EKOS POLITICS

ACCURATE POLLING, FLAWED FORECAST
AN EMPIRICAL RETROSPECTIVE ON ELECTION 41

By Frank Graves

1.0 INTRODUCTION: THE NATURE AND PURPOSE OF THIS TEST

"Mistakes are the portals of discovery" -James Joyce

As the dust settles on what was an extraordinary 41st Canadian election campaign, it may be worthwhile taking a more careful look back at the polls. While focussing on our own research, our observations are intended to have more general relevance to the debate about the role of polling in the democratic process. In fact, the research has important lessons on the shifting nature of our society which has important implications for the role of polling beyond the narrow yardstick of how well a final poll of voter intention resembles the final election result.

This exercise is not an attempt at apology or rationalisation. Our final election poll was over five points below the final results for the Conservative Party and our forecast of a Conservative minority was a mistake. The Conservative Party did considerably better than our final poll and went on to win a majority government. At the time, we committed to doing some hard thinking and testing to find out what the source(s) of the gap between our final polls and the election outcome was.

The ensuing exercise has provided some very interesting and surprising insights on why our final poll differed from the actual outcome. In the process of formally testing these questions, we uncovered some serendipitous findings which speak to some of the broader challenges confronting modern polling. We also found some important discoveries about the shifting nature of democracy and political participation which may have more important implications than merely assessing which poll came closest to the final outcome. The research shows a growing tension between the challenge of understanding all voting members of society and the more specific challenge of forecasting the outcome of the election.

The general view of the polls is that they performed rather poorly and no pollsters accurately and clearly predicted the final Conservative majority. This failure of forecast (a first for us), and the generally inauspicious connections between final polling and the actual election has been interpreted by some as further evidence that polls no longer "work". Others have speculated that the gap may well reflect the fact that voters altered their final choices as a strategic response to the polls themselves. Still others have noted that the difference between the poll and the final outcome may reflect the differences in the roughly 60 per cent who came out to vote and the broader population of all voters.

The objectives of our research are to test these three separate hypotheses:

1) Our polls were flawed either due to systematic sampling error or measurement error. In particular, did EKOS polling systematically understate Conservative support?

2) The second hypothesis was that there were final movements which occurred basically in the ballot booth (we polled until Sunday) and that these late shifts accounted for the final discrepancy. More specifically, the speculation was that enough of the residual Liberal voter support abandoned the Liberal Party and shifted to the Conservatives in order to forestall the chances of a NDP–led coalition government. Embedded in this hypothesis is also the notion that the polls themselves had a causal influence on the final shape of the election (as the knowledge of a possible NDP–led government was a surprise that would only have been evident through reading the polls). Notably, EKOS was the first and most consistent source of information on this NDP surge.

3) The final hypothesis was that the differences in the final polling and the election outcome were due to differential voter turnout between all eligible voters and the actual subpopulation of actual voters. More specifically, the idea here is that the Conservatives were much more successful in getting out their voters than other parties.

Of course, it is possible that the gap could be a product of a mixture of all three of these hypotheses. If the first hypothesis of survey bias is true, it has damning implications for our confidence in modern survey research. In our case, we took special care to model the entire population (including non-internet and cell only households). We also incorporated random sampling methods with careful call-backs, replacement, and weighting. If our methods were no longer capable of accurately modelling overall populations, this would be a severe blow to our credibility. Worse, it seems that there was no apparent advantage to having used comprehensive and random sampling as some of our competitors who had utilised non-randomly recruited panels or failed to bother with the growing number of cell only households had indeed been slightly closer to the final outcome. Our evidence leaves us comforted that it is still better to randomly and comprehensively sample but in so doing, we may have had a larger gap because we had a more accurate sample of the entire voting population which was more different from the final vote than the partial portions others were sampling.

If the second hypothesis of an eleventh hour shift is true, it poses at least two major challenges: first, a severe methodological challenge occurs in a world where the act of observing, recording, and reporting on public opinion measurably alters public opinion (and voting behaviour). The pollster becomes a coagent in the subject matter in ways that could make Heisenberg's uncertainty principle seem minuscule. It is already very difficult to accurately measure and model human attitudes behaviour. Does it become intractable when the reporting of polling is leading to and altering the very matter it seeks to record?

The second issue is one of ethics. If the reporting of polling is altering political outcomes, it that a desirable thing? Should the public be forced to make "purer" democratic choices in state of relative cerebral hygiene (when it comes to polling data) or should, for example, residual Liberal voters be entitled to shift allegiances to the Conservatives because polls suggest that to do otherwise might produce an NDP-led coalition government of which they disapprove?

