors across all register variables amount to .051, .070, and .066 for the panels of the three Providers, to .027 for the Selects post-electoral survey, and to .034 for the Selects panel (not shown in Figure 1).

Figure 1a shows the results for the register variables with the two *full* probability-based samples. They perform slightly better than in the main analysis. The average mean errors across the register variables amount to .019 for the full Selects post-electoral survey, and to .027 for the full Selects panel (not shown in Figure 1a).

Though the overall level of error is higher, the same trend emerges when using the maximum absolute bias instead of the mean absolute bias (Figure 1b).

The average maximum errors across the register variables amount to .107, .136, and .142 for the panels of the three providers, to .039 for the full Selects post-electoral survey, and to .058 for the full Selects panel (not shown in Figure 1b).

We then compare political variables using two main measures of political behaviour: turnout (Figure 2) and vote choice (Figure 3). Because the Selects 2015 post-electoral survey is the benchmark here, it is omitted from the figure.

In Figure 2, the panel from Provider 1 is closest to the reference survey when comparing the mean absolute turnout bias with the panels from the two other panel providers. It is more uneven than the Selects Panel, but the differences are relatively small. The panels from Providers 2 and 3 have higher errors overall and a few very high peaks, all related to income (the fifth and the last bars in each group). The average mean error across the political variables amount to .039, .057, and .055 for the panels of the three providers, and to .033 for the Selects panel (not shown in Figure 2).

The full Selects panel performs even better when simulations are not used (Figure 2a) and when comparing the maximum absolute bias (Figure 2b). For the full Selects panel, the average mean error across the political variables amounts to .017 (not shown in Figure 2a).

The average maximum errors across the political variables amount to .059, .084, and .085 for the panels of the three providers, and to .031 for the full Selects panel (not shown in Figure 2b).

Looking at the vote for the right-wing Swiss People's Party (SVP), we can draw similar conclusions (Figure 3). Provider 1's panel shows uneven results but overall low errors, whereas the two other panels show a few very high errors and a higher overall error. The highest errors are related to income (the fifth and the last bars in each group). Even though, overall, the level of error in the Selects Panel is comparable to that of the opt-in panels, the errors are distributed more evenly, as was already the case in the previous analyses. The average mean errors across the SVP variables amount to .040, .055, and .043 for the panels of the three providers, and to .031 for the Selects panel (not shown in Figure 3).

Again, the Selects panel improves slightly relative to the simulations when the full sample is used to calculate the mean absolute bias (Figure 3a). For the full Selects panel, the average mean error across the SVP variables amounts to .023 (not shown in Figure 3a). Likewise, in terms of the maximum absolute bias, the Selects panel outperforms the opt-in panels (Figure 3b), with the average maximum errors across the SVP variables being .057, .084, and .071 for the panels of the three providers, and .037 for the Selects panel (not shown in Figure 3b).

We note that one obvious issue in our comparison of probability-based surveys and data from the opt-in online panels is the different sample size in the two types of data sources. We tackled this issue by running one hundred simulations on the probability-based surveys using equally sized samples for all surveys. This had the expected impact of reducing the effect of the sample size on the results. Evidently, as we then calculate the mean from the simulations based on subsamples drawn from a larger overall sample, the larger surveys still had an advantage. As expected, however, when moving from the simulations to using the full samples of the probability-based surveys produced, the results were even more favourable compared to the opt-in panels. The same holds for maximum absolute bias analyses. We also ran the analyses with the data from the three opt-in panels pooled (analyses not shown) and found that the error is overall higher in every case than when using the panel from the best performing provider alone. This gives further evidence for the fact that the lower sample size in the opt-in panels is not the main reason behind the differences we observe. Comparing the maximum absolute bias instead of the mean absolute bias leads to similar results.

