Consortium on Electoral Democracy
Consortium de la démocratie électorale

# The 2019 Canadian Election Study: A Mode Comparison in Electoral Studies

Alvaro Pereira, University of Western Ontario
Laura Stephenson, University of Western Ontario

Report # 2024-05-15

This report was prepared for the Consortium for Electoral Democracy (C-Dem) using the 2019 Canadian Election Study (Stephenson, Harell, Rubenson and Loewen, 2020). The views expressed in this report are solely those of the authors.

**Table of Contents**

## Executive Summary

Election surveys provide information about voters' attitudes and behaviours during a given pre- and post-electoral campaign. Despite years of research about the benefits and downfalls of utilizing online, non-probability panels rather than more traditional probability methods, the question of which method is most appropriate for cost-conscious, rigorous research still needs to be answered. Online methods have clear advantages, including lower costs and the ability to interview more respondents quickly. However, there is still a hesitation to trust the accuracy of estimates returned from data gathered through non-probability methods. In this paper, we explore this question in detail by comparing a medium-sized probability telephone survey (n = 4,021) with a large-scale, online non-probability sample survey (n = 37,822), both conducted as part of the 2019 Canadian Election Study (CES). We focus on assessing which survey returns a more accurate estimate of Canadians' characteristics and attitudes, taking into account mode and sample size considerations. To facilitate fair and targeted comparisons, we use two techniques—propensity score matching and bootstrapping—to evaluate the issues of sampling design and mode separately. We then consider whether size alone is a valuable benefit in favour of the online mode. Our findings reveal that when considering the effect of mode alone, estimations from the online survey are more accurate in predicting intentions to participate and vote choice. When considering the effect of sampling, the results confirm the precision of the online estimates in a Canadian election study. We also find that, regarding attitudes, the online sample is more consistently unbiased when predicting support for immigration. We ultimately conclude that adopting an Internet-only survey for Canadian electoral studies is a wise choice.

**Keywords**

Survey Mode; Election Studies; Mode Comparison; Sampling; Immigration.

The recent expansion of online survey methods in social science research has raised the question: Is adopting Internet-only surveys in electoral studies a good choice that enhances accuracy and precision in the modern data collection environment? Do Internet samples have advantages over RDD phone surveys in the current context? And what are their disadvantages? Several studies have already investigated the way that advancements in the technologies of data collection influence individuals' responses and the quality of survey data (e.g., Tourangeau et al., 2000; Malhotra & Krosnick, 2007). In these studies, the promise of lower costs and increasing sample sizes from Internet surveys contrasts with their supposed lack of representativeness and unknown distributional properties, underscoring the potential to collect a biased sample (Bytzek & Bieber, 2006; Chang & Krosnick, 2009; Stephenson & Crête, 2011; Berinsky, 2017; Horwitz et al., 2019).

Ultimately, the findings are mixed (e.g., Sander et al., 2007; Malhotra & Krosnick, 2007; Stephenson & Crête, 2011). Although the rise in use and general acceptability of non-probability samples is unquestionable, more must be studied. We also agree with the 2013 AAPOR report that, given the rapid development of survey methods, "research evaluations of other methods of non-probability sampling from panels may have little relevance to the current methods being used." (Baker et al., 2013, p.4). Further, we are keenly aware that probability sampling suffers from significant drawbacks — for face-to-face, the steep costs are prohibitive for most studies; for telephone, the rapid increase in non-response and mobile use—such that reconsidering these issues is pressing.

This report compares the value of a large-scale, online survey from a non-probability active sampling panel with population quotas with a smaller, telephone, probability-sampled survey on three dimensions—mode, sampling, and size.[1] Our data in this report come from Canada, where the Canadian Election Study (CES) has used the online sampling mode for several years. Based upon analysis of the 2015 mixed-mode CES (Fournier et al., 2015), Breton et al. (2017) find no consistent differences across online and telephone surveys in the responses' quality and when estimating political attitudes and behaviours. They conclude that it is impossible to determine which survey mode better represents, describes, and infers Canadian voters' political attitudes, favouring a mixed-mode approach for the CES. However, despite the insightful comparisons, the generalizability of Breton et al.'s findings for today's research environment is unclear, as cell phones appeared only randomly in the 2015 CES. In recent years, cell phones have systematically replaced landlines as the core telephone type (e.g., Tessier, Bodet & Gélineau, 2014; Berinsky, 2017; Stephenson et al., 2021). Therefore, potential cofounders are unclear and whether Breton et al.'s analysis is a relevant comparison is uncertain.

Our report uses two samples collected as part of the 2019 CES (Stephenson et al., 2020a; 2020b). Addressing the issues of cell phones, the 2019 telephone study followed updated sampling practices to include mobiles and landlines in a 66:34 ratio (Stephenson et al., 2021). The telephone sample, therefore, better represents a model probability sample. The online study is very large, with over 37,000 respondents, which reflects one of the benefits of the online mode: size. However, it adds an element to be considered when conducting comparisons. To address this, we cross-validate the 2019 CES online study estimates using the telephone sample as a reference in three ways. First, we use propensity score

[1] The online sample was actively built by Qualtrics, with targets stratified by region and balanced on age and gender. Qualtrics also contacted respondents for the post-electoral survey.

matching, by which we reduce the original sample size based on the propensity of being interviewed by telephone, to create a matched sample out of the online study that enables us to isolate the impact of mode. Observations in the matched online sample have a balanced distribution of covariates with their correspondents in the telephone sample. Second, through bootstrapping, we resample the online sample with replacements to create a dataset comparable to the telephone study. We then use this sample to evaluate the influence of the sampling method on the estimates' comparability. Finally, we identify whether any benefits accrue from being able to gather (and afford) more respondents using online sampling methods. Considering individual aspects of online studies allows us to assess the value of online samples for election research more comprehensively. It will enable us to evaluate the quality and value of the online survey in terms of three characteristics—mode, sampling, and size—relative to previous studies using observational data (e.g., Breton et al., 2017).