These are very difficult questions and our contribution here is intended to be unremittingly grounded in empirical evidence and logic. In constructing one last survey before putting away our federal vote intention polling tools for some time, we constructed a series of critically falsifiable tests designed to help us understand what was wrong and what was right about our polling. Failure and learning are ingredients of progress but only if the opportunities are clearly seized.

2.0 WERE OUR POLLS BIASED?

Figure 2.1 – Final EKOS poll versus actual results

Q. If a federal election were held tomorrow, which party would you vote for?
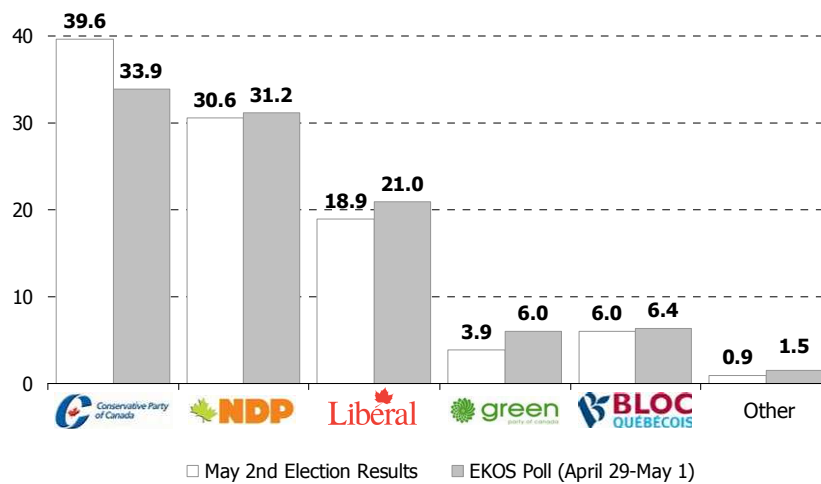
Note: The data on federal vote intention are based on decided and leaning voters only.

BASE: Decided voters; April 29-May 1, 2011 (n=2,876)

As shown in the chart above, our final poll showed a result which was quite off from the final outcome; while it was within the margin of error of the other final polls, it was significantly off from the better than expected Conservative results. In this section, we consider to what extent survey errors account for the discrepancy. In anticipation of our later review of the turnout hypothesis, we table the proposition that the application of margin of error from a sample of all eligible voters to the final population of the 60 per cent who actually voted is technically wrong in a non-trivial manner. This is particularly true when there are systematic and large differences between the population parameters of all eligible voters and those of the final actual voters. Unless the pollsters are specifically claiming that they were measuring actual rather than eligible voters, propinquity to the actual vote may not be a valid measure of polling accuracy. This is not to deny the importance of the separate challenge of forecasting the election result in which case we readily acknowledge our shortfall.

There are three main types of errors that can occur in survey research. First, there is the unavoidable random error commonly discussed under the rubric of margin of error. The gap between our final poll and the Conservative results was outside of the roughly two-point margin of error for our final poll. The second and third types of error are the more serious issues of systematic bias. This can be a product of both sampling bias (the methods under or overrepresented key groups) and measurement error (the method for asking the questions

caused error as in the case of a leading question asked before the vote intention question). We are very careful to avoid measurement errors by always asking the vote intention questions up front in a neutral and consistent manner. In the case of sampling, we take extra care to ensure that our samples included the entire (non-institutionalised, charter language speaking) population of voters. Evidence on declining response rates has shown that when good random sampling protocols are employed, the resulting samples continue to do a good job of modelling the broader populations.

Beyond the use of good sampling practices we conducted two specific tests to assess this hypothesis. The first test involved reviewing key measures of representativeness of our samples. The samples that we recruit using interactive voice response (IVR) tend to be closer to the population parameters than even our random computer assisted telephone interview (CATI) samples[1].

In addition to the usual tests of representativeness, we also collected vote behaviour in the last election as a control variable. The distribution of the sample on 2008 vote behaviour can be considered a good test of whether the sample is biased to one or another political party.

