

THE GREAT CANADIAN POLL-OFF

INITIATING AN OPEN-SOURCE APPROACH TO POLLING

By Frank Graves with Jeff Smith

In the same spirit that we offer our public opinion polling, data tables, and methodology as a forum for discussion, this document is offered as a draft and we welcome comments from those interested.

[Ottawa – March 29, 2011] – Recent articles by Joan Bryden, and the responses they have generated, have launched a useful, if mildly overwrought debate about the merits of contemporary polling, with a particular focus on the perils of political polling. There are many useful caveats and lots of points of consensus. Let's focus on the main points that have been raised:

1. Polling faces significant new challenges due to declining response rates and coverage issues where certain parts of the population are included in the sampling frames (listings of the entire population) the most significant new expression of this is the growing incidence of households which have abandoned land lines in favour of cell phones.
2. The symbiotic relationships between the media, the public and the pollsters has been compromised by a range of factors such as poor resourcing, declining methodological literacy in journalism, and a tendency to focus on the superficial with exaggerated and unduly pyrotechnical interpretation.
3. There is a broad tendency to ignore the guidelines of margin of error and to report spurious or substantively insignificant events.
4. Concerns that polling may distort the democratic process, either through leading voting trends or by compromising the inside political process.

Before turning to these four points, a couple of general observations are in order. Our main point is that polling can be done well. It is true that polling is constantly facing new and emerging challenges, but all of these problems are more or less solvable. It is not the case that all polls are of equal quality in meeting these challenges and the media often fail to adequately scrutinize the polls they publish. Expert, external, third-party sources are important, as is benchmarking. It is not enough for a firm to say their methods are good and point to their examples of success while ignoring failures and the expert consensus. In short, good quality polling is possible and exists. The problems are that the good is interlaced with the not so good and that the media are confused by differences which are methodological artefacts

In short, the overall depiction of media polling as error prone, unreliable, and corrosive to democracy is overstated and wrong. It would be better if it were more adequately resourced and if some media were more knowledgeable of the theories of probability and statistics. Much of the "confusion" is a product of comparing across polls of variable quality conducted over different time periods using alternative methods. For these reasons, we collect data using exactly the

same sampling and survey methods on a regular basis. Undoubtedly, more in-depth interviewing can provide more insightful and interesting analysis of the drivers of public opinion. The problem here is that there are rarely the resources or public attention spans to justify more exhaustive surveys. Some may feign disdain for the vulgar simplicity and irrelevance of the “horse race”, but without ambiguity, nothing is of more interest to politicians, media, pollsters, and the general public than the horse race (supported with a few judiciously chosen explanatory and background variables).

DISCUSSION:

The recent debate following the Joan Bryden articles launched an important contribution to the topic of how polls should be consumed and indeed whether in fact we would be better off with far fewer of them. Despite agreement with many of the things noted in this debate, I have some non-trivial points of clarification I would like to submit.

The impression that current Canadian polling is of limited utility and value is overstated. All of the experts consulted are seasoned veterans and they may have been a little florid in their depictions of current problems (I believe I used the term "dog's breakfast"). The sense that the whole field is a fetid miasma of error is simply untrue and unfair and fails to disentangle good from poor polling. In fact, as Eric Grenier at 308 has pointed out in his useful response to this article, the track record of polling isn't as egregiously flawed as one might surmise from the article¹. Independent experts in the United States have also repeatedly confirmed the continued positive track record of modern polling.

Finally, there is an underlying sense that the problems and challenges have become too daunting to justify the effort and that good media polling had become an oxymoron or at least an anachronism. Having contributed to this impression, let me try and provide an antidote by noting that there is a lot of value in current media polling despite the room for improvement. To suggest otherwise would fly in the face of the enormous public interest that the polls generate and would suggest that the audiences and media who cover this are basically foolish. Some of this rather disdainful view of the public is also contained in the final point noted in the first paragraph; that polling has become corrosive to the democratic process. I will choose to largely disagree with that point. Turning to the four points:

- 1. Polling faces significant new challenges due to declining response rates and coverage issues where certain parts of the population are included in the sampling frames (listings of the entire population) the most significant new expression of this is the growing incidence of households which have abandoned land lines in favour of cell phones.***

Have declining response rates and coverage issues rendered polling useless (or at the least tainted compared to the good old days)? This is one of the most misunderstood areas of contemporary polling. Yes the fact that portions of the public have become systematically unavailable because they have abandoned land lines in favour of cell phones and that the pollster is missing approximately one in three people (a socio-demographically predictable group) by polling on the internet, does indeed present formidable challenges. While declining response rates (and they have declined precipitously as the article notes) are serious issues, these issues have been overstated. Generally speaking, political polls do a better job of predicting political outcomes today than they did thirty years ago when response rates were much higher. As the Pew organization in the United States and other academic sources have shown, there is little

¹ See Grenier (2011)

difference in the results of well executed current practices (with proper call backs, replacement, and weight adjustment) and very expensive experimental alternatives which achieve the response rates of yore.

