The 2015 Canadian Election Study

Technical Documentation

The 2015 Canadian Election Study

Technical Documentation

David Northrup

Institute for Social Research
York University
August, 2016

Conditions of Release

All research based upon these data must include an acknowledgement such as the following:

> Data from the 2015 Canadian Election Study were provided by the Institute for Social Research, York University. The survey was funded by the Social Sciences and Humanities Research Council (SSHRC) and Elections Canada, and was completed for the Canadian Election Study Team of Patrick Fournier (Université de Montréal), Fred Cutler (University of British Columbia), Stuart Soroka (University of Michigan), and Dietlind Stolle (McGill University). Neither the Institute for Social Research, SSHRC, Elections Canada, nor the Canadian Election Study Team are responsible for the analyses and interpretations presented here.

Researchers are requested to forward a copy of any publications or scholarly papers to the Director Survey Research, Institute for Social Research, Victor Phillip Dahdaleh Building, 88 The Pond Road, York University, 4700 Keele Street, Toronto, Ontario, M3J 1P3 and to Dr. Patrick Fournier, Département de science politique, Université de Montréal, C.P. 6128, Succ. Centreville, Montréal QC, H3C 3J7.

Data acquired from the Institute for Social Research may not be re-disseminated outside the recipient institution.

Table of Contents

List of Tables

1 Study Description

The 42nd Canadian General Election was held on October 19, 2015. The Institute for Social Research completed the Canadian Election Study (CES) on behalf of the Canadian Election Study Team (Patrick Fournier, Université de Montréal; Fred Cutler, University of British Columbia; Stuart Soroka, University of Michigan; and Dietlind Stolle, McGill University). This Technical Report briefly outlines the design and conduct of the survey.

There were three phases of data collection. During the election campaign, telephone interviews were completed with 4,202 Canadian citizens. All respondents to the Campaign-Period Survey (CPS) were called after the election and asked to complete the Post-Election Survey (PES) and 2,988 (71%) did so. At the end of the PES, respondents were asked to provide their (postal) mailing address and 2,088 respondents (70%) did so. Of these PES respondents who provided an address, 1,289 (62%) returned a completed Mail-Back Survey (MBS) to ISR.

A modified random digit dialling (RDD) procedure was used to select telephone numbers for the CPS, and in households with more than one adult Canadian citizen the respondent was determined by using the birthday selection method. A rolling cross-sectional sample release was employed for the CPS. Interviewing commenced on Tuesday, September 8, 2015 (after the Labour Day long weekend) and concluded on the eve of the election on October 18, 2015. Interviews were completed every day except on Thanksgiving Day (October 12). A new sample of telephone numbers was released on each of the 40 days in the calling period during the election campaign. Each day's sample release was called multiple times over the 10-day 'clearance period' unless a completion was obtained. As a result, sample released for the first 31 days of calling was called on 10 consecutive days (unless a final outcome, such as a completion, not eligible result, etc., was obtained). Sample released on day 32 was called on nine days, sample released on day 33 for eight days, etc. Sample released the last day of the campaign was only called on that day.

Calling for the PES started on Tuesday, October 20, the day after the election, and all of the CPS respondents were called back within five days of the vote. Of course, not all respondents were available when first contacted and after nine days of calling 50 percent of the PES interviews were completed. (In 2011 it took 14 days to complete this proportion of the PES interviews.) Almost 90% of the calling was completed in the 31 days after the election but the final interviews were completed on December 23, 2015.

All telephone data collection was completed with Computer Assisted Telephone Interviewing (CATI). The Institute's CATI software is from the Computer-Assisted Survey Methods (CSM) program at the University of California, Berkeley.

The naming conventions for the variables in the data file indicate the survey source (CPS, PES, and MBS) as well as the last two digits of the survey year (15). For example, variables in the Campaign-Period Survey include the prefix CPS15, thus CPS15_INTDATE indicates the date of interview for the 2015 campaign period completions, PES15_INTDATE the date of interview for post-election interviews, etc.

2 SAMPLE DESIGN

2.1 Introduction

The sample component for the 2015 CPS was designed to represent the adult population of Canada defined as: Canadian citizens 18 years of age or older who speak one of Canada's official languages, English or French, and reside in private homes[1] in the ten Canadian provinces (thus excluding the territories). Because the initial survey (the CPS) was conducted by telephone, the small proportion of households in Canada without telephones were excluded from the sample population as were the households that did not have land lines.[2]

The CPS sample was not a simple random sample. The sample is both *clustered* and *stratified* and thus falls into the class of what are now commonly called "complex samples" – *clustered* because the probability of an adult member of the household being selected for an interview varies inversely with the number of people living in that household; *stratified* because the likelihood of being interviewed varied by province of residence (residents of the smaller provinces have a greater chance of being interviewed). In order to provide unbiased estimates it is necessary to correct for these unequal probabilities of selection.

2.2 Selection of Households

To select individual survey respondents a two-stage probability selection process was utilized. The first stage involved the selection of households by randomly selecting telephone numbers. The ideal sampling frame for the Campaign-Period Survey would have been a complete listing of all residential telephone numbers (both land lines and cell phones) in Canada. Unfortunately, such a listing does not exist. To select numbers ISR employs a modified form of random digit dialling. All telephone numbers in Canada consist of an area code, a "central office code" or exchange (the first three digits of the telephone number), and a suffix or "bank" (the last four digits of the number). A list of most telephone numbers in Canada can be constructed from CD-ROM versions of telephone books and other commercially available lists of telephone numbers. Numbers from these sources, which include unlisted numbers and some cell phone numbers, are included in the sampling frame. A computer is then used to generate a random sample of telephone numbers from this list.