Figure 2.2 – Tracking 2008 vote

Q. How did you vote in the federal election held in 2008?
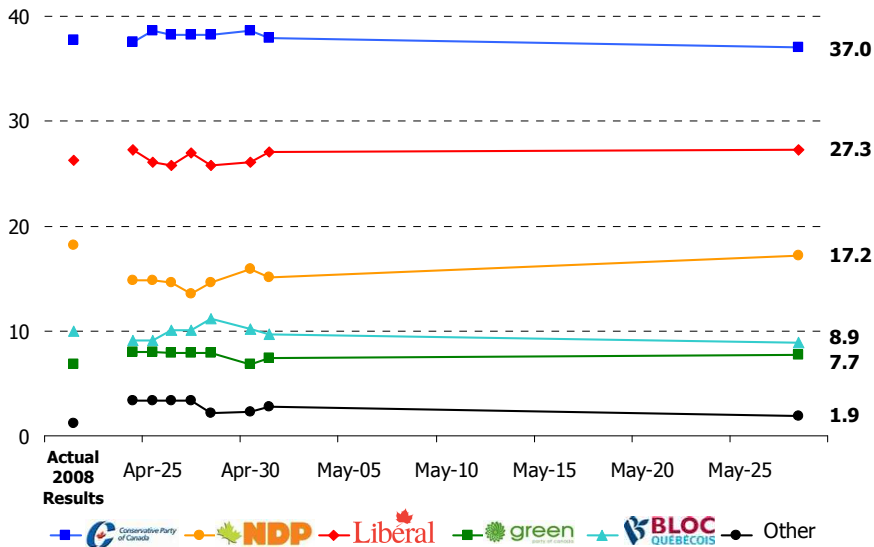
BASE: Those who voted in the 2008 election; most recent data point May 20-28, 2011 (n=2,195)

As one can see from the chart above, the final sample came very close to modelling the actual distribution of voters in 2008. This is a fairly demanding test and the fact that not only our final

---

[1] EKOS Research Associates, "Interactive Voice Response: The Past, the Present, and Into the Future – Presentation to The MRIA Ottawa Chapter", January 21, 2011

poll, but all of our polls throughout the election were within the margin of error of the 2008 election is one of the reasons that we had such high confidence in the integrity of the sample. It is possible that respondents could accurately report their 2008 vote but somehow be biased on intentions for 2011, but frankly, this is implausible. The 2008 reports do have the advantage of focussing responses on the voting population as opposed to all eligible voters. This is a crucial point which we will return to later but we note that our samples showed no evidence of underrepresenting Conservative voters.

We conducted a new primary test of sampling bias. Using our core methodology, we went back to the field and conducted a random sample of nearly 2,800 eligible voters. We wanted to probe on some specific questions related to the shift hypothesis but another equally important purpose was to see if the polling methods accurately reflected the actual 2011 results. If our methodology underrepresented Conservative voters, it should continue to do so after the election.

Figure 2.3 – May 2nd vote

Q. How did you vote in the most recent federal election, held on May 2nd?

☐ May 2nd Election Results   ▨ EKOS Poll (May 20-28)
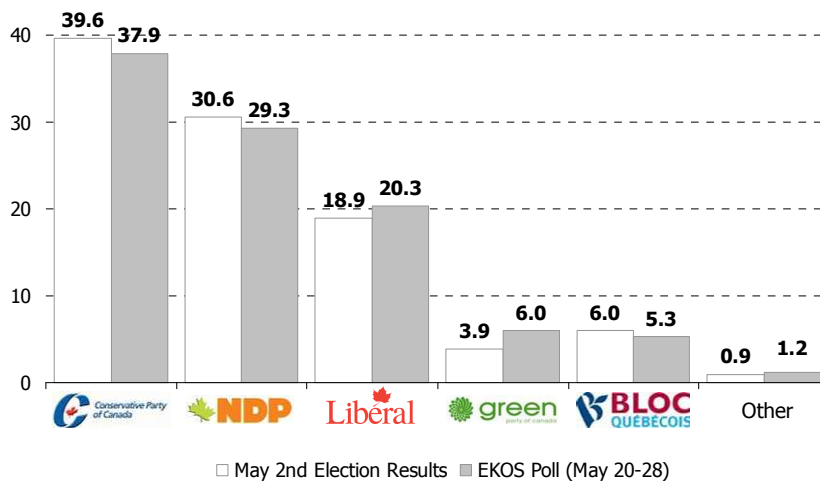
Note: The data is based on those who say they voted on May 2nd. Our survey also finds that 20.9% of respondents either did not respond or did not vote.

BASE: Those who voted on May 2nd; May 20-28, 2011 (n=2,209)

As one can see in Figure 2.3, the overall results of "how did you vote" are well within the margin of error of Election 41. The public stubbornly over remember their Green Party support as they consistently overstate their intentions to vote Green but everything else is very close to the actual result (more on the Green Party later). The main point here is that the methods passed the intersubjective repeatability test of generating the true outcome. Again, it is possible that the methods somehow "worked" on this test but failed in the final polls but that is frankly quite implausible.

