DRAFT

# EQUALITY, SECURITY AND COMMUNITY (ESC)
# WAVE 1, 1999, DATA COLLECTION

## TECHNICAL DOCUMENTATION

# EQUALITY, SECURITY AND COMMUNITY (ESC)
# WAVE 1, 1999, DATA COLLECTION

## TECHNICAL DOCUMENTATION

DAVID A. NORTHRUP

INSTITUTE FOR SOCIAL RESEARCH
YORK UNIVERSITY

2002

## CONDITIONS OF RELEASE

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

> Data from the first wave of the Equality, Security and Community (ESC) survey were provided by the Institute for Social Research, York University. The ESC project was funded by the Social Sciences and Humanities Research Council of Canada (SSHRC), grant number 412-97-0003, Project Director Dr Jonathan R Kesselman, University of British Columbia. The survey component of the ESC study was completed under the direction of Dr. Richard Johnston, UBC. Neither the Institute for Social Research, SSHRC, nor the ESC Research 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, Institute for Social Research, York University, 4700 Keele Street, Toronto, Ontario, M3J 1P3 and to Richard Johnston, Department of Political Science, University of British Columbia, 1866 Main Mall, Vancouver, British Columbia, V6T 1Z1.

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 Equality, Security and Community (ESC) survey was completed at the Institute for Social Research (ISR), York University. There are three sample components to the study: the National Sample (n= 4,101), the Urban Over Sample (n=1,051) and the British Columbia (BC) Resource Community Sample (n = 1,427). While the survey instrumentation and data collection strategies were the same, this documentation refers primarily to the first two sample components. Given that the same questionnaire was used and the Resource Community Sample was "piggy backed" on the first two survey components brief mention is made of the study as appropriate. This documentation, however, is not intended to provide the same level of detail for the Resource Community Sample. This documentation briefly outlines the conduct of the first survey wave in 1999/2000. The second survey wave is scheduled for Winter 2003.

The survey is one part of a large research effort. The larger research project deals with all aspects of the distribution of well-being in Canada and how this is changing over time. (More information about the project can be found by visiting the ESC web site at: http://www.arts.ubc.ca/cresp/outc.htm). The survey component of the ESC project is designed to provide information, for a sample of Canadians, on their social networks; well-being (including employment status, occupation, income, economic security, education, health, etc.); participation in civil society (group memberships, volunteer activities and charitable giving; voting, political party identification); their attitudes towards the role and efficacy of the state, support or opposition for government, and other existing or possible public policies. (A copy of the questionnaire is appended to this document.)

Random digit dialling (RDD) procedures were utilized to select the households and, within households, the most recent birthday selection method was used to identify respondents. At least 12 calls were made to all sampled numbers. Attempts were made to encourage refusers to participate in the survey by calling them at least once after they first refused. All of the interviewing was completed from ISR's centralized telephone facilities in Toronto using Computer Assisted Telephone Interviewing (CATI) techniques. The Institute's CATI software is from the Computer-Assisted Survey Methods Program (CSM) at the University of California, Berkeley.

# 2. SAMPLE DESIGN

## 2.1 National Sample

The National Sample for the ESC survey was designed to represent the adult population of Canada . More specifically, the intention was to sample Canadians 18 years of age or older who speak one of Canada's official languages, and reside in private dwellings[1] in the ten Canadian provinces (so *not* the three territories). Because the survey was conducted by telephone, the small proportion of households without telephones is excluded from the sample population.[2]

The study design, for the National Sample, called for the division of the country into five "regions":

1, Atlantic Canada (Newfoundland, Nova Scotia, Prince Edward Island and New Brunswick);
2, Quebec;
3, Ontario;
4, the Midwest (Manitoba and Saskatchewan); and
5, the West (Alberta and British Columbia).

Relative to their population, smaller provinces were over-represented in the sample. For multiple province "regions", equal numbers of telephone numbers were selected in each province (producing approximately equal numbers of completed interviews (see Table 2.1)). For example, the 800 cases in the Atlantic region were distributed as follows: 201 in Newfoundland, 195 in Prince Edward Island, 205 in Nova Scotia, and 198 in New Brunswick. The distribution is not exactly proportional due to small differences in the proportion of the sample numbers that were found to be households and the small variation in the response rate by province. The intent of over- representing the smaller provinces, of course, is to optimize comparisons between the five regions and provinces within region.

---

[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] Using their Household Inventory and Facilities and Equipment (HIFE) surveys, Statistics Canada estimates that two percent of the private households in Canada do not have a telephone (1997, Catalogue 52-203).

Because the sample distribution is not proportional to the population of the provinces, the data must be weighted before national estimates are derived. No household weight is required in comparisons between provinces. Weights are obtained by dividing the proportion of households (in Canada) in the province by the proportion of the households in the sample in the province. Ontario has the largest weight, 1.6733, as the province has 36 percent of Canada's households, but only 22 percent of the sample (Table 2.1). In preparing national estimates, each Ontario case counts for 1.6733 observations in the weighted data set. In other words, Ontario cases are "weighted up" to make the impact of the Ontario sample on national estimates an accurate reflection of Ontario's proportion of the number of households in Canada. Conversely, provinces where the weights are small, for example PEI (.0941), are those in which the proportion of the sample allocated to the province considerably exceeds the province's proportion of the national population. Such cases are "weighted down."