As well as household telephone numbers, this sample includes "not-in-service" and

[1] Interviews were not completed with respondents, who could not speak English or French well enough to complete the survey, were judged as having cognitive difficulties by the interviewer, and residents of old age homes, group homes, educational and penal institutions were excluded from the sample.

[2] Statistics Canada (www.statcan.gc.ca/daily-quotidien140623/dq140623a-eng.htm) estimates that just over one in five households (21%) in Canada are cell phone only and in an earlier release of the same survey they estimated 1.2 percent of the households in Canada do not have telephone service. Some cell phones are included in ISR RDD samples, but they comprise about 4-5% of the numbers. As a result, cell phone only households are under-represented in the data file.

"non-residential" telephone numbers. Typically, non-household numbers are identified the first time the interviewer calls. Most of the interviewer's subsequent efforts are then directed at encouraging an informant from the household to provide information about the number of adults living in the home, and after randomly selecting a respondent, completing the interview.

2.3 Selection of Respondents

The second stage of the sample selection process was the random selection of a respondent from the selected household. To be eligible for the interview the household member had to be an adult (18 years of age or older) and a Canadian citizen. If there was more than one eligible person in the household, the person with the next birthday was selected as the survey respondent.[3] The birthday selection method is used as it ensures a random selection of respondents and is a much less intrusive way to start an interview than asking about the number of people in the household, thus making it easier for the interviewer to secure the respondent's cooperation.

2.4 Household Weights

Weights, to correct for unequal probabilities of selection, facilitate the production of national estimates. Weights are provided for the CPS but not the other survey components as non-response to each successive survey increases the potential for an unrepresentative sample. The probability of an adult member of the household being selected for an interview varies inversely with the number of eligible adults living in that household. In a household with only one adult, this person has a 100 percent chance of selection; in a two-adult household each adult has a 50 percent chance of selection, and so on. Analyses based on unweighted estimates are therefore biased: members of one adult household are over-represented, and larger households with two or more adults are under-represented.[4]

The household weights were computed as follows. As shown in Table 2.1, for Newfoundland and Labrador, 39 interviews were completed in households in which there was one adult, 84 interviews were completed in households with two adults, 23 in households with three adults, and so on. Thirty-nine adults are represented by the 39 interviews in one-adult households, 168 adults are represented in the 'two-adult cluster,' and 69 in the 'three-adult cluster,' etc. In total, the 155 interviews in the province represent 314 adults.

[3] See O'Rourke and Blair, 1983, for a review of the birthday selection method.

[4] Weighting to correct for unequal probabilities of selection, stratification, and other factors in order to improve sample estimates is common in survey research. For a brief discussion of complex samples and weighting see Ornstein, A Companion to Survey Research, 2013, pp 77-80, for a more complete review Groves et al., 2009, chapter 10. Kish, 1965, specifically addresses the issue of weighting to correct for unequal probability of selection at the household level (p. 400) and suggests, unlike most survey researchers, that household weighting may not be necessary.

These 314 adults represent the 406,455 adults who live in Newfoundland and Labrador.[5] That is, we have interviewed 1 adult for every 1,294 people in the province (406,455/314 = 1,294, plus rounding). This figure (1,294) is the weight to use when data is weighted to the population size for the one-adult cluster in Newfoundland and Labrador. For respondents who live in two-adult households, the population weight is 2,589 (1,294 * 2), for three-adult households it is 3,883 (1,294 * 3), etc. The same procedure is used for each cluster (household size)/stratum (province) combination. These household weights are incorporated into the 'pop' and 'samp' weights provided in the data file.

Table 2.1 Calculation of Household Weights for Newfoundland and Labrador

# of adults in HH	**# of HHs each size**	**estimate of # adults**	**weight to population** (406,465)	**weight to sample** (155)
1 adult	39	39	1,294.44	0.49
2 adults	84	168	2,588.89	0.99
3 adults	23	69	3,833.33	1.48
4 adults	8	32	5,177.77	1.97
6 adults*	1	6	7,766.66	2.96
Totals	**155**	**314**		

* There were no 5 adult households in the province.

Survey data can be weighted to the population or to the original sample size. The weights required for weighting to sample size are also provided in Table 2.1. Because the distribution of the respondents for each household size is based on the survey results, it is an estimate. Continuing with the Newfoundland and Labrador example, the weight for one-adult households (1,294) is multiplied by the total number of interviews in Newfoundland and Labrador (155) divided by the total adult population (406,455) which is 0.49 [1,294*(155/406,455)]. For two-adult households the calculation is: (2,589)*(155/406,455) which is 0.99. The weight for each household size follows the same formula.

2.5 Provincial Sample Distribution and Provincial Weights

The adult population of the ten Canadian provinces and the distribution of the survey sample among the provinces are detailed in Table 2.2. In terms of the percentage of sample per

[5] Ideally the population numbers would be for 2015 and include only Canadian citizens but these numbers are not available so Statistics Canada numbers for the number of adults are used.

province, the design called for a slight over-representation of the smaller provinces and a corresponding under-representation in Ontario, Quebec, and Alberta.

