Microdata User Guide

The Canadian Survey of Experiences with Primary Health Care

2006-2007

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The Canadian Survey of Experiences with Primary Health Care (CSE-PHC) was conducted by Statistics Canada in January and February 2007 with the cooperation and support of the Health Council of Canada. This manual has been produced to facilitate the manipulation of the microdata file containing the survey results.

Any question 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

Health Council of Canada

Kira Leeb
90 Eglinton Ave East, Suite 900,
Toronto, Ontario, M4P 2Y3
Telephone: 416-453-8934
Fax: 416-481-1381
E-mail: kleeb@healthcouncilcanada.ca

2.0 Background

The Health Council of Canada (HCC) was created when the First Ministers' Accord on Health Care Renewal was signed in 2003. Their mandate is to report publicly on the progress of health care renewal in Canada. The Council's goal is to provide a system-wide perspective on health care reform to the Canadian public with a particular focus on issues related to the accountability and transparency.

Special Surveys Division was contacted by the HCC during the summer of 2006 to conduct this survey. The HCC identified key areas around access and use of health care, especially for people living with chronic conditions, for which national data was necessary. A questionnaire was developed in the fall of 2006 and collection began in January 2007.

3.0 Objectives

The main objectives of the survey are to collect data on issues relating to experiences with health care that impact Canadians. More specifically, the goal was to provide a picture of access and utilization of primary care as well as information on issues specific to Canadians living with chronic conditions and their experiences with the health care system. Ultimately, the data collected will provide information for the development of effective policies and strategies to help improve health care for all Canadians.

The data from this survey will provide a holistic perspective of Canadians' experiences with health care while identifying and raising awareness around issues that affect people living with chronic conditions. Finally, one of the ultimate goals of the survey is to help in decision-making about resources and provide baseline data to monitor change over time.

4.0 Concepts and Definitions

Since The Canadian Survey of Experiences with Primary Health Care is conducted over the telephone, an effort was made to use simple terminology throughout the questionnaire in order to minimize long complicated explanations of survey concepts. 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.

Primary Health Care refers to the main source of preventive as well as on-going or essential care people receive in their communities. They include regular medical doctors and family clinics. Often, this is the patient's first contact with the health care system.

Primary Care Provider was defined as a regular doctor or place where respondents go for health care.

Doctors are defined as medical doctors paid by provincial Medicare. All non medical doctors and those not covered under provincial Medicare systems were excluded.

5.0 Survey Methodology

The Canadian Survey of Experiences with Primary Health Care (CSE-PHC) was administered from January 16 to February 21, 2007 to a sub-sample of the people who participated to the Canadian Community Health Survey (CCHS) Cycle 3.1. Therefore its sample design is closely tied to that of the CCHS. The CCHS design is briefly described in the Sections 5.1 to 5.2.[1] Sections 5.3 and 5.4 describe how the CSE-PHC departed from the basic CCHS Cycle 3.1 design.

5.1 Canadian Community Health Survey Population Coverage

The CCHS data is collected from people aged 12 years and over living in private dwellings within the 10 provinces and three territories. Specifically excluded from the survey's coverage are residents of Indian Reserves and Crown land, full-time members of the Canadian Armed Forces, inmates of institutions and residents of isolated areas. The CCHS represents approximately 98% of the Canadian population aged 12 years and over.

5.2 Canadian Community Health Survey Sample Design

To provide reliable estimates to the 122 Health Regions (HR), and given the budget allocated to the CCHS (Cycle 3.1), a sample of 130,000 respondents was desired. The sample allocation strategy consisting of three steps, gave relatively equal importance to the HRs and the provinces. In the first two steps, the sample was allocated among the provinces according to their respective populations and the number of HRs they contain. In the third step, each province's sample was allocated among its HRs proportionally to the square root of the estimated population in each HR.

The CCHS used three sampling frames to select the sample of households: 50% of the sample of households came from an area frame, 49% came from a list frame of telephone numbers and the remaining 1% came from a Random Digit Dialling (RDD) sampling frame. For most of the health regions, 50% of the sample was selected from the area frame and 50% from the list frame of telephone numbers.

The CCHS used the area frame designed for the Canadian Labour Force Survey (LFS) as its primary frame. The sampling plan of the LFS is a multistage stratified cluster design in which the dwelling is the final sampling unit. In the first stage, homogeneous strata were formed and independent samples of clusters were drawn from each stratum. In the second stage, dwelling lists were prepared for each cluster and dwellings, or households, were selected from the lists.