Table 2.1: Sample Distribution Among the Provinces for the ESC National Sample and Calculation of the Provincial Weight Variable (NPROVWGT)

| Province | # of HHs* | % of HHs | # HHs in Sample | % HHs in Sample | Weight |
|---|---|---|---|---|---|
| Newfoundland | 187,406 | 1.72 | 201 | 4.90 | 0.3518 |
| PEI | 48,630 | 0.45 | 195 | 4.75 | 0.0941 |
| Nova Scotia | 344,779 | 3.17 | 205 | 5.00 | 0.6346 |
| NB | 272,915 | 2.51 | 198 | 4.83 | 0.5201 |
| Quebec | 2,849,149 | 26.21 | 804 | 19.61 | 1.3371 |
| Ontario | 3,951,326 | 36.36 | 891 | 21.73 | 1.6733 |
| Manitoba | 421,096 | 3.87 | 398 | 9.71 | 0.3992 |
| Saskatchewan | 375,740 | 3.46 | 404 | 9.85 | 0.3509 |
| Alberta | 984,275 | 9.06 | 406 | 9.90 | 0.9147 |
| BC | 1,433,533 | 13.19 | 399 | 9.73 | 1.3556 |
| Totals | 10,868,849 | 100.00 | 4,101 | 100.00 | |

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

## 2.2 Selection of Households

To select individual survey respondents, a two-stage probability selection process was necessary. The first stage involves the selection of households by randomly selecting telephone numbers. To select numbers ISR employs a 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 the other commercially available lists of telephone numbers. Numbers from these sources, as well as telephone numbers between or on either side of listed numbers are included in the sampling frame. For example, if the following two telephone numbers are found in a directory, (416) 651-8513 (416) 651-8518, then all numbers from (416) 651-8510 to (416) 651-8519 are included in the sample. A computer is then used to 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. In addition, concentrating on numbers that *are* listed or close to a listed number prevents the inclusion of large proportions of randomly generated numbers that are not numbers of actual households.

RDD samples include "not-in-service" and "non-residential" telephone numbers as well as household numbers (including unlisted household numbers). Typically, non-household telephone numbers are identified the first time the interviewer calls and most of the interviewer's subsequent efforts are then directed at encouraging an informant from a 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 an adult respondent (18 years of age or older) in the household. In households with more than one adult, 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 it is a much less

---

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

intrusive way to start an interview than the traditional "Kish method," which requires a listing of household residents or other methods based on a count of the adult men and women in the household. The less intrusive start makes it easier for the interviewer to secure the respondent's cooperation.

## 2.4 Household Weights for the ESC National Sample

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, she or he 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 are under-represented. As with the provincial weights, described above, 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]

Again, conventionally, some 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.

Following this strategy, the calculation of the household weights for the ESC National Sample is as follows. There were 4,101 households in the sample: 1,110 one-adult households, 2,268 two-adult households, 516 three-adult households, and so on. The weights are calculated as

---

[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.

[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.

follows. First, the total number of weighted cases is calculated using the number of eligible respondents in a household as the weight. For three-adult households, for example, the calculation is: 516 respondents times *3*, which gives 1,548 weighted members of three-adult households. In total there are 8,062 weighted cases. Second, the 8,062 weighted cases are adjusted down to the original sample size of 4,101 (weighted cases for each household size divided by the weighted sample size times the original sample size). For three-adult households the calculation gives 1,548 * (4,101 / 8,062) = 787.44 persons. Third, the weight for persons of each household size is calculated by dividing that desired number by the actual number of (in this case members of three-adult households), which is: 787.44/516 = 1.5260.

Table 2.2 Calculation of Household Weights for the ESC National Sample (NHHSWGT)

| HH Size | No. of HH's | Weighted Cases | Adjustment | Weight |
|---|---|---|---|---|
| 1 adult | 1,110 | 1,110 | 564.64 | 0.5087 |
| 2 adults | 2,268 | 4,536 | 2,307.38 | 1.0174 |
| 3 adults | 516 | 1,548 | 787.44 | 1.5260 |
| 4 adults | 173 | 692 | 352.01 | 2.0347 |
| 5 adults | 29 | 145 | 73.76 | 2.5434 |
| 6 adults | 4 | 24 | 12.21 | 3.0521 |
| 7 adults | 1 | 7 | 3.56 | 3.5608 |
| Totals | 4,101 | 8,062 | 4,101.00 | |

## 2.5 Weighting for National Estimates when using the ESC National Sample

When national estimates are required, the user is advised to correct for both the unequal probabilities of selection at the household stage and the unequal probabilities of selection based on province of residence. NATWGT (National Weight for the National Sample) is the product of the household weight (for the National Sample) and the provincial weight (for the National Sample) and should be used with the National Sample when national estimates are required.

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.

## 2.6 The Urban Over Sample

The Urban Over Sample was designed to increase the number of non-Charter respondents (people other than those with British or French ancestry), and Charter respondents (people of British or French ancestry) living in sampled areas with higher concentrations of non-Charter residents. All interviews for the Urban Over Sample were completed in Montreal (n = 251), Toronto (n = 400), and Vancouver (n = 400). Data collection for the Urban Over Sample overlapped with the last month of data collection for the National Sample and continued on for two months after the completion of data collection for the National Sample.