Table 2.2 Population, Sample Distribution, and Provincial Weights

Province	# pop	% of pop	# sample	% sample	Prov Weight
Nfld	406,455	1.65	155	3.69	0.44702
PEI	105,780	0.43	160	3.81	0.11270
Nova Scotia	729,545	2.96	163	3.88	0.76298
NB	582,625	2.36	164	3.90	0.60561
Quebec	5,996,930	24.33	797	18.97	1.28268
Ontario	9,439,960	38.30	1,534	36.51	1.04904
Manitoba	871,460	3.54	215	5.12	0.69097
Saskatchewan	734,250	2.98	214	5.09	0.58489
Alberta	2,515,180	10.20	240	5.71	1.78651
BC	3,267,345	13.25	560	13.33	0.99461
Totals	**24,649,530**	**100.00**	**4,202**	**100.00**	

Because the sample distribution is not exactly proportional to the population of the provinces, the data must be weighted before national estimates are derived (no province weight is required in comparisons between provinces). Weights are obtained by dividing the proportion of households in the population by the proportion of the households in the sample for that province. For example, Alberta has a weight of 1.78651. In preparing national estimates, each Alberta case counts for 1.78651 observations in the weighted data set; in other words, Alberta cases are "weighted up" so that the impact of the Alberta sample on national estimates is an accurate reflection of Alberta's proportion of the total number of households in Canada. Conversely, provinces where the weights are less than one, for example PEI cases, with a weight of 0.11270 are "weighted down." These provincial weights are incorporated into the 'pop' and 'samp' weights provided in the data file.

2.6 National Estimates

In order to produce national estimates it is advisable to correct for both the unequal probabilities of selection at the household stage and the unequal probabilities of selection based on province of residence. Essentially this means combining the household and province weight. Weights for the national estimates can total to the population or to the sample size. In the data file both weights are provided. The names of these two variables

are:

- WeightTOsampBYPopul_count_Prov_HHsize, and
- WeightBYPopul_count_Prov_HHsize.

As indicated by the variable names the first corrects for unequal probabilities of selection by province and household size and when invoked will produce figures that total to the sample of 4,202. The second corrects for the same unequal probabilities of selection but sums to the population.

2.7 Post Stratification and National Estimates

As is the case in almost all surveys younger people and, to a lesser extent, males are underrepresented in the CPS. The gender split for the survey, using unweighted data, is 45.9% for males and 54.1% for females, but in the population, for this age group (18 and older), the distribution is 48.4% for males and 51.6% for females. So males are slightly under-represented in the CPS. The under-representation of youth is much greater: those between 18 and 40 represent 36% of the population but only 18% of the interviews. Weights can also be used to count up the younger respondents (essentially they have a weight of 2, determined by dividing the population percent by the sample percent (36/18 = 2).

All four weights can be combined so that household size, province, gender and age are "corrected' in a single weight. Distributions for ages under 40, 41 to 65 and over 65 by gender for each province in the population is compared to each province in the sample. Dividing the population percent by the sample percent and adjusting for the household size produces these weights. As was the case for the weights that only adjust for province and household size, these weights can, when invoked, sum to the sample or the population. The variable WeightTOsampBYPopul_count_Prov_Age_Gender, as suggested by the variable name sums to the sample of 4,202, while populations counts are provided by using the weight WeightBYPopul_count_Prov_Age_Gender.

For SPSS users, weights can be used by either using the SPSS 'weight by' command or by using a CSAplan (which can be obtained from ISR by writing to isrnews@yorku.ca). For point estimates, such as percentages, means, correlations and regression coefficients, use of the 'weight by' command and use of CSAplan will provide the exact same numbers. Note that use of the weights that sum to the sample size with the SPSS weight by command is based on the original sample size and there is no accounting for sample design effects. Weighting in this manner, so that the weighted sample size is equal to the actual number of interviews results in *in*correct estimates of standard errors and, of course, incorrect significance tests. While these differences are typically very small, use of a CSAplan provides correct estimates of standard errors and, of course, correct significance tests.[6]

[6] For a review of the use of complex sample plans see Heeringa, West, and Berglund. 2010, who in chapter 4 review the use of complex weights and appendix A5 specifically address the use of the SPSS Complex Samples procedures.

2.8 Daily Sample Distribution for the Campaign-Period Survey

The importance of campaign dynamics in understanding election results has been documented by a number of researchers (Nevitte, Blais, Gidengil, and Nadeau, 2000; Holbrook, 1996; Blais and Boyer, 1996; Johnston, Blais, Gidengil, and Nevitte, 1996; Johnston, Blais, Brady and Crête, 1992; Bartels, 1988; and Brady and Johnston, 1987). By interviewing a cross section of Canadians each day (and including date of interview as a variable in the data set), it is possible to determine the impact of events during a campaign. Using data from the election survey, the analyst can determine if support for specific policy issues, predictions of the results of the election, or ratings of the Prime Minister or the opposition leaders varied, or remained constant, over the course of the election campaign. Similarly, utilization of a rolling cross section sample facilitates division of the campaign-period data set into temporal components.