The results of the panel from Provider 1 deserve more discussion here: It is interesting to note that these data are the least biased in the conditional distributions for voting behaviour, even though they are the most biased with regards to univariate distributions compared to official figures. In part, this could be because it is also closest to the reference

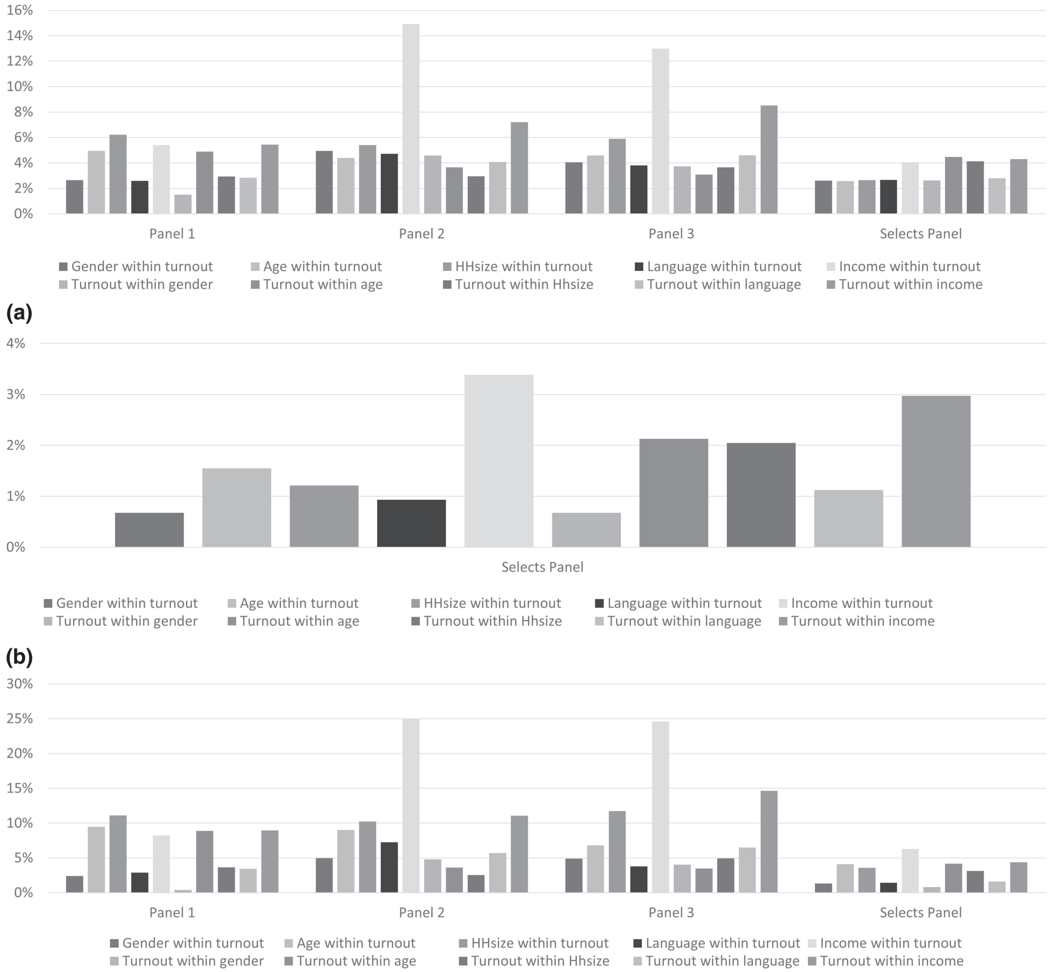

FIGURE 2 Mean Absolute Errors of the Conditional Distribution of The Turnout Variable. Full panels and N = 1000 for the probability sample, mean of 100 simulations. (a) Mean Absolute Errors of the Conditional Distribution of the Turnout Variable. Full probability sample. (b) Maximum Absolute Errors of the Conditional Distribution of the Turnout Variable. Full panels and full probability sample.

survey, with somewhat similar biases, underestimating non-voters and right-wing populist party voters. However, to calculate the conditional distributions for voting behaviour, all samples were weighted so that the weighted total represented the actual results of the election. Therefore, while in the case of the other providers panels, these weights distort the sociodemographic distribution *within* the variables of interest and the distribution of variables of interest *within* sociodemographic groups, in the data from Provider 1, the weighted data more closely resemble that of the reference survey. Assuming that the reference survey is correct, this means that data from this panel provider perform best both for descriptive statistics (after weighting) and for modelling political behaviour. In other words, they describe the sociodemographic characteristics of voters and non-voters, as well as voters from different parties, more accurately. In line with findings from the literature (e.g., Erens et al., 2014), the use of more complex quotas (e.g., including education) by the panel provider in this survey may have led to some improvement, although comparisons controlling for education were inconclusive (not presented here). In addition, models predicting voting

