The Effects of Survey Mode and Sampling in Belgian Election Studies
A Comparison of a National Probability Face-to-Face Survey and a Non-Probability Internet Survey*

Ruth Dassonneville[†], André Blais[‡], Marc Hooghe[§] and Kris Deschouwer[¶]

Abstract

National probability election surveys are more and more abandoned. Decreasing response rates and the escalating costs of face-to-face and telephone interviews have strengthened election scholars' reliance on non-probability internet samples to conduct election surveys online. In a number of countries, experiments with alternative ways of recruiting respondents and different interview modes have been well documented. For other countries, however, substantially less is known about the consequences of relying on non-probability internet panels. In this paper, we investigate the effects of survey mode and sampling method in the Belgian context. This is a particularly important and relevant case study because election researchers in Belgium can draw a sample of voters directly from the National Register. In line with previous studies, we find important differences in the marginal distributions of variables measured in the two surveys. When considering vote choice models and the inferences that scholars would draw, in contrast, we find minor differences.

Keywords: election study; Belgium; survey mode effects; representativeness; non-probability sample.

---

*This is a post-peer-review, pre-copyedit version of an article published in Acta Politica. The final authenticated version is available online.

[†]Département de science politique, Université de Montréal.

[‡]Département de science politique, Université de Montréal.

[§]Centre for Citizenship and Democracy, University of Leuven.

[¶]Department of Political Science, Vrije Universiteit Brussel.

1 Introduction

Practices in survey research, and in election studies more specifically are changing rapidly. Faced with declining response rates and increasing costs of face to face and telephone surveys based on probability samples, scholars of electoral research are more and more tempted to gather data online from non-probability samples. Compared to face-to-face or telephone probability surveys, online survey modes have a number of noteworthy advantages: they come at a low cost, rendering large sample sizes affordable and offering more statistical power; they allow for the inclusion of visual and audio information and stimuli for experimental designs. Furthermore, the fact that they are self-administered reduces the impact of a social desirability bias in participants' responses (Breton et al. 2017; Sanders et al. 2007).[1] It is therefore no surprise that online surveys are increasingly used for studying voting behavior.

Despite these advantages, online surveys and non-probability online surveys in particular have also met with severe criticism. Skepticism towards such samples relates, first, to their non-probability nature. Respondents in these surveys are not sampled randomly, but have opted in to participate to the survey or to some online panel. Developing a sampling frame to recruit respondents in a random manner online, it seems, is very hard (Couper 2000; Dillman 2000). Second, online surveys suffer from a 'coverage error' (Couper 2000). While the target population for an election study is the whole electorate, not everyone has internet access and—even more worrisome—there are systematic biases in who is online and who is not (Sanders et al. 2007; Simmons and Bobo 2015). As a result of these two issues—sampling error and coverage error—it is argued that it is highly unlikely that those participating in an online non-probability sample are representative of the electorate at large. Online samples generally overrepresent male citizens, respondents are younger and higher educated. Furthermore, samples of respondents in online non-probability election surveys tend to be more politically involved compared to respondents recruited by means of the more conventional probability sampling frames (Chang and Krosnick 2009; Schoen and Faas 2005). Most likely, these differences will have an effect on the relations that will be detected between variables (Hooghe et al. 2010).

Given these important limitations, a number of studies and reports have warned

1. Though it is important to point out that self-administered surveys come with problems as well, such as challenges related to measuring political knowledge in a reliable way (and preventing cheating) (Motta, Callaghan, and Smith 2016).

against non-probability online surveys, in particular when the goal is to estimate population values (Baker et al. 2010; Malhotra and Krosnick 2007). Societal changes as well as continuously developing methods and practices in survey research, however, lead scholars to regularly reassess and evaluate the advantages and disadvantages of online non-probability surveys. On the one hand, more conventional probability surveys suffer from declining response rates (Stern, Bilgen, and Dillman 2014). The initial response rate for the face-to-face component of the American National Election Study (ANES) has declined from well above 70% in the 1950s to a level of 50% for the 2016 ANES. Similar trends have been documented in other countries, regardless of whether election surveys take the form of face-to-face, telephone or mail surveys (as evident from trends in response rates in the British, Canadian and Australian Election Studies).[2] These rapidly declining response rates and the fact that nonresponse is not a random phenomenon imply that the representativeness of probability samples is in decline. While non-probability samples as well can suffer from high nonresponse rates (Manfreda et al. 2008), survey researchers start to doubt whether the advantages of relying on a probability sample are still worth the additional costs (Sanders et al. 2007). As Groves (2006, p. 668) indicates, '[u]nder some circumstances, probability samples with high nonresponse that are drawn from sparse sampling frames may lose out to nonprobability samples from rich sampling frames with powerful adjustment models.' Significant efforts are made by internet survey firms to improve the quality of non-probability online surveys. Such efforts include a stronger focus on the recruitment of panelists and sophisticated weighting schemes. While such weights can obviously be applied to probability samples as well, the implementation of such corrections means that the comparative advantage of probability over non-probability surveys is declining. That is, after weighting, both types of samples are more equal.

In summary, while declining response rates imply that the representativeness of probability samples is deteriorating, the reliability of online non-probability samples might be improving. As a result, scholars of electoral research and survey researchers more generally cannot take for granted that the more traditional approaches to sampling and surveying will result in higher quality data. Furthermore, given the fact that research budgets in several countries are in decline, it be-

2. For the British Election Studies, the response rate has dropped from 79% in 1963 (Crewe, Sarlvik, and Alt 1977) to 56% in 2015 (Fieldhouse et al. 2016); in Canada, the response rate has dropped from 63% in 1965 (Converse et al. 2002) to 37% in 2015 (Northrup 2016); in Australia, the response rate has declined from 63% in 1987 to 23% in 2015 (Cameron and McAllister 2016).

comes harder and harder to defend opting for a survey of somewhat better quality when the extra cost for such a survey is substantial. Internet surveys are no longer a 'promising alternative' (Chang and Krosnick 2009) in electoral research, they are now effectively used in election studies worldwide. Examples include the British Election Study Internet Panel (Fieldhouse et al. 2017), the French 2017 Election Study (Foucault 2017) and the Cooperative Congressional Election Study (Vavreck and Rivers 2008). Somewhat reassuringly, furthermore, comparisons based on parallel election surveys conducted in different modes point towards relatively minor differences, in particular when correlations and relations between variables are focused upon (Breton et al. 2017; Bytzek and Bieber 2016; Pasek 2016; Sanders et al. 2007; Stephenson and Crête 2010).

It would, however, be all too easy to take the results from a select number of countries[3] as sufficient evidence that election researchers in other countries (our focus is on the Belgian case) can safely make the transition to conducting their national election studies online without a cost in terms of the quality and representativeness of their data. Before transitioning to online survey modes, it is of foremost importance that scholars verify whether the same conclusions hold in their own countries (Sanders et al. 2007). Verifying whether differences between traditional probability samples and online non-probability samples are minor is important because the quality of online surveys is likely to differ strongly from one country to another. Their quality likely depends on the internet penetration in a country[4], and on the reliability and experience of the firms conducting internet surveys in a particular country. Furthermore, the comparison between probability and non-probability online samples crucially depends on the nature, quality and sampling method of the 'traditional' probability sample as well. The Belgian case is rather unique in this regard, as probability samples can be drawn from the National Register of persons, ensuring a high quality initial sample.

In this paper we compare estimates and inferences drawn from two parallel surveys that were in the field in the context of the 2014 regional, federal and European

3. i.e., Canada, Germany, Great Britain and the United States. See also Appendix A.

