Table of Contents

1.0 Introduction

The National Electronic Media Use Survey (NEMUS) was conducted by Statistics Canada between March 6th and April 4th, 1996. The project was sponsored by the Research Services Department, head office, at the Canadian Broadcasting Corporation (CBC). The purpose of the survey was to access the television viewing, radio listening and Internet use habits of Canadians.

This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

Any questions about the data set or its use should be directed to:

Statistics Canada
Jonathan Ellison
Special Surveys Division, Statistics Canada
5th floor, Jean Talon Building
Tunney's Pasture
Ottawa, Ontario K1A 0T6
(613) 951-5882 (phone)
(613) 951-0562 (fax)
ellijon@statcan.ca (e-mail)

Sponsor

Marilyn Goodwin
Senior Research Analyst
CBC Research
1500 Bronson Avenue
Ottawa, Ontario K1G 3J5
(613) 738-6560 (phone)
(613) 738-6858 (fax)
research@ottawa.cbc.ca (e-mail)

IT IS IMPORTANT FOR USERS TO BECOME FAMILIAR WITH THE CONTENTS OF THIS DOCUMENT BEFORE PUBLISHING OR OTHERWISE RELEASING ANY ESTIMATES DERIVED FROM THE MICRODATA FILE OF THE NATIONAL ELECTRONIC MEDIA USE SURVEY (**NEMUS)**.

2.0 Background

The Research Services Department at the Canadian Broadcasting Corporation (CBC), otherwise refered to in this document as the client, conducts an annual media use survey to assess the television viewing and radio listening habits of Canadians.

In 1995, the Special Surveys Division at Statistics Canada was approached to conduct this survey on behalf of the CBC for 1996 as a cost recoverable project, due to disappointing response rates (19 to 21 percent) experienced by the client during previous survey years. The client requested that Special Surveys Division use the questions from previous CBC Annual Media Surveys (AMS) to as great an extent as possible, in order to facilitate data comparability. The client further requested that the 1996 survey be modified to included a section related to the Internet use. Since 1994, the CBC has conducted the AMS survey as a telephone interview whereas during the previous seven years it had been conducted employing a mail out questionnaire.

Data collection commenced the week of March 6th and continued for approximately 4 weeks. The interviewing was done by Operations and Integration Division (OID) at Statistics Canada, using Computer Assisted Telephone Interview (CATI) collection. This survey was conducted using Random Digit Dialling (RDD) technology.

Telephone numbers were selected randomly by computer from all possible telephone numbers used in the provinces of Canada. Selection of sample took into consideration such characteristics as linguistic profile, size of population by province, as well as differences associated with unique large Census Metropolitan areas such as Montreal and Toronto. Residences were contacted, a roster of household members developed and the computer system randomly selected a respondent.

3.0 Objectives

The National Electronic Media Use Survey (NEMUS) is a national survey excluding the Yukon and Northwest Territories. The objectives of the survey were threefold in that it was to determine: 1) television viewing habits and preferences; 2) radio listening habits and preferences; and 3) level of use of the Internet of Canadians.

The goal of the survey was to collect approximately 1000 French and 1000 English responses in order to produce representative national statistics for Canadians 15 years and over. However this was later modified to include only Canadians 18 years and over to assist in the clients need for data comparibility with previously collection periods.

4.0
Concepts and Definitions

Since the National Electronic Media Use Survey NEMUS was conducted over the telephone, terminology that easy to understand was used throughout the questionnaire to avoid long explanations. Some standard concepts and definitions should be used in the analysis and interpretation of this data. The survey questions were designed with these definitions in mind.

1. The NEMUS was designed to gather information on specific as well as common aspects of French and English television viewing and radio listening habits. Service provision for these two markets are somewhat different and are therefore reflected in the survey design. However, where possible common questions were asked of both the English and French respondent.

   A capital E or F in the variable name was used to distiguish questions pertaining the two separate linguistic markets. For example, QE7a was asked only of respondents that stated that they listened to television most often in English which is reflected in the derived variable TV (Language of choice for televison viewing).

   Likewise, the variable RA (Language of choice for radio listening) determines the English/French flow of radio listening questions.

2. Analysts should pay particular attention to the “coverage” comment for each question which is located on the record layout in Chapter 13.

   Note: Appendix 1 contains a flow diagram of the questionnaire.

5.0 Survey Methodology

The National Electronic Media Usage Survey (NEMUS) was administered between March 6 and April 4, 1996 as a random digit dialling (RDD) survey, a technique whereby telephone numbers are generated randomly by computer.

5.1 Population Coverage

The original target population for the NEMUS was all persons 15 years of age and over living in Canada; however, the micro data file only presents those persons 18 years of age and over living in Canada with the following two exceptions:

1. Residents of the Yukon and Northwest Territories; and
2. Full-time residents of institutions.

The modification of the sample population presented in the micro data file was made at the request of the client after collection in order to facilitate data comparibility with that which had been collected employing the CBC Annual Media Survey.

Because the survey was conducted using a sample of telephone numbers, households (and thus persons living in those households) that do not have telephones were excluded from the sample population. People without telephones account for less than 3% of the target population. However, the survey estimates have been weighted to include persons without telephones.

5.2 Sample Design

The desired sample size was about 2,000 respondents across Canada, with the added requirement of 1,000 of the respondents being English-speaking and 1,000 being French-speaking. The initial sample sizes took into account the expected RDD hit rate (proportion of telephone numbers belonging to households) by stratum and the expected response rate.

5.3 Sample Design and Allocation

(Refer to section 5.2)

5.4 Sample Selection

The sample for the NEMUS was generated using a refinement of RDD sampling called the Elimination of Non-Working Banks (ENWB) method. Using ENWB, the first stage in selecting the sample was to attempt to identify all working banks (i.e. all banks with at least one residential telephone number). This set of working banks became the frame for the survey. A bank is defined as the first 8 digits of the 10-digit telephone number (including area code). Thus, all banks with only unassigned, non-working, or business telephone numbers are excluded from the survey frame. The information needed to assemble the frame came from various telephone companies across Canada.

