

On the Measurement of Public Opinion in the Age of Big Data*

Clifton van der Linden[†]

December 31, 2019

Abstract

Democracy is predicated on the idea that governments are responsive to the publics which they are elected to represent. In order for elected representatives to govern effectively, they require reliable measures of public opinion. Traditional sources of public opinion research are increasingly complicated by the expanding modalities of communication and accompanying cultural shifts. Diversification of information and communications technologies as well as a steep decline in survey response rates is producing a crisis of confidence in conventional probability sampling. An increasingly rich yet relatively untapped source of public opinion takes the form of extraordinarily large, complex datasets commonly referred to as Big Data. Artificial Intelligence, and machine learning in particular, offers new opportunities for addressing the challenges for statistical inference as they pertain to Big Data, not least of which is that these data typically take the form of non-probability sample. This paper argues that, under specific circumstances and given the application of machine learning techniques, certain types of non-probability sample may be capable of yielding reliable inferences about a population of interest. To demonstrate this argument, it analyzes the inferences derived from the most extraordinary probability and non-probability samples collected during the 2015 Canadian federal election campaign: the Canadian Election Study (CES) and Vote Compass, respectively. It uses the election outcome as a benchmark and models the observations collected from each sample to assess how accurately they are able to forecast the distribution of the vote.

Keywords: Public opinion, Election forecasting, Big Data

*This paper is an early draft. Please do not cite without permission.

[†]McMaster University Department of Political Science; e-mail: cliff.vanderlinden@mcmaster.ca

1 Introduction

How effective are non-probability samples at measuring public opinion? Conventional wisdom holds that only probability samples can be generalized to a population of interest such as to allow statistical inferences about said population. However, emergent modalities of communication are increasingly diverse and esoteric, compounding the potential for coverage error and non-response bias in probability samples (De Heer and De Leeuw, 2002; Keeter et al., 2006; Kohut et al., 2012; Holbrook et al., 2007; Steeh et al., 2001; Council et al., 2013).

The same advances in information and communication technologies that have significantly complicated the collection of probability samples may, however, bolster the potential for deriving representative inferences about a population of interest using non-probability samples. Specifically, emergent technologies have enabled the collection of non-probability samples of much greater size at faster rates and lower cost than conventional techniques for probability sampling.

Nevertheless, non-probability samples are widely considered to be inferior to probability samples in that respondents self-select, resulting in an inherently non-random sample. Though techniques such as raking (Battaglia et al., 2009), matching (Vavreck and Rivers, 2008), post-stratification weighting (Dever, Rafferty and Valliant, 2008; Gelman et al., 2007), or propensity score weighting (Lee, 2006; Lee and Valliant, 2009; Schonlau et al., 2009) are commonly applied to attempt to adjust for bias in non-probability samples purporting to make externally valid inferences, many public opinion researchers contend that statistical inference is impossible without probability sampling (Baker et al., 2013).

This paper argues that, under specific circumstances, certain types of non-probability sample may be capable of yielding reliable inferences about a population of interest. To demonstrate this argument, it analyzes the inferences derived from the most extraordinary probability and non-probability samples collected during the 2015 Canadian federal election campaign—the Canadian Election Study (CES) and *Vote Compass*, respectively. It uses the election outcome as a benchmark and models the observations collected from each sample

to assess how accurately they are able to forecast the distribution of the vote.

2 The case for non-probability sampling

Though non-probability samples are often dismissed as unscientific, recent scholarly inquiry into the suitability of non-probability sampling for statistical inference has challenged this perspective (Baker et al., 2013; Brick, 2011). Resistance to non-probability sampling is largely grounded in notions of sample randomness as the fundamental criterion for external validity, but this is arguably both a theoretically and practically tenuous position.

Statistical theory does not posit random sampling as a requisite condition for statistical inference, but rather the most generally accepted method. Smith (1983, p. 402) posits that post-stratification techniques applied to non-random samples can yield externally valid inferences so long as neither the known prior values nor the selection variable contains information beyond that in the post-stratifying variables. Thus, if the factors that determine the presence or absence of a member of a given population in a non-probability sample are uncorrelated with the variables of interest in a study, or if they can be fully controlled for by making adjustments to the sample, then externally valid inference is theoretically possible.

At the same time, the credibility of probability sampling as a technique which can reliably generate statistical inferences about a population of interest is increasingly subject to critique. As coverage error and non-response rates increase for probability samples, so too do questions as to whether such samples still meet the criteria for probability samples. Edgington (1966) argues that probability samples “rarely meet the assumption of random sample that conventional statistical hypothesis-testing procedures are generally believed to require.” Few if any sampling techniques available to public opinion researchers are capable of producing sampling frames which meet the criteria for random sample, specifically that every member of the population of interest has a non-zero probability of inclusion in their samples and that these probabilities are known. Moreover, non-response would have to be

zero unless one were to assume that non-response was uncorrelated with answers to the survey question of interest—a highly problematic assumption in most cases of conventional sampling techniques.