We repeat tests performed by Breton et al. (2017) in our analyses below. To preview our results, we find that the online estimations from the matched and bootstrapped samples are more accurate for electoral studies; online estimations more accurately predict political variables, such as intentions to vote choice and participation. In evaluating the surveys for measuring attitudes, we compare modes of support for immigration. Nativism is a socially reprehensible attitude shaped by social desirability (e.g., Hainmueller & Hopkins 2015), influencing how people respond to the (lack of) support for immigration across survey modes. The telephone mode underreports nativism, suggesting social desirability bias related to the mode itself. Finally, we directly address the issue of sample size. We demonstrate that the online sample performs more accurately in estimating the population parameter compared with benchmarks (mean square error) and with the correlational structure described in the literature (e.g., attitudes toward immigration) when we increase the online study's sample size, as expected, an inherent advantage of online surveys due to the lower costs. All these results suggest that the decision to invest in Internet-only surveys for future election studies is wise and should also be considered by other researchers looking to maximize the efficiency of their grant dollars.

***Survey Details and Panel Attrition***

We begin our analysis by examining the general characteristics across survey modes in the 2019 CES. *Table 1* summarizes the technical details of the survey from the Campaign Period Survey (CPS) and the Post-Election Reinterview Survey (PES).[2] The online sample size is almost ten times bigger than the telephone sample in the CPS and five times in the PES. In the telephone sample, the reinterview rate is 51.6%. By comparison, although the online sample has a much higher number of interviews completed, the reinterview rate is much lower, at 27.8%.

[2] We undertook an additional step in cleaning the data to remove individuals who expressed "don't know" on several key variables. Appendix A provides complete details. Ultimately, 4.5% of the total respondents (n = 1,705) were removed from the online sample and 1.5% (n = 54) were removed from the telephone sample, following this rule.

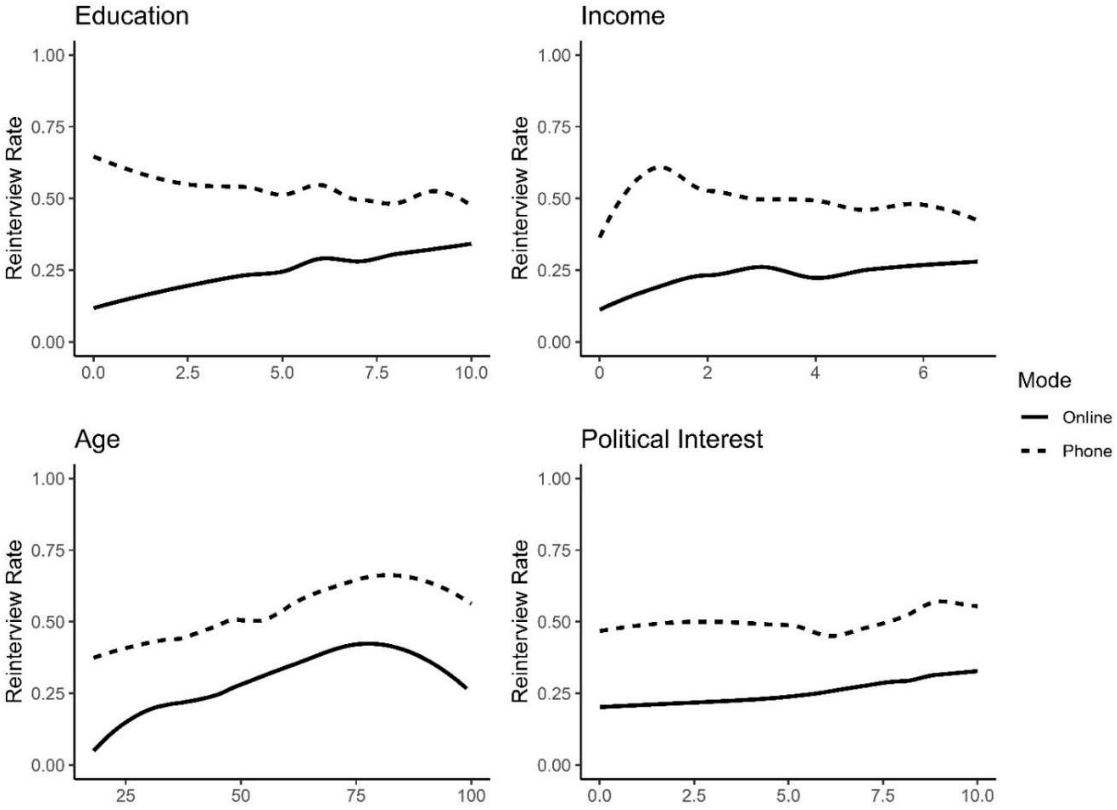
**Table 1: Survey Details**

| Survey | Completed Interviews | Data Collection Period | Duration (mins.) | | |
|---|---|---|---|---|---|
| | | | Mean | Median | Std Dev |
| **Telephone** | | | | | |
| **Campaign** | 3,967 | Sept 13 - Oct 21 | 20.07 | 19 | 6.23 |
| **Post** | 2,049 | Oct 24 - Nov 11 | 19.61 | 19 | 8.67 |
| **Internet** | | | | | |
| **Campaign** | 36,117 | Sept 10 - Oct 20 | 35.17 | 20.02 | 209.37 |
| **Post** | 10,049 | Oct 22 - Nov 21 | 46.59 | 23.17 | 266.70 |

*Source: 2019 Canadian Election Study – Online Survey and 2019 Canadian Election Study – Telephone Survey.*

To investigate these reinterview rates closely, we examine trends by survey mode on completed interviews in the PES surveys. *Figure 1* shows reinterview rates during the PES by survey mode—Internet or telephone—for each of the following four independent variables: i) educational level (0-10); ii) income categories (0-6); iii) political interest (0-10); and iv) age (in years).