It is critical to any analysis which includes date of interview as a continuous or contingent variable that the sociodemographic characteristics of the survey respondents do not systematically vary over time. Because 'easy-to-reach' respondents (people who are more often home and willing to do the interview when first contacted) have different characteristics than 'hard-to-reach' respondents (Groves, 1989), it is important that each day of interviewing includes a mix of easy and hard-to-reach people. Assume, for example, that educational achievement is found to covary with attitudes about a specific election issue such as the importance of creating jobs. If more of the interviews at the beginning of data collection were completed with respondents with lower levels of education (and if they were more supportive of job creation efforts as compared to paying down the debt), and if more of the interviews at the end of data collection were completed with respondents with high levels of education (and they were less supportive of job creation efforts), it would be possible to mistake a change in respondent characteristics for a change in attitudes.

Given the small sample for any one day, daily variation in the number of completed interviews is expected. However this variation is less pronounced when the number of completed interviews is averaged over a three- or five-day period. Variation in the number of interviews per day varies in part because some days, for example Fridays, tend to have lower co-operation rates and other days, such as Sundays, have higher rates. Other factors such as the weather ("nice" days have lower co-operation rates), the complement of interviewers working each shift (there is variation among interviewers in the response rates they obtain), and the number of days before the vote (all things being equal, co-operation increases the closer to the vote the interview attempt is made). There is an attempt to minimize the variation by controlling the amount of sample released each day for calling. Each day of sample release was, within provinces, divided into "sample replicates." Each sample replicate was a random sample of the day's release. Three to five replicates were released each day depending on the factors noted above.

2.9 Post-Election Survey Sample

The sample for the Post-Election survey was comprised of respondents to the CPS. At the end of the CPS, interviewers ensured that they had a first name or some other identifier (such as the respondent's initials or position in the household, e.g., mother). This information, as well as the gender, year of birth, and the respondent's telephone number, was recorded on a

"cover sheet." At the start of the PES, the cover sheets were put into a random order (shuffled) so that the time of the first call for the PES was not related to the date of interview, or the day of sample release during the CPS. The interviewer called and asked for the person by name or identifier. If there was any concern about reaching the correct person, the interviewer also checked age and gender.

2.10 Mail-back Sample

At the end of the Post-Election survey, respondents were asked to provide their postal address so they could be sent the mail-back survey. Mail-back information was provided by 70 percent of the PES respondents. More than half (1,289 or 62%) of the PES respondents who provided an address returned a MBS.

3 Data Collection

3.1 Introduction

A description of the data collection procedures is outlined in this section of the technical documentation. Computer Assisted Telephone Interviewing (CATI) was used for data collection and all interviewing was completed from ISR's in-house telephone call centre. Supervisors monitored (listened to) about 10 percent of interviewers' calls to coach interviewers in following procedures and to verify that the interviewers were reading questions and recording answers correctly.

3.2 Data Collection Procedures

3.21 CPS and PES Number of Calls

In order to maximize the chances of getting a completed interview from each telephone number, call attempts were made during the day and the evening - for both week and weekend days. In the CPS, typically between two and four call attempts were made each day (split between day and evening hours) during the first four days that a sample was released. Although over half of the interviews completed in the CPS required only one or two call attempts, five percent of the completed interviews required ten or more calls (Table 3.1).

Table 3.1 Call Attempts Required to Obtain Completed Interview

	CPS		PES	
Attempts	#	%	#	%
1	1,241	29	822	27
2	934	22	575	19
3	583	14	383	13
4	412	10	290	10
5	287	7	214	7
6-9	548	13	376	13
10-14	158	4	189	6
15 or more	39	1	139	5
Totals	**4,202**	**100**	**2,988**	**100**

More intensive call efforts were required to complete the PES as compared to the CPS. In the later, only five percent of the interviews took ten or more calls, while 11 percent of the PES interviews took ten or more call attempts. The variables CPS15_ATEMPTS and PES15_ATEMPTS identify the number of calls required to obtain a completion. Note that the larger number of interviews completed on the tenth or subsequent call attempts in the PES were crucial to achieving an acceptable re-interview rate. The more truncated calling efforts in the CPS results from the use of the rolling cross section sample release. Sample released on the last day of calling was only called on that day, sample released on the second last day of calling was only called for two days, etc.

3.22 CPS and PES Refusal Conversions

In addition to making numerous call attempts and spreading these attempts over day, evening and weekend time slots, efforts were made to "convert" refusers on both the CPS and PES. Respondents and/or households who refused to participate when initially contacted by an interviewer were contacted a second time in both surveys. (Often, several call attempts were required to reach refusers.) In the CPS, refusal conversion attempts had to be made within the 10 day calling period whereas in the PES the conversion attempts were typically made two or three weeks after the initial refusal.

The number of converted refusals in the CPS data file is 415 or almost ten percent of the completions (which is almost twice as many conversions as there were in the 2011 CPS). There were fewer converted refusals in the PES, 163, or 5.5% (which is almost 2% more than in the 2011 PES). Of course, the total number of refusals to the CPS was much larger than in the PES. CPS respondents were "cold called" and knew nothing about the survey in advance of the interviewer's calls whereas the PES respondents had completed the first survey and knew that an ISR interviewer would call back after the election. The likelihood of a successful conversion was much higher in the PES. Just over one-quarter of the initial PES refusals were converted but the percent for the CPS was 11%. As compared to the CPS, the higher conversion rate in the PES, like the greater number of calls to 'hard-to-reach' respondents, resulted in a higher response to the PES survey – 71% of the CPS respondents completed the PES. The variables CPS15_REFUSALS and PES15_ REFUSALS identify whether the interview was a "standard" completion or a "converted" refusal.

