Microdata User Guide

CANADIAN TOBACCO USE MONITORING SURVEY

CYCLE 1

FEBRUARY - JUNE 2003

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The Canadian Tobacco Use Monitoring Survey (CTUMS) was conducted by Statistics Canada from February to June 2003 with the cooperation and support of Health Canada. 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

Client Services
Special Surveys Division
Telephone: (613) 951-3321 or call toll-free 1 800 461-9050
Fax: (613) 951-4527
E-mail: ssd@statcan.ca

Elizabeth Majewski
Special Surveys Division
2nd floor, Main Building, Tunney's Pasture
Ottawa, Ontario K1A 0T6
Telephone: (613) 951-4584
Fax: (613) 951-0562
E-mail: elizabeth.majewski@statcan.ca

Health Canada

Murray Kaiserman
Office of Research, Surveillance and Evaluation
Tobacco Control Programme
Healthy Environments & Consumer Safety Branch
MacDonald Building, AL 3507C
123 Slater Street, Room A723
Ottawa, Ontario K1A OK9
Telephone: (613) 954-5851
Fax: (613) 954-2292
E-mail: Murray_Kaiserman@hc-sc.gc.ca

Judy Snider
Office of Research, Surveillance and Evaluation
Tobacco Control Programme
Healthy Environments & Consumer Safety Branch
MacDonald Building, AL 3507C
123 Slater Street, Room A716
Ottawa, Ontario K1A OK9
Telephone: (613) 957-0697
Fax: (613) 954-2292
E-mail: Judy_Snider@hc-sc.gc.ca

2.0 Background

Statistics Canada has conducted smoking surveys on an ad hoc basis on behalf of Health Canada since the 1960s. These surveys have been done as supplements to the Canadian Labour Force Survey and as random digit dialing telephone surveys.

In February 1994, a change in legislation was passed which allowed a reduction in cigarette taxes. Since there was no survey data from immediately before this legislative change, it was difficult for Health Canada or other interested analysts to measure exactly the impact of the change.

As Health Canada wants to be able to monitor the consequences of legislative changes and anti-smoking policies on smoking behaviour, the Canadian Tobacco Use Monitoring Survey (CTUMS) was designed to provide Health Canada and its partners/stakeholders with continual and reliable data on tobacco use and related issues.

Since 1999, two CTUMS files have been released every year: a file with data collected from February to June and a file with the July to December data. Additionally, there is also a yearly summary. The present file covers the period from February to June 2003.

3.0 Objectives

The primary objective of the survey is to provide a continuous supply of smoking prevalence data against which changes in prevalence can be monitored. This objective differs from that of the National Population Health Survey (NPHS) which collects smoking data from a longitudinal sample to measure which individuals are changing their smoking behaviour, the possible factors which contribute to change, and the possible risk factors related to starting smoking and smoking duration. Because the NPHS collects data every two years and releases the data about a year after completing the collection cycle, it does not meet Health Canada's need for continuous coverage in time, rapid delivery of data, or sufficient detail of the most at-risk populations, namely 15 to 24 year olds.

The Canadian Tobacco Use Monitoring Survey allows Health Canada to look at smoking prevalence by province-sex-age group, for age groups 15 to 19, 20 to 24, 25 to 34, 35 to 44 and 45 and over, on a semi-annual and annual basis. Data will continue to be collected on an on-going basis depending on availability of funds.

4.0 Concepts and Definitions

Since the Canadian Tobacco Use Monitoring Survey is conducted over the telephone, easy to understand terminology is 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.

Current Smoking Status

1) Daily smoker: A person who currently smokes cigarettes every day.

2) Non-daily smoker: A person who currently smokes cigarettes, but not every day.

3) Non-smoker: A person who currently does not smoke cigarettes.

4) Current smoker: A person who currently smokes cigarettes daily or occasionally.

Smoking History

1) Former smoker: A person who has smoked at least 100 cigarettes in his life, but currently does not smoke.

2) Experimental smoker: A person who has smoked at least one cigarette, but less than 100 cigarettes, and currently does not smoke cigarettes.

3) Lifetime abstainer: A person who has never smoked cigarettes at all.

4) Ever smoker: A person who is a current smoker or a former smoker.

5) Never smoker: A person who was an experimental smoker or who is a lifetime abstainer.

Smoking Prevalence

Proportion of population which smokes cigarettes at the current time.

True age versus roster age

Information about the respondent's age is obtained from two sources: from a household respondent who provided the age of all the household members (roster age), and later, at the end of the interview, directly from the individual respondent who is asked his/her birth date. The true age (DVAGE) is derived from the birth date, and only when the date is not available is the roster age used. Usually, there are some cases of discrepancy between the roster age and the true age. While roster age determines the flow of questions and the editing of the file, on the final edited file the age variables refer to true age.

5.0 Survey Methodology

The Canadian Tobacco Use Monitoring Survey was administered between February 1st and June 30th, 2003 as a Random Digit Dialing (RDD) survey, a technique whereby telephone numbers are generated randomly by computer. Interviewing was conducted over the telephone.

5.1 Population Coverage

The target population for the Canadian Tobacco Use Monitoring Survey was all persons 15 years of age and over living in Canada with the following two exceptions:

1) residents of the Yukon, Northwest Territories and Nunavut, and
2) full-time residents of institutions.

Because the survey was conducted using a sample of telephone numbers, households (and thus persons living in 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 Stratification

In order to ensure that people from all parts of Canada were represented in the sample, each of the ten provinces were divided into strata or geographic areas. Generally, within each province, a census metropolitan area (CMA) stratum and a non-CMA stratum was defined. In Prince Edward Island, there was only one stratum for the province. In Ontario, there was a third stratum for Toronto, and in Quebec, there was a third stratum for Montreal. CMAs are areas defined by the census and correspond roughly to the cities with populations of 100,000 or more.

5.3 Sample Design and Allocation

The sample design is a special two-phase stratified random sample of telephone numbers. The two-phase design is used in order to increase the representation in the sample of individuals belonging to the 15 to 19 and 20 to 24 age groups. In the first phase, households are selected using RDD. In the second phase, one or two individuals (or none) are selected based upon household composition.

Because the main purpose of the survey is to produce reliable estimates in all ten provinces, an ***equal number of respondents in each province*** is targeted. The target is to get responses from 5,000 individuals aged 15 to 24 and 5,000 individuals aged 25 and over across Canada, or 500 individuals in each age group per province. The initial sample size of telephone numbers depended upon the expected response rate and the expected RDD hit rate (proportion of sampled telephone numbers which are screened in as households). To achieve the required sample sizes, two adjustments to the standard RDD methodology were introduced. First, the probabilities of selection within the household were unequal and second, households with only persons aged 25 and over present were sub-sampled. It is estimated that a total of almost 130,000 telephone numbers per year will be needed to get the 20,000 respondents per year. This assumed a 75% response rate and about 20% of households having individuals aged 15 to 24; the hit rate varies substantially by province, with an expected overall average of about 40%.

The targets and required telephone numbers were adjusted accordingly.

5.4 Sample Selection

The sample for the Canadian Tobacco Use Monitoring Survey was generated using a refinement of RDD sampling called the Elimination of Non-Working Banks (ENWB). Within each province-stratum combination, a list of working banks (area code + next five digits) was compiled from telephone company administrative files. A working bank, for the purposes of social surveys, is defined as a bank which contains at least one working residential telephone number. Thus, all banks with only unassigned, non-working, or business telephone numbers are excluded from the survey frame.