4. According to the United Nations, by 2014, 85% of Belgians were internet users. This is substantially higher than what was the case 5 or 10 years before (70% and 54% respectively). For more information, see http://data.un.org/Data.aspx?d=WDI&f=Indicator_Code%3AIT.NET.USER.P2.

elections in Belgium. We compare the data from the PartiRep[5] face-to-face survey, that was based on a representative sample of citizens drawn from the National Register, with the data from the online non-probability survey that was fielded simultaneously by the Making Electoral Democracy Work (MEDW) project.[6]

To preview our findings, our analyses indicate noteworthy differences in the marginal distributions of some socio-demographic variables and suggest that the PartiRep probability survey more closely represents the population under study. Correcting for the underrepresentation of particular groups by means of weights, we still find substantial differences in the point estimates of a number of key political variables. When focusing on explanatory models and the inferences that scholars of electoral behaviour draw from such models, in contrast, differences appear to be fairly limited.

2 Why studying mode and sampling effects in Belgium?

Scholars increasingly rely on data gathered online through nonprobability sampling frames and electoral research is no exception in this regard. In several countries, the impact and consequences of relying on an online nonprobability survey instead of a traditional survey and sampling approach are fairly well documented. Using data from parallel surveys that were collected under different survey modes, previous work has studied differences in marginal distributions as well as how survey modes affect the relationships between variables and inferences.

Previous work on sampling and mode effects has mainly focused on two sets of contrasts. First, several studies offer insights on differences and similarities between face-to-face probability and online nonprobability surveys (Baker et al. 2010; Bytzek and Bieber 2016; Malhotra and Krosnick 2007; Sanders et al. 2007). Second, a large number of studies have studied sampling and mode effects while contrasting random digit dialing (RDD) telephone surveys and online nonprobablity surveys (Ansolabehere and Schaffner 2014; Breton et al. 2017; Bytzek and Bieber 2016; Pasek 2016; Stephenson and Crête 2010). Given the rapid decline in the number of adults that have a landline telephone connection (Keeter et al. 2007), election studies that rely on random digit dialing (RDD) to sample respondents in particular are likely

5. PartiRep stands for 'Participation and Representation', an Inter-University Attraction Pole that was funded by the Belgian Science Policy. More information on the project can be found at `www.partirep.eu/`.

6. More info on this research project can be found at `electoraldemocracy.com/`.

to face more important challenges in terms of the representativeness of their surveys than what holds for countries where sampling is done on the basis of a list of households, a national register or a voting register.

The studies cited offer valuable insights in the consequences of transitioning to online nonprobability surveys in a varied set of countries, including Canada, Germany, Great Britain, and the United States. Investigating the implications of relying on an online nonprobability sample instead of a traditional probability sample for the study of electoral behaviour in Belgium is not only a geographical extension of this previous work. More importantly, our focus on the Belgian case implies that we study a contrast in sampling approach that—upon our knowledge—has not been investigated before.

The survey mode that is traditionally relied on in the context of Belgian election studies, face-to-face and telephone interviews, corresponds to the 'traditional' approaches under investigation in previous work (Baker et al. 2010; Bytzek and Bieber 2016; Malhotra and Krosnick 2007; Sanders et al. 2007). However, Belgian election studies are unique because of their sampling approach, with a sample of individual respondents being drawn from the National Register. This Register is the authoritaritive source of accurate and up to date information on all citizens. Furthermore, voter registration is automatic in Belgium. As a result, relying on a sample from the National Register can be considered as the best possible source for constructing a probability sample of voters in Belgium.

3 Previous studies

As evident from the previous section, a large number of studies offer a systematic comparison of probability and nonprobability samples in the field of electoral research. We summarize the main findings here.

First, previous work has reported that when looking at the marginal distributions of variables that can be compared to some validated benchmark, such as reported votes, online nonprobability surveys tend to result in less accurate estimates compared to probability samples (Bytzek and Bieber 2016; Pasek 2016; Stephenson and Crête 2010), though there are exceptions (Breton et al. 2017).

Second, scholars comparing probability and nonprobability samples have reported important differences in terms of the political attitudes and political engagement of respondents in the two types of samples. A number of studies find that internet panels tend to be more knowledgeable and more interested in politics (Ansolabehere and Schaffner 2014; Schoen and Faas 2005). Others find—in contrast—that online samples suffer less from self-selection in terms of political involvement than what holds for an RDD sample (Breton et al. 2017).

Third, scholars have focused on differences in terms of the relations between variables (Alvarez, Sherman, and VanBeselaere 2003; Ansolabehere and Schaffner 2014; Breton et al. 2017; Bytzek and Bieber 2016; Malhotra and Krosnick 2007; Sanders et al. 2007; Stephenson and Crête 2010). Malhotra and Krosnick (2007) stress the differences—and their substantive importance—when comparing vote choice and participation models estimated on data from a probability or a nonprobability sample. Most other studies, however, focus on the large similarities in terms of inferences drawn from probability and nonprobability surveys (Ansolabehere and Schaffner 2014; Berrens et al. 2003; Breton et al. 2017; Bytzek and Bieber 2016; Sanders et al. 2007; Stephenson and Crête 2010). From a comparison of differences between Internet, mail and phone surveys, for example, Ansolabehere and Schaffner (2014, p. 301) conclude that 'researchers will not consistently get more accurate results, nor reach substantially different conclusions, when using one mode relative to another'.

In summary, the conclusions from earlier research on sampling and mode effects in electoral research are rather mixed. On the one hand, several studies have reported on significant and rather large differences between (face-to-face) probability surveys and online nonprobability surveys (Chang and Krosnick 2009; Malhotra and Krosnick 2007; Yeager et al. 2011). Malhotra and Krosnick (2007, p. 312), for example, conclude that 'results may differ considerably depending upon the mode/sampling method employed'. A number of recent papers in particular, however, have evaluated more positively the quality and usefulness of data from nonprobability online surveys (Ansolabehere and Schaffner 2014; Breton et al. 2017; Bytzek and Bieber 2016; Pasek 2016; Sanders et al. 2007; Stephenson and Crête 2010). Importantly, while these studies evaluate the usefulness of nonprobability surveys for electoral research rather positively, they generally find important differences in terms of the marginal distributions of variables measured in different types of surveys. While such differences are acknowledged, it is argued that inferences are largely similar across sampling approaches and survey modes and that it is explanatory models that are of most interest to students of electoral research. In this paper, therefore, we compare the PartiRep and MEDW-samples not just in terms of the descriptives and the distribution of variables, but we also look at the effect these differences have when explaining the vote choice.

While previous research has compared the quality of internet surveys with both RDD-samples and with face-to-face surveys[7], it is important to point out that the conclusions from these two sets of comparisons are equally mixed. When only focusing on the internet/face-to-face comparison—a comparison we pursue in this paper as well—we find both studies that highlight differences between internet and face-to-face surveys (Malhotra and Krosnick 2007; Schoen and Faas 2005) and studies that conclude that differences between both survey types are minor (Bytzek and Bieber 2016; Sanders et al. 2007).

4 The Two Belgian surveys

On 25 May 2014, elections for the federal, regional and European parliaments were held in Belgium (André and Depauw 2015). For studying electoral behavior and its determinants in this election, the Inter-University Attraction Pole PartiRep commissioned a reputed commercial company to administer a pre- and post-electoral survey among a representative sample of citizens eligible to vote in the two main regions of Belgium: Flanders and Wallonia (Deschouwer 2018). Simultaneously, the MEDW team collaborated with a different commercial survey company to field a two-wave online survey among voters in Belgium. Since the two studies were commissioned by overlapping teams of scholars, both studies on politics and elections resulted in a number of similarities in terms of the concepts measured and question wording, allowing for a comparison of responses in the two samples. The timing of both surveys, too, is similar.