However, the empirical evidence amassed by scholars has largely supported, with few exceptions (Braunsberger, Wybenga and Gates, 2007), arguments about the reliability of probability samples over non-probability samples (Yeager et al., 2011; Berrens et al., 2003; Brick, 2011; Chang and Krosnick, 2009; Malhotra and Krosnick, 2007). For example, Yeager et al. (2011) compare the estimates derived from a series of RDD telephone surveys and Internet surveys against benchmarks largely from “official government records or high-quality federal surveys with high response rates” (p. 712). They find that probability samples consistently yielded more accurate results. Setting aside that “administrative records are often incomplete and out of date, and typically the data are not missing at random” (Brick, 2011) and that the benchmark federal surveys used in the study were ostensibly conducted using sampling modes similar or equivalent to the RDD telephone surveys, it is noteworthy that the authors rely on comparable sample sizes between the probability and non-probability samples. This belies a nearly ubiquitous assumption among public opinion researchers that non-probability samples should be evaluated using sampling frameworks which are equivalent to those of probability samples.

Scholarly comparisons of probability and non-probability samples generally assert a false equivalence between the two. They assume that equivalent sampling frameworks can be used to empirically validate the accuracy of both probability and non-probability samples (in terms of their ability to accurately estimate a given variable of interest). While this makes for an ostensibly logical methodological control within a research design (e.g. matching sample sizes), it unreflexively transposes certain attributes and characteristics of probability samples onto non-probability samples.

Making the case for the plausibility of non-probability samples as having external validity should not by extension imply that the differences between probability and non-probability

samples are immaterial or inconsequential. On the contrary, there are good reasons to believe that the selection effects in most non-probability samples are more pronounced than they are in probability samples. In probability and non-probability samples of equivalent size, the former is by definition more likely to reflect the actual distribution of the population of the interest. In the context of a probability sample, assuming the sample to be approaching randomness, there is a threshold at which additional sample should produce negligible statistical power. If an equivalent threshold exists in a non-probability sample, it is almost certainly much higher than in a probability sample.

According to the central limit theorem, the distributions within two separate random samples should be roughly equivalent and should reflect the distributions within a population of interest. The same cannot be said of non-random sample. The distributions within two separate non-random samples of equivalent size will almost invariably be different from one another and also deviate from a given population of interest. However, depending on the sampling method, additional observations may contribute significantly to the sample composition, thus enhancing the potential for weighting techniques to arrive at more representative inferences about a given population.

Many non-probability sampling frameworks consistently reproduce systematic selection bias and thus would not realize any gains in external validity as a result of additional sample. Other frameworks, however, may result in the diversification of the sample composition as the size increases. No matter what the sample size, that respondents self-select into non-probability samples is constant, but it is conceivable that the pattern of self-selection—ergo, the non-response bias itself—is variable across time. In said cases, increasing the size of a non-probability sample may have the effect of reducing non-response bias.

The prospect that increased sample size may, in certain contexts, further diversify the composition of a non-probability sample also has the potential benefit of reducing coverage error. [Van de Kerckhove et al. \(2009\)](#) find little evidence of non-response bias even in probability samples with low response rates; they do, however, find indications of coverage

error even in cases where the coverage rate is above 80 per cent. Depending on the sampling framework being sufficiently broad and comprehensive, it is possible that non-response bias is equally inconsequential in certain non-probability samples. Moreover, certain non-probability sampling frameworks may vary in terms of their coverage of a population of interest as the sample size increases. In such instances, additional sample may contribute to a reduction in coverage error.

Sample sizes of approximately 1,000 respondents have long been the convention for probability samples used to make statistical inferences about a population of interest. This reflects a tacit consensus among most public opinion researchers that a peg of three per cent is a reasonable margin of error. But calculating a margin of error as the inverse of the square root of the sample size requires an assumption that the sample is randomly selected. Since non-probability samples do not meet this criterion, calculating a margin of error is, at best, more complex. And yet, in comparative analyses between probability and non-probability samples, there are few if any allowances for said complexity. One such allowance involves recognizing that the order in which non-random sample accrues is less uniform than that of random selection. By extension, a somewhat arbitrary cap of the first n observations collected inhibits certain non-probability sampling frameworks from amassing the requisite sample size to effectively control for obvious and non-obvious selection effects.

The upshot here is that the epistemic properties of probability samples should not be unreflexively applied to their non-probability counterparts. [Baker et al. \(2013, p. 99-100\)](#) observe that:

Unlike probability sampling, there is no single framework that adequately encompasses all of non-probability sampling non-probability sampling is a collection of methods rather than a single method, and it is difficult if not impossible to ascribe properties that apply to all non-probability sampling methodologies.

Accordingly, comparative analyses of probability and non-probability samples require attention to the variant properties between them.