Next, a systematic sample of banks (with replacement) was selected within each stratum. For each selected bank, a two-digit number (00 to 99) was generated at random. This 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 numbers (i.e. not currently or never in service), to have a chance of being in the sample.

Each telephone number in the sample was dialled to determine whether or not it reached a household. If the telephone number is found to reach a household, the person answering the telephone was asked to provide information on the individual household members. The ages of the household members were used to determine who, in the household, would be selected for the tobacco use interview. Proxy interviews were not accepted.

To ensure that enough people were reached in the younger age groups, the random selection was set up such that at least one person aged 15 to 19 or 20 to 24 would be selected within a household, if they exist. The reason for this is that about 76% of all households in Canada are made up of only people over 25 years of age; another 20% consist of people over 25 living with people in either the 15 to 19 or 20 to 24 age group; and only 4% of households contain no one aged over 25. If all ages were selected with equal probability and retained, the 25 and over age group would be over-represented with respect to the survey objectives. Thus, to save on the costs of additional interviews, some of the selected people in the 25 and over age group were screened out and did not receive the tobacco use interview. Two people were selected if more than one of the age groups 15 to 19, 20 to 24, and 25 and over were represented in the household. When two people in the same household were selected, they were always from different age groups. This ensured that there was no negative impact on the precision of the estimates by age group due to correlation within households. There was a small impact on the precision for the total estimates for all ages, but the sample size was sufficiently large so the impacts were minimal.

The detailed logic for the selection of individuals was as follows:

1) If everyone in the household is 15 to 19 then one person is selected at random.

2) If everyone in the household is 20 to 24 then one person is selected at random.

3) If everyone in the household is 25 and over then one person is selected at random; however, this selected person is retained for only a proportion of the cases.

4) If some household members are 15 to 19 and the rest are 20 to 24 then two people are selected at random, one from each age group.

5) If some household members are 15 to 19 and the rest are 25 and over then two people are selected at random, one from each age group; however, the person selected from the 25 and over age group is retained for only a proportion of the cases.

6) If some household members are 20 to 24 and the rest are 25 and over then two people are selected at random, one from each age group; however, the person selected from the 25 and over age group is retained for only a proportion of the cases.

7) If all three age groups are represented in the household, then two age groups are selected at random and then rule 4), 5), or 6) applies.

6.0 Data Collection

6.1 Questionnaire Design

The questionnaire design for this survey borrows heavily from the 1994 Survey on Smoking in Canada. Some questions have been added for consistency with international surveys which use the concept of smoking behaviour "in the last 30 days". The questionnaire used for the Canadian Tobacco Use Monitoring Survey during the Cycle 1 of 2003 contains several questions that were either somewhat modified or not asked in Cycle 2 of 2002. Specifications for valid ranges and inter-question consistency were incorporated into the computer-assisted telephone interviewing (CATI) application to the extent feasible. Additional consistency edits were done during the data processing phase.

6.2 Data Collection and Editing

The interviews were conducted every month, from February through June 2003.

Data were collected using computer-assisted telephone interviewing. The CATI system has a number of generic modules which can be quickly adapted to most types of surveys. A front-end module contains a set of standard response codes for dealing with all possible call outcomes, as well as the associated scripts to be read by the interviewers. A standard approach set up for introducing the agency, the name and purpose of the survey, the survey sponsors, how the survey results will be used, and the duration of the interview was used. We explained to respondents how they were selected for the survey, that their participation in the survey is voluntary, and that their information will remain strictly confidential. Help screens were provided to the interviewers to assist them in answering questions that are commonly asked by respondents.

The CATI application ensured that only valid question responses were entered and that all the correct flows were followed. Edits were built into the application to check the consistency of responses, identify and correct outliers, and to control who gets asked specific questions. This meant that the data was already quite "clean" at the end of the collection process.

Interviewers were trained on the survey content and the CATI application. In addition to classroom training, the interviewers completed a series of mock interviews to become familiar with the survey and its concepts and definitions. Every attempt is made to ensure that the same set of interviewers is used each month. This minimizes training and yields better and more consistent data quality.

The cases were distributed to four of the Statistics Canada regional offices. The workload and interviewing staff within each office was managed by a project manager. The automated scheduler used by the CATI system ensured that cases were assigned randomly to interviewers and that cases were called at different times of the day and different days of the week to maximize the probability of contact. There were a maximum of 17 call attempts per case; once the maximum was reached, the case was reviewed by a senior interviewer who determined if additional calls would be made.

7.0 Data Processing

The main output of the Canadian Tobacco Use Monitoring Survey are two "clean" microdata files, one for the household level information and one for the person level information. This chapter presents a brief summary of the processing steps involved in producing these files.

7.1 Data Capture

As the data was collected using computer-assisted telephone interviewing (CATI), there was no need for a separate data capture system since the information was entered in the Regional Offices systems directly by the interviewers during the interview.

7.2 Editing

The first stage of survey processing was to merge the monthly files into a single file. Any ”out-of-range” values on the data file were replaced with blanks. This process was designed to make further editing easier.

The first type of error treated was errors in questionnaire flow, where questions which did not apply to the respondent (and should therefore not have been answered) were found to contain answers. In this case a computer edit automatically eliminated superfluous data by following the flow of the questionnaire implied by answers to previous, and in some cases, subsequent questions.

The second type of error treated involved a lack of information in questions which should have been answered. For this type of error, a non-response or "not-stated" code was assigned to the item.

7.3 Creation of Derived Variables

A number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. Examples of derived variables include the average number of cigarettes smoked daily and the number of years the respondent smoked.

7.4 Weighting

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

The weighting phase is a step which calculates, for each record, what this number is. This weight appears on the microdata file, and **must** be used to derive meaningful estimates from the survey. For example, if the number of people in Canada who smoke daily is to be estimated, it is done by selecting the records referring to those individuals in the sample with that characteristic (Q010 = 1) and summing the weights entered on those records. A separate weight for households and persons is calculated every six months.

Details of the method used to calculate these weights are presented in Chapter 11.0.

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 Chapter 9.0 of this document.

Geographic Identifiers

Household File and Person File:

The survey master data files include explicit geographic identifiers for province, and stratum (census metropolitan area (CMA), non-CMA, Toronto or Montreal). The survey public use microdata files only contain an identifier for province.

Person File:

Starting with Cycle 1 of 2002, the master data file contains the first three digits of the respondent's postal code.

Household Age Composition

Household File and Person File:

Household age composition is available as the number of household members (capped at two) in the following age ranges: 0 to 14, 15 to 24, 25 to 44, and 45 and over.

Marital Status of Respondents

Person File:

The detailed marital status variable (six categories) is available on the master file only, while on the public use microdata file this variable has been grouped into three categories.

Other Modifications to Person File

A small number of records, below 20, had a demographic variable recoded to "not stated" to avoid potential identification of respondents resulting from an unusual combination of characteristics.

8.0 Data Quality

For the Canadian Tobacco Use Monitoring Survey (CTUMS), the response rates computed include the following:

The Telephone Resolved Rate and the Hit Rate apply to both the Household file and the Person file.

- Telephone Resolved Rate, where telephone numbers that were confirmed as residential, business or out-of-scope, were considered resolved.

- Hit Rate, where resolved telephone numbers that were confirmed as residential, had **valid household data** or **valid person data,** were considered to belong to a household.

This Household Response Rate applies only to the Household file.