An important difference between the two studies, however, is the target population. Due to the bilingual character of the Brussels capital region in Belgium, and the associated cost of face-to-face interviews in this region, the PartiRep survey did not target voters in Brussels, while the MEDW survey did. It could do so exactly because it was not a face-to-face survey. Conducting a face-to-face survey in a region where respondents speak one of two main languages in the country, while the interviewers only know which language that is when they make their

7. Ansolabehere and Schaffner (2014) include a mail survey in their comparison as well. For a full overview of the exact sampling- and mode-comparisons in the studies cited in this literature review, see the supplementary materials.

first contact, requires either perfectly bilingual interviewers or the need for sending another interviewer mastering the correct language after a first contact has already been made. Both possible solutions add to the cost of the survey, and make a face-to-face survey in a bilingual region extremely expensive. An online survey can simply prepare the questionnaire in each language, and let the respondents choose their language when they start the survey. For reasons of comparability, we will analyze differences and similarities between the PartiRep data and the MEDW-data for Flanders and Wallonia only, excluding Brussels for which there is no comparison possible. Missing Brussels means missing some 10% of the Belgian voting population.

4.1 The PartiRep face-to-face probability survey

The target population for the PartiRep survey consisted of voters living in Flanders (Dutch-speaking region) or Wallonia (French-speaking region). In order to limit non-response due to inabilities to participate, it was decided to limit the target population to citizens younger than 85. Respondents were selected from the National Register of persons, based on a geographically stratified sample of 274 sampling points. The number of sample points by district was proportional to its number of inhabitants. A total of 4511 addresses were used, and only the selected person living at that address was allowed to be surveyed.

By the end of the fieldwork for the first wave, 2019 interviews had been conducted (1001 in Flanders and 1018 in Wallonia), or a response rate of 45%. Interviews were conducted face-to-face, and interviewers made use of CAPI software. The average interview time was 68 minutes. During the face-to-face interview, respondents were recruited for the post-electoral wave of the survey. A total of 76% of all wave 1 respondents took part in this 20 minute telephone interview.

Considerable efforts were made to limit non-response for each of the two survey waves. Selected persons received an introduction letter before the start of the fieldwork. For the face-to-face surveys, interviewers were instructed to attempt contacting respondents at least four times at different moments in time over a period of at least two weeks (including during the weekend and after 6 PM). All participants to the first wave received another letter and a monetary incentive (€10) before the start of the second wave. During the field work for the second wave, interviewers made up to twelve attempts to contact respondents (PartiRep 2014).

4.2 The MEDW online non-probability survey

The MEDW survey used a stratified quota sampling method, with quotas for respondents' age, gender and level of education.[8] Respondents were recruited among active panelists of different panel providers that are active in the Flemish and Walloon region.[9] The quality of responses was verified to identify and exclude respondents who took too little time to fill out the questionnaire or who misrepresented themselves.

The fieldwork for the first wave resulted in a total of 2061 completed surveys (1026 in the Flemish region and 1035 in Wallonia) and the participation rate for this first wave is estimated to be about 7%. A total of 66% of these respondents also participated to the post-electoral survey wave. The average length was about 20 minutes for the pre-electoral survey and around 10 minutes for the survey that was fielded after the election.

Respondents received an invitation that included the link to the survey as well as a password.[10] Respondents that had not completed the survey received reminders and each of the panel providers offered incentives—usually points—for completing the surveys (MEDW 2014).

4.3 Comparing the two surveys

Table 1 summarizes some important information regarding both studies. As can be read from the table, the sample size of the pre-electoral wave was around 2000 for both studies, but attrition for the second wave was somewhat higher in the online MEDW survey. The higher response rate after an initial face-to-face survey is in line with previous work (Lynn 2015).

The initial response rate for the PartiRep survey was 45%, which is substantially higher than the estimated participation rate in the MEDW online survey (7%) but well below a level that was considered acceptable some decades ago (Curtin, Presser, and Singer 2005; Morton et al. 2012). Over time, response rates in surveys have dropped dramatically, and the Belgian case appears to be no exception in this

8. Quota for age were based on three broad age-categories; 18–34 years, 35–54 years and 55–99 years. The education quota as well were based on three categories; lower secondary education, upper secondary education and tertiary education.

9. For the Flemish region: GMI, HPOL and Toluna. In Wallonia: GMI, HPOL, Toluna and SSI.

10. Passwords were sent to panelists along with the URL to the survey. In this way, access to the survey was controlled and only selected panelists could participate.

regard.

Finally, Table 1 draws attention to an important advantage of online surveys; while the field work of the PartiRep survey took place over an extended period of time (eight weeks for wave 1 and two weeks for wave 2), the data for the online MEDW surveys could be gathered rather quickly, reducing the potential impact of contextual events on participants' responses and attitudes (Sanders et al. 2007).

Table 1: Summary information about the two surveys

Survey	Mode	Sample size	Field dates	Response rate	Retention rate
PartiRep, wave 1	Face-to-face	2019	20/3 to 17/5	45%	–
PartiRep, wave 2	Telephone	1532	16/5 to 1/7	34%	76%
MEDW, wave 1	Online	2061	13/5 to 21/5	7%	–
MEDW, wave 2	Online	1358	26/5 to 9/6	4%	66%

Note: The response rates for the the MEDW surveys are an estimation of the participation rate, based on the number of respondents willing to participate (including those participating while quota had already been met) on the total number of usable invitations.

It should be noted that in comparing the PartiRep face-to-face probability survey and the MEDW online non-probability survey, any differences can be driven by differences in the sampling strategy as well as by the different interview modes of the two studies. While we cannot disentangle the effects of both sources of variation, the study of these compound effects is relevant because this is exactly the choice set for electoral researchers in Belgium: That is, to continue with the traditional approach for conducting election surveys in Belgium or to transition towards online non-probability surveys (Bytzek and Bieber 2016).

In contrast to previous work that has compared probability and non-probability surveys in the field of electoral research, the Belgian probability sample is unique because of its reliance on the National Register. This approach can be considered the 'gold standard' for obtaining a high quality and representative sample. Comparing the data from an online non-probability sample with such a high-quality survey, we can ascertain that if differences between both survey approaches would be minor, a lack of difference is definitely not the result of a comparison with a sub-par probability sample.