The careful attention to the number and timing of call backs and refusal conversions is designed to increase the response rate, thereby improving sample representativeness. Many researchers have found that respondents who are 'hard-to-reach' and those who 'refused' have characteristics that are somewhat different from typical survey responders (Fitzgerald and Fuller, 1982). A cursory review of the 2015 data file indicates that 'hard to reach respondents' (those who were interviewed after more rather than fewer call attempts in the CPS) were younger and had higher levels of education, and in the PES the 'hard-to-reach' were more likely to report they were interested in the election and to have voted. Respondents completing interviews after an initial refusal were more likely to admit to being a non-voter and have lower levels of education.

3.23 Mail-back Survey

At the end of the PES, respondents were asked if they would be willing to provide an address so that a mail-back questionnaire could be sent to them. Twenty-nine percent of the PES respondents declined to provide an address and could not be included in the MBS component of the CES. The PES respondents who provided mailing addresses received up to five contacts encouraging them to complete and return the mail-back questionnaire.

The first contact included the questionnaire, a covering letter, and a postage-paid pre-addressed return envelope. The second was a reminder/thank you card (resembling a post card) sent one week after the first questionnaire package was sent. The first and second mail contacts were sent to all respondents. The mailings were staggered and sent every week at the start of the PES calling and somewhat less often near the end of calling. A second questionnaire (plus covering letter and return envelope) was sent only to

non-responders and typically was mailed about three weeks after the first reminder card. One week later the second reminder card was sent. Finally, telephone calls were made to all non-responders about two weeks after the last reminder card was sent.

3.3 Response and Re-Interview Rates

3.31 Campaign-Period Survey Response Rate

There are numerous ways to calculate response rates in survey research (Dillman, 2000; Smith, 1995; Groves, 1989; and Groves and Lyberg, 1988). The method used in this project is conservative; most other ways of calculating the response rate would produce inflated values. The response rate was defined as the number of completed interviews divided by the estimated number of eligible households times 100 percent.

Details on the calculation of the response rate for the 2015 CPS are as follows. Of the 14,976 telephone numbers included in the sample, 10,327 were identified as being eligible households (completions [n=4,202] + refusals [n=3,801] + call backs [n=2,324], see Table 3.2). Not eligible households (where the respondent was unable to speak English or French, was not physically or mentally healthy enough to complete the interview, was not a Canadian citizen, etc. [n=1,106]), and non-residential and not-in-service numbers ([n=2,279]) accounted for 3,385 of the telephone numbers. It was not possible to determine the eligibility status for 1,264 of the sample telephone numbers. For response rate calculations, it was assumed that the proportion of these 1,264 numbers which were eligible household numbers was the same as it was in the rest of the sample.

Table 3.2 Final Sample Disposition and Response Rate: 2015 CPS

Results	number	percent
completions	4,202	28
refusals	3,801	25
call backs	2,324	16
ill/aged/language problem/absent/not a citizen/died	1,106	7
not-in-service & non-residential	2,279	15
eligibility not determined	1,264	9
total sample of numbers	**14,976**	**100**
total number of households determined eligible	10,327	
total number of households determined not eligible	3,385	
household eligibility rate	.753136	
estimated number of eligible households	11,729	
response rate	37	

This proportion, or "household eligibility rate" was .753 (eligibles [10,327]/(eligibles [10,327] + not eligibles [3,385]) = .753136). The estimated total number of eligibles was then computed as 11,279 (10,327 + [.75 x 1,264] = 11,279). Dividing the number of completions (4,202) by the estimated number of eligibles (11,279) gives a final response rate of 37 percent. The rate is 4% lower than that achieved 4 years earlier for the 2011 CPS. The lower response rate is indicative of a similar trend in declining response rates at ISR and as reported by American survey researchers starting in the late 1990s. See reports by: Curtin, Presser and Singer, 2005; and Groves, Dillman, Eltinge and Little, 2002.

Many organizations would not include "eligibility not determined" numbers in the denominator for the response rate calculations on the argument that few of these numbers would be eligible households. (See: Groves and Lyberg, 1988 for a debate on this issue.) This version of the response rate, sometimes called a completion rate, calculated as completions/known eligibles is 41 percent (4,202/10,327). Other organizations calculate response rates as the number of completions over the number of completions plus refusals. This version of the response rate, which is sometimes known as the participation rate, is 53 percent (4,202/4,202+3,801).

Because of the rolling cross section sample design, numbers released for calling on the last days of the campaign survey get fewer calls and no refusal conversion attempts thus a lower response rate is achieved. The response rate for the sample released for the last three days of the campaign was 20 points lower than that obtained for the sample released in the first 31 days of the campaign. This sample release strategy helps to explain why the election survey has a lower response rate than most other RDD general population surveys completed at ISR (which typically range between 45% and 55%).

3.32 Post-Election Survey Re-Interview Rate

The PES re-interview rate in 2015, 71%, was lower than the 76% obtained for the RDD component in in 2011. About 70% of the non-response to the PES was accounted for by refusals and call backs. Illness or death of CPS respondents, never answered telephones (typically 20 or more calls), and changes in telephone numbers (PES respondents had their number changed and the new number was unlisted; the number was changed and the new number listed by the telephone company reached the wrong household; respondent left the household and those remaining in the household either could not or would not provide a new number) account for the remaining non-response to the PES.