Statistics Canada information was used to locate census tract (CTs) that had a high incidence of non-Charter people, defined as approximately three times the average for the city as a whole. Using Statistics Canada's CT postal code conversion files, we prepared a list of six-digit postal codes in these areas of high concentration of non-Charter people in each city. Telephone numbers associated with these postal codes are available from commercial sources (such as machine readable listings of telephone books and other commercially available sources that have provided postal codes for addresses listed in telephone books). A list of these numbers was compiled for each city and was used to locate households included in the Urban Over Sample.

Note that in the Urban Over Sample, as in the National Sample, interviews were only completed in English and French. Also, respondents to the Urban Over Sample included both Charter and non-Charter members. That is, while the households selected for the Urban Over Sample were in areas of higher concentration of non-Charter people, being from a Charter group did not preclude being a respondent in the Urban Over Sample. The Urban Over Sample is not a random sample of non-Charter people from each city, but rather a random sample of non-Charter and Charter members who are living in areas of higher concentration of non-Charter people.

## 2.7 Household Weights for the Urban Over Sample

Within households included in the Urban Over Sample interviewers again randomly selected the adult respondent who had the next birthday. As in the case for the National Sample, household size weights have been calculated. Household weights for each city: HHWT_MTL for Montreal,

HHWT_TOR for Toronto, and HHWT_VAN for Vancouver and the complete Urban Over Sample HHWT_OS, (all three places in the Over Sample) are included in the data file. When working only with the Urban Over Sample the analyst should use HHWT_OS weight; when working with only one city, the household weight for that city should be utilized.

## 2.8 Combining the National Sample and Urban Over Sample

If the researcher chooses to combine the National Sample and the Urban Over Sample (OS) in the same analytical procedure (for example using all the cases in a single regression equation) then the data must be weighted to take into account the probability of selection within each strata of the combined sample. This is accomplished by treating each of the three areas in the Urban Over Sample, as well as the remainder of the three provinces they are part of, and the remaining provinces, as 13 different strata and then correcting for inequities in the chances of selection in each strata. This is completed by dividing each strata's proportion of the households in Canada by each strata's proportion of the households in the sample. The figures are provided in Table 2.3.

Note that the weights for the three parts of the Urban Over Sample are quite small when combined with the National Sample cases: each case in the Montreal Urban Over Sample counts for about one-third of an observation, each case in the Toronto Urban Over Sample counts as about one-fifth of an observation, and each case in the Vancouver Urban Over Sample counts for about one-eighth of an observation (the exact weights, given in the Table, are .3160, .2291, and .1395, respectively).

When making national estimates, using the combined data set, the weight to correct for unequal distribution among the provinces and urban areas (NOPRVWGT–National and Over Sample provincial weight), needs to be combined with the household weight for the combined data set (NOHHSWGT). This weight is provided in the data set and is labelled as NONATWGT (National and Over Sample National Weight).

Table 2.3: Sample Distribution Among the Geographical Components in the Combined National and Urban Over Samples and Calculation of the Weight Variable NOPRVWGT

| Province | # of HHs | % of HHs | # HHs in Sample | % HHs in Sample | Weight |
|---|---|---|---|---|---|
| Newfoundland | 187,406 | 1.72 | 201 | 3.90 | 0.4420 |
| PEI | 48,630 | 0.44 | 195 | 3.79 | 0.1182 |
| Nova Scotia | 344,779 | 3.17 | 205 | 3.98 | 0.7972 |
| NB | 272,915 | 2.51 | 198 | 3.84 | 0.6534 |
| Quebec | 2,681,827 | 24.67 | 804 | 15.61 | 1.5811 |
| Montreal OS | 167,322 | 1.54 | 251 | 4.87 | 0.3160 |
| Ontario | 3,757,958 | 34.58 | 891 | 17.29 | 1.9992 |
| Toronto OS | 193,368 | 1.78 | 400 | 7.76 | 0.2291 |
| Manitoba | 421,096 | 3.87 | 398 | 7.73 | 0.5015 |
| Saskatchewan | 375,740 | 3.46 | 404 | 7.84 | 0.4409 |
| Alberta | 984,275 | 9.06 | 406 | 7.88 | 1.1492 |
| BC | 1,315,809 | 12.11 | 399 | 7.75 | 1.5632 |
| Vancouver OS | 117,724 | 1.08 | 400 | 7.76 | 0.1395 |
| Totals | 10,868,849 | 100 | 5,152 | 100.00 | |

Note that in the calculation of both the household and national weights the total number of observations in the sample – often termed the "weighted sample size" – is based on the original sample size, but we do not have a true random sample 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 "tricks" SPSS and SAS into thinking there are the "right" number of cases, and it provides researchers with a reasonable approximation of the precision of their sample. 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 and significance tests

with too many false positives. Of course, given the more unequal weights when the data is combined the design effects are very likely to be considerably larger.

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.