There is a counterintuitive quality to this finding but much of survey methodology is based on counterintuitive foundations. People are often shocked to find that the margin of error of a random sample of 1,000 is exactly the same for modelling a city of 100,000 as it would be for modelling a country the size of China. The key to representative sampling seems to be rooted more in the process of random selection than in the response rate (with the right practices and adjustments supplied). That is one of the reasons that the move to a voluntary census is such a poor idea – it introduces self selection factors which are largely immeasurable. Incidentally, the ability to produce sound parameter estimates even when there are more non-responders than responders is something which might be seriously compromised by the move to a voluntary long form (ethnicity and race measures come from that source for example).

On another note, it isn't a declining response rate among the more affluent which is growing (as noted in the article). In fact, the evidence clearly shows that the decline is most severe amongst the poorly educated and economically vulnerable.

Without belabouring the technical issues, it is fair to say the following: declining response rates, particularly among younger and economically vulnerable households, is a significant but manageable problem. The burgeoning non-coverage due to cell only households and doing internet surveys when roughly one in three never use the internet (although it might be in the household) are also serious issues. Both issues are, however, largely fixable, particularly in Canada where a much smaller portion of the population uses cell phones exclusively and where we can legally call them (in the United States, the cell only population is three times larger and there are serious restrictions on calling them).

The problem of the non-internet household is also fairly easily solved. This older, more vulnerable rural population continues can be randomly sampled using other methods, such as phone and mail (they are all in our *Prob/it* panel). Dual methods are possible blending internet and non-internet samples or "enabling" non-internet households with a terminal or other device (as Knowledge Networks does in the United States). Address-based sampling helps solve the limits of non-coverage due to no land line and our Interactive Voice Response (IVR) methods perform relatively better with some of the challenging younger populations (see Annex II).

The most popular solution has been the creation of opt-in panels of those who volunteer to join an internet panel in exchange for pretty trivial economic rewards. Unsurprisingly, these panellists reveal serious systematic biases when compared to the general public. Moreover the vast majority of respondents tend to be a very small portion of the original non-random sample (many of whom tend to be belong to many other "panels"). Randomly sampling from within these non-random panels doesn't constitute a random sample and weighting will help but not fix the problem. This is the clear conclusion of virtually all of the independent experts and third parties, most notably the American Association for Public Opinion Research (AAPOR) which, in a summary

of a major international review last year, noted that these samples do not constitute scientifically representative samples and should not be quoted with margin of error methods. More recent research reported by Gary Langer (2011) using census comparisons suggests that the problems may be even deeper than AAPOR thought.

Opt-in polls *may* do well in predicting certain political events, but they often don't do well and it isn't clear why they do well when they do. Some of these problems may be solved through time but there is little convincing scientific evidence to that effect to date. It makes little sense to ask individual companies to critique their own approaches. There is no serious academic or professional body which suggests that these approaches can be considered scientifically representative samples. They have useful applications but the tendency to say that they are as good as random methodologies is a claim limited to those selling them. There are non-response problems with Random Digit Dialling (RDD) methods, but the response rates are at least calculable and the literature is clear on their advantages over opt-in online methodologies.

In conclusion, good surveys sample using random selection and should cover the entire population. There have been growing challenges to these foundation principles but they remain largely solvable. The notion of random selection from the broader population can be emulated across a variety of media but if the sample doesn't use those approaches, the power of the law of large numbers and central limit theorem doesn't apply. Don't allow legitimate concerns with these problems to accept the notion that random and comprehensive sampling is an impractical and obsolete goal.

2. The symbiotic relationships between the media, the public and the pollsters has been compromised by a range of factors such as poor resourcing, declining methodological literacy in journalism, and a tendency to focus on the superficial with exaggerated and unduly pyrotechnical interpretation.