Each working bank was assigned to the proper province-stratum combination. Next, a systematic sample of banks was selected within each stratum. For each selected bank, a two digit number between 00 and 99 was generated at random. The random number was added to the bank to form a complete telephone number. This method allowed listed and unlisted residential numbers, as well as business and non-working phone numbers, to have a chance of being in the sample. At the same time it gave a much higher chance of reaching a residence than would be obtained if the last four digits of the number were randomly generated.

Each telephone number in the sample was dialled to determine whether or not it reached a household. For each household reached, an attempt was made to list all eligible household members and to sample one of these at random.

The random selection was set up such that all people in the household aged 15 or more had an equal probability of selection.

6.0 Data Collection

Data collection for the National Electronic Media Usage Survey used Computer-Assisted Telephone Interviewing (CATI). In a CATI application, the survey questions and response categories are programmed. The interviewer reads the question to the respondent and enters the respondent's answer. Thus, data collection and data capture occur simultaneously. The application is programmed to ensure that only valid answers can be entered, that the proper flow between questions is automatic, that the discrepancies between answers to related questions are passed through an edit and, if necessary, that the respondent is asked to verify or correct the response(s) in error.

6.1 Question Design

The survey was divided into two main sections: initial contact and NEMUS questions. The first section contained questions used to: (a) determine whether the telephone number belonged to a household, (b) determine the probability of that household being selected for the survey (i.e. number of other phone numbers belonging to the same household), (c) list the age, sex and marital status of each household member, and (d) randomly select one of the household members to complete the NEMUS questions.

The client requested that Special Surveys Division use the questions from previous CBC Annual Media Surveys (AMS), to as great an extent as possible, in order to facilitate data comparability.

6.2 Supervision and Control

All CATI interviewers are under the supervision of senior interviewers who are responsible for ensuring that interviewers are familiar with the concepts and procedures of the survey, and also for periodically monitoring their interviewers. Monitoring of the interviewers consisted of the supervisor listening to the telephone interview and watching the responses being entered into the CATI application. Any errors or problems were noted and immediately brought to the attention of the interviewer.

6.3 Data Collection Methodology

The "initial contact" section of the CATI application was used each time a different telephone number was dialled by the interviewer. If the telephone number belonged to a household, the household members were listed in order of age, from oldest to youngest. One household member aged 15 or more was selected at random by the computer.

The NEMUS Survey was then conducted with the selected person. If this selected person was not available to be interviewed at that time, an attempt was made to determine a convenient time to phone back to complete the interview. Because the survey included questions on behaviour, attitudes and knowledge, all interviews had to be conducted with the selected respondent only; no proxy reporting was accepted.

6.4 Collection Period

Interviews were conducted from Statistics Canada's head office (Operations and Integration Division). All interviews took place between 8:30AM and 10:00 PM local time (ie. respondent's time) on weekdays - the exception being between 12:00 noon and 7:00PM on Fridays and Sundays. No interviews took place on Saturdays. The collection period was between March 6th and April 4th, 1996.

7.0 Data Processing

The main output of the National Electronic Media Use Survey is a "clean" microdata file. This section presents a brief summary of the processing steps involved in producing this file.

7.1 Data Capture

No separate data capture step was needed for this survey. Computer-assisted interviewing means that the data collection and capture are combined into a single process.

7.2 Editing

Raw data was collected for 1,931 selected respondents. To accommodate most statistical packages, all blank fields were converted to a numeric value. Questions that were skipped because of a flow pattern in the questionnaire were assigned a code to indicate a "valid skip" or an imputed code was entered where applicable. Responses of "don't know" or "refused" were also assigned specific codes. Any question that was skipped because of a flow pattern associated with an earlier response of "don't know" or "refused" was assigned a value of "not stated", because it is uncertain whether or not the question would have applied to the person.

7.3 Creation of Derived Variables

After all numerical verification was completed, derived variables were created to accommodate user needs. These include demographoic variables as well as language of choice for television viewing and language of choice for radio listening. Derived variables for all parts of question 8 and question 9 are provided at the end of the microdata record layout.

7.4 Weighting

The principle behind estimation in a probability sample such as the NEMUS is that each person in the sample "represents", besides himself or herself, several other persons not in the sample. For example, in a 2% simple random sample of the population, each person in the sample represents 50 persons in the population.

The weighting phase is a step which calculates this number for each record. This weight must be used to derive estimates from the microdata file. For example, if the number of people in Canada who watch television is to be estimated, it is done by selecting the records referring to people with that characteristic (Q3A=1) and summing the weights of those records.

Details of the method used to calculate these weights are presented in Section 11.

7.5 Suppression of Confidential Information

It should be noted that the 'Public Use' microdata files described above differ in a number of important respects from the survey 'master' files held by Statistics Canada. These differences are the result of actions taken to protect the anonymity of individual survey respondents. Users requiring access to information excluded from the microdata files may purchase custom tabulations. Estimates generated will be released to the user, subject to meeting the guidelines for analysis and release outlined in Section 9 of this document.

Geographic Identifiers: The survey master data file includes geographic identifiers for province and stratum (CMA, non-CMA, Toronto, Montreal). The public use microdata file does not contain the identifiers for province and stratum. Because of the small sample size, estimates at the province and sub-provincial level would not be statistically reliable.

8.0 Data Quality

8.1 Response Rates

For the National Electronic Media Usage Survey, 7,995 telephone numbers were called. 4,229 of these were assumed to belong to households and 3,766 belonged to non-households. Of the 4,229 numbers, 831 were called repeatedly, but no contact was made; for the purpose of weighting and calculating response rates, we assume that these numbers belong to households. 1,287 of the households were non-responding because either they refused or the selected respondent could not be reached during the survey collection period. A complete interview was administered to the selected respondent in the remaining 2,011 households. However, due to changes in eligibility requirements (which were made after the survey was conducted), 80 respondents were dropped from the sample. The final sample, of respondents 18 years of age and over, totaled 1,931.

8.2 Survey Errors

The survey produces estimates based on information collected from and about a sample of individuals. Somewhat different estimates might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. as those actually used in the survey. The difference between the estimates obtained from the sample and those resulting from a complete count taken under similar conditions is called the sampling error of the estimate.