**Figure 1: Trends in Reinterview Rates by Mode**




*Notes: The functional form for each trend was calculated based on weighted local polynomial smoothing, with the degree of smoothing set to .5 in all calculations. Source: 2019 Canadian Election Study – Online Survey and 2019 Canadian Election Study – Telephone Survey.*

Except for education, all factors increase the recontact rates in each survey mode; that is, older, wealthier, and politically interested respondents have a higher probability of returning to the sample in the post-election survey in the online and telephone modes. Education is the only variable for which the return to the telephone sample decreases as individuals' educational level increases. As a result, highly educated respondents are less likely to participate in a recontact survey by telephone, and the increasing expansion of the Internet seems to close the gap between educational levels in the reinterview rates.

***Rendering the Sample Comparable***

A considerable difference in size between the telephone and online 2019 CES samples makes comparing them challenging. In observational studies, not accounting for size might mask similarities across estimations, increasing the occurrence of type I errors. We rely on two statistical techniques to examine the accuracy of estimates from the 2019 CES while accounting for different sizes. First, we reduce the online sample to a size similar to the telephone sample using propensity score matching, by which observations across modes are paired based on the probability of being treated as if in an experimental setting (Stuart, 2010).[3] In this report, the likelihood of interest is 'being surveyed by telephone.' The propensity score matching enables us to build a matched sample with a distribution across covariates similar to that of the telephone sample. In turn, the matched sample returns estimates from a dataset with covariates that are balanced with those of the telephone survey. Second, we sampled from the online study, with a large sample size, using bootstrapping and estimated based on the 2019 full online distribution (Chernick, 2011).[4] This procedure is designed to return the data characteristics from the survey itself, based on the distribution of the full-sized online sample and regardless of any assumption of the population (Mooney & Duval, 1993). In sum, bootstrapping returns estimates based on the characteristics of observations in the entire online sample, but with comparable size.

Due to the unknown properties and large online sample size, our goal with these two techniques is to compare estimates across survey modes fairly, disentangling the mode effect from covariate distribution (i.e., income categories, age, and gender) and sample sizes. Therefore, instead of selecting observations from the whole online sample, we resample through these two techniques to boost our confidence in the conclusions drawn from the mode comparison. We expect similarity across estimates gathered using these two techniques, such that any differences with telephone estimates can be interpreted as solely the product of the mode of delivery. In the analyses below, we present our results with the matched and bootstrapped samples to the full telephone sample. We address the benefit of sample size separately.

***External Cross-validation: Comparing to Benchmarks in Participation and Vote Choice***

Benchmarks help examine the accuracy of estimates across survey modes (Ansolabehere & Schaffner, 2014; Breton et al., 2017). We calculate the mean squared error (MSE), the distance between estimates from the benchmarks available from Elections Canada. We could identify benchmarks for turnout and vote share. The MSE returns an objective measurement of whether an estimate is more accurate in representing the population or predicting electoral results.[5] The more accurate the estimate, the lower the MSE is. We calculate the MSE for the political variables—previous and intended vote choice and participation in the 2015 and 2019 Canadian elections. These variables are particularly relevant for evaluating samples from election studies, as researchers often concentrate on explaining these electoral outcomes using the CES.

Our results suggest that the online mode is more accurate than the telephone mode with respect to the electoral benchmarks. The matched and bootstrapped samples' total MSE are .00525 and .00492, respectively. They are lower than the telephone's MSE, which is .01, but the difference is not statistically significant. The similarity in online sample MSE despite the different techniques suggests that the differences are not due to sampling or size. *Figure 2* displays the point estimates for vote share in the three major parties (i.e., Liberal, Conservative, and NDP in 2015 and 2019) and participation in two