- Household Response Rate, where households with ages provided for everyone in the roster and **valid household data,** were considered a response.

The Household Response Rate, Roster Completion Rate, Person Response Rate and Combined Household Response Rate apply only to the Person file.

- Household Response Rate, where telephone numbers that were confirmed as residential or **valid person data** existed with a completed roster, were considered a response.

- Roster Completion Rate, where households with a completed roster containing ages for each person in the roster, were considered a response.

- Person Response Rate, where records with ages provided for everyone in the roster had a corresponding household response record and **valid person data** exists, were considered a response.

- Combined Household Response Rate equals the Household Response Rate multiplied by the Person Response Rate.

Telephone Resolved Rate and Hit Rate by Province

Province	Total Number of Telephone Numbers Generated	Total Resolved	Telephone Resolved Rate (%)	Total Households	Hit Rate (%)
Newfoundland and Labrador	6,903	6,721	97.4	2,301	34.2
Prince Edward Island	6,077	5,802	95.5	2,399	41.3
Nova Scotia	6,817	6,564	96.3	2,699	41.1
New Brunswick	8,998	8,652	96.2	3,006	34.7
Quebec	6,005	5,916	98.5	3,029	51.2
Ontario	6,570	6,263	95.3	2,390	38.2
Manitoba	6,872	6,872	100.0	2,903	42.2
Saskatchewan	6,669	6,669	100.0	2,596	38.9
Alberta	5,370	5,370	100.0	2,559	47.7
British Columbia	5,995	5,717	95.4	2,601	45.5
Canada	**66,276**	**64,546**	**97.4**	**26,483**	**41.0**

8.1 Household Response Rates - February to June 2003

A **household respondent** must complete the roster with no age refusals, and valid household data must exist. There were 3,289 (12.4%) households that were non-responding.

Household Response Rate by Province

Province	Total Households	Responding Households	Household Response Rate (%)
Newfoundland and Labrador	2,301	2,092	90.9
Prince Edward Island	2,399	2,161	90.1
Nova Scotia	2,699	2,434	90.2
New Brunswick	3,006	2,711	90.2
Quebec	3,029	2,597	85.7
Ontario	2,390	2,148	89.9
Manitoba	2,903	2,494	85.9
Saskatchewan	2,596	2,180	84.0
Alberta	2,559	2,076	81.1
British Columbia	2,601	2,301	88.5
Canada	**26,483**	**23,194**	**87.6**

Household Response Rate by Survey Month

Survey Month	Total Households	Responding Households	Household Response Rate (%)
February	5,160	4,615	89.4
March	5,365	4,758	88.7
April	5,266	4,654	88.4
May	5,279	4,556	86.3
June	5,413	4,611	85.2
Total	**26,483**	**23,194**	**87.6**

8.2 Person Response Rates - February to June 2003

A **person respondent** has the following characteristics:

- The telephone number of the selected person belonged to a household.
- The roster was completed with no individual age refusals.
- The selected person was 15 years of age or older, when the survey was conducted.
- The age given in the roster was verified with the date of birth given by the selected person.
- The selected person answered the key questions on smoking habits, at minimum.

There were 12,890 households, in which, household data was collected but nobody was selected to continue with the CTUMS. (See Section 5.4 (Sample Selection), for more information.) Of the remaining households, 8,824 had one person selected while 1,564 had two people selected. There were only 158 (1.3%) people that refused to complete the survey and 1,235 (10.3%) other non-respondents.

Household Response Rate, Roster Completion Rate and Combined Household Response Rate by Province

Province	Total Households	Total Households with Rosters	Household Response Rate (%)	Households with Valid Roster Data	Roster Completion Rate (%)	Combined Household Response Rate (%)
Newfoundland and Labrador	2,301	2,155	93.7	2,100	97.4	91.3
Prince Edward Island	2,399	2,248	93.7	2,166	96.4	90.3
Nova Scotia	2,699	2,512	93.1	2,445	97.3	90.6
New Brunswick	3,006	2,785	92.6	2,714	97.5	90.3
Quebec	3,029	2,693	88.9	2,609	96.9	86.1
Ontario	2,390	2,189	91.6	2,152	98.3	90.0
Manitoba	2,903	2,543	87.6	2,505	98.5	86.3
Saskatchewan	2,596	2,226	85.7	2,191	98.4	84.4
Alberta	2,559	2,118	82.8	2,087	98.5	81.6
British Columbia	2,601	2,367	91.0	2,309	97.5	88.8
Canada	**26,483**	**23,836**	**90.0**	**23,278**	**97.7**	**87.9**

Person Response Rate by Province

Province	Total Persons Selected	Total Persons Responding	Person Response Rate (%)
Newfoundland and Labrador	1,204	1,024	85.0
Prince Edward Island	1,225	1,056	86.2
Nova Scotia	1,216	1,066	87.7
New Brunswick	1,467	1,285	87.6
Quebec	1,317	1,169	88.8
Ontario	1,043	963	92.3
Manitoba	1,243	1,108	89.1
Saskatchewan	1,145	993	86.7
Alberta	1,116	970	86.9
British Columbia	976	925	94.8
Canada	**11,952**	**10,559**	**88.3**

Person Response Rate by Survey Month

Survey Month	Total Persons Selected	Total Persons Responding	Person Response Rate (%)
February	2,369	2,126	89.7
March	2,486	2,238	90.0
April	2,387	2,110	88.4
May	2,328	2,038	87.5
June	2,382	2,047	85.9
Total	**11,952**	**10,559**	**88.3**

Target Number of Respondents and Person Response Rate by Age Group

Age Group	Target Number of Respondents	Total Persons Selected	Total Persons Responding	Person Response Rate (%)
15 to 19	2,500	3,165	2,791	88.2
20 to 24	2,500	2,718	2,345	86.3
25 and over	5,000	6,069	5,423	89.4
Total	**10,000**	**11,952**	**10,559**	**88.3**

8.3 Survey Errors

The estimates derived from this survey are based on a sample of households. 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 questionnaire 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 include extensive training of interviewers with respect to the survey procedures and computer-assisted telephone interviewing (CATI) application, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions and testing of the CATI application to ensure that range checks, edits and question flow were all programmed correctly.

8.4 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 because the interviewer was either unable to contact the respondent or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of households or individuals who responded to the survey to compensate for those who did not respond.

8.5 Partial Non-response

In most cases, partial non-response to the survey occurred when the respondent did not understand or misinterpreted a question, refused to answer a question, or could not recall the requested information. Partial non-response is indicated by codes on the microdata file i.e. refused, don't know.

8.6 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.

8.7 Measurement 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. This section of the documentation outlines the

measures of sampling error which Statistics Canada commonly uses and which it urges users producing estimates from this microdata file to use also.

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 (CV) 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 19.9% of Canadians are currently cigarette smokers, and this estimate is found to have standard error of 0.0074. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.0074}{0.199} \right) X \ 100 \ \% \ = \ 3.7\%$$

There is more information on the calculation of coefficient of variation in Chapter 10.0.

9.0 Guidelines for Tabulation, Analysis and Release

This chapter 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 files. 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 files 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 hundred 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 100, if the last two digits are between 00 and 49, they are changed to 00 and the preceding digit (the hundreds digit) is left unchanged. If the last two digits are between 50 and 99 they are changed to 00 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 100 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 100 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 Canadian Tobacco Use Monitoring Survey (CTUMS) 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.

