The 2004 Canadian Election Survey

Technical Documentation

The 2004 Canadian Election Survey

Technical Documentation

David A. Northrup

Institute for Social Research
York University
February 2006

Conditions of Release

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

> Data from the 2004 Canadian Election Survey were provided by the Institute for Social Research, York University. The survey was funded by the Social Sciences and Humanities Research Council of Canada (SSHRC) and Elections Canada, and was completed for the 2004 Canadian Election Team of André Blais (Université de Montréal), Joanna Everitt, University of New Brunswick, Patrick Fournier (Université de Montréal), Elisabeth Gidengil (McGill University), and Neil Nevitte (University of Toronto). Neither the Institute for Social Research, the SSHRC, Elections Canada nor the Canadian Election Survey 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 Associate Director, Institute for Social Research, The TEL Building, 88 The Pond Road, York University, 4700 Keele Street, Toronto, Ontario, M3J 1P3 and to André Blais, Départment de Politique Science, Université de Montréal, CP6128 Succ. Centreville, Montréal, 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

1.1 Introduction

As a result of Monday June 28th 2004 Canadian General Election the Liberal Party formed a Minority government. A set of surveys conducted during the campaign and after the election with a random sample of Canadians provide data that can be used to help explain the election outcome. As was the case in the last four Canadian Elections (1988, 1993, 1997 and 2000) the Institute for Social Research (ISR) at York University conducted pre- and post-election telephone surveys as well as a post-election mail back survey under the direction of the Canadian Election Team. This Technical Report outlines the design and conduct of the 2004 Canadian Election Study (CES) study. The design of the 2004 study largely replicates those of the previous four Canadian Election Studies completed at ISR.

The 2004 CES included three survey components: the Campaign-Period Survey (CPS), the Post-Election Survey (PES) and the Mailback Survey (MBS). Interviews in the CPS were collected over the 36 days of the campaign, from Sunday May 23 (the day the writ was dropped) to Sunday June 27, the day before the election on Monday June 28. The PES survey commenced one week after the election on July 5 and ran to Sunday September 19. The week gap between the election and the start of the PES was not typical of earlier CES where the PES typically started the day after the election. The delay was caused by uncertainty about who would form the government, depending on recounts, and the desire not to start the PES on the Canada Day holiday weekend. The MBS, started about one week after the start of the PES and questionnaires returned to ISR before the end of November were added to the data file.

A rolling cross sectional sample release was employed for the CPS. The sample selection methodology used in the 2004 Canadian Election Survey duplicated that used in previous Canadian Election Studies. Modified random digit dialling (RDD) procedures were utilized to select households, and, within households, the birthday selection method was used to select respondents.

The design of the CES included a longitudinal component as the CPS respondents were asked to complete the PES and respondents to the PES were asked to complete the MBS. There was no substitution of respondents for those who declined to participate in the second or third wave of the study. The number of completed interviews, the sample size, for the three studies was 4,323, 3,138 and 1,674 respectively. The response rate to the CPS was 55 percent and the reinterview rate for the PES was 73 percent and 53 percent of the PES respondents (representing 39 percent of the CPS respondents) completed the MBS.

Twenty percent of the CPS interviews were completed by the Hitachi Survey Research Centre at the University of Toronto. The remaining 80 percent of the CPS and all of the PES and MBS were

conducted at ISR. Because the Hitachi centre uses the same Computer Assisted Telephone Interviewing (CATI) software as ISR it was possible to send them an exact copy of the questionnaire. ISR also provided the sample used by Hitachi for the proportion of the CPS they completed. The selection of respondents, number of call attempts, the response rates and the daily distribution of the sample, was the same for the two survey houses through out the conduct of the CPS.

The Institute's CATI software is from the Computer-Assisted Survey Methods Program (CSM) at the University of California, Berkeley.

Easy-to-read copies of the questionnaire accompany this documentation. Complete copies of the CATI surveys are available at ISR upon request. (Contact Anne Oram oram@yorku.ca .) An explanation of the way in which CATI was used in the survey is included in Section 4 of this documentation.