For the purpose of the plan, each province is divided into three types of regions: major urban centres, cities and rural regions. Geographic or socio-economic strata are created within each major urban centre. Within the strata, dwellings are regrouped to create clusters. Some urban centres have separate strata for apartments or for census enumeration areas (EA) in which the average household income is high. In each stratum, six clusters or residential buildings (sometimes 12 or 18 apartments) are chosen by a random sampling method with a probability proportional to size (PPS), the size of which corresponds to the number of households. The number six was used throughout the sample design to allow a one-sixth rotation of the sample every month for the LFS.

The other cities and rural regions of each province are stratified first on a geographical basis, then according to socio-economic characteristics. In the majority of strata, six clusters (usually census EAs) are selected using the PPS method. Where there is low population density, a three-step plan is used whereby two or three primary sampling units (PSU), which normally correspond

[1] For a detailed description of the CCHS Cycle 3.1 sample design see the Public Use Microdata File guide, Catalogue no. 82M0013GPE.

to groups of EAs, are selected and divided into clusters, six of which are sampled. The final sample is obtained using a systematic sampling of dwellings.

5.3 The Canadian Survey of Experiences with Primary Health Care Population Coverage

The target population for the CSE-PHC is defined in the same way as for the CCHS Cycle 3.1, except that it is limited to people aged 18 and over on January 16, 2007. In addition, because the CSE_PHC is intended to represent the population of Canada at the beginning of 2007 but its sample is selected from the CCHS Cycle 3.1 respondents, who were interviewed between January and December 2005, people who joined the target population between the two surveys are excluded. This does not affect people who were not yet 18 at the time of the CCHS Cycle 3.1, since the latter included people aged 12 and over.

5.4 Person Sampling Strategy and Sample Size

Since the CSE-PHC requires only a small portion of the CCHS Cycle 3.1 sample, the decision was made to sample only from the CCHS area frame respondents. The main reasons for this were that the area frame provides nearly complete coverage of the population and that using a single frame makes it easier to calculate the sampling weights. In addition, the CSE-PHC is intended to produce estimates at the national level only. For that reason, the sample from the CCHS Cycle 3.1 area frame was stratified only for the 122 Health Regions, and the sample was allocated across the HRs in proportion to their percentage of the population aged 18 and over. The aim of this approach is to reduce the design effect observed for estimates from the CCHS area frame. In the CCHS, because of the estimation goals, the sampling fraction is much larger in some smaller HRs. Within each HR, the CCHS respondents were ranked according to the CCHS (and consequently the LFS) other stratification and selection variables and a systematic random sample was selected.

The sample size of the CSE-PHC is 3,800 persons. The table below shows the number of persons sampled in each province and territory.

Provinces and Territories	**Sample Size**
Newfoundland and Labrador	65
Prince Edward Island	20
Nova Scotia	113
New Brunswick	92
Quebec	903
Ontario	1,463
Manitoba	137
Saskatchewan	114
Alberta	372
British Columbia	506
Territories	15
Canada	**3,800**

6.0 Data Collection

An introductory letter and a pamphlet containing information on the Health Council of Canada (HCC) were mailed to respondents approximately one week before data collection began. Collection for The Canadian Survey of Experiences with Primary Health Care (CSE-PHC) was carried out from mid-January to the last week of February 2007. This was a paper/pencil survey collected over the telephone, meaning the interviewers read respondents the questions over the telephone and recorded answers to the questions manually on paper questionnaires.

A front-end module of the questionnaire included 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 developed. We explained to respondents how they were selected for the survey, that their participation in the survey was voluntary, and that their information would remain strictly confidential.

The survey manager met with senior staff responsible for collection to discuss issues and questions before the start of the training session. A description of the background and objectives as well as a detailed description of concepts and definitions particular to the CSE-PHC was provided for interviewers in their Interviewer Manual. A glossary of terms and a set of questions and answers was also included.

Interviewers were trained on the survey content through a classroom training session. In addition, the interviewers completed a series of mock interviews to become familiar with the survey and its concepts and definitions. Question and answer documentation was provided to the interviewers to assist them in answering questions that are commonly asked by respondents.