9.3 Definitions of Types of Estimates: Categorical and Quantitative

Before discussing how the CTUMS 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 from the microdata file for the CTUMS.

9.3.1 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 currently smoke cigarettes, or the proportion of daily smokers that have attempted to quit smoking 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.

Examples of Categorical Questions:

Q: In the past THIRTY DAYS, did you smoke any cigarettes?
R: Yes / No

Q: What prompted you to quit smoking?
R: Current health problems / Smoking-related illness or death of friend / Pregnancy / Advice from doctor / Concern for future health

9.3.2 Quantitative Estimates

Quantitative estimates are estimates of totals or of means, medians and other measures of central tendency of quantities based upon some or all of the members of the surveyed population. They also specifically involve estimates of the form $\hat{X} / \hat{Y}$ where $\hat{X}$ is an estimate of surveyed population quantity total and $\hat{Y}$ is an estimate of the number of persons in the surveyed population contributing to that total quantity.

An example of a quantitative estimate is the average number of cigarettes smoked, on Saturday, per person. The numerator $\left(\hat{X}\right)$ is an estimate of the total number of cigarettes smoked on Saturday, and its denominator $\left(\hat{Y}\right)$ is the number of persons who reported smoking on Saturday.

Examples of Quantitative Questions:

Q: Thinking back over the past 7 days, starting with yesterday, how many cigarettes did you smoke on Saturday?
R: |_|_| cigarettes

Q: At what age did you smoke your first cigarette?
R: |_|_| years old

9.3.3 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 $\hat{X} / \hat{Y}$ are obtained by:

a) summing the final weights of records having the characteristic of interest for the numerator $\left(\hat{X}\right)$,

b) summing the final weights of records having the characteristic of interest for the denominator $\left(\hat{Y}\right)$, then

c) divide estimate a) by estimate b) $\left(\hat{X} / \hat{Y}\right)$.

9.3.4 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 the total number of cigarettes smoked on Saturday, multiply the value reported in question Q090SAT (number of cigarettes smoked on Saturday) by the final weight for the record, then sum this value over all records with Q090SAT< 96 (all respondents who reported a value in this field).

To obtain a weighted average of the form $\hat{X} / \hat{Y}$, the numerator $\left(\hat{X}\right)$ is calculated as for a quantitative estimate and the denominator $\left(\hat{Y}\right)$ is calculated as for a categorical estimate. For example, to estimate the average number of cigarettes smoked on Saturday,

a) estimate the total number of cigarettes smoked on Saturday $\left(\hat{X}\right)$ as described above,

b) estimate the number of people $\left(\hat{Y}\right)$ in this category by summing the final weights of all records with Q090SAT < 96, then

c) divide estimate a) by estimate b) $\left(\hat{X} / \hat{Y}\right)$.

9.4 Guidelines for Statistical Analysis

The Canadian Tobacco Use Monitoring Survey is based upon a complex sample design, with stratification, 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. In order for survey estimates and analyses to be free from bias, the survey weights must 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 poor. Approximate variances for simple estimates such as totals, proportions and ratios (for qualitative variables) can be derived using the accompanying Approximate Sampling Variability Tables.

For other analysis techniques (for example linear regression, logistic regression and analysis of variance), a method exists which can make the variances calculated by the standard packages more meaningful, by incorporating the unequal probabilities of selection. The method rescales the weights so that there is an average weight of 1.

For example, suppose that analysis of all male respondents is required. The steps to rescale the weights are as follows:

1) select all respondents from the file who reported SEX = men;

2) calculate the AVERAGE weight for these records by summing the original person weights from the microdata file for these records and then dividing by the number of respondents who reported SEX = men;

3) for each of these respondents, calculate a RESCALED weight equal to the original person weight divided by the AVERAGE weight;

4) perform the analysis for these respondents using the RESCALED weight.

However, because the stratification and clustering of the sample's design are still not taken into account, the variance estimates calculated in this way are likely to be under-estimates.

The calculation of more precise variance estimates requires detailed knowledge of the design of the survey. Such detail cannot be given in this microdata file because of confidentiality. Variances that take the complete sample design into account can be calculated for many statistics by Statistics Canada on a cost recovery basis.

9.5 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimate from the Canadian Tobacco Use Monitoring Survey users should first determine the quality level of the estimate. The quality levels are *acceptable*, *marginal* and *unacceptable*. Data quality is affected by both sampling and non-sampling errors as discussed in Chapter 8.0. However for this purpose, the quality level of an estimate will be determined only on the basis of sampling error as reflected by the coefficient of variation as shown in the table below. Nonetheless users should be sure to read Chapter 8.0 to be more fully aware of the quality characteristics of these data.

First, the number of respondents who contribute to the calculation of the estimate should be determined. If this number is less than 30, the weighted estimate should be considered to be of unacceptable quality.

For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the estimate and follow the guidelines below. These quality level guidelines should be applied to weighted rounded estimates.

All estimates can be considered releasable. However, those of marginal or unacceptable quality level must be accompanied by a warning to caution subsequent users.

Quality Level Guidelines

Quality Level of Estimate	**Guidelines**
1) Acceptable	Estimates have: a sample size of 30 or more, and low coefficients of variation in the range of 0.0% to 16.5%. No warning is required.
2) Marginal	Estimates have: a sample size of 30 or more, and high coefficients of variation in the range of 16.6% to 33.3%. Estimates should be flagged with the letter M (or some similar identifier). They should be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimates.
3) Unacceptable	Estimates have: a sample size of less than 30, or very high coefficients of variation in excess of 33.3%. Statistics Canada recommends not to release estimates of unacceptable quality. However, if the user chooses to do so then estimates should be flagged with the letter U (or some similar identifier) and the following warning should accompany the estimates: "Please be warned that these estimates [flagged with the letter U] do not meet Statistics Canada's quality standards. Conclusions based on these data will be unreliable, and most likely invalid."

9.6 Release Cut-off's for the Canadian Tobacco Use Monitoring Survey – Household File

The minimum size of the estimates are specified in the table below by province for households. Estimates smaller than the minimum size given in the "Unacceptable" column must be flagged in the appropriate manner.

Table of Release Cut-offs – Household File

Province	Acceptable CV 0.0% - 16.5%	Marginal CV 16.6% - 33.3%			Unacceptable CV > 33.3%	
Newfoundland and Labrador	4,500 & over	1,000	to <	4,500	under	1,000
Prince Edward Island	1,000 & over	500	to <	1,000	under	500
Nova Scotia	6,500 & over	1,500	to <	6,500	under	1,500
New Brunswick	4,500 & over	1,000	to <	4,500	under	1,000
Quebec	54,000 & over	13,500	to <	54,000	under	13,500
Ontario	91,000 & over	22,500	to <	91,000	under	22,500
Manitoba	7,500 & over	2,000	to <	7,500	under	2,000
Saskatchewan	6,500 & over	1,500	to <	6,500	under	1,500
Alberta	25,500 & over	6,500	to <	25,500	under	6,500
British Columbia	31,000 & over	7,500	to <	31,000	under	7,500
Canada	**53,000 & over**	**13,000**	**to <**	**53,000**	**under**	**13,000**

9.7 Release Cut-off's for the Canadian Tobacco Use Monitoring Survey – Person File

The minimum size of the estimates are specified in the table below by province and age group. Estimates smaller than the minimum size given in the "Unacceptable" column must be flagged in the appropriate manner.