Errors which are not related to sampling may occur at almost every phase of a survey operation. Interviewers may misunderstand instructions, respondents may make errors in answering questions, the answers may be incorrectly entered on the computer and errors may be introduced in the processing and tabulation of the data. These are all examples of non-sampling errors.

Over a large number of observations, randomly occurring errors will have little effect on estimates derived from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Considerable time and effort was made to reduce non-sampling errors in the survey. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures included extensive training of interviewers with respect to the survey procedures and CATI application; monitoring of interviewers to detect problems of questionnaire design or

misunderstanding of instructions; and testing of the CATI application to ensure that range checks, edits, imputes and question flow were all programmed correctly.

8.2.1 Total Non-Response

Total non-response can be a major source of non-sampling error in many surveys, depending on the degree to which respondents and non-respondents differ with respect to the characteristics of interest. Total non-response occurred when the selected person could not be contacted or refused to participate in the survey. Total non-response was handled by adjusting the weight of individuals who responded to the survey to compensate for those who did not respond.

8.2.2 Partial Non-Response

Partial non-response to the survey occurred when the respondent refused to answer a question, or could not recall the requested information. Partial non-response is indicated by codes on the microdata file.

8.2.3 Coverage

As mentioned in Section 5.1 (Population Coverage), less than 3% of households in Canada do not have telephones. Individuals living in non-telephone households may have unique characteristics which will not be reflected in the survey estimates. Users should be cautious when analyzing subgroups of the population which have characteristics that may be correlated with non-telephone ownership.

8.2.4 Measures of Sampling Error

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error. The basis for measuring the potential size of sampling errors is the standard error of the estimates derived from survey results. However, because of the large variety of estimates that can be produced from a survey, the standard error of an estimate is usually expressed relative to the estimate to which it pertains. This resulting measure, known as the coefficient of variation (C.V.) of an estimate, is obtained by dividing the standard error of the estimate by the estimate itself and is expressed as a percentage of the estimate.

For example, suppose that, based upon the survey results, one estimates that 33.9% of Canadians who watch television are in the age group 18 - 34, and this estimate is found to have standard error of 2%. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{.0132}{.339} \right) \; x \; 100\% \; ' \; 3.9\%$$

8.3 Data Quality Note - Question 16

Caution should be exercised when using data related to “Interest in Specialty Channels” (QE16a to QE16m and QF16a to QF16h). Estimates of level of interest for the first listed specialty channels may be biased downwards. Data for the following questions (QE16b to QE16m and QF16b to QF16h) were suppressed for confidentiality reasons unrelated to the following discussion.

The order in which these questions were presented to the respondents was as indicated in the record layout and did not vary -- the questions were not rotated. Previous studies on this topic, conducted by other organizations, have variously rotated or not rotated these questions. Rotation refers to a procedure where different respondents are asked the list of questions concerned starting at differing points in the list and working around the list in a circular or random fashion until done.

A simple comparison of the results of these studies suggests there may be a bias associated with not rotating these questions. In particular, levels of interest expressed by respondents about the first specialty channels they were asked about in not rotated situations appear to be lower than those expressed for the same specialty channels in rotated situations. Unfortunately, several effects are confounded in the comparison of these results and so the evidence is inconclusive.

All of the non Statistics Canada studies suffer from possibly very high non-response bias due to non-response rates in excess of 80% (The non-response rate for the Statistics Canada study although much better at 52%, may also have led to non-response bias). There are question wording differences between the various studies. The observed differences also include effects due to temporal changes. There are context differences between the studies; in some these questions were the first ones asked while in others they were not the first questions and were preceded by differing questions.

A more recent study, by another organization, made a direct comparison of rotating these questions or not. In the test, the same questions were used in the same context and at the same point in time. The study covered only English speaking Canadians. The results are strongly suggestive of a bias

affecting the category CBC Newsworld, the first listed specialty channel.. For these two channels statistical tests of significance indicate higher proportions of the response “very interested” in the rotated group than in the not rotated group. This study again suffers from the weakness of a very high non-response rate in excess of 80% and associated non-response bias the magnitude of which is unknown but may be very high.

Using the CV Release Guidelines (see Chapter 9.0), the quality level of estimates from these questions for CBC Newsworld should be reported as marginal or unacceptable -- depending on the estimated coefficient of variation from the Approximate Sampling Variability Table (Chapter 10). Even if the estimated CV is below 16.5%, the quality level should be reported as marginal.

9.0 Guidelines for Tabulation, Analysis and Release

This section of the documentation outlines the guidelines to be adhered to by users tabulating, analysing, publishing or otherwise releasing any data derived from the survey microdata tapes. With the aid of these guidelines, users of microdata should be able to produce the same figures as those produced by Statistics Canada and, at the same time, will be able to develop currently unpublished figures in a manner consistent with these established guidelines.

9.1 Rounding Guidelines

In order that estimates for publication or other release derived from these microdata tapes correspond to those produced by Statistics Canada, users are urged to adhere to the following guidelines regarding the rounding of such estimates:

a) Estimates in the main body of a statistical table are to be rounded to the nearest thousand units using the normal rounding technique. In normal rounding, if the first or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is raised by one. For example, in normal rounding to the nearest 1000, if the last two digits are between 000 and 499, they are changed to 000 and the preceding digit (the thousands digit) is left unchanged. If the last digits are between 500 and 999 they are changed to 000 and the preceding digit is incremented by 1.

b) Marginal sub-totals and totals in statistical tables are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 1000 units using normal rounding.

c) Averages, proportions, rates and percentages are to be computed from unrounded components (i.e. numerators and/or denominators) and then are to be rounded themselves to one decimal using normal rounding. In normal rounding to a single digit, if the final or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is increased by 1.

d) Sums and differences of aggregates (or ratios) are to be derived from their corresponding unrounded components and

then are to be rounded themselves to the nearest 1000 units (or the nearest one decimal) using normal rounding.

e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released which differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s).

f) Under no circumstances are unrounded estimates to be published or otherwise released by users. Unrounded estimates imply greater precision than actually exists.