2. SAMPLE DESIGN

2.1 Introduction

The sample for the CPS was designed to represent the adult population of Canada: 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 survey was conducted by telephone, the small proportion of households in Canada without telephones were excluded from the sample population.[2]

2.2 Selection of Households

To select individual survey respondents for the CPS, 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 in Canada. Unfortunately, such a listing does not exist. To select numbers ISR employs a modified form of random digit dialling (RDD).

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, as well as blocks of telephone numbers between or on either side of listed numbers are included in the sampling frame. For example, if the following telephone number was found in a directory, (416) 651-8513 then all numbers from (416) 651-8500 to (416) 651-8599 would be included in the sample. A computer would then generate a random sample of telephone numbers from this list. Since unlisted numbers and numbers too new to be included in the directory are interspersed among valid numbers, this strategy provides a much better sample than one based on listed numbers alone.

---

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

[2] Statistics Canada (Residential Telephone Service Survey, Catalogue 56M0001X) estimates that 3.7percent of the private households in Canada do not have a residential telephone number. Almost equal numbers have either no telephone or a cell phone only. Most cell phones are not included in RDD samples.

As well as household telephone numbers, RDD samples include "not-in-service" and "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 begin an interview than more traditional methods that require the interview to obtain a list of all adult household residents. This less intrusive approach makes it easier for the interviewer to secure the respondent's cooperation.

2.4 Household Weights

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. 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 households are over represented, and larger households with two or more adults are under represented. Most practitioners of survey research "weight the data" in order to compensate for the unequal probabilities of selection (one adult households are given a weight of one, two adult households are given a weight of two, three adult households are given a weight of three, etc.).[4]

---

[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. See, for example: Lessler and Kalsbeek, 1992 Chapter 8; Kalton, 1983 Chapter 10; and Babbie, 1992 Chapter 5, 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.

Conventionally, most users of survey data wish to have the same number of observations in the weighted and unweighted data set.[5] This adjustment is made by determining the number of cases in each household size category that would have been in the sample, if an interview had been completed with each adult member of the household, and then dividing the sample among each household size category according to the proportion of interviews completed in each household size category.

In the campaign-period survey there were 4,323 households in the sample and 1,267 were one-adult households, 2,271 were two-adult households, 508 were three-adult households, etc. (Table 2.1 and variable NADULTS in the data set). The weights for each household are calculated as follows. First, the total number of weighted cases is calculated (number of cases times the number of adults in the household). For three-adult households the calculation is: 508 times 3 which gives 1,524 three-adult households in the weighted sample. In the campaign-period survey there are 8,513 weighted cases. Second, the 8,513 weighted cases are adjusted down to the original sample size of 4,323 (calculated as weighted cases for each household size divided by the weighted sample size times the original sample size). For three-adult households the calculation is: (1,524/8,513) * 4,323 = 773.90. Third, the weight for each household size is calculated (by dividing the adjustment to original sample size by the number of cases). For three-adult households the calculation is: 773.90/508 = 1.523435.

Table 2.1 Calculation of Household Weights (CESHHWGT)

Household Size	No. of Households	Weighted Cases	Adjustment	Weight
1 adult	1,267	1,267	643.3973	0.507812
2 adults	2,271	4,542	2,306.4800	1.015623
3 adults	508	1,524	773.9049	1.523435
4 adults	219	876	444.8429	2.031246
5 adults	47	235	119.3357	2.539058
6 adults	8	48	24.3750	3.046869
7 adults	3	21	10.6640	3.554681
Totals	4,323	8,513	4,323.0000	

[5] While such weights are common they do not include a downward adjustment in sample size to compensate for design effects. Another option used by some researchers is to "weight up" to the population.

Note that in the calculation of the household weights the total number of observations in the sample –the "weighted sample size" – is based on the original sample size, but we do not have a true random sample (as households were used to locate adults) 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, provides researchers with a very good approximation of the precision of their sample for point estimates (such as percentages, means, correlation coefficients (r), coefficients of determination (r^2), and so on). But, treating the sample as if it was a simple random sample of equal size results in *in*correct estimates of standard errors and, of course, incorrect significance tests. Worse, the errors are *down*wardly biased and so give a false sense of the precision of estimates as well as significance tests with too many false positives.