Table of Release Cut-offs – Person File

Province	Age Group	Acceptable CV 0.0% - 16.5%	Marginal CV 16.6% - 33.3%	Unacceptable CV > 33.3%
Newfoundland and Labrador	All	30,500 & over	8,000 to < 30,500	under 8,000
	15-19	5,500 & over	1,500 to < 5,500	under 1,500
	20-24	7,500 & over	2,000 to < 7,500	under 2,000
	25+	33,500 & over	9,000 to < 33,500	under 9,000
Prince Edward Island	All	8,000 & over	2,000 to < 8,000	under 2,000
	15-19	1,500 & over	500 to < 1,500	under 500
	20-24	2,000 & over	500 to < 2,000	under 500
	25+	7,500 & over	2,000 to < 7,500	under 2,000
Nova Scotia	All	58,000 & over	15,000 to < 58,000	under 15,000
	15-19	10,000 & over	3,000 to < 10,000	under 3,000
	20-24	11,000 & over	3,000 to < 11,000	under 3,000
	25+	61,000 & over	16,000 to < 61,000	under 16,000
New Brunswick	All	35,000 & over	9,000 to < 35,000	under 9,000
	15-19	6,000 & over	1,500 to < 6,000	under 1,500
	20-24	9,000 & over	2,500 to < 9,000	under 2,500
	25+	37,000 & over	9,500 to < 37,000	under 9,500
Quebec	All	347,500 & over	89,000 to < 347,500	under 89,000
	15-19	75,000 & over	21,000 to < 75,000	under 21,000
	20-24	83,000 & over	23,500 to < 83,000	under 23,500
	25+	407,500 & over	106,500 to < 407,500	under 106,500
Ontario	All	743,500 & over	193,500 to < 743,500	under 193,500
	15-19	140,500 & over	39,500 to < 140,500	under 39,500
	20-24	176,000 & over	52,000 to < 176,000	under 52,000
	25+	815,500 & over	216,500 to < 815,500	under 216,500
Manitoba	All	54,000 & over	14,000 to < 54,000	under 14,000
	15-19	12,000 & over	3,500 to < 12,000	under 3,500
	20-24	14,500 & over	4,000 to < 14,500	under 4,000
	25+	60,000 & over	15,500 to < 60,000	under 15,500
Saskatchewan	All	53,500 & over	14,000 to < 53,500	under 14,000
	15-19	12,500 & over	3,500 to < 12,500	under 3,500
	20-24	15,000 & over	4,500 to < 15,000	under 4,500
	25+	57,000 & over	15,000 to < 57,000	under 15,000
Alberta	All	170,500 & over	44,000 to < 170,500	under 44,000
	15-19	42,500 & over	12,000 to < 42,500	under 12,000
	20-24	51,500 & over	15,000 to < 51,500	under 15,000
	25+	191,000 & over	50,500 to < 191,000	under 50,500
British Columbia	All	288,500 & over	75,500 to < 288,500	under 75,500
	15-19	49,000 & over	14,000 to < 49,000	under 14,000
	20-24	96,500 & over	32,000 to < 96,500	under 32,000
	25+	312,500 & over	83,500 to < 312,500	under 83,500
Canada	**All**	**465,500 & over**	**116,000 to < 465,500**	**under 116,000**
	15-19	**89,500 & over**	**22,500 to < 89,500**	**under 22,500**
	20-24	**128,000 & over**	**33,000 to < 128,000**	**under 33,000**
	25+	**514,000 & over**	**128,500 to < 514,000**	**under 128,500**

10.0 Approximate Sampling Variability Tables

In order to supply coefficients of variation (CV) 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 CV 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 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 (usually the 75th percentile) to be used in the CV tables which would then apply to the entire set of characteristics.

The table below shows the conservative value of the design effects as well as sample sizes and population counts by province, which were used to produce the Approximate Sampling Variability Tables for the Canadian Tobacco Use Monitoring Survey (CTUMS) Household file.

Household File

Province	Design Effect	Sample Size	Population
Newfoundland and Labrador	1.30	2,092	197,409
Prince Edward Island	1.19	2,161	53,940
Nova Scotia	1.15	2,434	370,187
New Brunswick	1.10	2,711	293,187
Quebec	1.25	2,597	3,113,119
Ontario	1.19	2,148	4,553,928
Manitoba	1.20	2,494	431,024
Saskatchewan	1.01	2,180	382,897
Alberta	1.25	2,076	1,167,577
British Columbia	1.22	2,301	1,624,565
Canada	**2.77**	**23,194**	**12,187,833**

The table below shows the conservative value of the design effects as well as sample sizes and population counts by province and age group, which were used to produce the Approximate Sampling Variability Tables for the CTUMS Person file.

Person File

Province	Age Group	Design Effect	Sample Size	Population
Newfoundland and Labrador	All	2.07	1,024	440,404
	15-19	1.38	292	37,490
	20-24	1.49	216	36,611
	25+	1.41	516	366,303
Prince Edward Island	All	2.14	1,056	113,356
	15-19	1.33	294	10,435
	20-24	1.42	219	9,804
	25+	1.32	543	93,117
Nova Scotia	All	2.36	1,066	772,242
	15-19	1.55	302	64,126
	20-24	1.51	254	61,640
	25+	1.45	510	646,476
New Brunswick	All	2.10	1,285	617,513
	15-19	1.35	351	49,837
	20-24	1.70	274	49,536
	25+	1.38	660	518,140
Quebec	All	1.92	1,169	6,111,588
	15-19	1.51	279	453,260
	20-24	1.48	279	509,588
	25+	1.43	611	5,148,740
Ontario	All	2.15	963	9,809,759
	15-19	1.41	245	805,552
	20-24	1.45	189	800,484
	25+	1.59	529	8,203,722
Manitoba	All	1.92	1,108	905,744
	15-19	1.30	276	82,457
	20-24	1.50	243	78,345
	25+	1.40	589	744,942
Saskatchewan	All	1.98	993	785,351
	15-19	1.37	263	76,321
	20-24	1.55	211	71,280
	25+	1.39	519	637,750
Alberta	All	1.93	970	2,501,664
	15-19	1.50	243	230,472
	20-24	1.57	215	242,609
	25+	1.45	512	2,028,583
British Columbia	All	2.32	925	3,419,246
	15-19	1.61	272	274,769
	20-24	2.86	204	284,223
	25+	1.50	449	2,860,255
Canada	**All**	**5.35**	**10,559**	**25,476,868**
	15-19	**3.44**	**2,817**	**2,084,720**
	20-24	**3.98**	**2,304**	**2,144,120**
	25+	**3.67**	**5,438**	**21,248,028**

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. Since the approximate CV is conservative, the use of actual variance estimates may cause the estimate to be switched from one quality level to another. For instance a marginal estimates could become acceptable based on the exact CV calculation.

Remember: If the number of observations on which an estimate is based is less than 30, the weighted estimate should be considered *unacceptable* and should be flagged in the appropriate manner, 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 Coefficient of Variation Tables for Categorical Estimates

The following rules should enable the user to determine the approximate coefficients of variation from the Approximate 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 of Persons Possessing a Characteristic (Aggregates)

The coefficient of variation depends only on the size of the estimate itself. On the Approximate 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 of Persons 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 total upon which the proportion or percentage is based. 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 former smokers that quit for current health problems is more reliable than the estimated number of former smokers that quit for current health problems. (Note that in the tables the coefficients of variation decline in value when reading from left to right).