At most, for certain forms of non-probability sample, statistical power may be a function of sample size, which need not be constrained according to existing conventions for probability sample. But at the very least, further introspection is warranted in terms of whether the current framework for comparing probability and non-probability samples is appropriate. It is possible that the different properties of probability and non-probability samples warrant different theoretical and methodological considerations in their individual and comparative evaluations. Treating the two as equivalent, either in terms of the statistical techniques applied in weighting the sample or the sample size and composition required in order to claim external validity, overlooks their fundamental differences.

To that end, the following analysis relaxes the constraint of sample size equivalence between probability and non-probability samples. It speculates that, as the size of the non-probability sample increases, so too does the external validity of the estimates derived from the modelled data. It then compares those estimates with the results of a robust probability sample.

3 Data

In demonstrating the capacity of certain non-probability samples to facilitate reliable inferences about a population of interest, the case of the 2015 Canadian federal election is instructive. Two of the most formidable samples of their kind collected during and in relation to that particular election—the Canadian Election Study (CES) and *Vote Compass*—permit us to substantively interrogate several of the theoretical claims posited herein regarding comparisons between probability and non-probability samples.

The external validity of these two samples is evaluated by comparing the accuracy of each sample in terms of forecasting the outcome of the 2015 Canadian federal election. This is an admittedly imperfect test. Neither instrument was designed to act as a poll or specifically to forecast electoral outcomes. These data are nevertheless worthwhile sites of inquiry when

it comes to comparing probability and non-probability samples for a number of reasons. First, each is arguably the highest calibre sample of its kind generated during the election campaign—certainly in terms of sample size, but also in terms of the quality of the sampling framework. This allows us to compare exemplar probability and non-probability samples. Second, they each offer a high degree of transparency into the composition of the sample and the mechanics of the sampling framework. Notwithstanding their sampling mode, commercial polls are inconsistent at best in terms of making available raw sample, rarely if ever report weights or response rates, and have been subject to criticisms of herding (Sturgis et al., 2016; Whiteley, 2016). The absence of said information obscures details about the sample that contribute to its evaluation and the verification of reported results. Third, both samples—the CES via a rolling cross-section design and *Vote Compass* by virtue of continuous operation throughout the course of the 2015 Canadian federal election campaign—are not single point-in-time samples. As such, they are uniquely capable of monitoring campaign effects (Blais et al., 2000; Blais and Boyer, 1996; Johnston et al., 1996; Johnston, 1992; Johnston and Brady, 2002).

3.1 Canadian Election Study

A robust probability sample, the CES has functioned as Canada's response to the American National Election Studies (ANES) since 1965. Its inaugural principal investigators included John Meisel, Philip Converse, Maurice Pinard, Peter Regenstreif, and Mildred Schwartz (Kanji, Bilodeau and Scotto, 2012). As of 1997, the CES has also received support from Elections Canada, Canada's federal election commission.

The 2015 CES includes three modes of sample collection, including a rolling cross-sectional sample selected using a modified RDD procedure (the Campaign-Period Survey or CPS); a return-to-sample phone-based re-interview with respondents to the CPS after the election (the Post-Election Survey or PES); and a paper-based survey that respondents to the PES could opt to take part in (the Mail-Back Survey or MBS) (Fournier et al., 2015;

Northrup, 2016). The 2015 CES also included an online component, which involved recruiting panel respondents from a sample provider and directing them to an online questionnaire, so as to adhere to the principles of probability sampling.¹ The field dates, sample sizes, and response rates for each sampling mode are reported in Table 1.

Table 1: 2015 CES Sample Attributes

Mode	Field Dates	N	Response rate
Campaign-Period Survey (CPS)	2015-09-08 to 2015-10-18	4,202	37%
Post-Election Survey (PES)	2015-10-20 to 2015-12-23	2,988	71%
Mail-Back Survey (MBS)	N/A	1,289	61%
Online	N/A	7,412	N/A

Source: Northrup (2016), Author’s calculations.

The CES makes available for analysis an archetypal probability sample—one with an uncommonly large sample size and methodological transparency. Its use of multiple modes also allows comparison of live interviewer and online probability samples.

3.2 Vote Compass

A unique source of non-probability sample, *Vote Compass* is an online, survey-based instrument that purports to estimate a users alignment with the candidates running in a given election campaign.² Users respond to a questionnaire relating to their political views and are then presented with a series of visualizations that represent the distance between the user and each candidate. *Vote Compass* falls within a class of online instruments commonly referred to as Voting Advice Applications (VAAs) (Alvarez et al., 2014; Fossen and Anderson, 2014), although its stewards which include the author of this paper argue for its exclusion from this definition based on a particular set of attributes (van der Linden and Dufresne, 2017; Dufresne and van der Linden, 2015).

The *Vote Compass* questionnaire is election-specific and concentrated on the particular issues that delineate between the positions of the candidates for office in a given campaign.