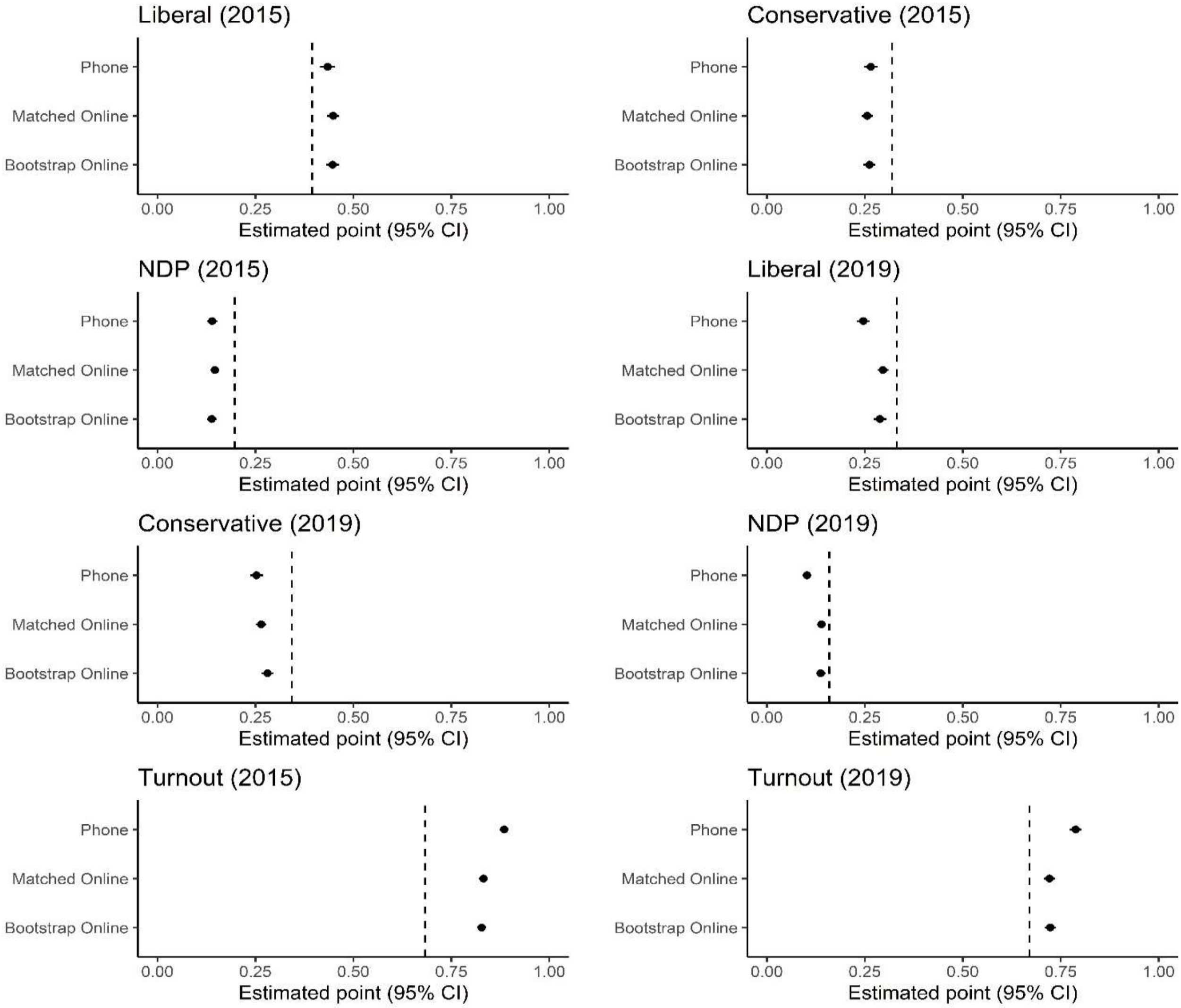
[3] Appendix B details the technique to select the best-matched sample after 10,000 matchings to avoid a 'greedy' trend in the computational procedure and a poor covariate balance. We present the selected sample's balanced performance and distance measure distribution across trials.

[4] Rather than comparing samples directly, we resample both modes with the same sample size (n = 3,967) and take the average of each variable 10,000 times (see Mooney, 1996).

[5] Weights are included in the MSE calculus and are available in the CES data. Appendix C details the weights used by each mode.

federal elections. The online mode estimates are closer to the vertical lines, indicating that the online samples are more accurate. For instance, the matched and bootstrapped estimates predict 13.9% and 13.7% of votes for the NDP in 2019, respectively, only three percentage points below the actual vote share (p<.05). By comparison, the telephone sample underestimates NDP performance by six percentage points in the 2019 federal elections. For turnout, the online estimations are much more accurate than the telephone estimates. In 2019, the online samples overestimated the turnout by 6 points, while the telephone sample overestimated the turnout in that election by 11 percentage points (p<.05).

**Figure 2: Comparing modes with benchmarks**




*Note: Point estimates for vote intention, previous turnout and expected turnout with 95% confidence intervals. Vote intention was collected from Sept. 10-Oct. 21, 2019. The vertical line corresponds to the benchmark—source: Elections Canada.*

Considering the estimates from matching and bootstrapping discussed above, we can infer that, relative to the telephone sample, the online mode is more accurate in estimating important political variables for electoral studies. Comparing these estimates with an external source (i.e., Elections Canada) to benchmark parameters, the online sample estimates outperform estimates from telephone samples for most electoral variables. In particular, online estimates predict vote share in forthcoming federal elections

(2019) and participation more accurately. These are critical to understanding Canadians' vote-making processes and behavioural models. Moreover, there are few differences between the matched and bootstrapped estimates, which suggests that these techniques return similar performance in their estimation accuracy.

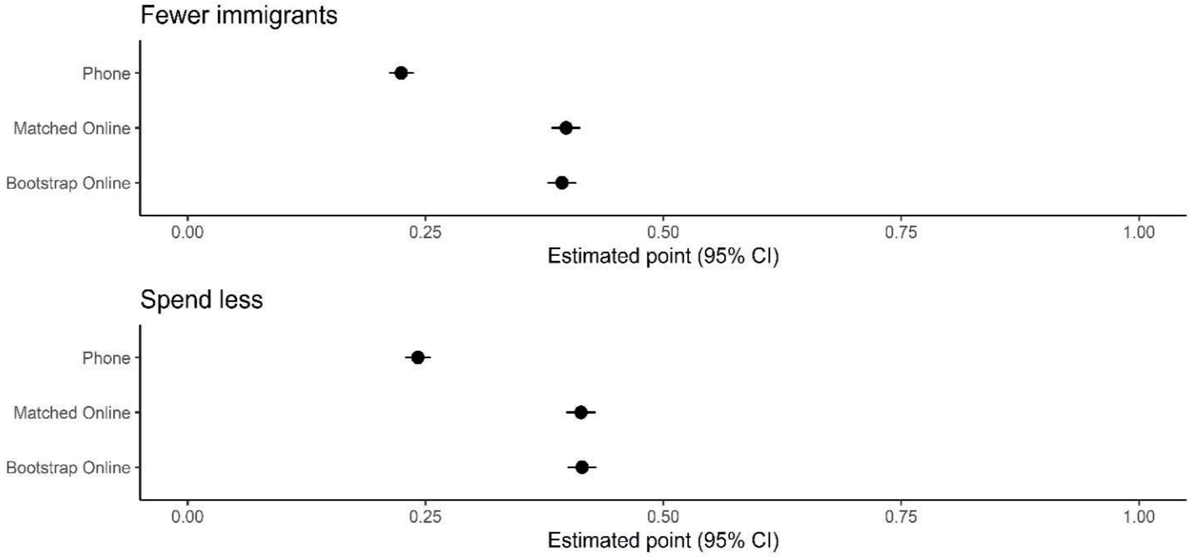
***Internal Cross-Validation: Attitudes Toward Immigration and Immigrants***