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the NEMUS was not self-weighting. When producing simple estimates, including the production of ordinary statistical tables, users must apply the proper sampling weight.

If proper weights are not used, the estimates derived from the microdata files cannot be considered to be representative of the survey population, and will not correspond to those produced by Statistics Canada.

Users should also note that some software packages may not allow the generation of estimates that exactly match those available from Statistics Canada, because of their treatment of the weight field.

9.2.1 Definitions of types of estimates: Categorical vs. Quantitative

Before discussing how the NEMUS data can be tabulated and analysed, it is useful to describe the two main types of point estimates of population characteristics which can be generated.

Categorical Estimates

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of people who watch television and the proportion of people who listen to the radio are examples of such estimates. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Quantitative Estimates

An example of a quantitative estimate is the total number of hours spent watching television in a typical week. Another is the average number of hours spent watching television in a typical week. For this average, the numerator is an estimate of the total number of hours spent watching television (by those people who answered yes to watching television), and its denominator is the number of persons who do watch television.

9.2.2 Tabulation of Categorical Estimates

Estimates of the number of people with a certain characteristic can be obtained from the microdata file by summing the final weights of all records possessing the characteristic(s) of interest. Proportions and ratios of the form X/Y are obtained by:

(a) summing the final weights of records having the characteristic of interest for the numerator (X),
(b) summing the final weights of records having the characteristic of interest for the denominator (Y), then
(c) dividing the numerator estimate by the denominator estimate.

9.2.3 Tabulation of Quantitative Estimates

Estimates of quantities can be obtained from the microdata file by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest. For example, to obtain an estimate of total number of hours spent watching television in a typical week, multiply the value reported in Q3B (number of hours spent watching television in a typical week) by the final weight for the record, then sum this value over all records with Q3A=1 (all respondents who reported watching television).

9.3 Guidelines for Statistical Analysis

The NEMUS is based upon a complex design, with stratification and multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used.

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures differ from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced

by the packages are correct, the variances that are calculated are almost meaningless.

For many analysis techniques (for example linear regression, logistic regression, analysis of variance), a method exists which can make the application of standard packages more meaningful. If the weights on the records are rescaled so that the average weight is one (1), then the results produced by the standard packages will be more reasonable; they still will not take into account the stratification and clustering of the sample's design, but they will take into account the unequal probabilities of selection. The rescaling can be accomplished by dividing each weight by the overall average weight before the analysis is conducted.

In order to provide a means of assessing the quality of tabulated estimates, Statistics Canada has produced a set of Approximate Sampling Variability Tables (commonly referred to as "C.V. Tables") for the NEMUS. These tables can be used to obtain approximate coefficients of variation for categorical-type estimates and proportions. See Section 10 for more details.

9.4 C.V. Release Guidelines

Before releasing and/or publishing any estimate from these microdata tapes, users should first determine the number of respondents who contribute to the calculation of the estimate. If this number is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the **rounded** estimate and follow the guidelines below.

Sampling Variability Guidelines

Type of Estimate	C.V. (in %)	Guidelines
1. Unqualified	0.0 - 16.5	Estimates can be considered for general unrestricted release. Requires no special notation.
2. Qualified	16.6 - 25.0	Estimates can be considered for general unrestricted release but should be accompanied by a warning cautioning subsequent users of the high sampling variability associated with the estimates. Such estimates should be identified by the letter Q (or in some other similar fashion).
3. Restricted	25.1 - 33.3	Estimates can be considered for general unrestricted release only when sampling variabilities are obtained using an exact variance calculation procedure. Unless exact variances are obtained, such estimates should be deleted and replaced by dashes (---) in statistical tables.
4. Not for Release	33.4 or greater	Estimates cannot be released in any form under any release OR circumstances. In statistical tables, such estimates should be deleted and replaced by dashes (--).

10.0 Approximate Sampling Variability Tables

In order to supply coefficients of variation which would be applicable to a wide variety of categorical estimates produced from this microdata file and which could be readily accessed by the user, a set of Approximate Sampling Variability Tables has been produced. These "look-up" tables allow the user to obtain an approximate coefficient of variation based on the size of the estimate calculated from the survey data.

The coefficients of variation (C.V.) are derived using the variance formula for simple random sampling and incorporating a factor which reflects the multi-stage, clustered nature of the sample design. This factor, known as the design effect, was determined by first calculating design effects for a wide range of characteristics and then choosing from among these a conservative value to be used in the look-up tables which would then apply to the entire set of characteristics.

The table on the next page shows the design effects, sample sizes and population counts which were used to produce the Approximate Sampling Variability Tables. The population numbers are the revised demographic

estimates based on the 1991 Census, including non-permanent residents; these Census counts are projected forward using data on births, deaths and migration. The population counts in the table are the demography projections for the population 18 years of age and over, for March 1996, the reference month for the survey weights.

Table of Design Effects

	Design Effect	Sample Size	Population
Canada	1.44	1,933	22,292,718

All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Estimates of actual variance for specific variables may be obtained from Statistics Canada on a cost-recovery basis. The use of actual variance estimates would allow users to release otherwise unreleaseable estimates, i.e. estimates with coefficients of variation in the 'confidential' range.

Remember: If the number of observations on which an estimate is based is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. This is because the formulas used for estimating the variance do not hold true for small sample sizes.

10.1 How to use the C.V. tables for Categorical Estimates

The following rules should enable the user to determine the approximate coefficients of variation from the Sampling Variability Tables for estimates of the number, proportion or percentage of the surveyed population possessing a certain characteristic and for ratios and differences between such estimates.

Rule 1: Estimates of Numbers Possessing a Characteristic (Aggregates)

The coefficient of variation depends only on the size of the estimate itself. On the Sampling Variability Table for the appropriate geographic area, locate the estimated number in the left-most column of the table (headed "Numerator of Percentage") and follow the asterisks (if any) across to the first figure encountered. This figure is the approximate coefficient of variation.