5 Representativeness: Comparison with validated information

5.1 Socio-demographics

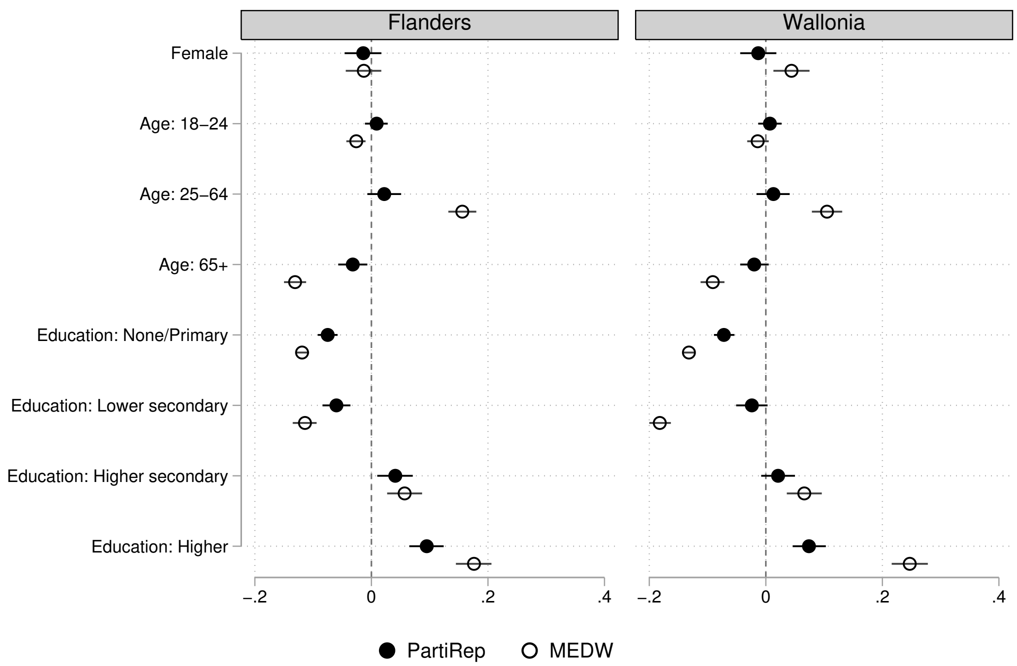
We first verify the representativeness of the two surveys by comparing the characteristics of survey respondents with validated benchmarks among the population (i.e., the voting age population in both regions). For a limited number of socio-demographic variables—gender, age and level of education—distributions can be compared to an approximation of the population parameters as provided by official government statistics. Obviously, the under- of overrepresentation of particular groups can be corrected for by means of weights. Nevertheless, it is important to verify how large the differences are between the raw data and the population, because large weights can distort findings when analyzing weighted data.[11]

Looking at the subsamples of respondents in the most populous Flemish region first, both datasets can be considered representative in terms of respondents' gender (see the left panel in Figure 1, estimates are reported in Appendix B). Differences between the population parameters and the characteristics of the samples are larger for age and level of education. More specifically, the older and the lower educated groups of the population are underrepresented in both surveys. It is important to note, however, that the differences with the population benchmarks are consistently smaller for the PartiRep probability survey than for the MEDW data. Overall, for the eight socio-demographic subgroups included in this comparison, the average difference between the estimate in the PartiRep sample and the population parameter is 0.044, while it is 0.099 for the MEDW sample. The mean squared error (MSE), referred to by Ansolabehere and Schaffner (2014, p. 291) as 'the most common measure of TSE [Total Survey Error]', for both surveys is 0.003 and 0.013 respectively. In summary, while distortions between the samples and the Flemish population are reasonably limited for the PartiRep date, errors are two to three times as large in the MEDW online nonprobability data.

The right panel in Figure 1 shows more pronounced differences between the two datasets for the Walloon subsamples (for detailed estimates see Appendix B).

11. This was the case for the U.S.C. Dornsife/Los Angeles Times Daybreak poll during the 2016 presidential elections in the United States, cfr. www.nytimes.com/2016/10/13/upshot/how-one-19-year-old-illinois-man-is-distorting-national-polling-averages.html

Figure 1: Comparing the samples with benchmarks for socio-demographics



Note: Difference between sample means and the benchmarks, with 95% confidence intervals. Estimates are based on unweighted data. Information on benchmarks is obtained from the Federal Ministery of Economic Affairs (`https://bestat.economie.fgov.be`).

In fact, differences between population benchmarks and the characteristics of the PartiRep sample are only significant for education—with the lowest educated being underrepresented and those who hold a higher education degree being overrepresented. Differences are more pronounced for the data from the MEDW survey. Figure 1 shows that with only one exception, the Walloon respondents in the MEDW survey differ significantly from the target population. This difference is also evident from estimates of the total survey in the Walloon subsamples. The average difference on the eight indicators included in Figure 1 is 0.031 in the PartiRep-data and 0.110 in the MEDW-data. The MSE-statistics for both datasets amount to 0.002 and 0.017. In summary, depending on what estimate is relied on, differences between the population parameters and the characteristics of the survey respondents are 3.5 to 8.5 times larger for the online nonprobability sample compared to the probability sample.

In short, an analysis of the representativeness of the raw data of the PartiRep- and MEDW-surveys—based on a comparison with known population parameters—suggests the PartiRep-sample matches the target population more closely. This is in line with what could be expected given the probability sampling frame of the PartiRep survey and its substantially higher response rate.

Thus far we have worked with the unweighted, raw data. However, as is well known, surveys routinely use population weights to correct or this kind of sampling errors. The availability of validated information regarding the socio-demographic characteristics of the target population allows for the creation of such weights that can be applied to correct for the underrepresentation of some groups in the survey samples. Both the PartiRep- and the MEDW-datasets include a basic socio-demographic weight.[12] As evident from Appendix C, applying these weights somewhat reduces the distance between samples and benchmarks, though errors remain. In what follows, we consistently apply socio-demographic weights, as in practice, researchers will use the weighted dataset for their analyses.

12. We apply the `FINALweightg` weight for the PartiRep dataset and the `WEIGHT1` weights for the MEDW-data. For the pre-electoral wave, the PartiRep weight varies between .55 and 3.19, with a mean of 1.00 and a standard deviation of .43. The MEDW basic socio-demographic wave for the pre-electoral wave has a minimum value of .54 and a maximum value of 3.71. Its mean is 1.00 with a standard deviation of .43.

5.2 The vote choice

The second type of benchmark that allows verifying the representativeness of the samples is the election result. For this comparison, we might expect larger differences, as we know that there might not just be sampling error, but that we could also expect non-sincere answers. For this comparison we compare marginal distributions of the datasets while employing the socio-demographic weights. We focus on two indicators of interest; electoral participation and the vote choice of those who turned out to vote. Regarding electoral participation, it is important to point out that the system of compulsory voting results in high turnout rates. For the 2014 election, the turnout level was 89.7 per cent (87.9 per cent in the Walloon region; 92.5 per cent in the Flemish region).

In Figure 2 we present the differences between marginal distributions for both datasets and the actual election results. Our focus is on votes for the regional Parliament elections. In line with what is generally observed in election studies (Selb and Munzert 2013), electoral participation is overestimated in both datasets. Given that voter turnout is compulsory in Belgium, turnout is already very high. For both datasets, however, the turnout estimates are significantly higher than the actual turnout level of 93% and 88% in the Flemish and Walloon regions. The estimates for blank and invalid voting—the only viable 'exit-option' in a compulsory voting country (Hooghe, Marien, and Pauwels 2011)—are the mirror image of what we find for turnout. That is, those casting blank and invalid votes are underrepresented in both the PartiRep and the MEDW data (detailed estimates are reported in Appendix D).

Looking at the marginal distributions for the reported votes in the Flemish region (left panel in Figure 2, it can be observed that the PartiRep data offer a fairly accurate reflection of the actual election results.[13] We find a vote share estimate that is significantly different from the election results for only one party; the vote share of the populist radical right party Flemish Interest is underestimated (the party obtained 5.9% of the votes, while the estimate based on the PartiRep data is 3.7%). The underestimation of support for a right-wing party is in line with previous work on vote intention surveys (Durand, Blais, and Larochelle 2004)

The differences between the official election results and respondents' reported voting behavior are somewhat larger for the MEDW data. Figure 2 illustrates that