¹See <https://ces-eec.arts.ubc.ca/english-section/surveys/> for details.

²See <http://www.votecompass.com> for details.

However, it consistently includes survey items that capture a range of sociodemographic attributes, which serve as weighting variables, as well as vote intention.

Vote Compass is typically operated in partnership with major news media organizations during election campaigns. Said media partners promote the initiative across their networks, which draws considerable audience. *Vote Compass* itself also contains numerous built-in features that encourage sharing across social media platforms, thus furthering its reach.

Though *Vote Compass* sample is self-selected by virtue of its distribution model (i.e. open online access), the size, attributes, and composition of the sample set it apart in a variety of ways from most of its non-probability counterparts. In many jurisdictions, the public opinion datasets collected by *Vote Compass* during election campaigns are several orders of magnitude larger than any other such sample on record. Respondents provide a rich battery of sociodemographic identifiers and have unique incentives to do so honestly—the result they receive depends on their responses.

Moreover, respondents ostensibly use *Vote Compass* because they are seeking an accurate reflection of how their views situate them in a given political landscape. In order to achieve this end, they must provide accurate representations of themselves. *Vote Compass* solicits a substantial amount of sociodemographic and behavioural information, including numerous variables that correspond with those included in population-level datasets such as the national census and General Social Survey. These allow for a rigorous weighting schema to be developed and applied to the *Vote Compass* sample. Given the sample size, weights can be developed not only for marginal distributions but, in my cases, at the level of interaction of two or more socio-demographic variables. Thus, instead of separately weighting the marginal distribution of gender and age in the population of interest, *Vote Compass* data can be weighted by cross sections of age and gender.

Finally, the composition of the *Vote Compass* sample may uniquely control for selection effects not readily addressed by demographic weights. Given the context in which *Vote Compass* operates, it would be reasonable to assume that participation skews towards politically

interested individuals. Indeed, in an analysis of *Vote Compass* data, Johnston (2017) finds that early respondents demonstrate particularly high levels of political interest. By the end of the campaign, however, average political interest among respondents declines by as much as 20 percent. Johnston (2017, p. 103) argues that motivational dynamics are at play:

By implication, early participants—less shy about their choice or more likely to have made one—are more interested in politics. They have less need of the tools immediate benefit, the compass itself. Late participants, evidently, are more likely to be making up an informational deficit. What they do not yet have, or do not want to reveal, is a party preference.

That the composition of the sample changes over the course of the campaign may result in the reduction of coverage bias over time and with the accrual of additional sample. Indeed, as Johnston (2017, p. 99) notes:

In any case, the correspondence of the VC to published commercial polls is striking. The aggregate of self-recruited VC participants brought themselves to roughly the same place as survey respondents were brought by the aggregate of commercial polls. Even the dynamics are similar, with certain telling exceptions.

Taking together the size, attributes and composition of the *Vote Compass* sample, further analysis is warranted as to how said sample compares, in terms of its external validity, with probability samples such as the CES.

4 Method

The result of the 2015 Canadian federal election serves as a benchmark of public opinion at a particular point in time. The degree to which the CES and *Vote Compass* can accurately estimate the distribution of opinion at said point in time is an indicator of their capacity to

Table 2: Vote intention variable design

	CES	Vote Compass
Vote intention	Which party do you think you will vote for?	If the Canadian federal election were to take place today, which party would you vote for?
Vote leaning (if vote intention is don't know / undecided)	Is there a party you are leaning towards?	Which party are you leaning toward?

render representative inferences about a given population of interest—in this case, Canadian eligible voters.

Although, for reasons previously articulated, using forecasts as tests of external validity of the CES and *Vote Compass* samples are but one of many possible modes of interrogation, it is nevertheless a worthwhile test to use actual outcomes as a benchmark rather than rely on comparisons with previous studies. Tests that compare sample estimates with previous sample estimates always run the risk of reproducing a systemic bias inherent to a given sampling framework or class of frameworks.

4.1 Vote share projections

Given that the sample size of most commercial polls does not permit prediction at the level of individual electoral districts—referred to as ridings in the Canadian context—the most common forecasting practice is to report overall vote share.

Vote share estimates are derived from both the CES and *Vote Compass* using the stated vote intention of the respondent and, where available, the vote leaning variable. Vote intention is captured using the survey items detailed in Table 2.

The construction of the vote intention variables is sufficiently similar between the CES and *Vote Compass* to compare results.

To evaluate the forecasting accuracy of the vote share projections from each sample, the

root-mean-square error (RMSE) is used, which is the square root of the average of squared differences between prediction and actual observation, given as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

Where n is the number of candidates whose vote share is forecasted, y is the actual observation and \hat{y} is the predicted value.

RMSE is arguably a more robust measure of forecast accuracy than MAE because the errors are squared before they are averaged, and thus increases with the variance of the frequency distribution of error magnitudes. Unlike MAE, RMSE effectively gives greater weight to larger errors.