3.33 Mail-back Re-Interview Return Rate

As noted above, 70% of PES respondents provided an address and of these 1,289 or 62% completed and returned the mail-back survey. This represents 43% of the PES respondents (not everyone provided an address) and 31% of the CPS respondents (those who competed only the CPS were not asked for an address as they did not compete the PES). These re-interview rates are a few points lower than 2011 where 46% of the PES respondents completed the MBS.

4 Questionnaire Issues and Data Processing

4.1 Introduction

Computer Assisted Telephone Interviewing (CATI) was used for data collection in the CPS and the PES. With CATI, interviewers read questions from a computer screen and enter answers directly into a series of computer files for processing. CATI software automates skip patterns so that interviewers do not have to determine what questions are asked, allows questions to be date stamped so they can be asked on certain set days, and provides a mechanism for systematically varying the order in which respondents receive questions or deliberate variations in question wording.

Note that most variables in the campaign-period survey include the prefix CPS. The prefixes PES, and MBS are used to indicate that the variable is from the post-election and mail-back, surveys, respectively. The digits '15' are added, as in CPS15_xxx to indicate the year of the survey. (Previous CES data files have been merged and use of the year allows differentiation when the same questions have been in multiple election studies.)

4.2 Assigning Missing Values

With some frequency, in both the CPS and PES surveys, whether or not a respondent is asked a question is conditional on answers to previous questions. For example, respondents who said they were unlikely or certain not to vote (CPS15_10) were not asked questions about who they were going to vote for (CPS15_11 and CPS15_12). These respondents have "missing data" for the questions they skipped. But they were asked who they would vote for if they decided to vote at CPS15_14. All other respondents have missing data for this question. In addition, respondents who, when asked which party they were going to vote for (CPS15_11), responded "don't know' were asked if they were leaning towards a party (CPS15_12) and all other respondents have missing data for this question.

The vote intention section includes the most complex conditional logic used in both the CPS and PES surveys. In examining the data, analysts need to be cognizant of the skip pattern logic and how it affects the missing data at any one question.

The most frequent source of non-response in the PES and MBS results from non-completion of the survey. Respondents who did not complete the PES or the MBS have missing data for all questions asked in these surveys.

By and large, with a little examination of either the frequency counts or the questionnaire, the reasons for having missing data are self-evident. There are times, however, when the reason for skipping questions is not quite as obvious. In the CPS, the order that respondents were asked to rate the parties and the leaders, was randomized for the four main parties (PC, Liberal, NDP and Bloc) and four main party leaders (the green party and its leader Elisabeth May were asked at the end of each randomization). When asked about any one specific party, a respondent who answered they did not know anything about any of the parties (questions CPS15_18 to CPS11_21) were not asked about the other parties. Because the order of the questions asking for party ratings was random, the number of cases assigned a missing value on these questions varies somewhat between questions. Similarly, respondents who say they do not know anything about any of the leaders were not asked to

rate the remaining leaders. Again, because of the randomization of the order in which the leader questions were asked, the number of cases assigned a missing value on these questions varies somewhat. The randomization of questions is detailed in Section 4.4

Respondents who indicated in the campaign-period survey that they were not working were not asked to describe their occupation or answer questions related to their job in the post-election survey (PES 11_91 to PES11_ 93).

4.3 Province Specific Questions

There were five questions (three in the CPS and two in the PES) which were only asked of respondents in Quebec. These were:

- rating the Bloc Québécois party (CPS15_21),
- rating the leader of the Bloc - Duceppe (CPS15_26),
- the need for a sovereigntist party in Quebec (CPS15_65c),
- view of how a possible Quebec separation would affect the 'language situation' in Quebec (PES15_39), and
- how a possible Quebec separation would affect the 'standard of living' in Quebec (PES15_40).

Respondents residing in a province other than Quebec have missing data for these questions.

4.4 Randomization in the CPS

The logical operators resident in CATI were used to randomize the order in which respondents received items in several sections of the two telephone questionnaires. Given that order effects have been identified in surveys, but are not always easy to predict (Schuman and Presser, 1981), the order randomization was designed primarily as a precautionary measure to limit the impact question order had on overall response. CATI was also used to vary the wording of questions. The importance of the way in which issues are framed in question wording has been recognized by survey researchers (Converse and Presser, 1986; and Schuman and Presser, 1981).

The software used at ISR makes it easy for users of the data to determine what effect, if any, the random order and variation in question wording had on response. To examine the effect of randomization the analyst must run cross tabulations of the questions of interest by the random number variables (RANDOM*X* , where *X* is the specific random number used for the question(s) of interest, along with the year and survey identifiers). The random numbers were created before interviewing commenced and were added to the data set as part of the sample record (along with telephone number, ID number, etc.). The range and value of each random number (i.e., a range of 2 with values 1 and 2 each of which was used about one-half of the time, or a range of 3 with values of 1, 2 and 3 with each used one-third of the time, etc.) can be determined by running a frequency count on the random number as each random number is a variable in the data set.