We therefore conclude the testing of survey bias hypothesis with a high level of confidence that the gap was not a product of faulty data or sample bias. As we shall see some of the rigour that we applied to ensuring the representation of some groups that are often systematically ignored in other sampling had the ironic impact of enlarging the gap between our final poll and the election. This will become clearer in testing of the differential turnout hypothesis.
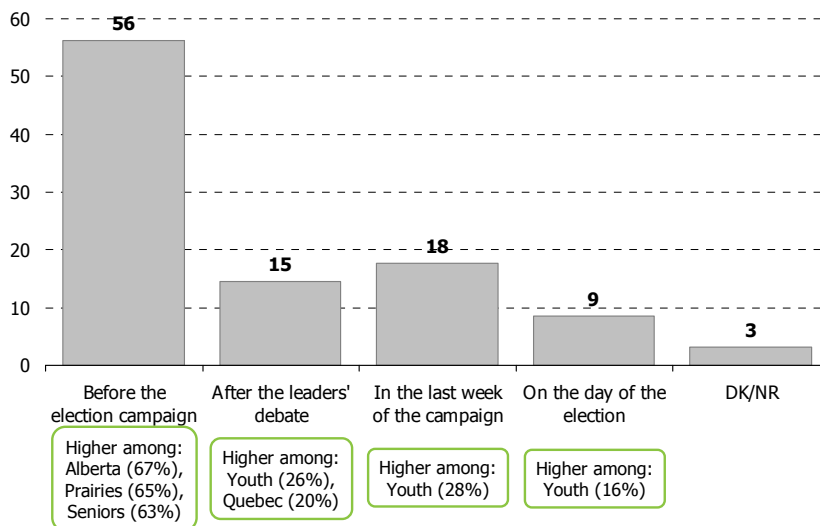
3.0 SECOND HYPOTHESIS: SOME VOTERS CHANGED ON FINAL DAY

An attractive theoretical explanation for the gap was that voters switched to the Conservatives in sufficient numbers to produce the final gap. The thinking was that this was a product of fear of an NDP–led coalition. Intertwined with this hypothesis is the direct implication that the polls were actually instrumental in supporting strategic voter shifts with the entire methodological and ethical issues attendant to that possibility.

We had two separate tests for this hypothesis. First, we explored when voters made their final choices and whether this was a change of mind. If voters did change their minds, we assessed where they had been positioned before the shift. In the case of those who had shifted to the Conservative Party, we asked whether or not fear of an NDP–led coalition was the crucial factor (along with the other options of giving the Conservatives a majority and keeping the economy on track).

Figure 3.1 – Timing of final decision

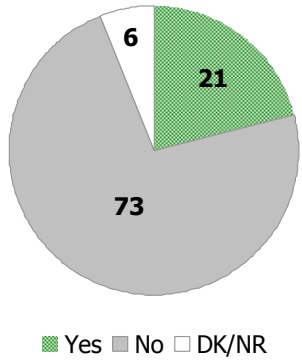*Q. When did you make your final decision regarding how you were going to vote?*

Before the election campaign	Higher among: Alberta (67%), Prairies (65%), Seniors (63%)
After the leaders' debate	Higher among: Youth (26%), Quebec (20%)
In the last week of the campaign	Higher among: Youth (28%)
On the day of the election	Higher among: Youth (16%)

Bar chart values: Before the election campaign = 56; After the leaders' debate = 15; In the last week of the campaign = 18; On the day of the election = 9; DK/NR = 3
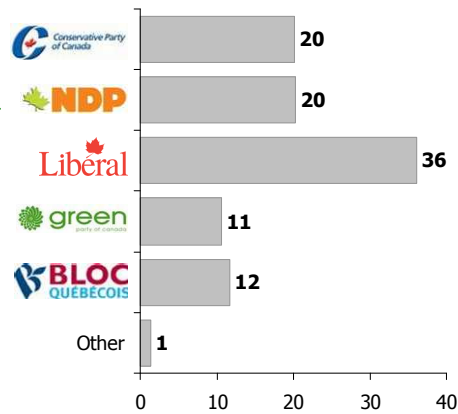
BASE: Those who voted on May 2nd; May 20-28, 2011 (n=2,203)

Figure 3.2 – Voter shifting

Q. *Were you originally leaning towards voting for a different party at the beginning of the election campaign?*

Q. *[IF YES] Which party were you leaning towards before you changed your mind?*

Conservative Party of Canada 20
NDP 20
Liberal 36
green 11
BLOC QUÉBÉCOIS 12
Other 1

Pie chart: 6, 21, 73

■ Yes ■ No □ DK/NR

BASE: Voters; May 20-28, 2011 (n=2,203) / Those who changed their mind; May 20-28, 2011 (n=414)

The evidence shows very clearly that the gap between the final polls and the election outcome was not due to last minute shifting. Whatever last minute shifting occurred was fairly picayune. Slightly less than 21 per cent of the overall 21 per cent of voters that shifted from their original choice did so on Election Day. Moreover, there was no clear pattern favouring the Conservatives so this factor simply could not explain the final gap.

Next, we asked shifters whether they shifted to the Conservatives because of concerns about an NDP-led coalition. To put a further nail in the late shift coffin, virtually no one shifted in fear of an NDP-led coalition and, in Ontario, the putative hotbed of this apocryphal late shift, there wasn't a single late shifter who did so because of this factor in our large Ontario sample.