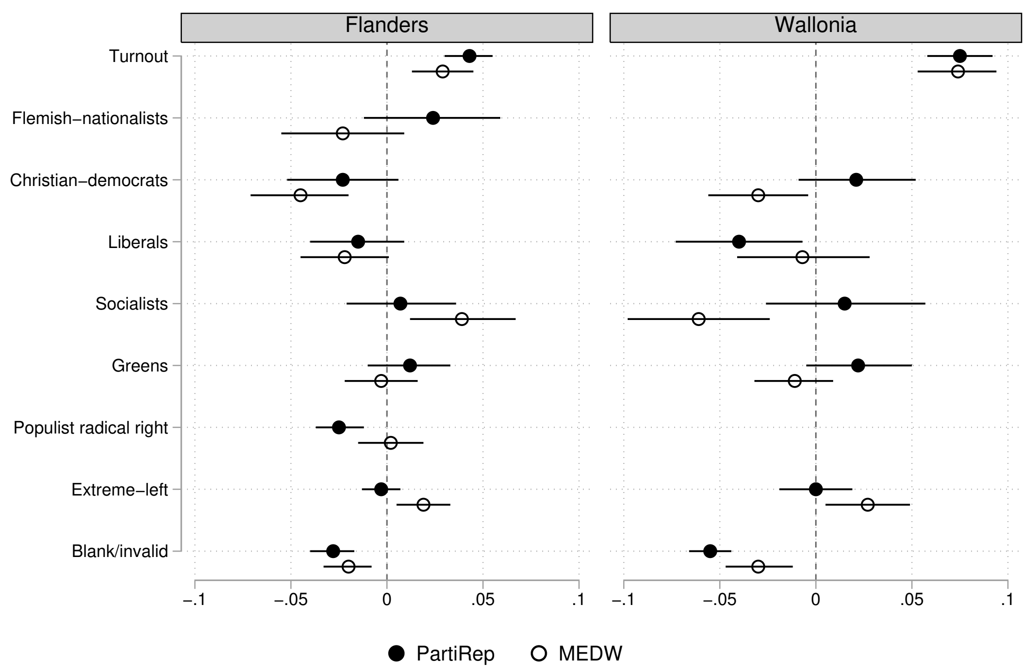
13. We compare the reported votes in the samples with the vote shares that parties obtained in the Flemish region, excluding voters in Brussels.

the online survey significantly underestimated the proportion of Christian-democrats and Liberals and significantly overestimated proportions of Socialists and voters of the extreme-left. This pattern of survey errors, with an overestimation of parties on the left and parties on the right being underestimated, corresponds to what has been observed in other countries as well (Bytzek and Bieber 2016; Mellon and Prosser 2017). While errors are somewhat larger in the MEDW data, it should be stressed that the online MEDW survey resulted in more accurate estimates of the vote share of the populist radical right party.

Summarizing the differences between the marginal distributions in the samples and the actual election results of the 2014 Flemish Parliament election (excluding votes cast in the Brussels capital region), nuances to some extent the differences that we have found. As evident from the TSE-statistics in Appendix D, the average difference between the estimates and the election results is about 2 percentage points for both datasets while the MSE is .001 for both the PartiRep and the MEDW samples. When we restrict these summary indicators to indicators of the vote choice only—excluding turnout—the average difference between the reported vote choice and the election results is .017 for the PartiRep data and .021 for the MEDW-data. Overall, while a number of differences are significant, and while we find more statistically significant deviations from the actual election results in the MEDW-sample, both surveys offered fairly reliable estimates of the vote shares that parties obtained in the 2014 Flemish regional elections. Thus far, therefore, the conclusion has to be that we do not find marked differences between both survey modes.

Turning to the Walloon samples, the right panel in Figure 2 shows that—in line with what we observed for the Flemish subsamples—both surveys overrepresent voter turnout and underrepresent the proportion of blank and invalid votes. In terms of the vote shares obtained by different parties, it can be observed that Liberal party voters are significantly underestimated in the PartiRep dataset. For all other parties, the estimated vote shares are not significantly different from the actual vote shares that parties obtained in the 2014 Walloon regional elections. The underestimation of the Liberals is in line with previous work that has indicated that right-wing parties tend to be underrepresented in vote intention surveys. Again, we find that differences are somewhat more pronounced when we consider the MEDW-data. The online survey significantly underestimates Christian-democrats and significantly overestimates the proportion of extreme-left voters. In addition,

Figure 2: Comparing reported vote choices with election results



Note: Difference between sample means and election results, with 95% confidence intervals. Socio-demographic weights are applied. Information on election results obtained from the Ministery of Internal Affairs (`https://verkiezingen2014.belgium.be`).

and in contrast to what is generally found in vote intention research, the online survey significantly underestimated the proportion of Socialists in the Walloon region.

Overall, however, the average difference between the estimates and the election results is fairly limited. The average difference is .033 for the PartiRep-survey and .034 for the MEDW-survey while the MSE is 0.002 for both datasets (see Appendix D). Excluding turnout, both samples estimated the vote shares of different parties with roughly the same amount of error (average difference of .026 in the PartiRep data and .028 in the MEDW sample).

Looking at the marginal distributions in terms of turnout and respondents' electoral behavior, both surveys overestimate turnout and underestimate blank and invalid votes. Furthermore, both surveys underrepresent some electorates and overrepresent the electorates of other parties. Assessing the extent to which the two surveys differed from the actual electoral results, we cannot conclude that one survey clearly outperformed the other in terms of the accuracy of the estimated electoral results.

6 Relations and inferences: explaining vote choice

While we find a number of noteworthy differences in distributions of variables between the two surveys, most analysts studying electoral behavior do not focus on marginal distributions but rely on the data to test hypotheses about the relation between variables. It is therefore important to verify whether scholars studying the determinants of particular political attitudes or behaviors would come to substantively the same conclusions irrespective of whether they relied on a face-to-face probability sample or an online nonprobability sample. Previous work on the differences between probability samples and non-probability online samples in the field of electoral research has mostly concluded that despite important differences in marginal distributions, causal inferences are remarkably similar for both types of survey data (Breton et al. 2017; Bytzek and Bieber 2016; Pasek 2016; Sanders et al. 2007; Stephenson and Crête 2010). In this section, we verify whether the same holds for the Belgian 2014 elections, and the PartiRep and MEDW-surveys in particular. Our focus is on explaining vote choice.

For evaluating the differences between the two datasets, we follow the approach of Malhotra and Krosnick (2007) and Stephenson and Crête (2010) and present a series of bivariate regression models. We add substantive predictors of vote choice and their interaction with a dummy variable identifying the surveys. Doing so, we can verify whether the relation between an independent variable and the vote choice differs significantly between the two surveys. As independent variables, we select variables that were consistently measured and that are theoretically related to vote choice in a Belgian context; we focus on the socio-demographic variables gender, age and levels of education. Furthermore, we look at the role of the ideological (left-right) distance between a respondent and a party, their evaluation of the state of the economy, the strength of their regional identity as well as the extent to which they indicate to like a party. All variables, with the exception of age, are rescaled to run from 0 to 1. Following Malhotra and Krosnick (2007), we present bivariate regressions, as the results of multivariate regression models strongly depend on the variables that are included in the models. This issue is of particular importance in the context of logistic regression models (Mood 2010), which is the estimation approach we employ for analyzing voters' party choices. As such, the analyses give insights in how the sampling and survey modes affect the effect of each of the independent variables, without suppressing effects by the inclusion of additional variables that scholars might or might not include in their models. We report the results of multivariate regression models that include all the independent variables simultaneously in Appendix E.