Given the sample size of most commercial polls, the ability to accurately forecast electoral outcomes in terms of vote share is constrained to national and regional geographies, beyond which the sample becomes unreasonably distorted.

4.2 District-level projections

Due to sample size constraints of conventional polls, results at the electoral district level are typically comprised of modelled estimates derived from vote share projections, rather than outright district-level projections. However, both the online wave of the CES and *Vote Compass* associated respondents with their electoral districts.³ This permits a more fine-grained analysis of forecast accuracy and arguably a more robust indicator of the representativeness of the sample. In the case of the 2015 Canadian federal election, it effectively means predicting electoral outcomes in 338 district-level races as opposed to one federal-level race.

Canadas parliamentary system is rooted in the Westminster tradition, wherein Parliament consists of the Crown and an upper and lower legislative Chamber. Members of the lower Chamber, or House of Commons, are individually elected to represent single electoral

³The CES does not make electoral district identifiers available for its phone survey respondents.

districts. As of the 2015 federal election, there were 338 electoral districts. Elections are determined using a single-member constituency, first-past-the-post or simple-plurality electoral system, wherein the candidate receiving the most votes in a district is elected to represent the constituents of that district. Thus, the determinant of electoral victory in the context of Canadian federal elections is not the share of the popular vote received by any given party, but rather how many seats that party's candidates are elected to represent.

To that end, the accuracy rate used to evaluate the CES and *Vote Compass* in terms of district-level projections is given as:

$$AccuracyRate = \frac{CorrectPredictions}{TotalSeats}$$

Both CES and *Vote Compass* data are modelled using multi-level regression and post-stratification (Dever, Rafferty and Valliant, 2008; Gelman et al., 2007). Weighting variables include past vote, sex, age, education, language, and riding.

5 Findings

The following analysis compares the weighted projections of the CES and *Vote Compass* to the actual outcome of the 2015 Canadian federal election. The summary statistics of the web mode of the CES, as well as *Vote Compass*, are reported in Table 3.

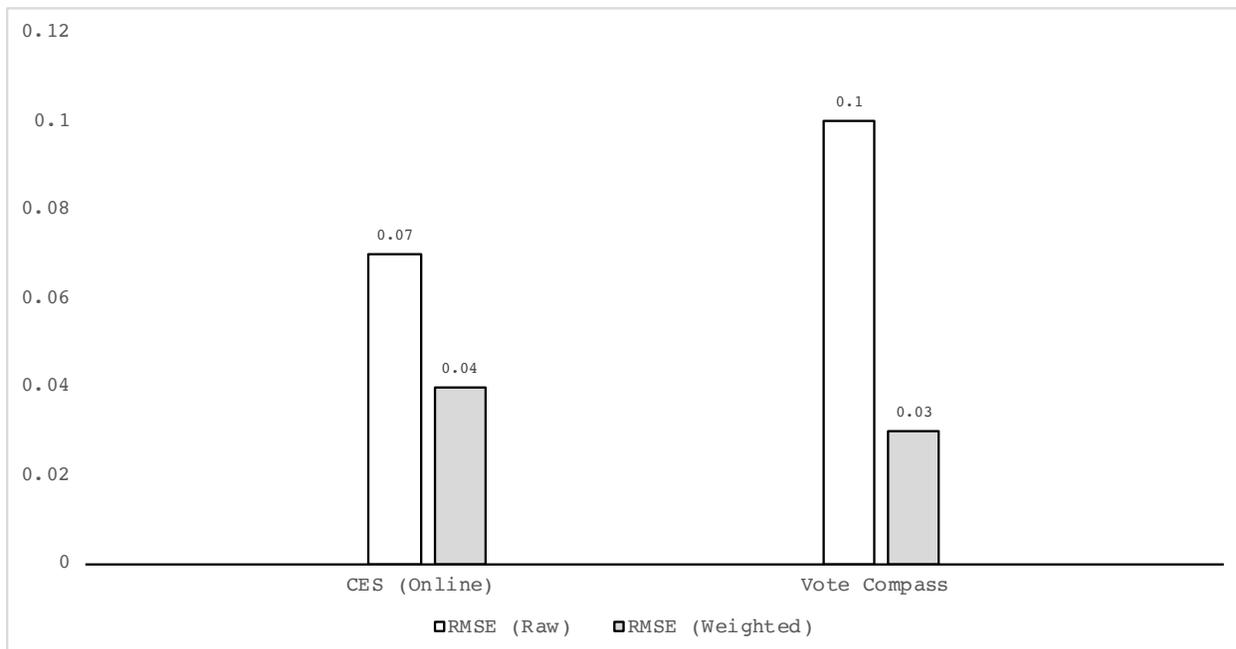
Table 3: Sample Summary Statistics

Source	Mode	N	Avg. by district	Min. by district
CES	Online	7,412	19	1
Vote Compass	Online	871,823	2,579	298

5.1 Vote Share

Figure 1 compares the predicted vote share estimated used each of the samples under analysis with the actual vote share distribution in the 2015 Canadian federal election. The RMSE by party is calculated using the vote intention measure in each of the samples.