Figure 3.3 – Biggest factor in switching (Conservative shifters only)

Q. What was the biggest factor in your final decision?

BASE: Those who switched to the Conservative Party during the campaign; May 20-28, 2011 (n=99)
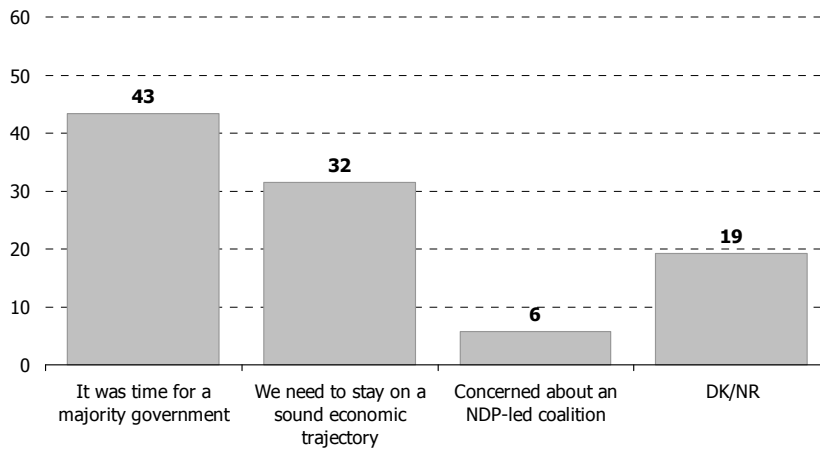
Our next test asked voters the degree to which they had been following the polls throughout the election and whether or not the polls were instrumental in determining their final voting decision. Using this evidence, we can get a pretty good handle on whether there was a final shift and what role the polls in general (and specifically the surprising polling finding that the NDP were in striking distance of forming and leading a coalition with the Liberal Party) had on final vote.

Figure 3.4 – Impact of public opinion polls

Q. *Throughout the election campaign, how closely did you follow public opinion polls?*

Q. *To what extent did public opinion polls affect your final decision on Election Day?*
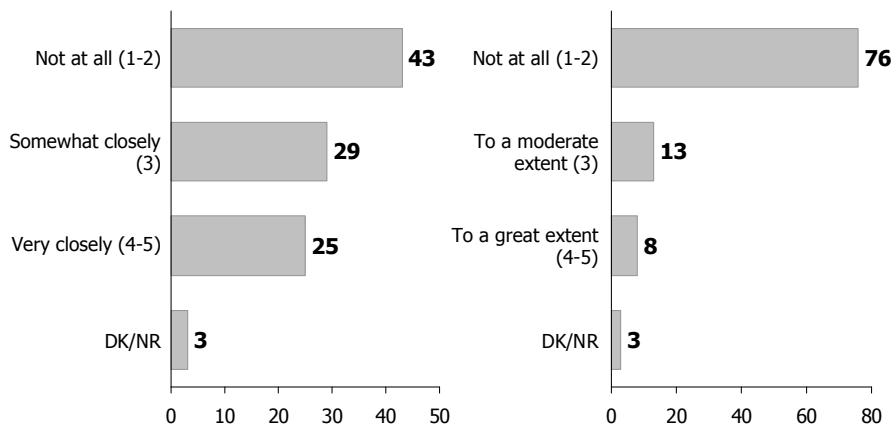
BASE: Those who voted on May 2nd; May 20-28, 2011 (n=2,203)

The test of poll attention and impacts also reinforces this conclusion and also suggests that the polls had no clear bearing on the final outcome of the election. Yes there were lots people following the polls. Anecdotally, the biggest consumers tended to be Liberal and Bloc supporters who must have watched the polls obsessively in the same manner that some cannot avert their eyes from a train wreck. Around one in five said the polls were at least a modest factor in their final decision although only four percent cited it as very important factor. When the patterns of shifting are examined for even the small group who said they shifted because of the polls, there was no clear pattern favouring any party and the late strategic shift can be safely dismissed as a spurious explanation for the gap. Perhaps even more importantly, the charge that the polls are now shaping what they purport to report on is not the case. The counterfactual hypothesis of what this election would have looked like if the polls hadn't been available to voters is that things would probably have been the same.