Rule 2: Estimates of Proportions or Percentages Possessing a Characteristic

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the numerator of the proportion or percentage. Estimated proportions or percentages are

relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, the proportion of people in Canada who watch television is more reliable than the estimated number of people in Canada who watch television. (Note that in the tables the coefficients of variation decline in value reading from left to right.)

When the proportion or percentage is based upon the total population of the geographic area covered by the table (i.e. if the denominator is equal to the total population), the coefficient of variation of the proportion or percentage is the same as the coefficient of variation of the numerator of the proportion or percentage. In this case, Rule 1 can be used.

When the proportion or percentage is based upon a subset of the total population (e.g. those in a particular sex or age group), reference should be made to the proportion or percentage (across the top of the table) and to the numerator of the proportion or percentage (down the left side of the table). The intersection of the appropriate row and column gives the coefficient of variation.

Rule 3: Estimates of Differences Between Aggregates or Percentages

The standard error of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard error considered separately. That is, the standard error of a difference ($\hat{d} = \hat{X}_1 - \hat{X}_2$) is:

$$s_{\hat{d}} = \sqrt{(\hat{X}_1 a_1)^2 + (\hat{X}_2 a_2)^2}$$

where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, and a_1 and a_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{d}$ is given by $s_{\hat{d}}/\hat{d}$. This formula is accurate for the difference between separate and uncorrelated characteristics, but is only approximate otherwise.

Rule 4: Estimates of Ratios

A ratio is an estimate taking the form $\hat{R} = \hat{X}_1/\hat{X}_2$ where $\hat{X}_1$ and $\hat{X}_2$ are both quantities estimated from the survey. In the case where the numerator ($\hat{X}_1$) is a subset of the denominator ($\hat{X}_2$), the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example, to the case where the denominator is the number of people who watch television and the numerator is the number of those people who watch television for 20 hours or more per week.

The ratio of the number of females who watch television for 20 or more hours per week as compared to the number of males who watch television for 20 or more hours per week is an example where the numerator is not a subset of the denominator. In this case, the standard deviation of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by R. That is, the standard error of a ratio ($\hat{R} = \hat{X}_1/\hat{X}_2$) is:

$$s_{\hat{R}} = \hat{R}\sqrt{a_1^2 + a_2^2}$$

where a_1 and a_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{R}$ is given by $s_{\hat{R}}/\hat{R}$. The formula will tend to overstate the error, if $\hat{X}_1$ and $\hat{X}_2$ are positively correlated and understate the error if $\hat{X}_1$ and $\hat{X}_2$ are negatively correlated.

Rule 5: Estimates of Differences of Ratios

In this case, Rules 3 and 4 are combined. The cv's for the two ratios are first determined using Rule 4, and then the cv of their difference is found using Rule 3.

10.1.1 Examples of using the C.V. tables for Categorical Estimates

The following 'real life' examples are included to assist users in applying the foregoing rules.

Example 1: Estimates of Numbers Possessing a Characteristic (Aggregates)

Suppose that a user estimates that 21,308,228 people in Canada watch television. How does the user determine the coefficient of variation of this estimate?

(1) Refer to the c.v. table for CANADA.

(2) The estimated aggregate (21,303,228) does not appear in the left-hand column (the 'Numerator of Percentage' column), so it is necessary to use the figure closest to it, namely 20,000,000.

(3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 0.9%.

(4) So the approximate coefficient of variation of the estimate is 0.9%. The finding that there were 21,308,000 people (after rounding) in Canada who watch television is publishable with no qualifications.

Example 2: Estimates of Proportions or Percentages Possessing a Characteristic

Suppose that the user estimates that 7,225,201/21,308,228 = 33.9% of Canadians who watch television are in the age group 18-34. How does the user determine the coefficient of variation of this estimate?

(1) Refer to the c.v. table for CANADA.

(2) Because the estimate is a percentage which is based on a subset of the total population (Canadians who watch television), it is necessary to use both the percentage (33.9%) and the numerator portion of the percentage (7,225,201) in determining the coefficient of variation.

(3) The numerator, 7,225,201 does not appear in the left-hand column (the 'Numerator of Percentage' column) so it is necessary to use the figure closet to it, namely 7,00,000. Similarly, the percentage estimate

does not appear as any of the column headings, so it is necessary to use the figure closest to it, 35.0%.

(4) The figure at the intersection of the row and column used, namely 3.9%, is the coefficient of variation to be used.

(5) So the approximate coefficient of variation of the estimate is 3.9%. The finding that 33.9% of Canadians who watch television are in the age group 18-34 can be published with no qualifications.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 4,447,395/11,345,699= 39.2% of females said they watched 20 or more hours of television per week, while 3,630,394/10,947,019 = 33.2% of males said they watched 20 or more hours of television per week. How does the user determine the coefficient of variation of the difference between these two estimates?

(1) Using the c.v. table for CANADA in the same manner as described in Example 2 gives the c.v. of the estimate for females as 5.0%, and the c.v. of the estimate for males as 5.2%.

(2) Using Rule 3, the standard error of a difference ($\hat{d} = \hat{X}_1 - \hat{X}_2$) is:

$$s_{\hat{d}} = \sqrt{(\hat{X}_1 a_1)^2 + (\hat{X}_2 a_2)^2}$$

where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, and a_1 and a_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

That is, the standard error of the difference $\hat{d} = (.392-.332) = .060$ is:

$$s_{\hat{d}} = \sqrt{[(.392)(.050)]^2 + [(.332)(.052)]^2}$$

$$= \sqrt{(.00038) + (.00029)}$$

$$= .026$$

(3) The coefficient of variation of $\hat{d}$ is given by $s_{\hat{d}}/\hat{d} = .026/.060 = 0.433$.

(4) So the approximate coefficient of variation of the difference between the estimates is 43.3%. This estimate can be not be released under any

circumstances and should be deleted and replaced by dashes.