## 2.9 British Columbia Resource Community Sample

Seven places were included in the BC Resource Community sample: Nanaimo, Powell River, Port Alberni, Squamish, Tahsis, Youbou and Chemainus. The geographic boundaries of these communities, as defined by geographic areas in the Canadian census, were provided by the research group. Within these communities a sample of telephone numbers was drawn from several sources, including telephone directories and mailing lists that were purchased by sample providers. Interviewers asked informants (adults who answered the telephone) the location/ address of the telephone number and determined (sometimes with the help of the respondent) if they were located in the communities. Once a household was determined to be eligible all other aspects of the sample design followed that of the national study (random selection of respondent, provision of household weights, etc.).

Different funders and investigators were responsible for the BC Resource Communities Sample. As a result, the sample for the BC Resource Communities has not been combined with the other two sample components.

# 3. DATA COLLECTION

## 3.1 Introduction

Interviewing was completed from ISR's centralized CATI (Computer Assisted Telephone Interviewing) facilities. Each supervisory station is equipped with a video display terminal that reproduces an image of the interviewer's screen and a telephone connection to each interviewer station. This allows supervisors to monitor (listen to) interviewers' calls and visually verify that interviewers record respondents' answers correctly. Just over 10 percent of the interviewing was monitored by supervisory personnel. The data collection procedures used were the same for all three survey components

## 3.2 Data Collection Procedures

In order to maximize the chances of getting a completed interview from each sample number, call attempts were made during the day and the evening - for both week and weekend days. A minimum number of 14 calls were made to each telephone number, of which at least ten were made during evening and weekend hours. Close to one-fifth of the interviews were completed the first time the interviewer called. Although almost two-thirds of the interviews took five or fewer call attempts to complete, close to one-fifth of the completed interviews required ten or more call attempts (Table 3.1). There was some variation in the number of calls required to complete interviews among the three sample components. More effort was required to complete interviews for the Urban Over Sample (26 percent of the interviews required 10 or more call attempts) and fewer were required for the BC Resource Communities (11 percent required 10 or more), while the National Sample calling effort was more or less the average of the other two sample components (18 percent).

Households who refused to participate in the survey were contacted a second time. Interviews were completed in 16 percent of households that refused the first time they were contacted after the initial refusal. In total, in all three sample components 896 interviews were completed on the second or subsequent contact after the initial refusal. (The variable "REFUSALS" identifies "converted" refusal completions.) The conversion rate was somewhat higher for the Urban and BC Samples (17 and 16 percent respectively), than for the National Sample (12 percent).

Table 3.1 Number of Call Attempts: National and Urban Over Sample

| Number of Call Attempts | National Sample Number | National Sample Percent | Urban Sample Number | Urban Sample Percent | BC Sample Number | BC Sample Percent |
|---|---|---|---|---|---|---|
| one | 740 | 18 | 146 | 14 | 313 | 22 |
| two | 646 | 16 | 141 | 14 | 253 | 18 |
| three | 476 | 12 | 96 | 9 | 184 | 13 |
| four | 399 | 10 | 95 | 9 | 139 | 10 |
| five | 312 | 8 | 82 | 8 | 116 | 8 |
| six to nine | 800 | 19 | 213 | 20 | 262 | 18 |
| ten to fourteen | 425 | 10 | 142 | 13 | 100 | 7 |
| fifteen to twenty | 229 | 5 | 73 | 7 | 42 | 3 |
| more than 20 | 74 | 2 | 63 | 6 | 18 | 1 |
| total | 4,101 | 100 | 1,051 | 100 | 1,427 | 100 |

The careful attention to the number and timing of callbacks and attempts to convert initial refusals maximizes the response rate and the representativeness of the sample. Many researchers have found that respondents who are "hard to reach" or "refusers" have characteristics that are somewhat different from typical survey responders (Dunkelberg and Day, 1973; Steech,1981; and Fitzgerald and Fuller, 1982). Whether the respondent refused during the initial contact, the number of call attempts, the number of times the telephone was answered, and other variables that describe the data collection process are included as part of the data set.

## 3.3 Response Rate

There are numerous ways to calculate response rates in survey research (Groves, 1989; Groves and Lyberg, 1988; Wiseman and Billington, 1984; Frey, 1983; and Dillman, 1978). The method used in this project was conservative: most other ways of calculating response rates would produce higher values. The response rate was defined as the number of completed interviews divided by the estimated number of eligible households times 100 percent. A response rate of 51 percent was obtained for all three sample components. The response rate was 50 and 51 percent respectively for the National and Urban Over Samples and 55 percent for the BC Resource Communities Sample.

Details on the calculation of the response rate for the National Sample are as follows. Of the 10,935 telephone numbers included in the sample, 8,107 were identified as being eligible households (Table 3.2). Not eligible households (respondents were unable to speak English or French, were not healthy enough to complete the interview, nonresidential and not in service numbers, etc.) accounted for 2,663 of the telephone numbers. It was not possible to determine the eligibility status for 195 of the sample telephone numbers as every call attempt resulted in either a ring no answer (RNA) or a busy signal.