4.41 Party and Leader Ratings

Respondents in the CPS were asked to rate the four main parties (PCs, Liberals, NDP and in Quebec the Bloc Quebecois) in CPS15_18- CPS15_21 and party leaders (Harper, Trudeau, Mulcair, and in Quebec Gilles Duceppe) in CPS15_23- CPS11_26 on a 0 to 100 scale where 0 meant they 'really disliked' and 100 meant they 'really liked' the party/leader. As in previous versions of the CPS, the order in which a respondent was asked to rate the political parties and their leaders was randomized. Because the randomization was for four leaders/parties there were 24 different possible orders (the product of 4*3*2*1). Each respondent was randomly assigned a random number which was a four-digit string of numbers (see variable itemseq). The item sequence numbers were comprised of the digits 1, 2, 3, and 4 and ranged from 1234 to 4321 (where the digit 1 referred to Harper/ Conservative Party, 2 was for Trudeau /Liberal Party, etc. The order of the four digits in the string determined the order in which the respondent was asked to rate the parties and their leaders. During the execution of the survey, CATI went to look at the first digit of the four digit string and then followed the code as constructed in the same way as the code was used for leader ratings. For example, when itemseq= 1234, the leaders (and parties) were presented to respondents in the following order: Conservative, Liberal, NDP, Bloc. When itemseq= 4321 the order was: Bloc, NDP, Liberal, Conservative. The same random number (itemseq) was used for both parties and leaders.

Given the small number of respondents receiving each of the possible sequences of questions, the randomization is precautionary (a frequency count for itemseq shows about four percent of respondents were assigned to each order). There is some evidence that ratings on a scale are relative to the first rating given by the respondent. For example, if leader "A" is given a 40 and leader "B" is liked more, they will be given a number higher than 40; if they are liked less, a number lower than 40 will be given. Thus, the first rating acts as an anchor point from which respondents adjust up or down as they are asked their ratings for other leaders. Because the exact placement of the first rating may have more variance than the ratings that follow, randomizing the order in which the leaders are ranked will minimize this effect.

Respondents were also asked for ratings of the Green Party but this rating was not part of the randomization and was always the last party rated (that is, after respondents provided ratings for the Liberals, Conservatives, NDP and Bloc). Ratings for Elisabeth May, the Green Party leader, were also requested after ratings had been given for the other four leaders. Respondents who volunteered they did not know *any* of the leaders well enough to rate them were not asked for ratings and non-Quebec respondents were not asked to rate Duceppe, thus accounting for missing values at the leader ratings questions.

4.42 Attention to Issues

In the Campaign Period Survey, respondents were asked how much attention they personally paid to six issues: health care, welfare, education, the environment, crime and justice, and defence (CPS15_2 - CPS15-7). When RANDOM1= 1, the order was as noted in the preceding sentence. When RANDOM1= 2, the order in which the questions were asked was:

welfare, education, the environment, crime and justice, defence, and health care.

When RANDOM1= 3 the order was:

> education, the environment, crime and justice, defence, health care and education, etc.

Running a crosstab of attention paid to health care (CPS15_2) by RANDOM1 will show that the number of times health care was asked first, second, third, etc. was between 16.0% and 17.2% so about an equal number of respondents were assigned to each of the six question order sequences. In addition, there is some variation in the percentage of respondents who said they pay 'a lot' of attention to health care according to when the health care question was asked relative to the other issues in the list of six (i.e., when it was asked first, second, third, etc.). On average 79.8% of respondents said they paid 'a lot' of attention to health care but when health care was the first issue asked of respondents only 72.9% said they paid 'a lot' of attention. Conversely, when health care was the third or fourth issue asked about the percentage of respondents who said they paid 'a lot' of attention was 84.0% and 80.3% respectively (*p =.000*). This pattern, however, was not found for any of the other five issues. There was no other issue whre the percent of respondents saying they paid 'a lot' of attention to the issue was lowest when the issue was the first of the six presented to respondents. While there is variation in the response to each of the six issues by placement, and this variance is statistically significant, there is no consistent pattern and this suggests that the variation is random.

4.43 Spending Cuts

Respondents were asked a series of questions about government spending. The areas of spending covered in the order experiment were the same as those used for the attention questions (health care (CPS15_32), welfare (CPS15_33), education (CPS15_34), the environment (CPS15_35), crime and justice (CPS15_36) and, defence (CPS15_37). The order in which the respondent was asked about spending in these areas was determined by RANDOM2.

4.5 Randomizations in the PES

4.51 Satisfaction with Harper Government

In the PES, survey respondents were asked how satisfied they were with the Harper government's performance in dealing with five issues. With the exception of welfare, which was dropped from the list, these were the same issues which respondents were asked about in the 'attention' and 'spending' in the CPS. In the PES, RANDOM5 was used.

4.52 Dealing with Economic Problems

Survey item PES15_57 also includes a subtle wording variation about "the BEST way to deal with major economic problems." When RANDOM2= 1 respondents' answer options were: "***more*** government involvement, or leave it to the private sector" and when it was 2 they could answer 'government involvement, or leave it to the private sector'.

The final split sample on wording was for PES15_59A, the federal party identification question. The first version of the question asked respondents if they usually ***think of themselves*** as a Liberal, Conservative, etc. The second version asked respondents if they ***identified with*** the Liberals, Conservatives, etc. When RANDOM3 was 1 respondents were asked the first version of the question, when it was 2 they got the second version.

4.6 Coding of Open-Ended Questions and "Other Specify" Options

4.61 Open-Ended Questions

There were only two open-ended questions, asked of all respondents, in the two telephone surveys. Respondents were asked about the most important issue to them personally in the election in both the campaign (CPS15_1) and post-election surveys (PES15_2).

In the PES, respondents who reported they were working for pay in the CPS were asked to describe their occupation.