Researchers should consider the use of a statistical package that takes proper account of weights (such as STATA) or the use of procedures in other packages that treat these data appropriately (e.g. UNIANOVA in SPSS) when analysing the data. Another, somewhat less desirable alternative, which would produce reasonably accurate standard errors, would be to compute design effects due to weighting, for a variety of survey items and multiply the weights by a factor that reduces the weighted sample size to a value corresponding to the actual precision of estimates.

2.5 Provincial Sample Distribution

The distribution of Canadian households and survey sample among the provinces, as well as the Provincial Weights (CESPWGT) are detailed in Table 2.2. In terms of the percentage of sample per province, the design called for a slight over representation of the eight smaller provinces and a corresponding under representation reduction in Ontario and Quebec. For example, Alberta has 9.02 percent of the households in the country, but 10.11 percent of the households in the sample. Conversely, Ontario which has 36.17 percent of Canada's population has only 30.49 percent of the sample.

Because the sample distribution is not proportional to the population size (pps) 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 province by the proportion of the households in the sample for that province. For example, Ontario has a weight of 1.1864 (36.17/30.49). In preparing national estimates, each Ontario case counts for 1.1864 observations in the weighted data set; in other words, Ontario cases are "weighted up" so that the impact of the Ontario sample on national estimates is an accurate reflection of Ontario's proportion of the total number of households in Canada. Conversely, provinces where the weights are less than one, for example Alberta (.8928), are "weighted down." Caveats about the effect of weighting on the variance estimates noted above apply here as well.

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. CESNWGT (National Weight) is the product of the household weight and the provincial weight and should be used with the National Sample when national estimates are required.
ROCNWGT allows for "national" estimates excluding the province of Quebec. As in the case for the National Weight, this weight is the product of a household weight and provincial weight, but these weights are computed by excluding the province of Quebec cases for the computations.

Table 2.2: Sample Distribution and Calculation of the Provincial Weight Variable (CESPWGT)

Province	# of HHs*	% of HHs	# HHs Sample	% HHs Sample	Weight
Nfld	185,495	1.71	120	2.78	0.6159
PEI	47,960	0.44	115	2.66	0.1662
Nova Scotia	342,590	3.16	112	2.59	1.2188
NB	271,155	2.50	116	2.68	0.9314
Quebec	2,882,030	26.56	1,048	24.24	1.0957
Ontario	3,924,515	36.17	1,318	30.49	1.1864
Manitoba	419,385	3.87	217	5.02	0.7700
Saskatchewan	372,820	3.44	214	4.95	0.6941
Alberta	979,175	9.02	437	10.11	0.8928
BC	1,424,640	13.13	626	14.48	0.9068
Totals	10,849,765	100.00	4,323	100.00	

* Statistics Canada, 1997. Dwellings and Households: The Nation. Ministry of Industry, Science and Technology, Catalogue No. 93-111, pp 78-89.

Although the weights are provided as part of the data set, users must specify the weights they wish to use in the appropriate programming language before analysing the data. If weights are not invoked the tabulations produced will be for unweighted data. Because the weights include fractions that are rounded and missing values vary by item, there will be minor variations in the number of cases for different analytical procedures and subsets of the data.

2.7 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 sets into temporal components. For example, analysts can divide the campaign-period data into before and after the leaders' debates, four nine day periods, the beginning, middle and end of the campaign, etc.

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; Hawkins, 1975; and Dunkleberg and Day, 1973), 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, the 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 interviewers per day varies in part, because some days, for example Friday tend to have lower co-operation rates and other days, such as Sunday, have higher rates. Other factors such as the weather (“nice” days have lower co-operation rates), the compliment 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 the 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.

2.8 Post-Election and Mailback Samples

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 sex and year of birth of the CPS respondent, 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.

At the end of the post-election survey, respondents were asked to provide their address so they could be sent the mailback survey. Mailback information was provided by 75 percent of the PES respondents. This number is down from the 79 percent obtained in the 2000 PES.

Separate weights were not prepared for the PES and MBS data sets. The re-interview rates are reasonably high and sample attrition between the surveys was not associated with household size or province.