I have little to quarrel with this second key point. I am not so sure about the rampant proliferation of polling firms on the media front. We have the occasional new entrants (like Abacus) but the field is dominated with a remarkably stable list of polling firms and pollsters who made their entrances thirty years ago. Environics, Decima, Angus Reid, and EKOS have all been around in various incarnations for some time. Nanos has been on the national polling stage for many years. The list of "go to" pollsters for media commentary is actually very short. The idea that there are a plethora of shoddy polls out there is also not really my view. I have some issues with the non-probability online polls, but I see very few "shoddy" polls. What I do see are shortcomings in things like sample sizes, questionnaire length, and the frequency and timing of media polls. As David Colleto of Abacus has argued, we need a greater level of transparency and disclosure for all polls in public domain². Question wordings and locations can be crucial, as can the sampling methods.

² See Colleto (2011)

John Wright made the reasonable observation that in the aftermath of the Rob Ford campaign, the media did not provide adequate resources to track the life cycle of the campaign and I believe that Nik Nanos made a similar observation on this same campaign. It may be more newsworthy to poll during the middle stages of a campaign than to measure the late stage outcomes. Often failures in polling come not because the polls were inaccurate, but because things changed in the late stages.

The Toronto mayoralty race is interesting for some additional reasons. EKOS conducted a late poll which showed Ford in a commanding lead. It was heavily questioned by the media who had basically stopped polling and wanted to stick with their narrative that the race was "too tight to call". Interesting drama, but not true. We actually did a follow-up poll using our IVR methods and found that the race was basically over³. In addition to the fact that we polled later and caught the final surge, we also polled the entire population, including the older and less educated voters who were key to Ford's victory but part of that group that don't do internet polls. It may also have been the case that the recorded voice methods minimised a social desirability bias where "closet" Ford supporters would acknowledge their choice to a computer and ballot booth, but were less likely to do so in the case of a live interviewer.

A crucial point here is that if voters want to vote "strategically" on the basis of where the candidates stand, they are entitled to accurate feedback from the media and pollsters. In this case, they really didn't get it and the fault was more that of the media who liked the tight horse race story and didn't want to pony up for late polling. Frustratingly, the Globe and Mail ran a post-mortem on the failure of polling, conveniently ignoring our spot on final predictions even though they had run our poll in late stages (with some scepticism). The Star also ran our poll with high scepticism and made no post-campaign acknowledgment. Full credit goes to Jane Taber of the Globe and Mail and the Toronto Sun for running clear post-mortems including our successful prediction.

I do agree with my fellow pollsters that there has been a decline in the level of statistical literacy among the media. In the past, large media had their own internal political polling experts (like Elly Alboim and Paul Adams at the CBC). This expertise now lies in the research departments, which are often preoccupied with the tasks of understanding audiences and rating wars.

The issue of whether media pollsters exaggerate or invent non-shifts which are artefacts of random variation is another point that several of my colleagues made. This certainly does happen to some extent, but it is more a product of media and audiences discussing these than the pollster. In our case, we are very careful to only discuss shifts which we believe are statistically and substantively significant. If we stray into reasoned conjecture or statistical evidence which is marginal, we will clearly say so. We almost never comment on week to week statistical fluctuations in regions (as some have suggested) unless there is a clear pattern which fits the longer time series, statistical evidence that this is highly unlikely to be a product of random

³ See Annex II for an overview of the IVR methods.

variation and better still a plausible external cause (best measured directly in the survey), or wherever there is a very obvious external event which concomitantly covaried.

This provides a good segue to discuss the much misunderstood notion of margin of error and statistical significance.

3. There is a broad tendency to ignore the guidelines of margin of error and to report spurious or substantively insignificant events.

Anyone who has read this far in our turgid response to the Joan Bryden article is familiar with the notion of margin of error (MOE). I am going to try to embellish that understanding with a few additional points that need to be understood about the whole issue of "significance" and its lesser known methodological sibling "power". I believe there is some misunderstanding about what constitutes a real finding and what is a spurious or irrelevant finding. I believe that there is a problem with over interpreting effects which are simply statistical artefacts. I also think that the discussion in the article leads to an overly conservative and frankly incorrect understanding of MOE and significance.