When the proportion or percentage is based upon the total population of the geographic area covered by the table, the CV of the proportion or percentage is the same as the CV 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 $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ is:

$$\sigma_{\hat{d}} = \sqrt{\left(\hat{X}_1\alpha_1\right)^2 + \left(\hat{X}_2\alpha_2\right)^2}$$

where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, and α_1 and α_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 $\sigma_{\hat{d}} / \hat{d}$. This formula is accurate for the difference between separate and uncorrelated characteristics, but is only approximate otherwise.

Rule 4: Estimates of Ratios

In the case where the numerator is a subset of the denominator, 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 smokers and the numerator is the number of daily smokers.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of daily smokers as compared to the number of non-smokers, 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 $\hat{R}$. That is, the standard error of a ratio $\left(\hat{R} = \hat{X}_1 / \hat{X}_2\right)$ is:

$$\sigma_{\hat{R}} = \hat{R}\sqrt{\alpha_1^{\,2} + \alpha_2^{\,2}}$$

where α_1 and α_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 $\sigma_{\hat{R}} / \hat{R}$. 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 Coefficient of Variation Tables for Categorical Estimates

The following examples based on the 2002 Annual data are included to assist users in applying the foregoing rules.

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

Suppose that a user estimates that during the reference period 5,414,335 persons were current smokers (DVSST1 = 1) in Canada. How does the user determine the coefficient of variation of this estimate?

1) Refer to the Person coefficient of variation table for CANADA – All Ages.

Canadian Tobacco Use Monitoring Survey 2002 - February to December - Person File

Approximate Sampling Variability Tables for Canada - All Ages

NUMERATOR OF PERCENTAGE	ESTIMATED PERCENTAGE													
('000)	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	197.2	196.3	195.3	192.3	187.1	181.9	176.4	170.8	165.0	159.0	152.8	139.5	108.0	62.4
2	139.4	138.8	138.1	135.9	132.3	128.6	124.8	120.8	116.7	112.5	108.0	98.6	76.4	44.1
3	113.8	113.3	112.7	111.0	108.0	105.0	101.9	98.6	95.3	91.8	88.2	80.5	62.4	36.0
4	98.6	98.1	97.6	96.1	93.6	90.9	88.2	85.4	82.5	79.5	76.4	69.7	54.0	31.2
5	88.2	87.8	87.3	86.0	83.7	81.3	78.9	76.4	73.8	71.1	68.3	62.4	48.3	27.9
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
75	********	22.7	22.5	22.2	21.6	21.0	20.4	19.7	19.1	18.4	17.6	16.1	12.5	7.2
80	********	21.9	21.8	21.5	20.9	20.3	19.7	19.1	18.5	17.8	17.1	15.6	12.1	7.0
85	********	21.3	21.2	20.9	20.3	19.7	19.1	18.5	17.9	17.2	16.6	15.1	11.7	6.8
90	********	20.7	20.6	20.3	19.7	19.2	18.6	18.0	17.4	16.8	16.1	14.7	11.4	6.6
95	********	20.1	20.0	19.7	19.2	18.7	18.1	17.5	16.9	16.3	15.7	14.3	11.1	6.4
100	********	19.6	19.5	19.2	18.7	18.2	17.6	17.1	16.5	15.9	15.3	13.9	10.8	6.2
125	********	17.6	17.5	17.2	16.7	16.3	15.8	15.3	14.8	14.2	13.7	12.5	9.7	5.6
150	********	16.0	15.9	15.7	15.3	14.8	14.4	13.9	13.5	13.0	12.5	11.4	8.8	5.1
200	********	13.9	13.8	13.6	13.2	12.9	12.5	12.1	11.7	11.2	10.8	9.9	7.6	4.4
250	********	12.4	12.4	12.2	11.8	11.5	11.2	10.8	10.4	10.1	9.7	8.8	6.8	3.9
300	********	********	11.3	11.1	10.8	10.5	10.2	9.9	9.5	9.2	8.8	8.1	6.2	3.6
350	********	********	10.4	10.3	10.0	9.7	9.4	9.1	8.8	8.5	8.2	7.5	5.8	3.3
400	********	********	9.8	9.6	9.4	9.1	8.8	8.5	8.3	8.0	7.6	7.0	5.4	3.1
450	********	********	9.2	9.1	8.8	8.6	8.3	8.1	7.8	7.5	7.2	6.6	5.1	2.9
500	********	********	8.7	8.6	8.4	8.1	7.9	7.6	7.4	7.1	6.8	6.2	4.8	2.8
750	********	********	********	7.0	6.8	6.6	6.4	6.2	6.0	5.8	5.6	5.1	3.9	2.3
1000	********	********	********	6.1	5.9	5.8	5.6	5.4	5.2	5.0	4.8	4.4	3.4	2.0
1500	********	********	********	********	4.8	4.7	4.6	4.4	4.3	4.1	3.9	3.6	2.8	1.6
2000	********	********	********	********	4.2	4.1	3.9	3.8	3.7	3.6	3.4	3.1	2.4	1.4
3000	********	********	********	********	********	3.3	**3.2**	**3.1**	3.0	2.9	2.8	2.5	2.0	1.1
4000	********	********	********	********	********	********	2.8	2.7	2.6	2.5	2.4	2.2	1.7	1.0
5000	********	********	********	********	********	********	**2.5**	2.4	2.3	2.2	2.2	2.0	1.5	0.9
6000	********	********	********	********	********	********	********	2.2	2.1	2.1	2.0	1.8	1.4	0.8
7000	********	********	********	********	********	********	********	********	2.0	1.9	1.8	1.7	1.3	0.7
8000	********	********	********	********	********	********	********	********	********	1.8	1.7	1.6	1.2	0.7
9000	********	********	********	********	********	********	********	********	********	********	1.6	1.5	1.1	0.7
10000	********	********	********	********	********	********	********	********	********	********	1.5	1.4	1.1	0.6
12500	********	********	********	********	********	********	********	********	********	********	********	1.2	1.0	0.6
15000	********	********	********	********	********	********	********	********	********	********	********	********	0.9	0.5

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

2) The estimated aggregate (5,414,335) 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 5,000,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 2.5%.

4) So the approximate coefficient of variation of the estimate is 2.5%. The finding that there were 5,414,335 (to be rounded according to the rounding guidelines in Section 9.1) current smokers in the reference period is publishable with no qualifications.

Example 2: Estimates of Proportions or Percentages of Persons Possessing a Characteristic

Suppose that the user estimates that 2,865,929 / 12,436,728 = 23.0% of men currently smoke in Canada in the reference period. How does the user determine the coefficient of variation of this estimate?

1) Refer to the Person coefficient of variation table for CANADA (see above). The CANADA level table should be used because it is the smallest table that contains the domain of the estimate, all men in Canada.

2) Because the estimate is a percentage which is based on a subset of the total population (i.e. men), it is necessary to use both the percentage (23.0%) and the numerator portion of the percentage (2,865,929) in determining the coefficient of variation.

3) The numerator, 2,865,929, 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 3,000,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 25.0%.

4) The figure at the intersection of the row and column used, namely 3.1% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 3.1%. The finding that 23.0% of men currently smoke can be published with no qualifications.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates that 2,548,406 / 12,814,359 = 19.9% of women currently smoke in Canada, while 2,865,929 / 12,436,728 = 23.0% of men currently smoke in Canada. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the Person CANADA coefficient of variation table (see above) in the same manner as described in Example 2 gives the CV of the estimate for women as 3.2%, and the CV of the estimate for men as 3.1%.