Example 4: Estimates of Ratios

Suppose that the user estimates that 13,279,900 Canadians are satisfied or very satisfied with their cable service. while 21,363,392 Canadians are satisfied or very satisfied with their telephone service. The user is interested in comparing the estimats in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

(1) First of all, this estimate is a ratio estimate, where the numerator of the estimate ($=\hat{X}_1$) is the number of Canadians who are satisified or very satisfied with their cable service. The denominator of the estimate ($=\hat{X}_2$) is the number of Canadians who are satisfied or very satisfied with their telephone service.

(2) Refer to the c.v. table for CANADA.

(3) The numerator of this ratio estimate is 13,279,900. The figure closest to it is 12,500,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 2.0%.

(4) The denominator of this ratio estimate is 21,363,392. The figure closest to it is 20,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 0.9%.

(5) So the approximate coefficient of variation of the ratio estimate is given by Rule 4, which is,

$$a_{\hat{R}} \doteq \sqrt{a_1^2 + a_2^2}$$

where a_1 and a_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

That is,

$$a_{\hat{R}} \doteq \sqrt{(.020)^2 + (.009)^2}$$

$$\doteq 0.022$$

The obtained ratio of Canadians satisfied or very satisfied with their cable service versus their telephone service is 13,279,900/

21,363,392, which is 1:1.6. The coefficient of variation of this estimate is 2.2%, which is releasable with no qualifications.

10.2 How to use the C.V. tables to obtain Confidence Limits

Although coefficients of variation are widely used, a more intuitively meaningful measure of sampling error is the confidence interval of an estimate. A confidence interval constitutes a statement on the level of confidence that the true value for the population lies within a specified range of values. For example a 95% confidence interval can be described as follows:

If sampling of the population is repeated indefinitely, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.

Using the standard error of an estimate, confidence intervals for estimates may be obtained under the assumption that under repeated sampling of the population, the various estimates obtained for a population characteristic are normally distributed about the true population value. Under this assumption, the chances are about 68 out of 100 that the difference between a sample estimate and the true population value would be less than one standard error, about 95 out of 100 that the difference would be less than two standard errors, and about 99 out 100 that the differences would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

Confidence intervals for an estimate, $\hat{X}$, are generally expressed as two numbers, one below the estimate and one above the estimate, as ($\hat{X}$-k, $\hat{X}$+k) where k is determined depending upon the level of confidence desired and the sampling error of the estimate.

Confidence intervals for an estimate can be calculated directly from the Approximate Sampling Variability Tables by first determining from the appropriate table the coefficient of variation of the estimate $\hat{X}$, and then using the following formula to convert to a confidence interval CI:

$$CI_X = [\hat{X} - t\hat{X}a_{\hat{X}}, \hat{X} + t\hat{X}a_{\hat{X}}]$$

where $a_{\hat{X}}$ is the determined coefficient of variation of $\hat{X}$, and

t = 1 if a 68% confidence interval is desired
t = 1.6 if a 90% confidence interval is desired
t = 2 if a 95% confidence interval is desired
t = 3 if a 99% confidence interval is desired.

Note: Release guidelines which apply to the estimate also apply to the confidence interval. For example, if the estimate is not releasable, then the confidence interval is not releasable either.

10.2.1 Example of using the C.V. tables to obtain confidence limits

A 95% confidence interval for the estimated proportion of Canadians who watch television who are in the age group 18-34 (from Example 2) would be calculated as follows.

$\hat{X}$ = 33.9% (or expressed as a proportion = .339)

t = 2

$a_{\hat{X}}$ = 3.9% (.039 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

CI_X = {.339 - (2) (.339) (.039), .339 + (2) (.339) (.039)}

CI_X = {.339 - .026, .339 + .026}

CI_X = {.313, .365}

With 95% confidence it can be said that between 31.3% and 36.5% of Canadians who watch television were between the ages of 18 and 34.

10.3 How to use the C.V. tables to do a t-test

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The sample estimates can be numbers, averages, percentages, ratios, etc. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

Let $\mathbf{X_1}$ and $\mathbf{X_2}$ be sample estimates for 2 characteristics of interest. Let the standard error on the difference $\hat{X}_1 - \hat{X}_2$ be $s_{\hat{d}}$.

If $t = \frac{\hat{X}_1 - \hat{X}_2}{s_{\hat{d}}}$ is between -2 and 2, then no conclusion about the difference between the characteristics is justified at

the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. That is to say that the characteristics are significant.

10.3.1 Example of using the C.V. tables to do a t-test

Let us suppose we wish to test, at 5% level of significance, the hypothesis that there is a difference between the proportion of males and females who watch 20 or more hours of television per week. From Example 3, the standard error of the difference between these two estimates was found to be =.026. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{s_{\hat{d}}} = \frac{.392 - .332}{.026} = \frac{.060}{.026} = 2.31$$

Since t = 2.31 is more than 2, it must be concluded that there is a significant difference between the two estimates at the 0.05 level of significance.

10.4 Coefficients of Variation for Quantitative Estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since most of the variables for the NEMUS are primarily categorical in nature, this has not been done.

As a general rule, however, the coefficient of variation of a quantitative total will be larger than the coefficient of variation of the corresponding category estimate (i.e., the estimate of the number of persons contributing to the quantitative estimate). If the corresponding category estimate is not releasable, the quantitative estimate will not be either.

Coefficients of variation of such estimates can be derived as required for a specific estimate using a technique known as pseudo replication. This involves dividing the records on the microdata files into subgroups (or replicates) and determining the variation in the estimate from replicate to replicate. Users wishing to derive coefficients of variation for quantitative estimates may contact Statistics Canada for advice on the allocation of records to appropriate replicates and the formulae to be used in these calculations.

10.5 Release cut-off's for the NEMUS

The minimum size of the estimate at the provincial, regional and Canada levels are specified in the table below. Estimates smaller than the minimum size given in the "Not Releasable" column may not be released under any circumstances.

Table of Release Cut-Offs

	Unqualified	Qualified	Restricted	Not releasable
CANADA	593,500 and higher	262,500 to 593,500	149,000 to 262,500	under 149,000

10.6 C.V. Tables

The C.V. tables to be used for the analysis of data from the NEMUS are given on the following pages.