A few points are important to understand. The MOE is not a constant. Traditionally, MOE estimates represent a conservative or "worst case" estimate (for example, let's use the familiar plus or minus 3.1%, 19 times out of 20 for a random sample of 1,000). This estimate, however, is based on a 50/50 split (e.g., if the Conservatives and the Liberals were to each come in at 50%). It's very rare to be looking at 50% and more often, the numbers are much smaller – such as the NDP's support of around 14% in our last poll. As the numbers move away from 50%, however, the margin of error decreases (in the case of the NDP, the MOE is only around 2.2%).

There is also an error in the test of significance for a lead. It is not simply adding the worst case scenario MOE to each of the percentages for the two leaders and then seeing that it could be a very large number. First of all, it isn't the worst case and should be adjusted to the actual values. Second, the MOE for a lead isn't simply the sum of the MOEs. While it would be correct to do so when there are only two possible survey responses, it is not correct when there are more than two choices (in the case of our federal vote intention poll, we offer six options – the five parties and "other"). How much difference this will make depends on the number of responses that fall outside the two categories of interest (the article uses the example of the Conservatives and Liberals), but as a rule of thumb, it is safe to multiply the MOE of the worst case situation of a 50/50 split by 1.6.

For example, the article says that if the Conservatives lead the Liberals 35% to 30%, the MOE (which is 2.0% in this case) means that Liberal support could be as low as 28% (or as high as 32%) and the Conservatives could be as high as 37% (or as low as 33%). While this part is true, it is incorrect to add the MOEs together and suggest that the lead ranges from 1 to 9 points. Using our rule of thumb, one would conclude that a lead of 3.2% is statistically significant. In fact, the article's example of a 5-point lead would be statistically significant at a confidence level of 99%.

More recently, there has been a fundamental rethinking of the continued relevance of margin of error⁴. Thoughtful experts are now calling on a shift to thinking of likelihood, rather than rejoicing at 0.049 and ignoring 0.051.

4. Concerns that polling may distort the democratic process, either through leading voting trends or by compromising the inside political process.

I have a profound disagreement with this last point. To argue that media polling is harmful to democracy is frankly nonsense. Without it, we leave it to the occluded machinations of the back room political pollsters and leave the media and public to speculate in a state of ignorance or worse, to rely on the uniformed speculation of the punditocracy. Media polling is an essential mirror of public judgement on their political choices which elevates debate and disciplines politicians to at least keep some connection to the preferences of the entire society; it is largely reflective of opinion – not causal. It has many weaknesses, but it is still a very valuable public service which should be strengthened – not abandoned.

Even if it is true that people are voting strategically by taking factoring poll results into their decisions (in a previous study, about 30% of respondents tell us they do), so what? Voters use many different factors in determining their final selection and vote rationally based on their understanding of how their vote could be best strategically “spent”. Strategic voting is a perfectly reasonable exercise of one’s democratic franchise. It is therefore imperative that both the media and the pollsters ensure that vote intention feedback is of the highest possible standard.

⁴ See Siegfried (2011)

Acknowledgements:

While any errors in this piece are the author's responsibility, I had very useful input and advice from Paul Adams of Carleton University and Nicki Doyle from Halifax.

BIBLIOGRAPHY:

American Association for Public Opinion Research; "[AAPOR Report on Online Panels](#)"; March 2010

American Association for Public Opinion Research; "[An Evaluation of the Methodology of the 2008 Pre-Election Primary Polls](#)"; April 13, 2009

Blumenthal, Mark; "[Update: How Accurate Were the Benchmarks?](#)"; Pollster.com; December 17, 2009

Colleto, David; "[The Quality of Public Opinion Research in Canada](#)"; Abacus Data; February 13, 2011; Accessed on March 28, 2011 at <http://tinyurl.com/4bh94a5>

Chang, Linchiat; Krosnick, Jon; "[National Surveys via RDD Telephone Interviewing versus the Internet](#)"; Public Opinion Quarterly, Vol. 73, No. 4, Winter 2009, pp. 641–678; 2009

Crassweller, Anne et al; "[Between Random Samples and On Line Panels, Where is the Next Lily Pad?](#)"; ESOMAR On Line Research Conf; 2008

Crassweller, Anne et al; "[In Search of Readers, A Brave New World](#)"; World Readership Research Symposium Paper; March 2009

Grenier, Eric; "[Should We Be Wary of Poll?](#)"; ThreeHundredEight.com; February 13, 2011; Accessed on March 28, 2011 at <http://tinyurl.com/4fd8m4q>