4.62 Most Important Issue (CPS and PES Questionnaires)

In both the campaign-period and post-election surveys, respondents were asked to identify the issue which was most important to them personally in the election. Most respondents provided a single response, and codes were developed for the more common "double answers" – for example, 'health care and education'. If a respondent provided more than one response that could not be coded into a single category, the first response was coded (unless it was not codeable and then the second response was used). To the extent possible, the same set of codes used in 2015 were based on those used in 2011, 2008, 2006 and 2004 in order to facilitate between-survey comparisons.

4.63 Occupation Coding

In the Campaign-Period Survey respondents were asked their employment status. In the Post-Election Survey respondents who reported they were employed in the CPS were asked for details about their occupation and this information was used to assign a National Occupation Code (NOC codes). There are multiple versions of NOC and, as was the case in 2011, 2008 and 2006, and 2004, the HRDC version was employed for the coding. Copies of the codes are available both in book form and on line (http://www5.hrsdc.gc.ca/NOC/English/NOC/2011/Welcome.aspx). Respondents who only completed the Campaign-Period survey as well as respondents who were not working or refused to provide their occupation all have missing values (i.e. no NOC code).

4.64 Other Specify Answers

Many of the questions allowed for a response other than those provided to respondents by interviewers. For example, questions about vote intentions or voting behaviour included an option for the respondent to provide an answer other than the ones read by the interviewer. When respondents gave these 'other' answers interviewers were asked to write out - specify - how the respondents answered. For every question, these 'other specify' answers were reviewed and, when possible, coded into existing response categories. For a very small

number of questions a new answer category was created, and for these questions the data file will have response options that are not listed in the questionnaire).

Respondents could also write out answers in the MBS and these text answers were also reviewed and treated in the manner as other specify comments in the telephone surveys.

4.65 Geographic Variables

For those respondents who provided a postal code, their postal code was run through Statistics Canada's Postal Code Conversion File (PCCF). The postal code is used to provide a range of geographic/location data such as Census Metropolitan Area codes (variable CMA_15 in the data file), Census Tract (CT_15), Dissemination Area (DA_15), Latitude and Longitude (LAT_15 and LONG_15), etc. Details about the conversion file and the variables provided from the conversion can be found by a Goggle Search (PCCF documentation) or by visiting the following address:
http://ivt.crepuq.qc.ca/recensements/recensement2006/documentation/92-153-g2011002-eng.pdf/pccf documentation

References

Bartels, Larry M. 1988. Presidential Primaries and the Dynamics of Public Choice. Princeton: Princeton University Press.

Blais, André and Martin Boyer. 1996. "Assessing the Impact of Televised Debates: The Case of the 1988 Canadian Election." British Journal of Political Science, 26:143-64.

Brady, Henry E. and Richard Johnston. 1987. "What's the Primary Message: Horse Race or Issue Journalism?" In Gary R. Orren and Nelson P. Polsby, eds., The New Hampshire Primary and Nomination Politics. Chatham, New Jersey: Chatham House.

Converse, Jean M. And Stanley Presser. 1986. Survey Questions: Handcrafting the Standardized Questionnaire. Beverley Hills: Sage.

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

Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method* (2nd edition). New York: John Wiley and Sons.

Groves, Robert M. and Lars E. Lyberg. 1988. "An Overview of Nonresponse Issues in Telephone Surveys." In *Telephone Survey Methodology*, Editors: Robert M. Groves, Paul P. Biemer, Lars E. Lyberg, James T. Massey, William L. Nicholls II and Joseph Waksberg. New York: John Wiley and Sons, 191-212.

Groves, Robert M. 1989. *Survey Errors and Survey Costs*. New York: John Wiley and Sons.

Groves, Robert M., Don A. Dillman, John L. Eltinge and Roderick J. A. Little. 2002. *Survey Nonresponse*. New York: John Wiley & Sons

Groves, Robert M., Floyd J, Fowler, Jr., Mick P. Couper, James M. Lepkowski. Eleanor Singer, and Roger Tourangeau. 2009. Survey Methodology. 2nd edition. Hoboken, New Jersey: John Wiley and Sons.

Heeringa, Steven G., Brady T. West, and Patricia A. Berglund. 2010. *Applied Survey Data Analysis*. London: Chapman and Hall.

Holbrook, Thomas M. 1996. *Do Campaigns Matter*? Thousand Oaks, California: Sage.

Johnston, Richard, André Blais, Elisabeth Gidengil, and Neil Nevitte. 1996. *The Challenge of Direct Democracy*. McGill-Queen's University Press: Montreal.

Johnston, Richard, André Blais, Henry E. Brady, and Jean Crête. 1992. *Letting the People Decide: Dynamics of a Canadian Election.* McGill-Queen's University Press: Montreal.

Kish, Leslie. 1965. *Survey Sampling*. New York: John Wiley and Sons.

Nevitte, Neil, André Blais, Elisabeth Gidengil, and Richard Nadeau. 2000. *Unsteady State: The 1997 Canadian Federal Election*. Toronto: Oxford.

Ornstein, Michael. 2013. *A Companion to Survey Research.* London: Sage.

O'Rourke, D. and J. Blair. 1983. "Improving Random Respondent Selection in Telephone Surveys." *Journal of Marketing Research*, 20, 428-432.

Schuman, Howard and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys*. Orlando: Academic Press.

Smith, Tom W. 1995. "Trends in Non-Response Rates." *International Journal Of Public Opinion Research,* 7 (2), 157-70.