Figure 1: Vote Share RMSE



In terms of overall RMSE, weighted vote intention outperforms its unweighted counterparts. There is a much greater difference between the raw and weighted versions of RMSE for *Vote Compass*. This suggests a larger selection bias in the raw *Vote Compass* sample than in either of the CES probability samples. However, once weights are applied, the RMSE for *Vote Compass* is actually lower than that of the online CES, which indicates that the *Vote Compass* weights are relatively effective in controlling for the sample bias.

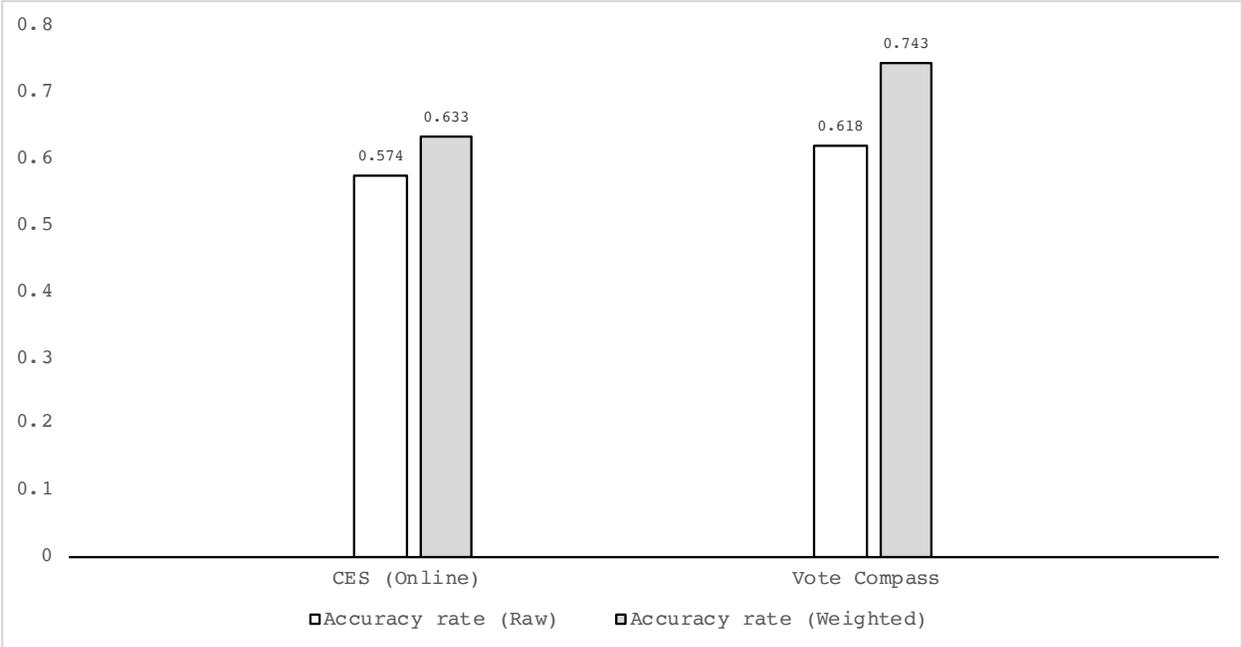
5.2 District-level projections

Although the phone-based version of the CES does not contain district-level indicators, such information is available in the online version of the CES. As online respondents to the CES

are selected via an RDD phone-based recruitment process, it serves as a probability sample for purposes of comparison with *Vote Compass*. It stands to reason that, as a probability sample, the selection of respondents to the CES should be randomly distributed across electoral districts. However, it should be taken into that, although random, the distribution will not be uniform across all 338 districts because the population of Canadian ridings varies substantially.⁴

The external validity of *Vote Compass* data at the district level can be evaluated using the accuracy rate between the CES online survey and *Vote Compass*, the results of which are reported in Figure 2.

Figure 2: Riding projection accuracy rates



The weighted vote intentions derived from *Vote Compass* produce the highest accuracy rate, correctly identifying 74.3 percent of the outcomes across Canadas 338 electoral districts. Weights continue to have little effect on the CES estimates. Both the raw and weighted *Vote Compass* data yield a more accurate forecast than either the raw or weighted CES data.

⁴According to census estimates, the average number of residents per electoral district is 99,034. The electoral district with the smallest population is Labrador with 26,728 residents. The district with the largest population is Brantford—Brant with 132,443 residents.

6 Discussion

This paper has offered theoretical justification and empirical validation in favour of the argument that certain non-probability samples can produce reliable and valid inferences about a population of interest. This is demonstrated by comparing the CES and *Vote Compass* in terms of their respective ability to accurately forecast the outcome of the 2015 Canadian federal election.