Kish, Leslie; "[Survey Sampling](#)"; John Wiley & Sons, Inc., New York, 1965

Langer, Gary; "[Study Raises New Questions for Opt-in Online Data](#)"; ABC News; March 21, 2011; Accessed on March 28, 2011 at <http://tinyurl.com/4mydaxq>

Langer, Gary; "[Survey Reporting Standards](#)"; ABC News; Survey Quality Conference; Harvard University Program on Survey Research; April 17, 2009

Pasek, Josh; Krosnick, Jon; "[Measuring Intent to Participate and Participation in the 2010 Census and Their Correlates and Trends: Comparisons of RDD Telephone and Non-probability Sample Internet Survey Data](#)"; Statistical Research Division, U.S. Census Bureau; December 28, 2010

Pew Research Center Publications; "[Perils of Polling in Election '08](#)"; June 25, 2009; Accessed on March 28, 2011 at <http://tiny.cc/w8q8n>

Rivers, Douglas; "[Second Thoughts About Internet Surveys](#)"; Pollster.com; September 6, 2009

Scheuren, Fritz; "[What is a Survey?](#)"; American Statistical Association; Accessed on March 28, 2011 at <http://tinyurl.com/4dmzkay>

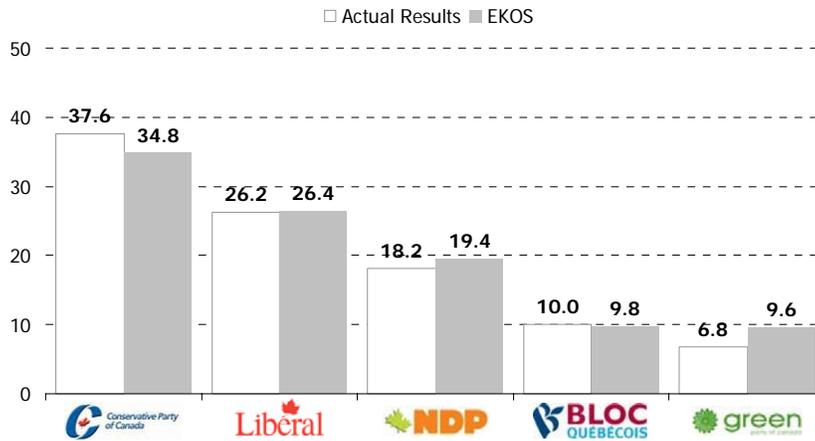
Siegfried, Tom; "[Odds Are, It's Wrong](http://tinyurl.com/yz22ldq)"; Science News; Accessed on February 16, 2011 at <http://tinyurl.com/yz22ldq>

Taylor, Humphrey; "[Social Desirability Bias - How Accurate were the Benchmarks?](#)"; Pollster.com; October 27, 2009

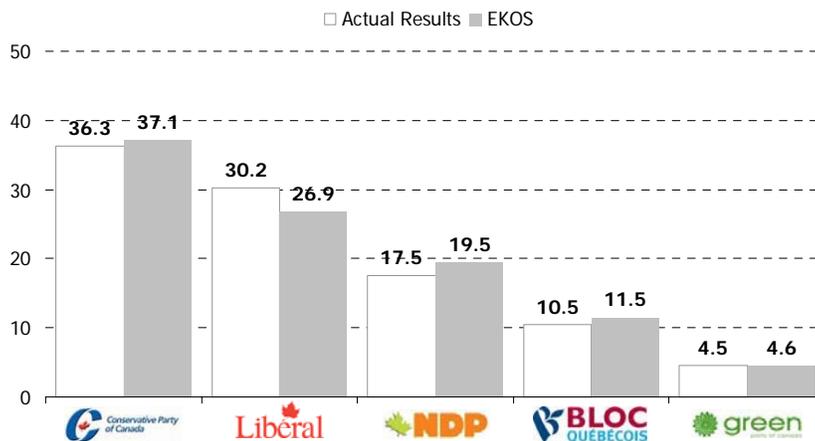
Yeager, David; Krosnick, Jon; "[Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples](#)"; August 2009

ANNEX I – EKOS RESULTS VERSUS ELECTION RESULTS:

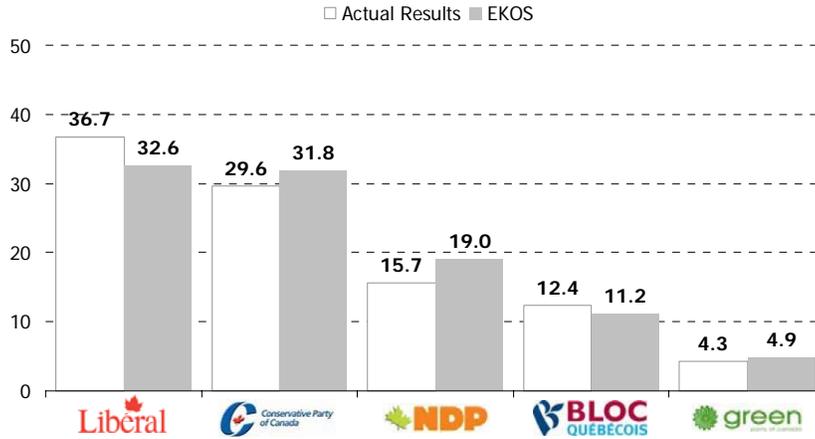
2008 Election



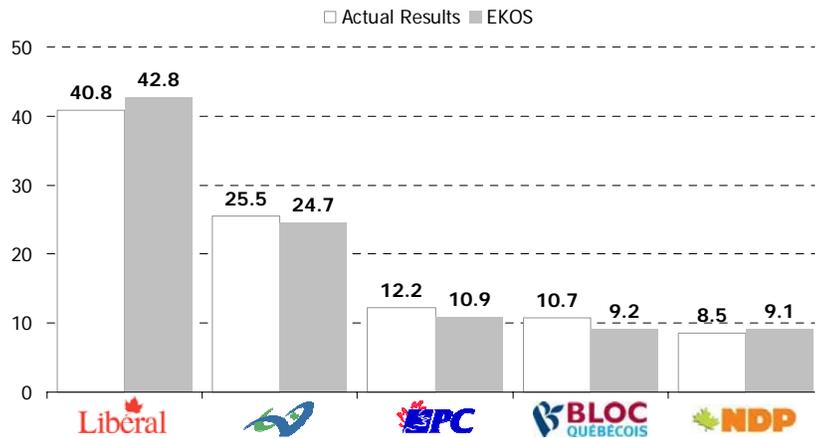
2006 Election



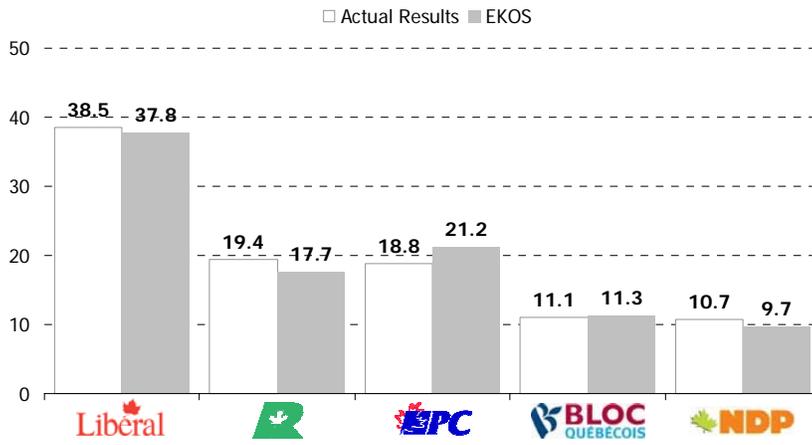
2004 Election



2000 Election



1997 Election



ANNEX II – PRESENTATION TO THE MARKETING RESEARCH AND INTELLIGENCE ASSOCIATION:



**INTERACTIVE VOICE RESPONSE:
THE PAST, THE PRESENT, AND
INTO THE FUTURE**

**Presentation to: The MRIA, Ottawa Chapter
January 21, 2011**

Dialing In

Current State of Survey Research

- » Massive changes in polling and survey research
- » Biggest is drift from live CATI to online methods
 - Fast, inexpensive
 - Self-administered
 - Multi-media capabilities
- » BUT non-probability online methods lack representativeness
 - Non-coverage issues
 - Not randomly selected
- » Another option: Interactive Voice Response (IVR) for both data collection and probability panel construction

Interactive Voice Response (IVR)

Advantages and Disadvantages of IVR

- » Advantages
 - + Perhaps closest to national population
 - + Avoids social desirability
 - + Cost-effective
 - + Higher reliability due to large sample sizes