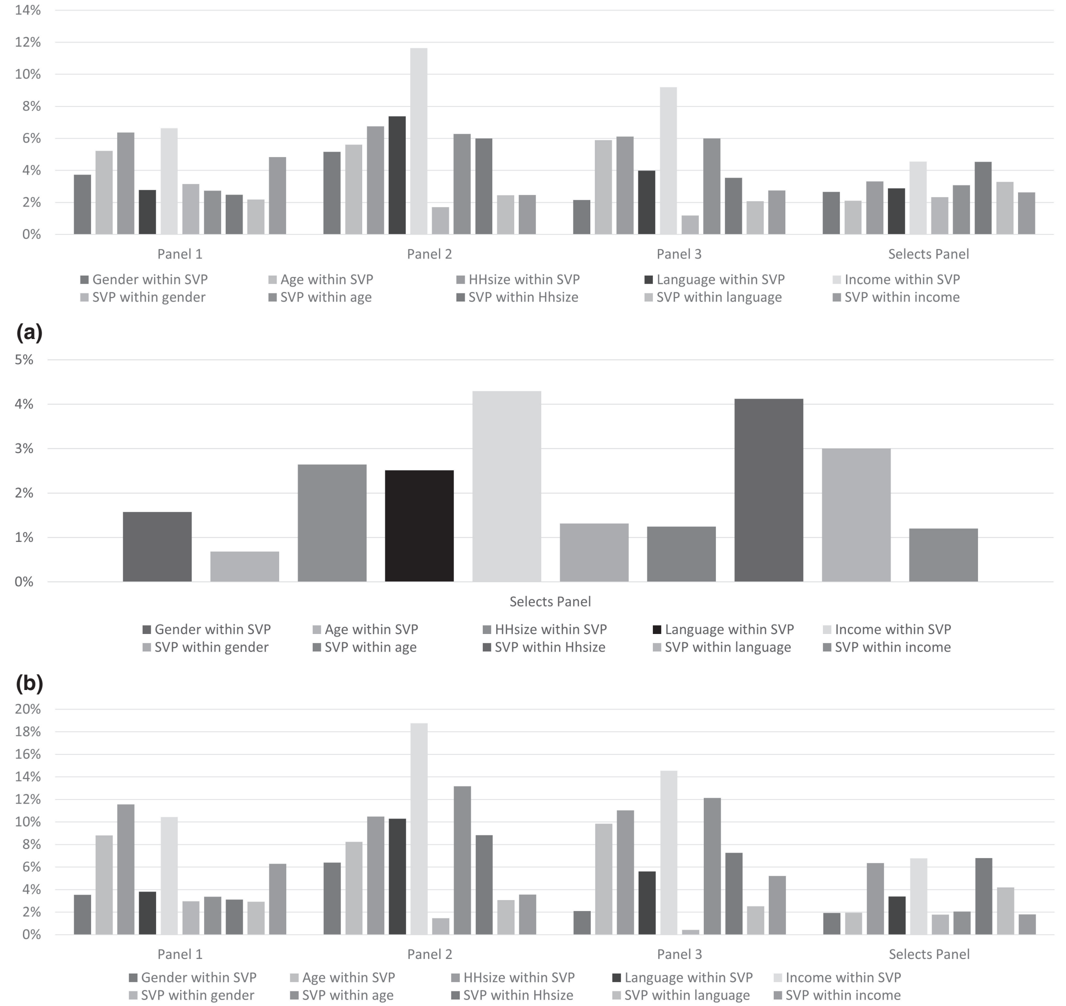

FIGURE 3 Mean Absolute Errors of the Conditional Distribution of the SVP Voting Variable. Full panels and N = 1000 forthe probability sample, mean of 100 simulations. (a) Mean Absolute Errors of the Conditional Distribution of the SVP Voting Variable. Full probability sample. (b) Maximum Absolute Errors of the Conditional Distribution of the SVP Voting Variable. Full panels and full probability sample.

behaviour run on the reference survey data and Provider 1's panel should yield results that are more similar than when the other panel providers' data are used.

DISCUSSION

We find evidence that there are specificities to the participants in opt-in panels that make them an inadequate representation of the general population. Our conditional distribution analyses in turn are very apt at showing how, in addition to the better-known shortcomings related to univariate statistics, opt-in panels are also an unreliable source of data to describe relationships between variables, e.g., how different groups of people behave or think, such as for which party older people are more likely to cast their vote.