4.0 TURNOUT, TURNOUT, TURNOUT

The remaining hypothesis turns out to be by far the key explanation of the gap between the election results and our final poll. We can document this in a number of ways, although not perfectly. The combination of the failure of the first two hypotheses to explain the gap, coupled with clear evidence of just how powerful the turnout factor was, leads us to conclude that the prime explanation of the gap was differential turnout. This conclusion leaves open the question of what to do about this. There are very large questions about what constitutes best polling practice and even larger questions about the very nature of democracy and Canadian society and how this may be changing. On a narrower basis, if the hypothesis of differential turnout is in fact the explanation of the gap between last polls and the final result, the yardstick of measuring polling accuracy on the basis of comparing a final poll of eligible voters to the final outcome with actual voters is flawed. The error is one of improper forecasting, not inaccurate polling.
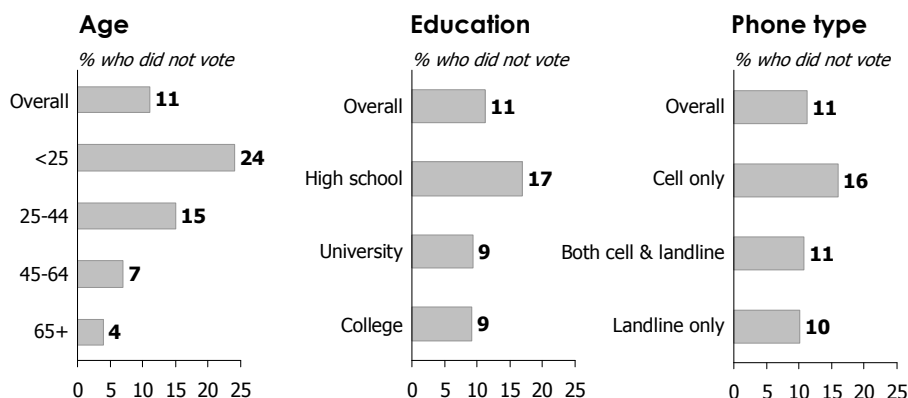
Figure 4.1 – Estimated non-voter support
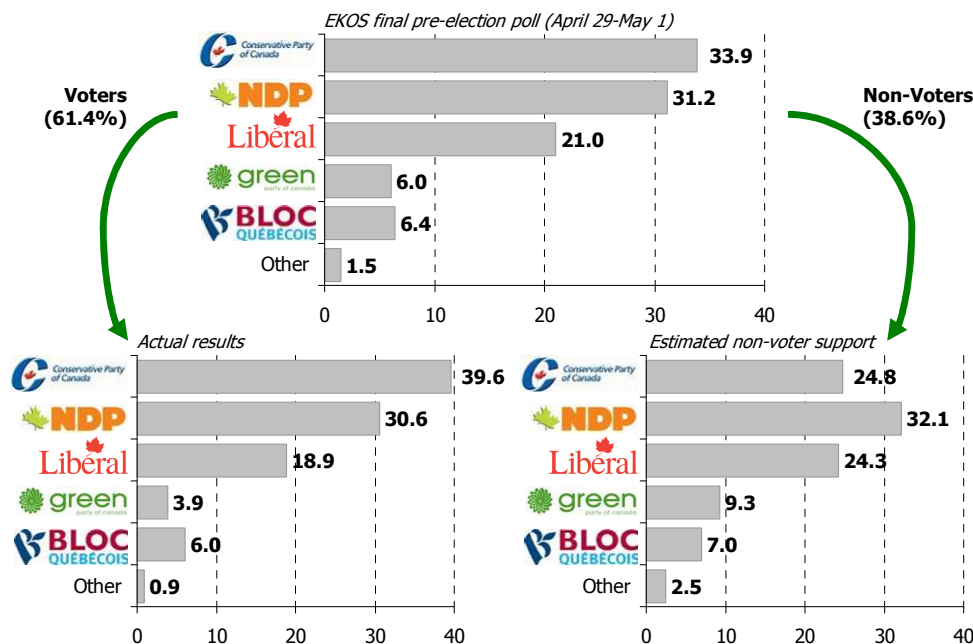
Figure 4.2 – Non-voters by key demographics

Q. How did you vote in the most recent federal election, held on May 2nd?

Age
% who did not vote

Overall	11
<25	24
25-44	15
45-64	7
65+	4

Education
% who did not vote

Overall	11
High school	17
University	9
College	9

Phone type
% who did not vote

|	
Overall	11
Cell only	16
Both cell & landline	11
Landline only	10

BASE: Canadians; May 20-28, 2011 (n=2,724)

What is the evidence that differential turnout (e.g. more Conservative supporters among actual voters than among eligible voters) is true? Figure 4.1 shows what the hypothetical composition of the nonvoting eligible population would have to be to make the entire sample unbiased. As we shall see there is strong indirect evidence that this hypothetical distribution of vote intention amongst the non-voters is highly plausible. Consider the following lines of evidence in support this hypothesis.