For response rate calculations, it was assumed that the proportion of these 195 numbers that were eligible household numbers was the same as it was in the rest of the sample. This proportion, or "household eligibility rate," was .75 (eligibles [8,107]/(eligibles [8,107] + not eligibles [2,663) = .75). The estimated total number of eligibles was then computed as 8,254 (8,107 + [.75 * 195] = 8,254). Dividing the number of completions (4,101) by the estimated number of eligibles (8,254) gives a final response rate of 50 percent.[6]

Table 3.2 Final Sample Disposition for the National ESC Survey

| Result | Number | Percent |
|---|---|---|
| completions | 4,101 | 38 |
| refusals | 3,168 | 29 |
| callbacks | 838 | 8 |
| subtotal eligible households | 8,107 | 74 |
| ill/aged/language problem/absent | 840 | 8 |
| not in service and non-residential | 1,793 | 16 |
| subtotal not eligible | 2,633 | 24 |
| never answered/always busy | 195 | 2 |
| total sample | 10,935 | 100 |

[6] Many survey firms use less strict methods to calculate response rates. A common method is to divide the number of completions over the number of completions plus refusals. Using this procedure the response rate for the survey was 56 percent.

The same procedures were used to calculate the response rates for the other two sample components.

The response rate are somewhat lower than that typically obtained for ISR surveys over the last two to three years. While it is difficult to be certain, most of the lower response rate likely result from the length of the questionnaire (20 minute surveys have higher response rates than surveys that average close to 30 minutes–as did the ESC National survey) and the more general nature of the research endeavour. Interviewers have a more difficult time obtaining cooperation for studies that have a general theme, like social capital, than ones that are more specific, for example, problems with the health care system. In addition, the lower rate is more typical of the declining rates being reported by US researchers over the last two decades (Dunkleberg and Day, 1973; Steech, 1981; Goyder, 1987; and Smith, 1995). *Survey Research*, a Newsletter for researchers in academic-based survey research centres has recently devoted three issues (volumes 29 through 31) to problems of declining response rates. Dillman, who has produced two of the more widely used reference guides to survey research, has cautioned that our expectations of what response rates can be obtained must be brought into line with recent survey results (Dillman, 2000, 8).

In the National Sample there was some variation in response rate by province. The rates were highest in Atlantic Canada (Newfoundland and PEI at 58 percent, Nova Scotia and New Brunswick at 61 and 62 percent respectively) and lowest in the larger and more urban provinces of Quebec (45 percent), Ontario (47 percent) and BC (44 percent). The Prairie provinces tended to be in the middle with rates of 49, 54, and 52 for Manitoba, Saskatchewan and Alberta respectively.

As indicated above, the response rate for the Urban Sample (51 percent) was very close to that of the National Sample (50 percent). But note, more intensive calling and refusal conversion efforts were required to get the same response rate in the Urban Sample as was achieved for the National Sample. The response rate was higher in Montreal (60 percent), than in Toronto and Vancouver (51 and 46 percent respectively).

# 4. DATA PROCESSING

## 4.1 Questionnaires

In CATI, interviewers record respondents' answers directly into a machine readable data file so there is no separate data entry step. In addition, when creating the data set, a number of variables that describe the data collection process (such as the date of the interview, the number of calls, whether or not there was a refusal before completion, the length of the interview, etc.) are attached to the data set.

An "easy-to-read" version of the questionnaire, that includes all of the questions and response categories, as well as an explanation of the way in which the questionnaire was delivered in CATI is appended to this documentation. In addition, a list of variables, grouped by topic area is provided with this documentation.

Questions/variables have been given descriptive names that are different from that used in the CATI questionnaire. For example, question A3 in the original CATI code ("Are you presently married, living with a partner...") has been re-labelled as "mstatus1." The question names used in the easy-to-read questionnaire are the same as the variable names in the data set. The re-naming along with the provision of the easy-to-read questionnaire makes it much easier for the analyst to utilize the data.

The questionnaire used in the BC Resources Communities was nearly identical to that used in the other surveys. However, as noted earlier, the data file for the BC Resource Communities is not combined with the other data and the variables have retained the names/labels used in the CATI coding of the questionnaire. (Upon request, the CATI version of the questionnaire (which includes both English and French text), a data file for the BC Resource Communities and CATI to SPSS cross reference chart can be obtained from ISR (by contacting ISR's data archivist Anne Oram via e-mail at: oram@yorku.ca).

## 4.2 Randomization of Question Order

The logical operators resident in CATI were used to randomize the order in which respondents received items in several sections of the questionnaire. There are two places where question order effects have been routinely found in surveys (Schuman and Presser, 1981, 23-56). First,

when respondents are asked about a list of related issues the answers given to the first question often act as an anchor, or an initial frame of reference point, that help determine the response to subsequent questions. If, for example, respondents were asked to score a number of well- known films on a 100 point scale, their answer to the first film rated will often be considered when they rate the second and subsequent films. So if the first film, A, was scored as a 75 and the second film (B) is one they liked more it will be scored higher, perhaps an 80. However, if the order they were asked about the films was reversed, so they were asked about film B first and they scored it a 75, then when asked to rate film A, they would need to score it lower, perhaps a 70.

Second, when respondents are asked questions that are related and can be seen as having a consistency or fairness aspect, the answer to the first may sway how the respondent answers the second. If we ask respondents if they think government should support professional sports by giving them tax breaks on the salaries they pay their athletes, a respondent may answer in the negative. If we then asked about tax breaks for a popular local team they may feel constrained to answer "no" even if they would indeed support the tax breaks for the local team. The respondent may feel trapped and their response may have as much to do with a need to be fair or consistent in their answers as it does with their opinion or view. Given that order effects have been identified in surveys, but are not always easy to predict (Schuman and Presser, 1981, 24 and Sudman, Bradburn, and Schwarz, 1996, 80) it is sensible, when we have reason to be concerned, to randomize the order of questions as a precautionary measure. Then at the analysis stage, the user can determine what impact, if any, question order had on response.