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

$$\sigma_{\hat{d}} = \sqrt{\left(\hat{X}_1\alpha_1\right)^2 + \left(\hat{X}_2\alpha_2\right)^2}$$

where $\hat{X}_1$ is estimate 1 (men), $\hat{X}_2$ is estimate 2 (women), and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

That is, the standard error of the difference $\hat{d}$ = 0.230 – 0.199 = 0.031 is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{\left[(0.230)(0.031)\right]^2 + \left[(0.199)(0.032)\right]^2} \\ &= \sqrt{(0.00005) + (0.00004)} \\ &= 0.009\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.009 / 0.031 = 0.290.

4) So the approximate coefficient of variation of the difference between the estimates is 29.0%. This estimate is considered marginal and Statistics Canada recommends the estimate be flagged with the letter M (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimate.

Example 4: Estimates of Ratios

Suppose that the user estimates that 237,261 women currently smoke in the age group 15 to 19, while 220,511 men currently smoke in the age group 15 to 19. The user is interested in comparing the estimate of women versus that of men 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 women currently smoking in the age group 15 to 19. The denominator of the estimate ($\hat{X}_2$) is the number of men currently smoking in the age group 15 to 19.

2) Refer to the Person coefficient of variation table for CANADA – 15 - 19.

Canadian Tobacco Use Monitoring Survey 2002 - February to December - Person File

Approximate Sampling Variability Tables for Canada - 15-19

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	95.8	95.3	94.9	93.4	90.9	88.3	85.7	83.0	80.2	77.3	74.2	67.8	52.5	30.3
2	67.7	67.4	67.1	66.0	64.3	62.5	60.6	58.7	56.7	54.6	52.5	47.9	37.1	21.4
3	********	55.0	54.8	53.9	52.5	51.0	49.5	47.9	46.3	44.6	42.9	39.1	30.3	17.5
4	********	47.7	47.4	46.7	45.5	44.2	42.9	41.5	40.1	38.6	37.1	33.9	26.2	15.2
5	********	42.6	42.4	41.8	40.7	39.5	38.3	37.1	35.9	34.6	33.2	30.3	23.5	13.6
6	********	38.9	38.7	38.1	37.1	36.1	35.0	33.9	32.7	31.5	30.3	27.7	21.4	12.4
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
95	********	********	********	9.6	9.3	9.1	8.8	8.5	8.2	7.9	7.6	7.0	5.4	3.1
100	********	********	********	9.3	9.1	8.8	8.6	8.3	8.0	7.7	7.4	6.8	5.2	3.0
125	********	********	********	********	8.1	7.9	7.7	7.4	7.2	6.9	6.6	6.1	4.7	2.7
150	********	********	********	********	7.4	7.2	7.0	6.8	6.5	6.3	6.1	5.5	4.3	2.5
200	********	********	********	********	**6.4**	6.2	6.1	5.9	5.7	5.5	5.2	4.8	3.7	2.1
250	********	********	********	********	********	**5.6**	5.4	5.2	5.1	4.9	4.7	4.3	3.3	1.9
300	********	********	********	********	********	5.1	4.9	4.8	4.6	4.5	4.3	3.9	3.0	1.7
350	********	********	********	********	********	********	4.6	4.4	4.3	4.1	4.0	3.6	2.8	1.6
400	********	********	********	********	********	********	4.3	4.1	4.0	3.9	3.7	3.4	2.6	1.5
450	********	********	********	********	********	********	********	3.9	3.8	3.6	3.5	3.2	2.5	1.4
500	********	********	********	********	********	********	********	3.7	3.6	3.5	3.3	3.0	2.3	1.4
750	********	********	********	********	********	********	********	********	********	********	2.7	2.5	1.9	1.1
1000	********	********	********	********	********	********	********	********	********	********	********	2.1	1.7	1.0
1500	********	********	********	********	********	********	********	********	********	********	********	********	********	0.8

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

3) The numerator of this ratio estimate is 237,261. The figure closest to it is 250,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 5.6%

4) The denominator of this ratio estimate is 220,511. The figure closest to it is 200,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 6.4%.

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

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^{\,2} + \alpha_2^{\,2}}$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

That is:

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0.056)^2 + (0.064)^2} \\ &= \sqrt{0.003136 + 0.004096} \\ &= 0.085\end{aligned}$$

The obtained ratio of women currently smoking in the age group 15 to 19 versus men currently smoking in the age group 15 to 19 is 237,261 / 220,511 which is 1.08 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 8.5%, which is releasable with no qualifications.

10.2 How to Use the Coefficient of Variation 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 of 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 $\left(\hat{X} - k, \hat{X} + k\right)$ 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 $(CI_{\hat{x}})$:

$$CI_{\hat{x}} = \left(\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}}\right)$$

where $\alpha_{\hat{x}}$ 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 = 2.6 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 Coefficient of Variation Tables to Obtain Confidence Limits

A 95% confidence interval for the estimated proportion of men who currently smoke (from Example 2, Section 10.1.1) would be calculated as follows:

$\hat{X}$ = 23.0% (or expressed as a proportion 0.230)

t = 2

$\alpha_{\hat{x}}$ = 3.1% (0.031 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.230 - (2) (0.230) (0.031), 0.230 + (2) (0.230) (0.031)}

$CI_{\hat{x}}$ = {0.230 - 0.014, 0.230 + 0.014}

$CI_{\hat{x}}$ = {0.216, 0.244}

With 95% confidence it can be said that between 21.6% and 24.4% of men currently smoke.

10.3 How to Use the Coefficient of Variation 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 $\hat{X}_1$ and $\hat{X}_2$ be sample estimates for two characteristics of interest. Let the standard error on the difference $\hat{X}_1 - \hat{X}_2$ be $\sigma_{\hat{d}}$.

If $t = \dfrac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$

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 difference between the estimates is significant.

10.3.1 Example of Using the Coefficient of Variation 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 men who currently smoke and the proportion of women who currently smoke. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.009. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.230 - 0.199}{0.009} = \frac{0.031}{0.009} = 3.44 .$$

Since t = 3.44 is greater 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 Coefficient 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 Canadian Tobacco Use Monitoring Survey 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. For example, the coefficient of variation of the total number of cigarettes smoked on Saturday would be greater than the coefficient of variation of the corresponding proportion of current smokers. Hence, if the coefficient of variation of the proportion is not releasable, then the coefficient of variation of the corresponding quantitative estimate will also not be releasable.

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 Coefficient of Variation Tables - Household File

Refer to CTUMS2003_C1_HH_CVTabsE.pdf for the coefficient of variation tables for the Household file for Cycle 1 of 2003.

10.6 Coefficient of Variation Tables - Person File

Refer to CTUMS2003_C1_PR_CVTabsE.pdf for the coefficient of variation tables for the Person file for Cycle 1 of 2003.

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. One weight was calculated for each household and a separate weight was calculated and provided on a different file, for each person.

The weighting for the annual file of the Canadian Tobacco Use Monitoring Survey consisted of several steps:

- calculation of a basic weight,
- adjustments for non-response,
- an adjustment for selecting one or two persons in the household,
- dropping out-of-scope records and finally
- an adjustment to make the populations estimates consistent with known province-age-sex totals from the Census projected population counts for persons 15 years and over.