NATIONAL ELECTRONIC MEDIA USAGE SURVEY
Approximate Sampling Variability Tables for CANADA

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE												
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	407.3	405.5	403.4	397.2	386.6	375.7	364.5	352.9	340.9	328.5	315.6	288.1	223.2	128.9
2	288.0	286.7	285.3	280.9	273.4	265.7	257.7	249.5	241.1	232.3	223.2	203.8	157.8	91.1
3	235.2	234.1	232.9	229.3	223.2	216.9	210.4	203.8	196.8	189.7	182.2	166.4	128.9	74.4
4	203.6	202.7	201.7	198.6	193.3	187.8	182.2	176.5	170.5	164.3	157.8	144.1	111.6	64.4
5	182.1	181.3	180.4	177.6	172.9	168.0	163.0	157.8	152.5	146.9	141.2	128.9	99.8	57.6
6	166.3	165.5	164.7	162.1	157.8	153.4	148.8	144.1	139.2	134.1	128.9	117.6	91.1	52.6
7	153.9	153.2	152.5	150.1	146.1	142.0	137.8	133.4	128.9	124.2	119.3	108.9	84.4	48.7
8	144.0	143.4	142.6	140.4	136.7	132.8	128.9	124.8	120.5	116.2	111.6	101.9	78.9	45.6
9	135.8	135.2	134.5	132.4	128.9	125.2	121.5	117.6	113.6	109.5	105.2	96.0	74.4	43.0
10	128.8	128.2	127.6	125.6	122.3	118.8	115.3	111.6	107.8	103.9	99.8	91.1	70.6	40.8
11	122.8	122.3	121.6	119.8	116.6	113.3	109.9	106.4	102.8	99.1	95.2	86.9	67.3	38.9
12	117.6	117.0	116.5	114.7	111.6	108.5	105.2	101.9	98.4	94.8	91.1	83.2	64.4	37.2
13	113.0	112.5	111.9	110.2	107.2	104.2	101.1	97.9	94.6	91.1	87.5	79.9	61.9	35.7
14	108.9	108.4	107.8	106.2	103.3	100.4	97.4	94.3	91.1	87.8	84.4	77.0	59.7	34.4
15	105.2	104.7	104.2	102.6	99.8	97.0	94.1	91.1	88.0	84.8	81.5	74.4	57.6	33.3
16	101.8	101.4	100.9	99.3	96.6	93.9	91.1	88.2	85.2	82.1	78.9	72.0	55.8	32.2
17	98.8	98.3	97.8	96.3	93.8	91.1	88.4	85.6	82.7	79.7	76.6	69.9	54.1	31.3
18	96.0	95.6	95.1	93.6	91.1	88.6	85.9	83.2	80.4	77.4	74.4	67.9	52.6	30.4
19	93.4	93.0	92.5	91.1	88.7	86.2	83.6	81.0	78.2	75.4	72.4	66.1	51.2	29.6
20	91.1	90.7	90.2	88.8	86.4	84.0	81.5	78.9	76.2	73.5	70.6	64.4	49.9	28.8
21	88.9	88.5	88.0	86.7	84.4	82.0	79.5	77.0	74.4	71.7	68.9	62.9	48.7	28.1
22	86.8	86.4	86.0	84.7	82.4	80.1	77.7	75.2	72.7	70.0	67.3	61.4	47.6	27.5
23	********	84.5	84.1	82.8	80.6	78.3	76.0	73.6	71.1	68.5	65.8	60.1	46.5	26.9
24	********	82.8	82.3	81.1	78.9	76.7	74.4	72.0	69.6	67.1	64.4	58.8	45.6	26.3
25	********	81.1	80.7	79.4	77.3	75.1	72.9	70.6	68.2	65.7	63.1	57.6	44.6	25.8
30	********	74.0	73.7	72.5	70.6	68.6	66.5	64.4	62.2	60.0	57.6	52.6	40.8	23.5
35	********	68.5	68.2	67.1	65.3	63.5	61.6	59.7	57.6	55.5	53.4	48.7	37.7	21.8
40	********	64.1	63.8	62.8	61.1	59.4	57.6	55.8	53.9	51.9	49.9	45.6	35.3	20.4
45	********	60.4	60.1	59.2	57.6	56.0	54.3	52.6	50.8	49.0	47.1	43.0	33.3	19.2