## 4.21 The Wallet Order Experiment

Respondents were asked to imagine they lost a wallet or purse that contained two hundred dollars and the likelihood it would be returned if it was found by four different people: someone who lived close by, a clerk at the nearest grocery store, a police officer and a complete stranger (see variables/survey items trust_w1 to trust_w4). The order that these four "finders of the wallet" were asked in the interview was controlled by random number 10 (RANDOM10). When RANDOM10 was 1 the order was as indicated above (someone close by, clerk, police and stranger). When RANDOM10 was 2 the order was clerk, police, stranger and someone close by. When 3: police, stranger, someone close by and clerk; and when four: stranger, someone close by, clerk and police officer. Running a cross tabulation (contingency table) of each of the "trust-wallet" questions by RANDOM10 will allow users of the data to determine the order in which the

questions were asked. For example, a crosstab of trust_w1 by RANDOM10 will indicate the extent to which responses to having the wallet found by someone who lived close by (variable trust_w1) varied according to when this question was the first, second, third or fourth of the trust-wallet questions asked.

Like all of the other random numbers used to determine question order or question wording experiments, random number 10, which was used for the wallet-trust questions is a variable in the data set (see RANDOM10). The value of each random number was assigned (randomly) to each case prior to the start of data collection, and like the telephone number read into the CATI file so it could be accessed within the CATI program during the interview.

## 4.22 Economic Security

The complete section of items about economic security (seven questions: econs_1 to econs_7) was asked either relatively early in the questionnaire, after the "trust-wallet" order experiment or very close to the end of the questionnaire immediately before the place of birth questions and the final demographic questions. When RANDOM11 was 1, it was asked early and when it was 2, it was asked near the end of the questionnaire.

## 4.23 Group Orientations

Respondents were asked to rate seven different groups on a 0 to 100 scale. The order in which the groups were asked was determined by the value of RANDOM14. The order in which the seven groups were asked about when RANDOM14 was 1, 2 or 3 is indicated below:

| RANDOM14 = 1 | RANDOM14 = 2 | RANDOM14 = 3 |
|---|---|---|
| unions | business community | *federal govt* |
| business community | *federal govt* | provincial govt |
| *federal govt.* | provincial govt | regional govt |
| provincial govt. | regional govt | the court system |
| regional govt. | the court system | the local police |
| the court system | the local police | unions |
| the local police | unions | the business community |

The order when RANDOM14 was 4, 5, 6, and 7 just follows the pattern indicated above. Each of the seven groups in the list was the first one asked of respondents about one seventh of the time, the second one asked about one seventh of the time, etc. Note the pattern is the same for each

value of the random number, so the federal government item always follows the business community item unless the federal government item is the first one asked. In the same manner, the local police question was last when RANDOM14 was 1, second last when RANDOM14 was 2, third last when RANDOM14 was 3 etc. The same format was used in all of the randomizations used in the study.

## 4.24 Ease of Getting Unemployment Benefits and the Size of the Benefits

The order in which these two questions (unemp_2 and unemp_3) were asked depended on the value of RANDOM4. When it was one respondents were first asked about how easy or hard they thought it was to get unemployment benefits and then asked if they think the size of the benefits are too high or too low. When the RANDOM4 was 2 the order the questions were asked was reversed.

## 4.25 Random Order of Response Categories in the Political Party Identification Questions

Respondents were asked if they identified with a political party at the federal (see question/ variable polpar1) and provincial level (see question/ variable polpar8). Respondents were asked if they thought of themselves as Liberal, Conservative, NDP, Reform (and Bloc Quebecois in Quebec). The order in which the first three parties were read to respondents was independently randomized for both the federal (RANDOM8) and provincial (RANDOM9) party identification questions. That is, respondents were randomly assigned a value for RANDOM8 and RANDOM9, and the value of these two numbers was independent. The order of presentation for each value of RANDOM8 and RANDOM9 was as follows:

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| Liberal | Liberal | Conservative | Conservative | NDP | NDP |
| Conservative | NDP | Liberal | NDP | Conservative | Liberal |
| NDP | Conservative | NDP | Liberal | Liberal | Conservative |

In the data set the variables polpar1 and polpar8 gives the overall response to the questions. The derived variables polpar1A to polpar1F and polpar8A to polpar8F detail the results of the questions for each of the different response option orders.

## 4.3 Randomization of Question Wording

On occasion, variation in wording can help users of survey data better understand the meaning respondents ascribe to words, phrases or the intent of the question. The use of split ballot experiments where respondents are randomly assigned to different questions has a long history in survey research and has been successfully used by a number of researcher on large Canadian data sets (see, for example, Ornstein, 1989; Bryant, Gold, Northrup and Stevenson, 1990; Sniderman, Fletcher, Russell and Tetlock, 1996; Johnston, Blais, Brady and Crête 1992; and Johnston, Blais, Gidengel, and Nevitte, 1996). There were a number of these experiments in the ESC questionnaire.