We present a series of models for explaining voting for the main parties in both regions (Flemish-nationalists, Christian-democrats, Liberals and Socialists in the Flemish region and Christian-democrats, Liberals and Socialists in the Walloon region).[14]

Table 2 presents the results for the four main parties in the Flemish region. The significant main effects are in line with expectations about the determinants of voting behavior in general and the Flemish region in particular (Dassonneville and Stiers 2018; Deschouwer 2018). That is, we find a strongly negative and significant effect of left-right distance, implying that as the ideological distance between a voter and a party increases, her probability of voting for this party decreases. Furthermore, we find that respondents who evaluated the state of the economy more positively were significantly more likely to vote for the Christian-democratic

14. We do not estimate models explaining the vote for some of the smaller parties, as the datasets only included small numbers of respondents who voted for green, populist radical right and extreme-left parties. It is important to note that Figure 2 indicated that differences in the reported vote choices are somewhat larger for radical parties. As a result, not investigating the determinants of voting for these parties might lead us to underestimate differences between the two surveys.

party—the party of the incumbent Minister President of the Flemish government. Evaluating the economy more positively also significantly increased the likelihood of voting for the Liberals. While the Liberals were not a member of the governing coalition at the regional level, they were in government at the federal level. Similarly, we find that evaluating the state of the economy more positively significantly decreased the probability of voting for the Flemish-nationalist party, despite the fact that the party was member of the incumbent regional government (it was, however, in opposition at the federal level). We also find that being more strongly attached to the Flemish region significantly increases the probability of voting for the Flemish nationalists and significantly decreases the likelihood of a vote for the Socialist party. Not surprisingly, giving a party a higher rating on the like/dislike scale significantly increases the probability of voting for this party. Finally, a limited number of socio-demographic variables appear to be significantly related to voting for one of the four main parties; older respondents are significantly more likely to vote for the Flemish nationalists while a higher level of education significantly decreases the probability of casting a vote for the Socialist party. In summary, these main effects are line with previous work on the determinants of the vote choice in Belgium.

More importantly, out of a total of 28 interaction-terms, only five are significant at conventional levels.[15] Furthermore, four of these five significant interaction effects relate to the like/dislike scales by means of which respondents could indicate the extent to which they like a particular political party. We suspect that these differences are partly driven by question format effect. More specifically, the visual presentation of these scales in the context of the MEDW questionnaires appears to have resulted in more variation along the 0-10-scale, while respondents to the PartiRep-questionnaire were more likely to use the scores 0, 5 and 10 for rating a party. Formally examining the equality of the distributions of respondents' answers on these like/dislike scales by means of Kolmogorov-Smirnov test confirms that distributions in the two samples are indeed different.

The results in Table 2 include five significant interaction effects. It could, however, be argued that we engage in multiple testing—increasing the probability that we will find significant effects by mere chance. One way to correct for this issue is

15. When estimating the bivariate models without applying the socio-demographic weight, six interaction terms are significant (results available from the authors). Weighting thus only has a marginal impact when focusing on explanatory models.

to apply stricter significance levels, that is, to apply a Bonferroni correction.[16] Three out of 28 interaction terms are significant under this stricter test.

16. For a desired significance level of .05, we divide .05 by the number of tests. In this case, .05/28 results in a p-value threshold of .002 (Gelman, Hill, and Yajima 2012).

Table 2: Comparison of the two surveys in logistic regression models explaining vote choice, Flemish region

Party	Independent variable	Main effect	Online survey	Interaction	Constant	N	Pseudo R^2
Christian-democrats	Female	.164	.051	-.426	-1.525***	1536	.003
	Age	.011	-.277	.003	-1.985***	1536	.033
	Education	-.029	-.526	.535	-1.424***	1536	.003
	Left-right distance	-35.679**	-.022	-4.499	-.912***	1450	.030
	Economic evaluation	.567*	-.299	.337	-1.177***	1495	.013
	Flemish identity	.644	-.437	.420	-1.945***	1517	.006
	Like/dislike party	66.898***	-4.159***	**47.783****	-5.593***	1518	.280
Flemish-nationalists	Female	-.207	-.252	.088	-.474***	1536	.003
	Age	.014**	.245	-.009	-1.266***	1536	.007
	Education	-.227	-.649*	.652	-.430*	1536	.004
	Left-right distance	-48.513***	.088	-8.907	.308*	1450	.096
	Economic evaluation	-.713**	-.073	-.570	-.164	1495	.027
	Flemish identity	3.182***	.031	-.272	-3.100***	1517	.055
	Like/dislike party	68.482***	-2.218**	21.363*	-4.521***	1523	.418
Socialists	Female	.394	.312	-.105	-1.919***	1536	.007
	Age	-0.000	.116	.003	-1.709***	1536	.003
	Education	-1.043**	.268	.076	-1.088***	1536	.019
	Left-right distance	-31.256**	.604*	-13.848	-1.014***	1450	.051
	Economic evaluation	.049	-.420	1.311**	-1.753***	1495	.025
	Flemish identity	-1.943***	0.145	.498	-.260	1517	.025
	Like/dislike party	70.571***	-5.344***	**65.376*****	-5.923***	1518	.393
Liberals	Female	-.056	-.231	.351	-1.870***	1536	.002
	Age	-.011	-.458	.008	-1.379***	1536	.003
	Education	.795	.407	-.695	-2.431***	1536	.005
	Left-right distance	-20.303*	-.024	-8.592	-1.554***	1450	.014
	Economic evaluation	.749*	.310	-.475	-2.370***	1495	.006
	Flemish identity	-.148	.606	-.913	-1.782***	1517	.005
	Like/dislike party	64.502***	-4.610***	**56.072*****	-5.637***	1515	.306

Note: Estimates of logistic regression analyses explaining voting for Christian-democrats, Flemish-nationalists, Socialists and Liberals in the Flemish region. Dependent variables were coded 1 if respondents reported to have voted for the party and 0 if they reported to have voted for another party (either one of the parties included in our analyses or a smaller party). Non-voters are excluded. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Interaction coefficients in bold are still significant after applying a Bonferroni correction for multiple testing.

Next, we look at the results of the bivariate vote choice models for the Walloon region. Estimates of these analyses are presented in Table 3. Overall, the predictors of interest appear to be somewhat less powerful for explaining vote choice in Wallonia compared to what we found for the Flemish region. We find strong and significant effects of the like/dislike scales and some significant effects of the socio-demographic variables. Economic evaluations and the strength of respondents' identification with the Walloon region, in contrast, are not significantly correlated to voting for one of the three main parties (Socialists, Liberals and Christian-democrats). Finally, for the ideological distance between a respondent and a party, we find the expected negative effect—though for the Socialist party this effect is not significant when considering the pooled sample (i.e., no significant main effect).

Of most interest for this paper, however, are the interaction effects—that give an indication of whether effects and effect sizes are dependent on sampling and mode effects. The estimates in column five in Table 3 indicate that six out of 21 interaction-effects attain statistical significance.[17] For this region, significant interaction effects are not restricted to the effect of the like/dislike scales. Table 3 gives indications of sampling and mode effects for variables such as ideological distance, economic evaluations, regional identity as well socio-demographics. These differences, however, are not observed consistently across the three parties that are considered in Table 3—though it will not surprise the reader to see more differences for the Socialist party, the party for which reported vote choices in both samples differed most. Applying the Bonferroni-correction for multiple testing to the estimates for the Walloon sample[18], only two out of the 21 interaction-effects included in Table 3 attain significance.

17. Compared to a series of bivariate models on unweighted data, there is no improvement in the number of significant interaction terms. (Results available from the authors.)