We can also cross-validate the survey modes using variables for which we do have theoretical expectations rather than benchmarks. Although we might not have an objective measure for comparison (as the MSE above), scholars assess differences based on the correlational structure and their accordance with the literature (e.g., Ansolabehere & Schaffner, 2014; Breton et al., 2017). We examine the online and telephone survey modes through a model of attitudes toward immigration, in which we evaluate whether the modes' estimations are consistent with the expectations of the literature about these attitudes. Due to the absence of live interviewers, online surveys are deemed less susceptible to social desirability bias in measurements of attitudes (e.g., Kreuter et al., 2008; Heerwegh, 2009). By social desirability, we mean that respondents provide insincere but acceptable responses to questions with clearly acceptable answers (Kuklinski et al., 1997). Online surveys, since they are self-administered, eliminate the contact between interviewers and interviewees, which can reduce the desirability of responses. Therefore, we expect to find differences in responses about immigrant attitudes across survey modes.

Nativism is considered by many to an attitude that is counter to current norms. Consequently, responses about immigration are often shaped by social desirability (e.g., Hainmueller & Hopkins, 2015). Providing socially desirable responses makes attitudes toward immigrants more positive, although arguably insincere. Analyzing attitudes toward immigration, then, is a good case for studying differences in responses across survey modes. We expect the telephone survey will consistently underestimate nativism (or report less support toward immigration). In contrast, in the online mode we expect nativism to be consistently higher.

We begin by exploring support for admitting fewer immigrants (Fewer immigrants = 1; Otherwise = 0) and preferences related to less government spending on immigration (Spend less = 1; Otherwise = 0). *Figure 3* presents the result of point estimates across the survey modes. On average, the telephone sample reports more support for admitting immigrants to Canada and spending on immigration than the online estimates. About 22% of respondents in the telephone sample support fewer immigrants in Canada, and 23% support less spending by the Canadian government. By comparison, in the online estimates, 39.8% in the matched sample and 39.4% in the bootstrapped sample support admitting fewer immigrants. Regarding spending less on immigration, the values are 41.3% and 41.5% of respondents, respectively. These results suggest that the telephone mode consistently underestimates nativism or attitudes against immigration among Canadian voters.

**Figure 3: Attitudes on admission and spending on immigration**




*Note: Point estimates and 95% confidence intervals. Source: Canadian Election Study, 2019.*

In addition, we estimate two OLS models in agreement with the following 5-point scale statements about immigrants: i) "Immigrants are generally good for Canada's economy;" ii) "Canada's culture is generally harmed by immigrants;" iii) "Immigrants increase crime rates in Canada." We reverse the order of items (ii) and (iii) so that for each dependent variable, 0 means *strongly disagree* and four indicates *strongly agree* with a pro-immigrant position. In the first model, we pool the telephone and matched samples and add a variable for survey mode. In the second model, we perform a bootstrapped regression (with 10,000 iterations) with randomly selected (with replacement) subsets of observations from the full-size online sample, also with a variable indicating survey mode.[6] In these two models, we include controls (i.e., age, education level, volunteerism, and unemployment).[7]

According to the expectations in the literature, respondents should express less agreement with pro-immigration statements in the online mode relative to the telephone mode. *Table 2.1* reports the estimates per model using the matched online sample. As expected, the coefficients are statistically significant across statements (p.<.001) and negative, suggesting that respondents in the online sample express less agreement with each statement than respondents in the telephone sample.

---

[6] We randomly select observations from the whole online sample. Then, we bootstrapped the regression and averaged the coefficients.

[7] Due to the dependent variable's missingness, the sample size in each model varies. For instance, in the matched online sample, 2,856 respondents failed to answer at least one of the three statements about immigration, while in the bootstrapped sample, 1,255 failed to answer the question about immigrants not increasing crime in Canada.

**Table 2.1: Coefficient of Mode in the OLS Regression on Attitudes Towards Immigrants**

| | Good for the economy | Not harmful for the Canadian culture | Not increasing crime |
|---|---|---|---|
| **Mode** | -.381***<br>(.035) | -.558***<br>(.043) | -.521***<br>(.041) |
| **N-respondents** | 3,950 | 3,947 | 3,864 |

* p < 0.05, ** p < 0.01, *** p < 0.001

*Table 2.2* shows that the estimates from the bootstrapped regression are consistent with the previous model—the coefficients for survey mode are significantly different (p.<.001) and negative, which means that the self-administered survey respondents express attitudes less favourable toward immigrants than when surveyed by interviewers on the telephone.