### 4.31 Use of the List Experiment in Measuring Support for Affirmative Action

There were four versions of the "group list question" and each respondent was asked one version of the question. The first or base version of the question (when RANDOM13 was 1) asked respondents to tell the interviewer how many of the following **four** things made them angry or upset. The four things included: Canadian hockey teams moving to the US, the amount of money we pay in taxes, genetically modified foods, and the weak condition of the Canadian armed forces. When RANDOM13 was 2, respondents were delivered a version of the questionnaire where they were asked about **five** things. The fifth thing inserted into the middle of the list was "special access for Native Canadians to land and resources." In the third and fourth versions of the question there also were five things and four were the same as the base question. The additional issue in the third version of the questionnaire was "giving special consideration in hiring and promoting to minorities" and in the fourth was: "giving special consideration in hiring and promoting to women." By comparing the number of things that upset people or make them angry in the second, third, and fourth version of the question to the first or base version we have one measure of the number of Canadians who are upset or angry by these types of affirmative action. The design of the group list question borrows from the work of Sniderman and Carmines (1997, 41-53).

### 4.32 Views Towards the Canada Pension Plan

There were three versions of this question (see RANDOM3). In the first version (when RANDOM3 was 1) respondents were asked if they thought *Canada* would be better off if the plan was shut down and Canadians were able to invest their own money. In the second version (when

RANDOM3 was 2) respondents were asked if they thought *Canadians* would be better off. In the last version of the question (when RANDOM3 was 3), respondents were asked if *they* would be better off if the plan was shut down and they were able to invest the money themselves.

### 4.33 Access to Health Care

Respondents were assigned to one of two versions of this question. The first version read:

> Which is closer to your own view. ONE, everyone should have equal access to health care, even if that means waiting for treatment, or TWO, if you can afford it you should be able to buy faster access to health care.

The second had just a minor variation in wording and referred to "if you are willing to pay" rather than "if you can afford it." When RANDOM2 was 1 respondents received the first version, when 2 the second.

### 4.34 Inter-Personal Trust

There were four questions in the survey dealing with inter-personal trust. The questions were as follows:

| | |
|---|---|
| trust_1 | Generally speaking, would you say that most people can be trusted or that you can not be too careful in dealing with people? |
| trust_2 | People CAN be trusted until they prove otherwise (agree/disagree). |
| trust_3 | Generally speaking, most people CAN be trusted (agree/disagree). |
| trust_4 | Generally speaking, you can NOT be too careful in dealing with people (agree/disagree). |

On the national version of the questionnaire trust_1 was asked half of the time (when RANDOM5 was 1) and when trust_1 was not asked respondents were either asked trust_3 (when RANDOM6 was 1) or trust _4 (when RANDOM6 was 2). Trust_2 was always asked, but when either trust_3 or trust _4 were asked (when RANDOM5 was 2) the question was asked late in the questionnaire (when RANDOM7 was 2) just before question values_3 (which is just prior to the final section of demographics asked at the end of the survey) or early in the questionnaire when RANDOM7 was 1.

In the Urban Over Sample all four of the questions were asked of every respondent. But the order in which the questions were asked varied and was determined by RANDOM5[7]. When RANDOM5 was 1 the order was trust_1, trust_2, trust_3 and trust _4. The order of the trust questions, in the Urban Over Sample, when RANDOM5 was 2: trust_1, 2, 4, 3 and when it was 3 the order was: trust_3, 4, 1, 2 and when it was 4 the order was trust _4, 3, 1, 2. For all values of RANDOM5 (in all four orders utilized) there was a significant gap between the first two and the last two questions. The first two questions were always asked early in the survey and the last two were always asked quite late the survey just prior to the final demographic questions. In addition, five questions about government policy were always asked between the third and fourth trust questions.

## 4.4 Questions Unique to the Urban Over Sample

Two questions were only asked in the Urban Over Sample: lifesat1 and lifesat2. The first is often know as the Bradburn happiness scale and reads: "All things considered, would you say you are very happy, happy, or not happy at all." The second life satisfaction question, used in the World Values Survey reads: "On a scale from 1 to 10 where 1 means dissatisfied and 10 means satisfied, all things considered, how satisfied are you with life these days?" Both questions were asked at the very end of the survey.

## 4.5 Problems of Question Interpretation

There were two places in the questionnaire where, during the course of creating the data set concerns about how respondents understood the question were uncovered.

### 4.51 Weeks of Employment

Question emp_22 read: "During the past 12 months for how many weeks did you get paid." Interviewers were provided instructions to include vacation pay, maternity leave, illness etc., when the respondent was still employed. This question (from the Statistics Canada Labour Force Survey) was mis-interpreted by some respondents. The intent was to measure the number of weeks the person was employed but some respondents understood the question to ask how

[7] In the national sample RNS had values of 1 and 2. In the Urban Over Sample, it had values of 1, 2, 3, 4. This can be seen by running a crosstab of RNS by sample type.

many works they worked and did not count vacation, maternity leave and other paid forms of leave and did not ask or seek clarification from the interviewer. If the respondent said 46 weeks the interviewer had no reason to doubt the answer. Nevertheless, in reviewing the data and reviewing the questionnaire with interviewers it became clear that the question was sometimes misinterpreted. In the follow-up question (emp_23) number of weeks looking for work the respondent would say "no weeks as I was on paid vacation." Or, in the next question (item emp_24) which asked respondents who were not working and not looking for work what they were doing, and the respondent said "they were on a six week paid vacation". When possible, and based on the respondent's sequence of answers, responses were repaired (about 100 interviews in the combined national and urban sample components).