18. For a .05-level and 21 tests, the p-value threshold is .05/21, or .002.

Table 3: Comparison of the two surveys in logistic regression models explaining vote choice, Walloon region

Party	**Independent variable**	**Main effect**	**Online survey**	**Interaction**	**Constant**	**N**	**Pseudo R^2**
Socialists	Female	.551**	-.265	-.255	-.872***	1274	.016
	Age	.021***	-.110	-.006	-1.615***	1274	.022
	Education	-1.982***	-.622	.513	.559*	1274	.064
	Left-right distance	-2.978	.791**	**-60.086*****	-.517**	1221	.068
	Economic evaluation	-.133	-.872***	**.958***	-.521**	1240	.017
	Walloon identity	.666	-1.350*	1.161	-1.102**	1258	.024
	Like/dislike party	57.040***	-3.613***	**41.529****	-4.086***	1261	.353
Liberals	Female	-.635**	-.014	.368	-.798***	1274	.010
	Age	-.012*	-.920*	.023**	-.524	1274	.007
	Education	1.530***	.388	-.401	-2.095***	1274	.030
	Left-right distance	-63.019***	.405	-21.145	.074	1221	.127
	Economic evaluation	.220	.252	-1.134	-1.200***	1240	.002
	Walloon identity	-.544	.680	-.699	-.678*	1258	.012
	Party like/dislike	60.090***	-1.960*	25.787*	-4.376***	1252	.327
Christian-democrats	Female	-.004	-.392	-.107	-1.437***	1274	.008
	Age	.007	.888	-.029**	-1.766***	1274	.017
	Education	.529	-.640	.240	-1.771***	1274	.014
	Left-right distance	-41.358***	-.756**	27.189	-.830***	1221	.035
	Economic evaluation	.133	-.482	.087	-1.476***	1240	.008
	Walloon identity	.531	-.685	.331	-1.856***	1258	.011
	Like/dislike party	57.369***	-2.836*	32.466	-4.867***	1246	.224

Note: Estimates of logistic regression analyses explaining voting for Socialists, Liberals and Christian-democrats in the Walloon region. Dependent variables were coded 1 if respondents reported to have voted for the party and 0 if they reported to have voted for another party (either one of the parties included in our analyses or a smaller party). Non-voters are excluded. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Interaction coefficients in bold are still significant after applying a Bonferroni correction for multiple testing.

Combined, the results presented in Table 2 and Table 3 include 49 interaction terms that allow verifying whether the relation between theoretically relevant indicators[19] and vote choice differs significantly depending on whether effects were measured by means of a face-to-face probability sample or an online nonprobability sample. Of these 49 interaction terms, less than a fourth (eleven) are significant at conventional levels. As evident from the results in Appendix E, when including all variables considered here simultaneously in multivariate logistic regression models, the proportion of significant interaction terms is about the same as what we found for the bivariate models presented here (fourteen on a total of 49 interaction effects). And again, these differences stem mostly from the like/dislike-ratings, a measure for which we suspect question format effects to influence results.[20]

Our results on the role of sampling and mode effects on inferences when explaining vote choice[21] point to a number of significant interaction effects. In terms of the proportion of significant differences on the total number of tests, we find roughly the same amount of differences as previous work in the British and Canadian contexts (Breton et al. 2017; Sanders et al. 2007; Stephenson and Crête 2010). Applying a Bonferroni correction, the number of significant differences is further reduced to five on a total of 49 estimated effects (or about 10%) and one might therefore conclude that—overall—researchers studying voting behavior in Belgium would not arrive at strongly different conclusions depending on the sampling approach and survey mode used for collecting their data.

7 Discussion and conclusion

The results of the analyses presented here are largely in line with the findings of earlier work on differences between probability and nonprobability samples in the

19. In a supplementary analysis, reported in Appendix F, we have also verified whether the impact of political interest on vote choice differs in the two samples. For none of the seven parties (four parties in the Flemish region and three parties in the Walloon region) the interaction term – survey × political interest – is significant at conventional levels.

20. For the Walloon sample as well Kolmogorov-Smirnov tests indicate that the distributions of respondents' answers on the party like/dislike scales differ significantly between the two samples – without a single exception.

21. Unfortunately, we do not dispose of a large number of alternative dependent variables for which we could investigate whether the two surveys would lead to different conclusions regarding what explains it. Given the overall high turnout, and overreported turnout, explaining reported turnout is not an option. We did pursue an additional analysis to explain hypothetical turnout under voluntary voting rules. These supplementary analyses did not reveal strong differences between the two surveys that could not be attributed to wording differences.

field of electoral research. That is, we find important differences in point estimates between the two samples, and additional analyses reported in the supplementary materials (Appendix G) suggest that there are important differences when comparing the point estimates and distributions of non-validated measures (e.g., political interest or national identity) between the two surveys. Despite these differences, we find that inferences drawn from explanatory models are similar for the two datasets.

This high level of correspondence between the two surveys is remarkable given the point of reference with which the online nonprobability sample is compared here: a probability sample drawn from the National Register. Our analyses certify the high quality data that this sampling approach results in, as the PartiRep data are consistently closer to validated benchmarks than the MEDW online nonprobability data.

When focusing on inferences and explanatory models—the type of analysis that is central in most published work in the field of electoral research—it appears that differences between the two surveys are modest. We find a small number of significant differences in terms of the vote choice models. Furthermore, the most substantial difference that we did observe, about the relationship between party like/dislike and vote choice, is most likely due to differences in question wording rather than on survey mode as such.

We find differences between the PartiRep and the MEDW-samples, and given that point estimates differ quite strongly a shift towards nonprobability online surveys will disrupt comparisons drawn from repeated cross-sections of electoral studies in Belgium. In the best possible world, with unlimited resources for survey research, scholars should continue relying on samples from the National Register and conduct face-to-face surveys. Resources are limited, however, and so the question is whether we can still afford to conduct face-to-face surveys.

If scholars are not primarily interested in comparisons over time, however, it is not obvious that one sampling or survey mode should be preferred over the other. In that regard, an element that we have not touched upon, yet, is that the price to reach a respondent can easily be eight times as high in face-to-face surveys than in online surveys.[22] The cost by interview was the very reason why it was decided not

22. Although it also has to be noted that the face-to-face interview can be much longer than the standard on-line survey. For that exact reason, it is not straightforward to compare the cost by respondent of the PartiRep-survey with the cost of the MEDW-survey.

to cover Brussels in the PartiRep-survey, which was not a problem for the online survey that the MEDW-project fielded. At a time when science budgets are increasingly being limited, this is a crucial consideration. The evidence presented in this paper for the Belgian context suggests that in these circumstances online surveys are a valuable alternative to more traditional survey modes.

The comparison that we pursued here is rather unique, as we compare a non-probability sample to a probability sample that is of exceptional quality—and that can be labeled as the 'gold standard' for its sampling frame. Such a comparison should in principle be favorable to the probability sample, but we find that differences between these two surveys are fairly limited, at least when we focus on explanatory models. These results are encouraging for anyone who relies on data from online non-probability samples. We are careful, however, not to generalize our results, because every country (even every survey) is different. As more electoral researchers conduct their surveys online, we should continue to verify the reliability and quality of online data, and the move towards conducting parallel studies with different survey modes certainly creates many opportunities for performing such comparisons.

References

Alvarez, R. Michael, Robert P. Sherman, and Carla VanBeselaere. 2003. Subject Acquisition for Web-Based Surveys. *Political Analysis* 11 (1): 23–43.

André, Audrey, and Sam Depauw. 2015. A Divided Nation? The 2014 Belgian Federal Elections. *West European Politics* 38 (1): 228–237.

Ansolabehere, Stephen, and Brian F. Schaffner. 2014. Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison. *Political Analysis* 22 (3): 285–303.

Baker, Reg, Stephen J. Blumberg, Michael J. Brick, Mick P. Couper, Melanie Courtright, J. Michael Dennis, Don Dillman, et al. 2010. AAPOR Report on Online Panels. *Public Opinion Quarterly* 74 (4): 711–781.

Berrens, Robert P., Alok K. Bohara, Hank Jenkins-Smith, Carol Silva, and David L. Weimer. 2003. The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples. *Political Analysis* 11 (1): 1–22.