50	********	57.3	57.1	56.2	54.7	53.1	51.5	49.9	48.2	46.5	44.6	40.8
31.6	18.2											
55	********	54.7	54.4	53.6	52.1	50.7	49.1	47.6	46.0	44.3	42.6	38.9
30.1	17.4											
60	********	52.3	52.1	51.3	49.9	48.5	47.1	45.6	44.0	42.4	40.8	37.2
28.8	16.6											
65	********	50.3	50.0	49.3	48.0	46.6	45.2	43.8	42.3	40.8	39.2	35.7
27.7	16.0											
70	********	48.5	48.2	47.5	46.2	44.9	43.6	42.2	40.8	39.3	37.7	34.4
26.7	15.4											
75	********	46.8	46.6	45.9	44.6	43.4	42.1	40.8	39.4	37.9	36.4	33.3
25.8	14.9											
80	********	45.3	45.1	44.4	43.2	42.0	40.8	39.5	38.1	36.7	35.3	32.2
25.0	14.4											
85	********	44.0	43.8	43.1	41.9	40.8	39.5	38.3	37.0	35.6	34.2	31.3
24.2	14.0											
90	********	42.7	42.5	41.9	40.8	39.6	38.4	37.2	35.9	34.6	33.3	30.4
23.5	13.6											
95	********	41.6	41.4	40.8	39.7	38.5	37.4	36.2	35.0	33.7	32.4	29.6
22.9	13.2											
100	********	40.5	40.3	39.7	38.7	37.6	36.4	35.3	34.1	32.9	31.6	28.8
22.3	12.9											
125	********	36.3	36.1	35.5	34.6	33.6	32.6	31.6	30.5	29.4	28.2	25.8
20.0	11.5											
150	********	33.1	32.9	32.4	31.6	30.7	29.8	28.8	27.8	26.8	25.8	23.5
18.2	10.5											
200	********	28.7	28.5	28.1	27.3	26.6	25.8	25.0	24.1	23.2	22.3	20.4
15.8	9.1											
250	**********	25.5	25.1	24.5	23.8	23.1	22.3	21.6	20.8	20.0	18.2	
14.1	8.2											
300	**********	23.3	22.9	22.3	21.7	21.0	20.4	19.7	19.0	18.2	16.6	
12.9	7.4											
350	**********	21.6	21.2	20.7	20.1	19.5	18.9	18.2	17.6	16.9	15.4	
11.9	6.9											
400	**********	20.2	19.9	19.3	18.8	18.2	17.6	17.0	16.4	15.8	14.4	
11.2	6.4											
450	**********	18.7	18.2	17.7	17.2	16.6	16.1	15.5	14.9	13.6		
10.5	6.1											
500	**********	17.8	17.3	16.8	16.3	15.8	15.2	14.7	14.1	12.9		
10.0	5.8											
750	**********	14.5	14.1	13.7	13.3	12.9	12.4	12.0	11.5	10.5		
8.2	4.7											
1000	**********	12.6	12.2	11.9	11.5	11.2	10.8	10.4	10.0	9.1		
7.1	4.1											
1500	**********	10.0	9.7	9.4	9.1	8.8	8.5	8.2	7.4			
5.8	3.3											
2000	**********	8.6	8.4	8.2	7.9	7.6	7.3	7.1	6.4			
5.0	2.9											
3000	**********	6.9	6.7	6.4	6.2	6.0	5.8	5.3				
4.1	2.4											
4000	**********	5.8	5.6	5.4	5.2	5.0	4.6					
3.5	2.0											
5000	**********	5.0	4.8	4.6	4.5	4.1						
3.2	1.8											
6000	**********	4.4	4.2	4.1	3.7							
2.9	1.7											
7000	**********	3.9	3.8	3.4								
2.7	1.5											
8000	**********	3.5	3.2									
2.5	1.4											
9000	**********	3.0										
2.4	1.4											
10000	**********	2.9										
2.2	1.3											
12500	**********											
2.0	1.2											

```
  15000     ********************************************************************************************************
   1.8     1.1
                              2         0         0         0         0
**************************************************************************************************************
0.9
```

NOTE: FOR CORRECT USAGE OF THESE TABLES PLEASE REFER TO MICRODATA DOCUMENTATION

11.0 Weighting

For the microdata file, statistical weights were placed on each record to represent the number of sampled persons that the record represents.

The weighting for the National Electronic Media Usage Survey consisted of several steps: calculation of a basic weight, an adjustment for non-response, an adjustment for selecting one person in the household, and finally, an adjustment for post-stratification to Sex-Age Group population totals at the Canada level.

11.1 Weighting Procedures for the NEMUS

Basic Weight

With the Elimination of Non-Working Banks method of RDD, each telephone number within a Province-Stratum has an equal probability of selection. This probability is equal to T_s/T_p, where T_s= number of telephone numbers sampled in the Province-Stratum, and T_p= number of possible telephone numbers in the Province-Stratum. Note that the number of possible telephone numbers for a Province-Stratum is equal to the number of working banks for that Province-Stratum multiplied by 100. Each telephone number in the sample was assigned a basic weight equal to the inverse of its probability of selection. That is, for a telephone number in Province-Stratum j, the basic weight W1 is defined as:

$$W1(j) = \frac{T_p(j)}{T_s(j)}$$

Non-Response Adjustment

After calculating the basic weight, all telephone numbers corresponding to non-households were dropped from the sample. For the remaining telephone numbers, the basic weights of households which responded were adjusted to represent non-responding households. Note that telephone numbers which were not resolved (i.e. not determined if they belonged to a household or not) were assumed to be non-responding households in the weighting. A household was considered as responding if the selected person responded to the interview. The non-response adjustment factor for a household in Province-Stratum j was calculated as H(j)/RH(j), where H(j) = {sum of the basic weights of all households in j}, and RH(j) = {sum of the basic weights of the responding households in j}. The non-response adjusted weight W2 was calculated by multiplying the basic weight for the responding households by the non-response adjustment factor:

$$W2(j) = W1(j) \left(\frac{H(j)}{RH(j)} \right.$$

Non-responding households were then dropped from further weighting procedures.

Multilple Telephone Adjustment

Weights for households with more than one (different) telephone number were adjusted downwards to account for the fact that such households have a higher probability of being selected. The weight for each household was divided by the number of distinct residential telephone numbers that serviced the household (NTel). That is, W4 = W3/NTel.

Selected Person Weight

A person weight was then calculated for each person who responded to the survey. The probability of an individual i being selected in a given household was 1/HHSIZE. The inverse of the probability of selection (that is, HHSIZE) was used as the selected person adjustment factor. The initial person weight (W5) was obtained by multiplying W4 by the selected person adjustment factor.

Canada-Sex-Age Group Adjustments

The next weighting step was to adjust the weights to agree with Census projected Canada-Sex-Age Group distributions. Counts were obtained for March 1996 for the following age groups: 18-34, 35-49, 50-64, 65+. The adjustment factor for Canada-Sex-Age Group k was defined as Ck/Sk, where Ck = {Census population projection for CSA k}, and Sk = {sum of weights W5 for persons in the sample in CSA k}. The adjusted weight W6 was calculated as W5 multiplied by the adjustment factor.

It should be noted that persons living in households without telephone service are included in these projections even though such persons were not sampled.

Calibration Estimation Adjustments

The weights for each respondent were adjusted by an iterative process using a calibrated estimation procedure. This procedure ensured that estimates produced for a Canada-Sex-Age Group would agree with the population totals for that Canada-Sex-Age Group. This adjustment was made by using a two-stage iterative weighting procedure, each time using the weight obtained from the previous step, until the set of estimates agreed with the Census projections. The final statistical weight can be found in the "WEIGHT" field on the microdata file. Note that this field has an implied decimal and should be read as (99999999V9999) where V represents the location of the implied decimal place.

12.0 Questionnaire

13.0
Record Layout and Univariates