11.1 Weighting Procedures for the Household and Person Files

1. Calculate telephone weight

Each telephone number in the sample was assigned a basic weight, W_1, equal to the inverse of its probability of selection.

$$W_1 = \left(\frac{Total\ number\ of\ possible\ sampled\ telephone\ numbers\ \ in\ province - stratum}{Number\ of\ sampled\ telephone\ numbers\ in\ province - stratum} \right)$$

There were 66,276 telephone numbers in the sample with assigned weights.

2. Adjust for non-resolved telephone numbers

There were 1,730 telephone numbers that were not resolved, leaving 64,546 resolved telephone numbers. The unresolved telephone numbers were not determined to belong to a household, business or out-of-scope. Each telephone number had a flag indicating whether it was expected to be a residential, business, or unknown type of telephone number. The adjustment for the unresolved telephone numbers was done within province-stratum and this expected line type.

For each province-stratum-expected line type,

$$W_2 = W_1 * \left(\frac{\sum W_1\ for\ resolved\ telephone\ numbers + \sum W_1\ for\ unresolved\ telephone\ numbers}{\sum W_1\ for\ resolved\ telephone\ numbers} \right)$$

3. Remove out-of-scope telephone numbers

Telephone numbers corresponding to businesses, out-of-service numbers, or out-of-scope numbers, such as cottage telephone numbers, were dropped after the non-response adjustment for telephone non-response had been applied. Note that if **household or person** data existed then the telephone number was assumed to be a household. There were

38,063 out-of-scope telephone numbers and 26,483 telephone numbers belonging to a household.

4. Adjust for non-response of number of telephone lines in the household

The number of telephone lines in the household was calculated. If the number of different telephone lines within the household could not be calculated **but household or person data** existed, then it was imputed as one in order to retain good data. After imputation, there were 2,647 telephone numbers that were still missing the number of lines. Thus, there were 23,836 households with the number of lines calculated or imputed. The adjustment was done within province-stratum.

$$W_3 = W_2 * \left(\frac{\sum W_2 \ for\ households\ with\ number\ of\ lines + \sum W_2 \ for\ households\ mis\sin g\ number\ of\ lines}{\sum W_2 \ for\ households\ with\ number\ of\ lines} \right)$$

5. Calculate household weight with multiple telephone lines adjustment

Weights for households with more than one telephone line (with different telephone numbers) 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 lines (up to a maximum of 4) that serviced the household.

$$W_4 = \left(\frac{W_3}{Number\ \ of\ \ in-scope\ \ telephone\ \ lines\ \ in\ the\ household} \right)$$

11.2 Weighting Procedures for the Household File

6. Adjust for non-responding households

Household respondents responded to the questions on their smoking habits. If these questions were not sufficiently answered, perhaps refused or only partially answered, then the household was considered a non-respondent. There were 642 non-respondents. Thus, 23,194 in-scope household weights were used and adjusted within province-stratum.

$$W_5 = W_4 * \left(\frac{\sum W_4 \ for\ household\ respondents + \sum W_4 \ for\ household\ non-respondents}{\sum W_4 \ for\ household\ respondents} \right)$$

7. Adjust to known external household stratum totals

An adjustment was made to the household weights on records within each province, stratum and month, in order to make household estimates consistent with known external household counts. The adjustment factor for province-stratum-month (P-S-M) was defined as:

$$W_6 = W_5 * \left(\frac{Known\ external\ household\ count\ in\ P-S-M}{\sum W_5 \ for\ responding\ households\ in\ the\ sample\ in\ P-S-M} \right)$$

The household weights, W_6, obtained after this step, were considered final and appear on the household microdata file.

11.3 Weighting Procedures for the Person File

6. Adjust for non-responding households

On the Person file, there are only those records for which household roster was completed with no age refusals. There were 558 non-respondents. Thus, 23,278 in-scope household weights were used and adjusted within province-stratum.

$$W_5 = W_4 * \left(\frac{\sum W_4 \ for\ household\ respondents + \sum W_4 \ for\ household\ non-respondents}{\sum W_4 \ for\ household\ respondents} \right)$$

7. Calculate group weight

All of the in-scope responding households with completed rosters (i.e. no missing ages) were assigned group weights. From the roster, three flags were assigned to indicate the presence of a person in the following age groups: 15 to 19, 20 to 24, and 25 and over. If one or two age group categories were represented then an individual was selected from each age group present (i.e. the probability of selection of the age group was 1). Thus, the weight was not inflated. However, if three age groups were represented, then two people were selected, so the probability of selecting the age group is 2 out of the 3 groups. Thus, the weight is inflated by its inverse.

If 1 or 2 age groups were represented then $W_6 = W_5$.

If all 3 age groups were represented then $W_6 = W_5 * 3/2$.

8. Remove households with no selected persons

There were 12,890 households where no one was selected to continue with the tobacco use survey or a selected person was not retained because of sub-selection of individuals. These households were dropped because they had no person level data. About 70% of selected respondents aged 25 and over were screened out. There were 10,388 households with selected persons. There were 8,824 households with one person selected and 1,564 with two people selected.

9. Assign household weights to selected persons

These 8,824 + 2(1,564) = 11,952 selected persons are associated with in-scope responding households and keep the corresponding weight, W_6.

10. Calculate selected person sub-weight

All in-scope individuals were assigned weights. The weight is inflated by the number of people within the selected age group and the inverse of the sub-sampling factor.

$$W_7 = W_6 * \left(\frac{Number\ of\ individuals\ in\ selected\ age\ group}{Sub-sampling\ factor} \right)$$

The sub-sampling factor was 1 for age groups 15 to 19 and 20 to 24. The sub-sampling factor was pre-assigned for the 25 and over age group and varied from 23.0% to 30.0%, depending on the province.

11. Adjust for non-responding individuals

The Person file includes records of individual respondents who completed the questions on smoking habits and gave a date of birth corresponding to the age given in the roster. There were 1,393 non-respondents.

Thus, 10,559 in-scope individual weights were used and adjusted within province, age groups derived from the roster (15 to 19, 20 to 24, 25 to 44, 45 to 64, 65 and over) and sex.

$$W_8 = W_7 * \left(\frac{\sum W_7 \ for\ person\ respondents + \sum W_7 \ for\ person\ non-respondents}{\sum W_7 \ for\ person\ respondents} \right)$$

12. Adjust to external totals

An adjustment was made to the person weights in order to make population estimates consistent with external population counts for persons 15 years and older. This is known as post-stratification. The following external control totals were used:

1) Monthly population totals for each province-stratum, and

2) For Cycle 1 and Cycle 2:
   population totals by province, sex and the following age groups: 15 to 19, 20 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 and over. These totals were averaged over the survey period.

   For the Annual Summary:
   population totals by province, sex and the following age groups: 15 to 19, 20 to 24, 25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69 and 70 and over. These totals were averaged over the survey period.

The method called generalized regression (GREG) estimation was used to modify the weights to ensure that the survey estimates agreed with the external totals simultaneously along the two dimensions.

The person weights obtained after this step were considered final and appear on the person microdata file.

12.0 Questionnaire

Refer to CTUMS2003_C1_QuestE.pdf for the English questionnaire used in Cycle 1 of 2003.

13.0 Record Layouts with Univariate Frequencies

13.1 Record Layout with Univariate Frequencies – Household File

Refer to CTUMS2003_C1_HH_CdBk.pdf for the record layout with univariate counts for the Household file for Cycle 1 of 2003.

13.2 Record Layout with Univariate Frequencies – Person File

Refer to CTUMS2003_C1_PR_CdBk.pdf for the record layout with univariate counts for the Person file for Cycle 1 of 2003.