First of all, we have a sample of those who didn't vote in our final sample. While non-voting is typically understated due to both sampling and measurement issues (even Statistics Canada finds nearly 80 per cent of one of its recent surveys claimed to have voted in 2008) the patterns of demographic correlates are highly revealing. The incidence of claimed non-voting closely resembles the real patters available from other sources such as the Canadian Election Survey. The likelihood of actually voting is much higher among seniors and declines progressively with age. Notably we see a very similar pattern of much higher support for the Conservatives among senior voters which declines progressively with age. This factor alone is probably the biggest factor which explains why the population of all eligible voters has a lower incidence of Conservative supporters than the final vote. It would also clearly explain the surface paradox of why our samples could accurately model the 2008 vote yet "miss" the 2011 vote. Clearly the incidence of Conservative supporters is higher in the population of actual voters than it is in the population of all voters. These patterns also apply (in weaker terms to those of lower socioeconomic status who were both less likely to vote and less likely to support the Conservatives and weaker still to women).

There has been considerable discussion about the role of the growing portion of the population who no longer maintain wired landline connections and now rely exclusively on cell phones. In the United States, this number is near a third of the population. In Canada, the number is a little over 10 per cent but growing quickly. Moreover, there are systematic correlates to those who have abandoned landlines; they tend to be under 40 and of moderate to lower socioeconomic status. In our sampling, we ensure that we have a robust sample of cell only households.

Figure 4.3 – Cell phone only versus cell/landline households

Q. If a federal election were held tomorrow, which party would you vote for?
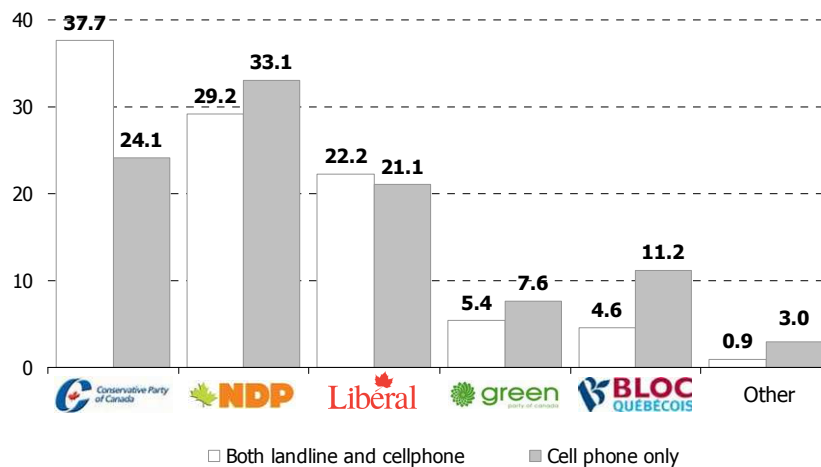
Note: The data on federal vote intention are based on decided and leaning voters only.

BASE: Decided voters; April 29-May 1, 2011 (n=1,836)

As shown in Figure 4.3, this methodological strength actually increased the distance between our final results and the actual election. Why? Once again there is a significant positive association between the likelihood of not voting and the likelihood of not supporting the Conservatives. Cell phone only respondents were around 14 points less likely to cite the Conservative Party as their vote intention. Moreover, as seen in Figure 4.2 cell only respondents were nearly 50 per cent more likely to be non voters than other respondent types. We can readily see how the discrepancies between Conservative support and likelihood of voting in both younger and cell only respondents could make the hypothetical distribution of vote intention in the nearly 40 per cent of non-voters closely resembles Figure 4.1. Given this evidence, the likelihood is that the levels of Conservative support among the entire population of eligible voters were much closer to our final finding of 34 per cent than the 39 per cent support among those who voted.

Whereas it would seem self-evident that sound random sampling which covers the entire population is a better approach, it is clear that, in this case, the coverage of (for example) cell phone only households increased the distance between our final poll and the election results. It is also notable how well the cell phone only sample provides an example of how including eligible

but non-voting members of the public can increase the distance between a poll of eligible voters and the final electoral outcome. The relative Conservatives weakness in the cell only population is quite similar to the hypothetical estimate of non voter support for the Conservative in Figure 4.1. It is also notable how the cell only sample is higher on Green Party support, although our gaps there suggest that both intentions and Green voting may well also be inflated by social desirability bias.

5.0 PARSING ERRORS AND CONSIDERING A BETTER YARDSTICK FOR ASSESSING POLLING QUALITY

Thinking about the evidence and the lessons to be drawn from our recent polling of the last election, some issues are now much clearer to us. At the top of the list is disentangling the two separate challenges of: (i) forecasting the election result and (ii) modelling the population of all eligible voters. Having made an error on (i), we ask the reader not to confound it with the second challenge. In fact, the evidence fairly clearly shows that in cases where the (i) and (ii) populations are significantly different, a less rigorous measurement of the entire population of voters may appear to have been more accurate polling. It may have been, by happenstance, closer to the final result but few could argue that by systematically excluding large and systematically different parts of the voting population (e.g., cell only households and non-internet households), we are doing a better job of modelling all voters.