By most measures, *Vote Compass* is able to either match or exceed the accuracy of the CES, particularly so at more fine-grained geographies such as electoral districts. One must appreciate that *Vote Compass* has an enormous advantage over the CES in terms of its sample size, but this is precisely the point. In a more controlled comparative analysis of probability and non-probability samples, *Vote Compass* data would likely be less performant relative to the CES. The comparative framework for such an analysis would require similar sample sizes and like weights. But such parameters impose a false equivalence between probability and non-probability samples and limit the potential of unconventional sources of public opinion data. Absent such constraints, non-probability samples such as *Vote Compass* demonstrate external validity on par with probability samples such as the CES.

If *Vote Compass* were, for example, to restrict its sample size or weighting schema to match that of the CES, its results would likely be less accurate as the selection effects in the sample would be substantially greater. The raw estimates from *Vote Compass* indicate that the unweighted data is subject to a more substantial selection bias than the CES, which one would expect given the sampling framework. But with the full sample and a robust weighting schema tailored to that sample, the *Vote Compass* data has consistently produced estimates of electoral outcomes that have been as or more accurate than those of the CES.

Of course, the findings presented herein require further replication to confirm the robustness of the theoretical claims attached to them. But the intended contribution is not simply to argue for the external validity of certain non-probability samples; it is rather to make the case for a reconsideration of the framework that is used to evaluate such samples.

Not all non-probability samples are equal. They have myriad different properties and thus cannot be expected to behave like a probability sample. In fact, the absence of a universal framework for non-probability sampling means that we cannot expect all non-probability samples to be alike.

The variability of non-probability sample inhibits us from extrapolating the case of *Vote Compass* to non-probability samples more broadly. Not all non-probability samples have sufficient breadth and depth to make representative inference possible. Moreover, since margins of error cannot be readily calculated for non-probability samples in the same way they can be for probability sample, there is no readily available, reliable measure to indicate the capacity of a given non-probability sample for representative inference.

In the absence of such a measure, the size, attributes, and composition of a non-probability sample must be carefully evaluated before endeavouring to derive generalizable inferences. A useful heuristic is to estimate known values within the population of interest but which are not present in the sample and then observing whether said values can be accurately predicted. In the case of *Vote Compass*, predicting the distribution of vote share in the 2015 Canadian federal election serves this end, but other possibilities may include predicting certain census values.

Given the lack of standardized tests to establish the reliability and validity of certain non-probability samples over others, considerations should be applied with great care. While non-probability samples can provide opportunities to learn more about a population or subpopulation of interest, not all non-probability sample can reliably produce these sorts of inferences. Even other VAAs do not necessarily capture the size, attributes and composition of sample to make representative inferences about a given population. Care must be taken not to falsely ascribe the properties of certain non-probability samples to others.

At the same time as technological advancements continue to make probability sampling more difficult to achieve in practice, they also make available new opportunities to construct non-probability samples that are capable of producing representative inferences. In order to

leverage this opportunity, a new framework is needed for the evaluation of non-probability sample as a means to establish the veracity and the robustness of its estimates.

References

- Alvarez, R Michael, Ines Levin, Peter Mair and Alexander Trechsel. 2014. "Party preferences in the digital age: The impact of voting advice applications." *Party Politics* 20(2):227–236.
- Baker, Reg, J Michael Brick, Nancy A Bates, Mike Battaglia, Mick P Couper, Jill A Dever, Krista J Gile and Roger Tourangeau. 2013. "Summary report of the AAPOR task force on non-probability sampling." *Journal of Survey Statistics and Methodology* 1(2):90–143.
- Battaglia, Michael P, David Izrael, David C Hoaglin and Martin R Frankel. 2009. "Practical considerations in raking survey data." *Survey Practice* 2(5):1–10.
- 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.
- Blais, André and M Martin Boyer. 1996. "Assessing the impact of televised debates: The case of the 1988 Canadian election." *British Journal of Political Science* 26(2):143–164.
- Blais, André, Neil Nevitte, Elisabeth Gidengil and Richard Nadeau. 2000. "Do people have feelings toward leaders about whom they say they know nothing?" *The Public Opinion Quarterly* 64(4):452–463.
- Braunsberger, Karin, Hans Wybenga and Roger Gates. 2007. "A comparison of reliability between telephone and web-based surveys." *Journal of Business Research* 60(7):758–764.
- Brick, J Michael. 2011. "The future of survey sampling." *Public Opinion Quarterly* 75(5):872–888.
- 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.
- Council, National Research et al. 2013. *Nonresponse in social science surveys: A research agenda*. National Academies Press.
- De Heer, W and E De Leeuw. 2002. "Trends in household survey nonresponse: A longitudinal and international comparison." *Survey nonresponse* p. 41.
- Dever, Jill A, Ann Rafferty and Richard Valliant. 2008. Internet surveys: Can statistical adjustments eliminate coverage bias? In *Survey Research Methods*. Vol. 2 pp. 47–60.
- Dufresne, Yannick and Clifton van der Linden. 2015. Digital technology and civic engagement: the case of Vote Compass. In *Canadian Election Analysis, 2015: Communication, Strategy, and Democracy*, ed. Thierry Giasson and Alexander J Marland. Vancouver: UBC Press pp. 114–116.
- Edgington, Eugene S. 1966. "Statistical inference and nonrandom samples." *Psychological Bulletin* 66(6):485.
- Fossen, Thomas and Joel Anderson. 2014. "What's the point of voting advice applications? Competing perspectives on democracy and citizenship." *Electoral Studies* 36:244–251.
- Fournier, Patrick, Fred Cutler, Stuart Soroka and Dietlind Stolle. 2015. "The 2015 Canadian Election Study." [dataset].