Breton, Charles, Fred Cutler, Lachance Sarah, and Alex Mierke-Zatwarnicki. 2017. Telephone versus Online Survey Modes for Election Studies: Comparing Canadian Public Opinion and Vote Choice in the 2015 Federal Election. *Canadian Journal of Political Science* 50 (4): 1005–1036.

Bytzek, Evelyn, and Ina E. Bieber. 2016. Does Survey Mode Matter for Studying Electoral Behaviour? Evidence from the 2009 German Longitudinal Election Study. *Electoral Studies* 43:41–51.

Cameron, Sarah M., and Ian McAllister. 2016. *Trends in Australian Political Opinion: Results from the Australian Election Study 1987-2016.* Canberra: Australian National University.

Chang, Linchiat, and Jon A. Krosnick. 2009. National Surveys via RDD Telephone Interviewing versus the Internet. Comparing Sample Representativeness and Response Quality. *Public Opinion Quarterly* 73 (4): 641–678.

Converse, Philip, John Meisel, Maurice Pinard, Peter Regenstreif, and Mildred Schwartz. 2002. *Canadian National Election Study, 1965. ICPSR 7225 [Data file].* Ann Arbor (MI): Inter-University Consortium for Political and Social Research [distributor].

Couper, Mick P. 2000. Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly* 64 (4): 464–494.

Crewe, Ivor, Bo Sarlvik, and James Alt. 1977. Partisan Dealignment in Britain 1964-1974. *British Journal of Political Science* 7 (2): 129–190.

Curtin, Richard, Stanley Presser, and Eleanor Singer. 2005. Changes in Telephone Survey Nonresponse over the Past Quarter Century. *Public Opinion Quarterly* 69 (1): 87–98.

Dassonneville, Ruth, and Dieter Stiers. 2018. Electoral Volatility in Belgium (2009-2014). Is there a Difference between Stable and Volatile Voters. *Acta Politica* 53 (1): 68–97.

Deschouwer, Kris, ed. 2018. *Mind the Gap. Political Participation and Representation in Belgium.* Lanham (MD): Rowman & Littlefield.

Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method.* New York: Harper & Row.

Durand, Claire, André Blais, and Mylène Larochelle. 2004. The Polls-Review. The Polls in the 2002 French Presidential Election: An Autopsy. *Public Opinion Quarterly* 68 (4): 602–622.

Fieldhouse, Ed, Jane Green, Geoffrey Evans, Herman Schmitt, Cees van der Eijk, Jonathan Mellon, and Chris Prosser. 2016. *British Election Study, 2015: Face-to-Face Post-Election Survey. [data collection].* UK Data Service. SN: 7972, `http://dx.doi.org/10.5255/UKDA-SN-7972-1`.

———. 2017. *British Election Study Internet Panel Waves 1-13 [datafile].* DOI `10.15127/1.293723`.

Foucault, Martial. 2017. *L'enquête électorale française (www.enef.fr).* Paris: CEVIPOF SciencesPo.

Gelman, Andrew, Jennifer Hill, and Masanao Yajima. 2012. Why we (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness* 5 (2): 189–211.

Groves, Robert M. 2006. Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly* 70 (5): 646–675.

Hooghe, Marc, Sofie Marien, and Teun Pauwels. 2011. Where Do Distrusting Voters Turn if There is No Viable Exit or Voice Option ? The Impact of Political Trust on Electoral Behaviour in the Belgian Regional Elections of June 2009. *Government and Opposition* 46 (2): 245–273.

Hooghe, Marc, Sara Vissers, Dietlind Stolle, and Valérie-Anne Mahéo. 2010. The Potential of Internet Mobilization: An Experimental Study on the Effect of Internet and Face-to-Face Mobilization Efforts. *Political Communication* 27 (4): 406–431.

Keeter, Scott, Courtney Kennedy, April Clark, Trevor Tompson, and Mike Mokrzycki. 2007. What's Missing from National Landline RDD Surveys? The Impact of the Growing Cell-Only Population. *Public Opinion Quarterly* 71 (5): 772–792.

Lynn, Peter. 2015. Alternative Sequential Mixed-Mode Designs: Effects on Attrition Rates, Attrition Bias, and Costs. *Journal of Survey Statistics and Methodology* 1 (2): 183–205.

Malhotra, Neil, and Jon A. Krosnick. 2007. The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples. *Political Analysis* 15 (3): 286–323.

Manfreda, Katja Lozar, Jernej Berzelak, Vasja Vehovar, Michael Bosnjak, and Iris Haas. 2008. Web Surveys Versus Other Survey Modes: A Meta-Analysis Comparing Response Rates. *International Journal of Market Research* 50 (2): 269–291.

MEDW, Making Electoral Democracy Work. 2014. *Making Electoral Democracy Work. Belgium Regional, National and European Election–Brussels, Flanders and Wallonia.* Technical report. Montréal.

Mellon, Jonathan, and Christopher Prosser. 2017. Missing Nonvoters and Misweighted Samples. Explaining the 2015 Great British Polling Miss. *Public Opinion Quarterly* in press.

Mood, Carina. 2010. Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review* 26 (1): 67–82.

Morton, Susan M.B., Dinusha K. Bandara, Elizabeth M. Robinson, and Polly E. Atatoa Carr. 2012. In the 21st Century, What is an Acceptable Response Rate? *Australian and New Zealand Journal of Public Health* 36 (2): 106–108.

Motta, Matthew P., Timothy H. Callaghan, and Brianna Smith. 2016. Looking for Answers: Identifying Search Behavior and Improving Knowledge-Based Data Quality in Online Surveys. *International Journal of Public Opinion Research* 29 (4): 575–603.

Northrup, David. 2016. *The 2015 Canadian Election Study.* Toronto: Institute for Social Research, York University.

PartiRep. 2014. *PartiRep Voter Panel Survey 2014.* Technical report. Brussels.

Pasek, Josh. 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.

Sanders, David, Harold D. Clarke, Marianne C. Stewart, and Paul Whiteley. 2007. Does Mode Matter For Modeling Political Choice? Evidence From the 2005 British Election Study. *Political Analysis* 15 (3): 257–285.

Schoen, Harald, and Thorsten Faas. 2005. When Methodology Interferes with Substance: the Difference of Attitudes towards E-Campaigning in Online and Offline Surveys. *Social Science Computer Review* 23 (3): 326–333.

Selb, Peter, and Simon Munzert. 2013. Voter Overrepresentation, Vote Misreporting, and Turnout Bias in Postelection Surveys. *Electoral Studies* 32 (1): 186–196.

Simmons, Alicia D., and Lawrence D. Bobo. 2015. Can Non-Full-Probability Internet Surveys Yield Useful Data? A Comparison with Full-Probability Face-to-Face Surveys in the Domain of Race and Social Inequality Attitudes. *Sociological Methodology* 45 (1): 357–387.

Stephenson, Laura B., and Jean Crête. 2010. Studying Political Behavior: A Comparison of Internet and Telephone Surveys. *International Journal of Public Opinion Research* 23 (1): 24–55.

Stern, Michael J., Ipek Bilgen, and Don A. Dillman. 2014. The State of Survey Methodology: Challenges, Dilemmas, and New Frontiers in the Era of the Tailored Design. *Field Methods* 26 (3): 284–301.

Vavreck, Lynn, and Douglas Rivers. 2008. The 2006 Cooperative Congressional Election Study. *Journal of Elections, Public Opinion and Parties* 18 (4): 355–366.

Yeager, David S., Krosnick Jon A., Chang LinChiat, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, and Rui Wang. 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.