- » Disadvantages
 - Higher non-response
 - Survey must be shorter
 - Some design limits
 - Reputation
 - Intrusiveness

Accuracy of Results

- » Rob Ford election as recent illustration of IVR success
 - Coverage issue (older, vulnerable)
 - Social desirability issues (closet supporters)

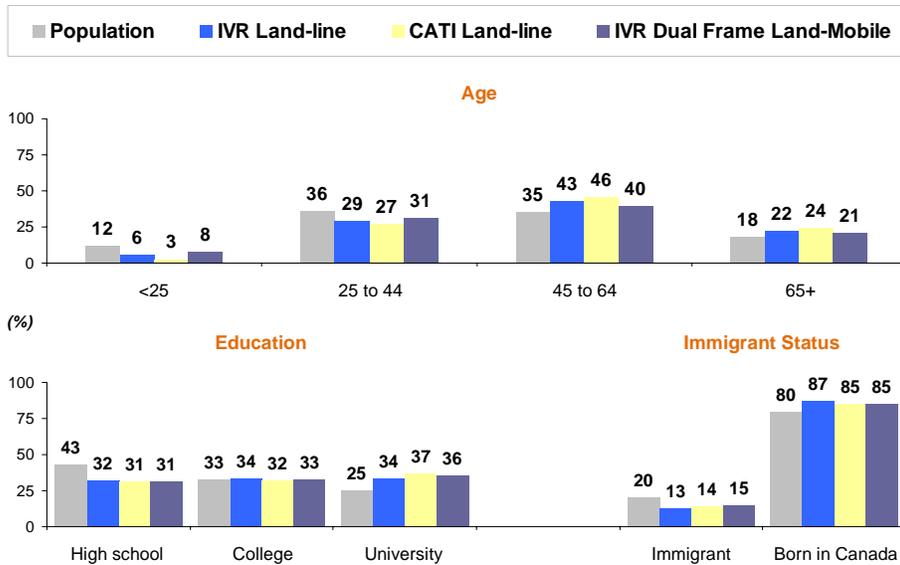
- » 2008 federal election (EKOS results most accurate of RDD polls – see <http://www.sfu.ca/~aheard/elections/results.html>)

- » U.S. mid-terms (mixed results – clear lessons)

- » IVR is good for short polls and for population seeding a panel, BUT live follow-up is crucial:
 - Verification
 - Explanation
 - Create dossier of key demographics



Demographic Analysis – IVR vs. CATI



Response Rate – IVR vs. CATI

	IVR Landline Sample	Live Interviewer CATI Landline Sample
Final Disposition		
Unused	0	0
A Invalid numbers	14,395	4,829
B Unresolved	5,707	8,083
C Non-responding (unknown eligibility)	0	0
D Ineligible (language barrier)	2,200 (estimated)	1,129
E Non-responding (eligible)	37,922	12,009
F Completed interviews	1,976	3,009
TOTALS	60,000	29,059
Response Rate		
Empirical Method $(f+d)/(b+d+e+f+c)$	8.7%	17.08%



Conclusions

Considerations

- » IVR in and of itself is not a polling method
 - Still need rigorous sampling, callbacks, etc.
 - But provides excellent equilibrium of cost and quality in appropriate circumstances
- » Hugely neglected area (largely reputational), but will improve
- » IVR is vastly superior to non-probability online polls or even live CATI omnibus (purpose built IVR versus uncertain context of omnibus)
- » Response rate issue overstated; non-response about half live CATI

Capabilities

- » In the United States, when pollsters had the accuracy of their results tested during the 2008 primary and election year, the two companies employing IVR were rated first and eighth among more than forty companies by the leading website monitoring polling in the U.S.
- » AAPOR has indicated:
"...the use of either computerized telephone interviewing (CATI) techniques or interactive voice response (IVR) techniques made no difference to the accuracy of estimates..." (in U.S. pre-primary polls)
- » The Pew Research Center has reported:
"...the mean error among IVR polls [in the 2008 U.S. election] was slightly lower than among those with live interviewers."

Emerging Issues

- » Education and civic literacy/interest bias
- » Better sampling and weighting
- » No long form census?
- » The cell phone only household
- » Broader credibility of scientific samples/evidence based research



For more information:

Frank Graves
EKOS Research Associates
fgraves@ekos.com
(613) 235-7215
www.ekos.com