There are other issues to be considered here as well. The focus on the gap between the final polls and the election outcome reduces the role of the pollster to the monitor of the horserace. In some respects, the final polls are the least important for anything other than assigning the vanity points to pollsters who had the smallest gaps to final results. As we have argued, in cases where there is a profound systematic difference between the actual and eligible voters, this may be a very poor assessment of polling accuracy in any traditional scientific sense. If we find that we come closer by ignoring large subpopulations then is the role of the pollster only to forecast the final result? In addition, if there is a widening gulf between the voting and non-voting populations, where do we hear the voice of the non-voter? Can we just say that this large group – nearly 40 per cent of all voters and over half of under 50 Canada can be safely ignored? What if the factors producing non-voting are not simply "laziness and apathy"? What if the lack of voting is linked to alienation and conscious political strategies designed to suppress the interest of those voters? What if there is a mutually reinforcing tendency to further weaken "next Canada's" interest in federal government by virtue of a federal agenda which systematically undervalues and deemphasises their interests and values and emphasises the interests and values of its constituency? Does this become a particularly troubling problem at a point where our highly unusual demographics have produced a voter whose median age is around 60? Are we fashioning the future Canada in the image of those who are disengaging or those poised to inherit the positions of authority and influence in short order?

Another issue is whether good polling should be charting the evolution of the campaign and noting the forces at play in producing the twists and turns throughout the campaign. Elections are rare periods of collective public judgement which should serve to elucidate important things about the societies in which they unfold. These are important roles for the pollster which are not merely connected to the closeness of final results.

It is important as well to understand what will happen on Election Day and this analysis is not intended to diminish our shortcomings on that front. We do, however emerge with a sounder knowledge of how to forecast. Our (regrettably) unreported commitment index did an excellent job of identifying the final turnout. It also showed that claimed certainty of voting is a next to

useless indicator whereas other measures of firmness and emotional engagement ('enthusiasm" appears to be much better predictors).

Below are the results of our final election polls applied to the top 60 per cent most committed voters. We also cite the items used to construct the index created which predicted voter turnout:

- how easily respondents select their voting preference
- the self-rated level of enthusiasm
- the expressed intensity of attachment
- the self-expressed likelihood of actually voting
- current choice versus declared past vote
- the respondent's willingness to consider other parties

Ruefully, these results, which were on my desk on Sunday and were nearly spot-on to the final results, were never released. Despite having tracked this throughout the campaign, I lost nerve due to inconsistency between the overall index and "absolutely certain to vote" (which turned out to be quite unreliable).

Figure 5.1 – EKOS commitment index versus actual results

Q. If a federal election were held tomorrow, which party would you vote for?
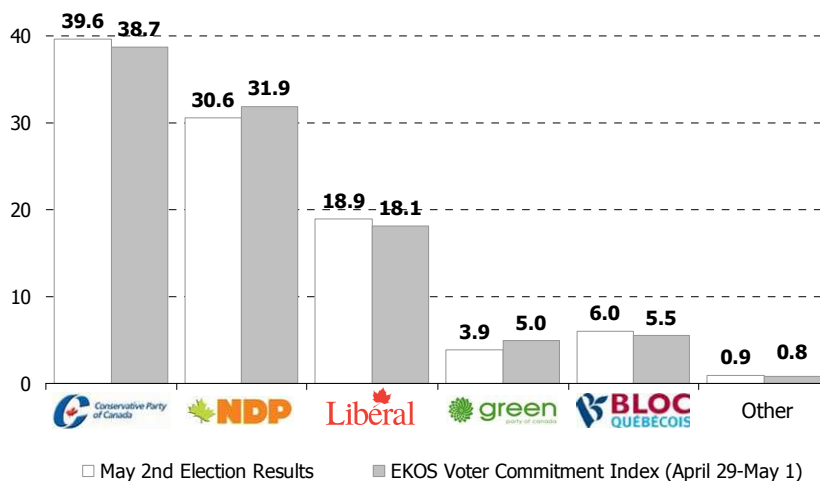
Note: The data on federal vote intention are based on decided and leaning voters only.

BASE: Decided voters; April 29-May 1, 2011 (n=2,876)

In closing, these are very complex and difficult questions. The Conservative Party of Canada won a hard fought and legitimate majority mandate working within all of the rules of the current electoral system. There are however, profound issues looming about how to register and take into account the huge portions of Canadian society who are soon to be the mainstream and who have systematically opted out of the electoral process. This may be a far greater challenge to

polling and democracy than the somewhat suspect polling sweepstakes of who came closest to the final vote outcome.