### 4.52 Voluntary Associations

Respondents were asked a set of questions about the number of associations to which they belonged (volas_1 to volas_6). Specifically, they were asked about service clubs, recreational groups, political groups, youth orientated groups, cultural groups and organizations that help people. Examples were given for each of these types of groups. In question volas_7 respondents were asked if they belonged to any other groups or organizations that they had not been asked about. Respondents answering affirmatively were asked the number of additional groups and then to identify the group (volas_9). When group or organization names were given, it was clear that many belonged in the categories that had already been asked and some of these respondents had answered affirmatively to the initial question. It is not clear if respondents were thinking about other groups or double counting.

## 4.6 Occupational Classification

Respondents who were currently working (including self-employed), respondents who worked in the last 12 months, and respondents who were retired were asked about their current, last or former occupation (emp_13). The description of their occupation, recorded as open-ended text by the interviewer, was coded into a 4-digit occupation category using Statistics Canada's "Standard Occupational Classification, 1980." For example, respondents who described their occupation as a high school teacher were assigned a code of 2733. Those who described their occupation as a homemaker were assigned a value of 9994; those who described their occupation as being a student were assigned 9995, disabled a 9996, retired a 9997, don't know a 9998 and if the respondent refused to answer, or provided an answer that was not codeable, the variable was assigned a 9999.

The codebook for the 1980 occupation classifications is contained in this section. Appended to each occupation is a socio-economic index score. These indices are commonly referred to as "Blishen Scores" and are based on the male labour force population who reported an occupation in the 1981 Canadian Census. The development of the scale is reported in Blishen, Carroll and Moore (1987). (See variable blish81r, where “r” refers to respondent.)

CCDO codes and Blishen scores were also assigned for the respondent’s father’s occupation, when this was provided at question emp_25. (See variable blish81f where “f” stands for father.)

Another well-known socio-economic index was developed by Pineo, Porter and McRoberts (1985), based on the 1971 Canadian Census. This index was updated in 1985 to reflect the 1981 Census. The variables pinporr and pinporf are the Pineo-Porter scores for the respondent and father respectively.

## References


Babbie, Earl. 1992. *The Practice of Survey Research* (6th Edition). Belmont, California: Wadsworth.

Blishen, Bernard, Bill Carroll and Catherine. Moore. 1987. "The 1981 Socioeconomic Index for Occupations in Canada." *Canadian Review of Sociology and Anthropology*, 24 (4), 465-488.

Bryant, Alan, Marc Gold, David A. Northrup and Michael H. Stevenson. 1990. Public Support for the Exclusion of Unconstitutionally Obtained Evidence. *Supreme Court Law Review,* 1, 555-87.

Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley and Sons.

2000. *Mail and Internet Surveys: The Tailored Design Method.* New York: John Wiley and Sons.

Dunkelberg, William C. and George S. Day. 1973. "Nonresponse Bias and Callbacks in Sample Surveys." *Journal of Marketing Research*, 10, 160-168.

Fitzgerald, Robert and Linda Fuller. 1982. "I Hear You Knocking But You Can't Come In: The Effects of Reluctant Respondents and Refusers on Sample Survey Estimates." *Sociological Methods and Research*, 11 (1), 3-32.

Frey, James H. 1983. *Survey Research by Telephone.* Beverly Hills: Sage Publications.

Goyder, John. 1987. *The Silent Minority: Nonrespondents on Sample Surveys*. Boulder, Colorado, Westview Press.

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.

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.

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

Kalton, Graham. 1983. *Introduction to Survey Sampling*. Newbury Park, California: Sage.

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

Lessler, Judith T. and William D. Kalsbeek. 1992. *Nonsampling Errors in Surveys*. New York: John Wiley and Sons.

Ornstein, Michael. 1989. *AIDS in Canada: Knowledge, Behaviour and Attitudes of Adults*. Ottawa, Federal Centre for AIDS at Health and Welfare Canada.

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

Pineo, P., J. Porter and H. McRoberts. 1985. "Revisions of the Pineo-Porter-McRoberts Socio-economic Classification of Occupations for the 1981 Census", *Program for Quantitative Studies in Economics and Population Report*. No. 125. McMaster University.

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.

Sniderman, Paul M., Joseph F. Fletcher, Peter H. Russell and Philip E. Tetlock. 1996. *The Clash of Rights*. New Haven, Conn., Yale University Press.

Sniderman, Paul M. and Edward G. Carmines. 1997. Reaching Beyond Race. Cambridge, Mass., Harvard University Press.

Steech, Charlotte G. 1981. "Trends in Nonresponse Rates, 1952-1979." *Public Opinion Quarterly,* 45 (1) 40-57.

Sudman, Seymour, Norman M. Bradburn and Norbert Schwarz. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco, Jossey-Bass.

*Survey Research* Vol 29-31.

Wiseman, Frederick and Marilyn Billington. 1984. "Comment on a Standard Definition of Response Rates." *Journal of Marketing Research*, 21, 336-338.