The data collection was conducted by specialized staff at the Statistics Canada head office in Ottawa. To manage and facilitate the collection process a computer-assisted telephone interview (CATI) automated scheduler system was used (BLAISE) to ensure that cases were assigned randomly to interviewers and that they were called at different times of the day, and different days of the week, to maximize the probability of contact.

The average interview time was estimated to be 22 minutes. However, the length of the interviews varied depending on the circumstances of the respondent. For example, the average interview time was estimated to be 30 minutes for a respondent with chronic conditions and 12 minutes for those without chronic conditions. The overall average was closer to 30 minutes than the 22 originally estimated.

Completed questionnaires were sent for imaging and data capture.

6.1 Questionnaire Design

The Health Council of Canada had created a draft questionnaire containing over 120 questions. This draft questionnaire was created by combining questions from various surveys conducted in the United States, Canada and other Commonwealth countries. The draft version was reworked to harmonize concepts, definitions and reference periods. The new version was created to reflect the research goals and objectives of the HCC and contained original questions not in the draft. The length was dramatically reduced and the flow of the interview was improved. The redesign questionnaire was translated by Official Languages and Translation Division and tested in conjunction with Statistics Canada's Questionnaire Design and Review Centre (QDRC) using face to face interviews in both official languages. The testing was conducted with respondents from various age groups and ethnic backgrounds. A portion of the test group was comprised of people diagnosed with chronic conditions. Further changes to the questionnaire were implemented based on the results of the questionnaire testing process.

6.2 Supervision and Quality Control

The team of interviewers was under the supervision of senior interviewers responsible for ensuring that everyone was familiar with the concepts and procedures of the survey. Periodical monitoring of interviewers and the review of completed documents was done in accordance with collection protocol.

7.0 Data Processing

The main output of The Canadian Survey of Experiences with Primary Health Care (CSE-PHC) is a “clean” microdata file. This chapter presents a brief summary of the processing steps involved in producing this file.

The first phase of error detection was done during the data collection. At that stage, the interviewer's supervisors reviewed the completed questionnaires. Observed inconsistencies were discussed with the interviewer who conducted the interview and the respondent was called back if required.

The second phase of error detection was conducted during data processing which was made up of many steps. The first was a general clean-up of the data to accomplish the following goals:

1) remove duplicate records from the file,
2) verify the collected data against the sample file,
3) identify missing records, and,
4) create a response file.

The editing phase of the data processing included top-down flow edits to clean up any paths that may have been mistakenly followed during the interview. This step was followed by analyzing frequency distributions to identify anomalies, for example missing or invalid categories or unusual frequencies.

7.1 Data Capture

Interviewers read respondents the questions over the telephone and recorded the answers to the questions manually on paper questionnaires. Completed questionnaires were sent for imaging and data capture and processed at head office in Ottawa.

Some editing is done directly at the time of the interview. The response data are subjected to further edit and imputation processes once they arrive at the processing center in head office.

7.2 Editing

The first stage of survey processing undertaken at head office was the replacement of any “out-of-range” values on the data file 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 Coding of Open-ended Questions

There were no open-ended questions on this survey.

7.4 Imputation

Imputation is the process that supplies valid values for those variables that have been identified for a change either because of invalid information or because of missing information. The new values are supplied in such a way as to preserve the underlying structure of the data and to ensure that the resulting records will pass all required edits. In other words, the objective is not to reproduce the true microdata values, but rather to establish internally consistent data records that yield good aggregate estimates.

We can distinguish between three types of non-response. Complete non-response is when the respondent does not provide the minimum set of answers. These records are dropped and accounted for in the weighting process (see Chapter 11.0). Item non-response is when the respondent does not provide an answer to one question, but goes on to the next question. These are usually handled using the “not stated” code or are imputed. Finally, partial non-response is when the respondent provides the minimum set of answers but does not finish the interview. These records can be handled like either complete non-response or multiple item non-response.

Since the data collected on this survey dealt with respondents’ individual experiences with the Health Care system, no imputation was done.

7.5 Creation of Derived Variables

There were a few instances where sex and age recorded on the file differed from what was on the Canadian Community Health Survey (CCHS) Cycle 3.1 file. This may be the result of non-sampling error. In order to maintain consistency and to facilitate data analysis, both sex and age have been derived. In both cases the variables were derived using CCHS Cycle 3.1 information since it was deemed to be more accurate.