- Gelman, Andrew et al. 2007. "Struggles with survey weighting and regression modeling." *Statistical Science* 22(2):153–164.
- Holbrook, Allyson, Jon A Krosnick, Alison Pfent et al. 2007. "The causes and consequences of response rates in surveys by the news media and government contractor survey research firms." *Advances in telephone survey methodology* pp. 499–528.
- Johnston, Richard. 1992. *Letting the people decide: Dynamics of a Canadian election*. Stanford University Press.
- Johnston, Richard. 2017. "Vote Compass in British Columbia: insights from and about published polls." *Journal of Elections, Public Opinion and Parties* 27(1):97–109.
- Johnston, Richard, André Blais, Elisabeth Gidengil and Neil Nevitte. 1996. *Challenge of Direct Democracy: The 1992 Canadian Referendum*. McGill-Queen's Press-MQUP.
- Johnston, Richard and Henry E Brady. 2002. "The rolling cross-section design." *Electoral Studies* 21(2):283–295.
- Kanji, Mebs, Antoine Bilodeau and Thomas J Scotto. 2012. *The Canadian Election Studies: Assessing Four Decades of Influence*. UBC Press.
- Keeter, Scott, Courtney Kennedy, Michael Dimock, Jonathan Best and Peyton Craighill. 2006. "Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey." *International Journal of Public Opinion Quarterly* 70(5):759–779.
- Kohut, Andrew, Scott Keeter, Carroll Doherty, Michael Dimock and Leah Christian. 2012. "Assessing the representativeness of public opinion surveys." *Washington, DC: Pew Research Center* .
- Lee, Sunghee. 2006. "Propensity score adjustment as a weighting scheme for volunteer panel web surveys." *Journal of official statistics* 22(2):329.
- Lee, Sunghee and Richard Valliant. 2009. "Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment." *Sociological Methods & Research* 37(3):319–343.
- 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.
- Northrup, David. 2016. The 2015 Canadian Election Study Technical Documentation. Technical report Institute for Social Research, York University.
- Schonlau, Matthias, Arthur Van Soest, Arie Kapteyn and Mick Couper. 2009. "Selection bias in web surveys and the use of propensity scores." *Sociological Methods & Research* 37(3):291–318.
- Smith, TMF. 1983. "On the validity of inferences from non-random sample." *Journal of the Royal Statistical Society. Series A (General)* pp. 394–403.
- Steeh, Charlotte, Nicole Kirgis, Brian Cannon and Jeff DeWitt. 2001. "Are they really as bad as they seem? nonresponse rates' at the end of the twentieth century." *Journal of Official Statistics* 17(2):227.
- Sturgis, Patrick, Baker Nick, Callegaro Mario, Fisher Stephen, Green Jane, Will Jennings, Kuha Jouni, Lauderdale Ben and Smith Patten. 2016. "Report of the Inquiry into the 2015 British general election opinion polls."
- Van de Kerckhove, Wendy, Jill M Montaquila, Priscilla R Carver and J Michael Brick. 2009. "An Evaluation of Bias in the 2007 National Household Education Surveys Program: Results from a Special Data Collection Effort. NCES 2009-029." *National Center*

for Education Statistics .

- van der Linden, Clifton and Yannick Dufresne. 2017. “The curse of dimensionality in Voting Advice Applications: reliability and validity in algorithm design.” *Journal of Elections, Public Opinion and Parties* 27(1):9–30.
- Vavreck, Lynn and Douglas Rivers. 2008. “The 2006 cooperative congressional election study.” *Journal of Elections, Public Opinion and Parties* 18(4):355–366.
- Whiteley, Paul. 2016. “Why Did the Polls Get It Wrong in the 2015 General Election? Evaluating the Inquiry into Pre-Election Polls.” *The Political Quarterly* 87(3):437–442.
- Yeager, David S, Jon A Krosnick, LinChiat Chang, 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.