**Table 2.2: Statistics for 10,000 Bootstrapped OLS Regression on Attitudes Toward Immigrants**

| | Good for the economy | Not harmful for the Canadian culture | Not increasing crime |
|---|---|---|---|
| **Mode** | -.335***<br>(.034) | -.511***<br>(.041) | -.564***<br>(.039) |
| **N-respondents** | 5,412 | 5,412 | 5,412 |

* p < 0.05, ** p < 0.01, *** p < 0.001

The relevance of attitudes to electoral behaviours is one of the most essential elements to be drawn from election studies. Such attitudes, however, rarely have benchmarks that scholars can use to determine whether estimates from diverse samples are accurate. In an attempt to cross-validate our samples, we have compared results from the 2019 CES online and telephone samples in terms of correlates of attitudes about immigration. The results are consistent with the literature, which argues that differences in attitudes found using self-administered surveys are due to the absence of social desirability bias; the telephone mode underestimates negative opinions about immigration.

***Sample size and accuracy***

A last word should be given to the sample size difference. So far, we have resampled from the whole survey to make it more comparable with the telephone survey. However, one of the most striking aspect of the 2019 CES is that the online sample is a large-scale collection of 37,822 observations. According to Stephenson et al. (2021), this large sample enables researchers to explore cleavages and nuances within political groups, opening up opportunities for scholars to take a novel perspective in exploring attitudes and behaviours. They find, for instance, that permanent residents are more likely to support the Liberal Party in 2019. This large sample is possible because technology makes data collection less expensive online. The last question to answer when comparing the online and telephone samples is whether the large-scale sample provides a benefit compared to our previous analyses' findings.

In short, we find similar or better results to those reported above when using the full-size sample. First, the full-size online sample MSE for voting and participation is .003, which is more accurate than all other estimations. Second, for attitudes toward immigration and immigrants, the full-size online sample consistently reports higher levels of nativism and lower support for immigrants.

This evidence is another benefit of using online data collection. Lower costs and larger sample sizes do not go hand in hand with probability samples. However, online data collection does not reveal any substantial reduction in quality. There is no clear benefit from using probability sampling; if anything, the mode effects are a more significant detriment to research. In addition, large-scale online data collection will enable researchers to explore subgroups with more accurate estimates in electoral studies. We contend that moving to an online-only data collection can be a wise decision with little to no detriment and significant benefits.

***Conclusion***

Survey researchers continuously incorporate new data collection technologies, and online non-probability sample data has quickly become commonplace. In this analysis, and using different techniques, we consider whether the adoption of online sampling has been too hasty. We have evaluated estimates from the 2019 CES online and telephone samples in a comparable manner. Our analyses pitted estimates from a modern probability-sample telephone survey with three versions from an online nonprobability sample—one from a comparable-size sample derived from propensity score matching to the parameters of the probability survey, another of comparable size conducted with bootstrapping techniques, and a third that utilized a large-scale online sample without any adjustments. Ultimately, we find that the estimates derived from the online sample are more often accurate or unbiased compared to those from the telephone survey. Online estimations are more accurate for vote share and turnout, central concerns in election studies. Although not all studies will have the size advantage of the 2019 Canadian Election Study, our analyses demonstrate that there are benefits to the online mode itself that make it worthwhile for survey research.

## References


Ansolabehere, S., & Schaffner, B. F. (2014). Does survey mode still matter? Findings from a 2010 multi-mode comparison. *Political Analysis*, 22(3), 285-303.

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J. & Tourangeau, R. (2013). Report of the AAPOR Task Force on Non-Probability Sampling. https://aapor.org/wp-content/uploads/2022/11/NPS_TF_Report_Final_7_revised_FNL_6_22_13-2.pdf

Berinsky, A. J. (2017). Measuring public opinion with surveys. *Annual review of political science*, pp. 20, 309-329.

Breton, C., Cutler, F., Lachance, S., & Mierke-Zatwarnicki, A. (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/Revue canadienne de science politique*, 50(4), 1005-1036.

Bytzek, E., & Bieber, I. E. (2016). Does survey mode matter for studying electoral behaviour? Evidence from the 2009 German Longitudinal Election Study. *Electoral Studies*, 43, 41-51.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly*, 73(4), 641-678.

Chernick, M. R. (2011). *Bootstrap methods: A guide for practitioners and researchers*. John Wiley & Sons.

Fournier, P., Cutler, F., Soroka, S., Stolle, D. (2015). *The 2015 Canadian Election Study.* [dataset]. https://ces-eec.arts.ubc.ca/english-section/surveys/.

Hainmueller, J., & Hopkins, D. J. (2015). The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants. *American Journal of Political Science*, 59(3), 529-548.

Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: an experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, 21(1), 111-121.

Horwitz, R., Lesser, V., Marken, S., Mathiowetz, G. N., Opsomer, J., Steiger, W. D., & Sterrett, W. D. (2019). Report of the AAPOR Task Force on Transitions from Telephone Surveys to Self-Administered and Mixed-Mode Surveys.

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political analysis*, 27(4), 435-454.

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys the effects of mode and question sensitivity. *Public opinion quarterly*, 72(5), 847-865.

Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial attitudes and the" New South". *The Journal of Politics*, 59(2), 323-349.

Malhotra, N., & Krosnick, J. A. (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.

Mooney, C. Z. (1996). Bootstrap statistical inference: Examples and evaluations for political science. *American Journal of Political Science*, 570-602.

Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A nonparametric approach to statistical inference* (No. 95). Sage.

Sanders, D., Clarke, H. D., Stewart, M. C., & Whiteley, P. (2007). "Does Mode Matter For Modeling Political Choice? Evidence From the 2005 British Election Study." *Political Analysis* 15(3): 257-285.