In addition to the higher overall levels of bias in the distributions, what is most striking are the uneven results within and between opt-in panels. This shows particularly effectively the uncertainty involved in this type of data collection compared to probability-based data collection and the long tradition and research behind its methods. Thus, while offering a faster and cheaper option, opt-in online panels do not have the same level of established means to enhance quality that has been established during a long tradition in probability-based survey research. Given the at times vast disparities we find in the conditional distributions using the different datasets, there are, in our view, significant risks that these disparities translate into unreliable estimates when running multivariate analyses on opt-in panel data.

As opposed to probability-based data, where the sample size limit is mainly defined by the available budget, another problem of opt-in online panels is the fact that the sample size is limited by problems meeting given quotas. For most panel providers we contacted, even 1,000 interviews with the required parameters was at the limit of what could be provided. In addition to sample quality issues, collecting data from a large number of opt-in online panel respondents is thus a challenge, given the restricted number of suitable providers available. Especially in a small country like Switzerland, larger sample sizes might only be possible by combining the panels of various providers, which in turn further decreases the control over the sampling process and increases the risk of multiple responses by the same person. In addition, while online opt-in panels rely on the web as sole mode, probability-based data allows for mixing modes, even though this has obvious consequences in terms of costs. In this research, we used a push-to-web with telephone and paper follow-up as the main reference survey. We know from previous research that adding additional modes usually improves the sociodemographic representation of surveys (Lipps & Pekari, 2021). Not having the possibility of adding or mixing modes to improve sample representativeness or data quality in general thus represents a further limitation of online opt-in panels.

CONCLUSION

In this study, we were able to use relatively rare high quality sociodemographic register information, and, rarer still, income register data, which grants us a particularly reliable basis for our analyses. Though we use benchmark data that is highly reliable overall, the income data had some drawbacks. Indeed, it does not include all possible income sources and is thus less suited for some types of persons, especially self-employed and retired individuals. We have taken these issues into account in our analyses to the best of our abilities, but it remains that income is the variable where the conditional error was the highest and the only one where the two probability-based surveys differed strongly. This result can be partly explained by the way income is measured in the administrative data and therefore the results regarding income need to be interpreted with caution. Nevertheless, we believe we have been able to add convincing evidence, based on a unique dataset, to the well-established result that opt-in online panels are unreliable in producing adequate population estimates and that results vary widely between panel providers. We also add evidence to the less studied issues opt-in panels pose to studying the relationship between variables and multivariate analyses. Finally, we were able to provide a case study for the Swiss context.

ACKNOWLEDGEMENT
Open access funding provided by Universite de Lausanne.

DATA AVAILABILITY STATEMENT
The data used in this paper are available on request from the authors.

ORCID

Nicolas Pekari https://orcid.org/0000-0002-8406-9130
Oliver Lipps https://orcid.org/0000-0001-9865-2311
Caroline Roberts https://orcid.org/0000-0002-9819-1903
Georg Lutz https://orcid.org/0000-0001-5851-2955

REFERENCES

American Association for Public Opinion Research (AAPOR) (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 9th edition. Alexandria, VI: AAPOR.

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature* (Comment), *567*: 305–307.

Ansolabehere, S., & Rivers, D. (2013). Cooperative Survey Research. *Annual Review of Political Sciences 16*: 307–329.

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., et al. (2010). Research Synthesis: AAPOR Report on Online Panels. *Public Opinion Quarterly, 74*(4): 711–781.

Brüggen, E., Van den Brakel, J., & Krosnick, J.A. (2016). Establishing the Accuracy of Online Panels for Survey. *Statistics Netherlands, 11*(04): 43.

Callegaro, M., Villar, A., Yeager, D., & Krosnick, J. A. (2014). *A critical review of studies investigating the quality of data obtained with online panels based on probability and nonprobability samples1*. Online panel research*:* A data quality perspective, 23–53.

Cornesse, C., Blom, A.G., Dutwin, D., Krosnick, J., De Leeuw, E.D., Legleye, S., et al. (2020). A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. *Journal of Survey Statistics and Methodology 8*(1): 4–36.

Dutwin, D., & Buskirk, T. D. (2017). Apples to Oranges or Gala versus Golden Delicious? Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples. *Public Opinion Quarterly, 81*(S1): 213–239.