Stephenson, L. B., & Crête, J. (2011). Studying political behavior: A comparison of Internet and telephone surveys. *International Journal of Public Opinion Research*, 23(1), 24-55.

Stephenson, L. B.; Harell, A.; Rubenson, D.; Loewen, P. J. (2020). "2019 Canadian Election Study – Online Survey", https://doi.org/10.7910/DVN/DUS88V, Harvard Dataverse, V1

Stephenson, L. B.; Harell, A.; Rubenson, D.; Loewen, P. J. (2020). "2019 Canadian Election Study – Telephone Survey", https://doi.org/10.7910/DVN/8RHLG1, Harvard Dataverse, V1

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. Statistical science: *a review journal of the Institute of Mathematical Statistics*, 25(1), 1.

Tessier, C., Bodet, M. A., & Gélineau, F. (2014). L'efficacité de l'échantillonnage passif pour obtenir un portrait représentatif de l'électorat: Le cas de Vote au pluriel–Québec. *Canadian Journal of Political Science/Revue canadienne de science politique*, 47(4), 697-715.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. London: Cambridge Univ. Press.

**Appendix A: Data quality**

Poor data quality can directly contribute to misleading conclusions. For instance, respondents often do not pay attention to questions in the online surveys or accelerate the end of the survey with "don't know" responses. By examining these responses in key questions, researchers can clean data and improve its quality.

The 2019 CES dataset has already been cleaned following standard procedures, but that does not include looking at "don't know" responses. We, therefore, took the additional step of examining responses to feeling thermometers (3 major parties and their 3 leaders in 2019), one demographic (education), one attitudinal (democratic satisfaction), and one retrospective evaluation (economic retrospection) variable.

Following Breton and colleagues (2017), we remove observations that report "don't know" in all feeling thermometers or in two of three variables described earlier. *Table A1* describes the percentage of "don't know" responses for each condition. These observations amount to 1,705 respondents in the online sample and 54 observations in the telephone sample.

**Table A1: Data Quality (%)**

| **Survey** | DKs in leader feelings | DKs in party feelings | DKs in two of three questions |
|---|---|---|---|
| **Online** | 2.8 | 2.4 | 0.7 |
| **Telephone** | 1.1 | 0.6 | 0.2 |

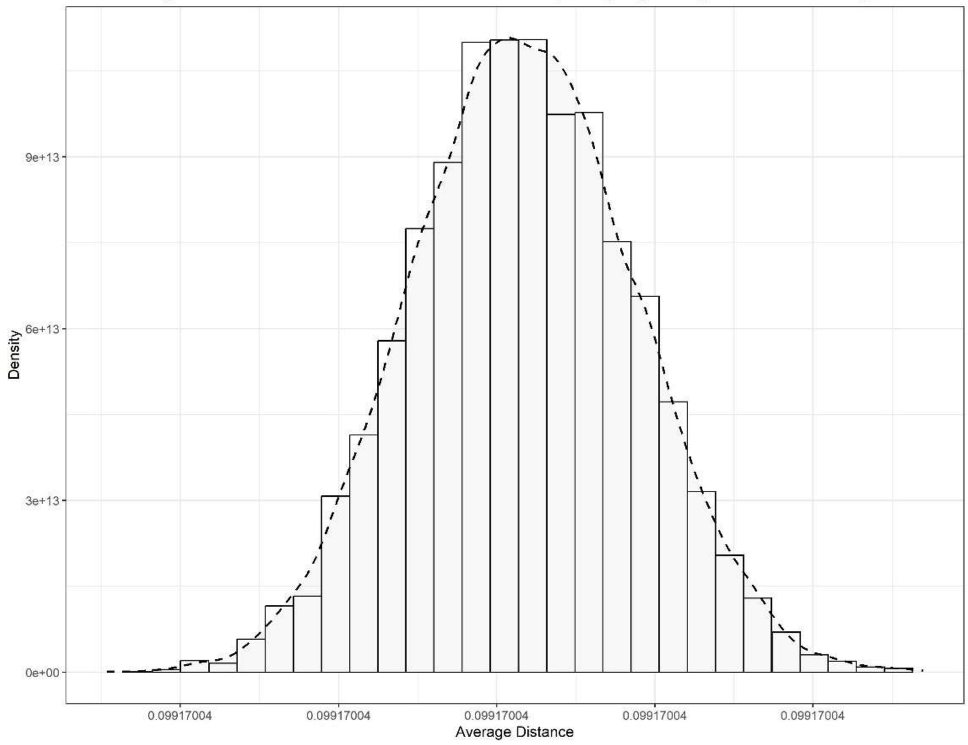
## Appendix B: Propensity score matching

The main goal of propensity score matching (PSM) is to reduce the pretreatment effects in observational data. Given the observed baseline covariates, the PSM is the conditional probability of receiving a treatment. Thus, after matching, we expect that the distribution of covariates will be balanced across treatment levels. In this paper, the treatment is being surveyed by telephone.

We use the nearest neighbour method employed by the *MatchIt* package in R. This method, though, is called "greedy;" that is, as the algorithm finds in the data the first observation that matches with another observation, it stops looking for another matching. Therefore, the nearest neighbour method is suboptimal in terms of finding the best matching pairs. To address this, we began by reordering the original sample about 10,000 times. For each trial, we ran a PSM using age, gender, educational level, language, volunteerism, religion, employment status, and income category with their original scale in the Canadian Election Study. We calculated the distance of each observation score from the propensity score of being treated; in this particular case, surveyed by telephone.