Erens, B., Burkill, S., Couper, M.P., Conrad, F., Clifton, S., Tanton, C., et al. (2014). Nonprobability web surveys to measure sexual behaviors and attitudes in the general population: a comparison with a probability sample interview survey. *Journal of medical Internet research, 16*(12): e276.

ESOMAR (2012). 28 Questions to Help Buyers of Online Samples. ESOMAR Guidelines. ESOMAR, Amsterdam. Online: https://www.esomar.org/what-we-do/code-guidelines/28-questions-to-help-buyers-of-online-samples [accessed: 29.1.2021].

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology*. Hoboken, NJ: John Wiley & Sons.

Johnston, R., & Brady, H.E. (2002). The rolling cross-section design. *Electoral Studies*, *21*(2): 283–295.

Lipps, O. & Pekari, N. (2021). *Sequentially mixing modes in an election survey*. Survey Methods: Insights from the Field. Retrieved from https://surveyinsights.org/?p=15281.

Luiten, A., Hox, J., & de Leeuw, E. (2020). Survey nonresponse trends and fieldwork effort in the 21st century: Results of an international study across countries and surveys. *Journal of Official Statistics, 36*(3): 469–487.

Lutz, G. (2016). *Élections fédérales 2015*. Participation et choix électoral. Selects, FORS.

MacInnis, B., Krosnick, J.A., Ho, A.S., & Cho, M.J. (2018). The accuracy of measurements with probability and nonprobability survey samples: Replication and extension. *Public Opinion Quarterly, 82*(4), 707–744.

Malhotra, N., & Krosnick, J.A. (2007). The effect of survey mode and sampling on inferences about political attitudes and behaviour: Comparing the 2000 and 2004 ANES to Internet surveys with nonprobability samples. *Political Analysis*, *15*(3): 286–323.

Pasek, J. (2016). When will nonprobability surveys mirror probability surveys? Considering types of inference and weighting strategies as criteria for correspondence. *International Journal of Public Opinion Research, 28*(2): 269–291.

Pasek, J., & Krosnick, J.A. (2020). Relations Between Variables and Trends Over Time in Rdd Telephone and Nonprobability Sample Internet Surveys. *Journal of Survey Statistics and Methodology, 8*(1): 37–61.

Pennay, D., Neiger, D., Lavrakas, P.J., Borg, K., Mission S., & Honey, N. (2018). The Online Panels Benchmarking Study: a total survey error comparison of findings from probability-based surveys and non-probability online panel surveys in Australia. *ANU Centre for Social Research & Methods (CSRM) 2/2018*, The Australian National University.

Steenbergen, M. (2014). *Optimal Allocation of the SELECTS Sample*. Mimeo, University of Zürich.

Sturgis, P., Kuha, J., Baker, N., Callegaro, M., Fisher, S., Green, J., Jennings, W., et al. (2018). An Assessment of the Causes of the Errors in the 2015 UK General Election Opinion Polls. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 181*: 757–781.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., & Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly, 75*(4): 709–747.

[Dataset] Selects: Panel / Rolling cross-section study - 2015. Distributed by FORS, Lausanne, 2016. www.selects.ch https://doi.org/10.23662/FORS-DS-828-3.

[Dataset] Selects: Post-electoral study - 2015. Distributed by FORS, Lausanne, 2016. www.selects.ch https://doi.org/10.23662/FORS-DS-726-5.

AUTHOR BIOGRAPHIES

Nicolas Pekari is the Head of Data Collection and Analysis at FORS and a PhD student at the University of Lausanne. His main research topics are survey methodology and political behavior.

Oliver Lipps is the head of the methodological research program at FORS and lecturer at the Institute of Sociology at the University of Bern. He is interested in aspects of methods of data collection and analyses, but also in the research of substantive topis.

Caroline Roberts is a senior lecturer and researcher in survey methodology and quantitative research methods at the Institute for Social Sciences, University of Lausanne. Her research focuses on the impact of survey design features on different types of survey error, particularly in mixed mode surveys, and on improving the measurement of social attitudes.

Georg Lutz is FORS Director and professor of political science at the University of Lausanne. He focuses on social science methods, political behaviour and political institution as well as Swiss politics in research and teaching.

How to cite this article: Pekari, N., Lipps, O., Roberts, C., & Lutz, G. (2022). Conditional distributions of frame variables and voting behaviour in probability-based surveys and opt-in panels. *Swiss Political Science Review*, *00*, 1–16. https://doi.org/10.1111/spsr.12539