*Figure B1* reports the distribution of the distances after 10,000 trials. Although the distances look pretty close, we can find the closest distance. We use the minimal distanced order and produce a matched online sample with 3,967 observations.

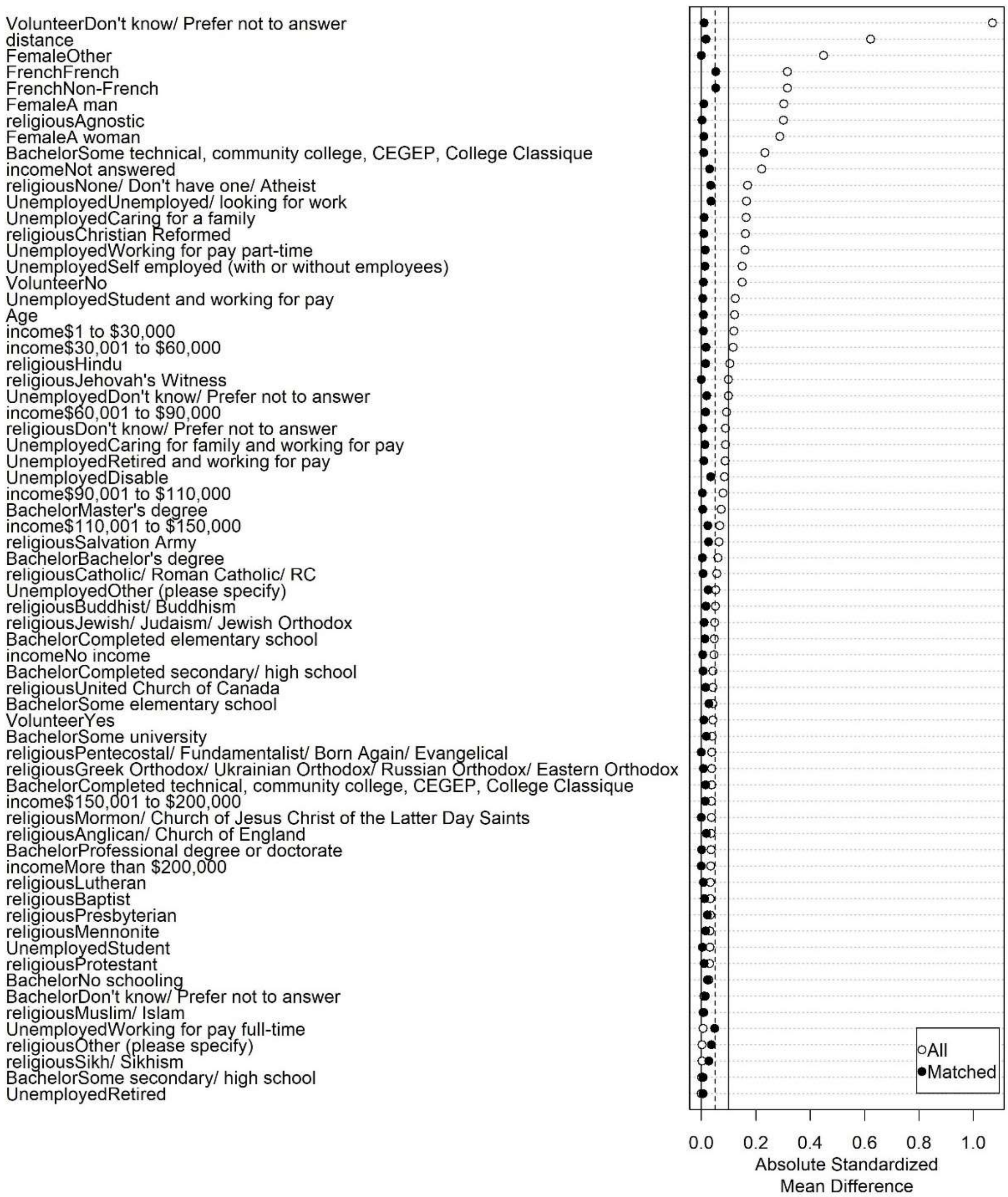
Figure B1: Distribution of distances in 10,000 propensity score matchings




*Note: The matchings were performed by the MatchIt package in R. We selected the telephone as the treatment and the online survey mode as the control, so we had a larger control group. We create a vector with the position of each row in the original data and reorder it. We select the lowest average distance produced by each iteration.*

King and Nielsen (2019) suggest that randomly pruning the data by PSM might produce imbalanced rather than balanced results. *Figure B2* displays the summarized balance across covariates used within the matching model. Note that there are imbalances across covariates in the original data, from age to volunteerism. After matching, the covariates look balanced with all absolute standardized mean differences (SMD) within a threshold (0.10). This result suggests that PSM using the CES produces a good balance in the covariate distribution across treatments. The largest SMD is 0.0525 for francophones. The largest maximum and mean Empirical CDF is gender. However, we conclude that the PSM improved the balance between survey modes.

**Figure B2:**
**Covariate Balance**



**Appendix C: Weighting**

Survey weights correct the sample distribution based on the reference population. In the 2019 CES, naturally, each survey mode has its own weighting procedure. In the online survey, the weights are built in an iterative raking process, which creates weights according to survey (i.e., CPS and PES) and a set of variables, as province, gender, age, and educational level. The 2016 Census informed these variables' distributions. Respondents from the territories do not have weights, thus, we removed these observations from the analyses.

In the telephone sample, the weights are simpler. They are calculated by dividing the sample proportion for province and ownership metrics. In the PES, the weights are similar. However, they account for individuals who previously responded to the CPS survey.