7.6 Weighting

The principle behind estimation in a probability sample such as the CSE-PHC 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 individuals who would definitely or probably recommend their primary care provider to a friend or relative is to be estimated, this would be done by selecting the records referring to those individuals in the sample with that characteristic and summing the weights entered on those records.

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

7.7 Suppression of Confidential Information

It should be noted that the only microdata file produced for the CSE-PHC is a share file. There is no “Public Use” Microdata File (PUMF) nor is there a “master” file held by Statistics Canada. The share file contains data for all respondents who agreed to share their data with the Health Council of Canada as well as those who agreed to allow Statistics Canada to link their survey data to the CCHS Cycle 3.1. It should be noted that linked data, in accordance with Statistics Canada confidentiality policies, is not included on the share file. Consequently, linked data is not shared with the Health Council of Canada. Since the share/link rate was very high, over 94%, it was felt that the creation of a master file was not warranted. All of the personal identifier information has been removed from the share file. This includes names, telephone numbers, street addresses and postal codes. 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.

8.0 Data Quality

8.1 Response Rates

A total of 3,800 people were selected to take part in the CSE-PHC. Of that number, 24 were no longer in the CSE-PHC's target population (for example, due to death). Of the 3,776 eligible people, 2,194 responded to the survey, for an overall response rate of 58.1%. The table below contains a summary of the CSE-PHC response rates by province.

Provinces and Territories	**CCHS Cycle 3.1 Selected Person**	**Potential Survey Respondents**	**CSE-PHC Respondents**	**Response Rate (%)**
Newfoundland and Labrador	65	65	38	58.5
Prince Edward Island	20	20	11	55.0
Nova Scotia	113	111	74	66.7
New Brunswick	92	91	56	61.5
Quebec	903	899	504	56.1
Ontario	1,463	1,456	897	61.6
Manitoba	137	136	82	60.3
Saskatchewan	114	113	56	49.6
Alberta	372	371	204	55.0
British Columbia	506	499	267	53.5
Territoires	15	15	5	33.3
Canada	**3,800**	**3,776**	**2,194**	**58.1**

8.2 Survey Errors

The estimates derived from this survey are based on a sample of persons. 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 were taken 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 the use of highly skilled interviewers, extensive training of interviewers with respect to the survey procedures and questionnaire, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions, procedures to ensure that data capture errors were minimized, and coding and edit quality checks to verify the processing logic.

8.2.1 The Frame

Because the CSE-PHC was a supplement to the Canadian Community Health Survey (CCHS) Cycle 3.1 (the area frame only) which was based on the Labour Force Survey (LFS), the quality of sample variables on the frame was very good as was the coverage. Note that the LFS frame excludes about 2% of all households in Canada. Therefore, the CSE-PHC frame also excludes the same proportion of households in the same geographical area. It is unlikely that this exclusion introduces any significant bias into the survey data.

It is important to note that the CSE-PHC interview took place 13 to 24 months after the CCHS Cycle 3.1 interview. For some people selected for the CSE-PHC, there was no telephone number in the sample frame, and for others, the number was out of date. Tracing was carried out in an effort to find a telephone number, but because of time and resource constraints, 557 of the 3,800 people selected could not be contacted.

8.2.2 Data Collection

Interviewer training consisted of reading the Interviewer's Manual and becoming familiar with the survey material. A description of the background and objectives of the survey was provided, as well as a glossary of terms and a set of questions and answers. The original collection period of January 16 to February 2, 2007, was extended to February 21, 2007.

Although the CSE-PHC was collected on paper, appointments and call backs were managed through the BLAISE software. All the interviewers were already familiar with this system.

8.2.3 Non-response

A major source of non-sampling errors in surveys is the effect of non-response on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. In the case of CSE-PHC there was little partial non-response because respondents tended to complete the questionnaire once they started the interview. 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 individuals who responded to the survey to compensate for those who did not respond. See Chapter 11.0 for more details on weighting adjustments for non-response. No imputation was done for partial non-response.

8.2.4 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 38.3% of Canadians were diagnosed or treated by a health care professional for at least one of the chronic conditions listed on the survey and this estimate is found to have a standard error of 0.012. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{0.012}{0.383}\right) X \ 100\ \% = 3.1\%$$

There is more information on the calculation of coefficients 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, analyzing, 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 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 Survey of Experiences with Primary Health Care (CSE-PHC) was not self-weighting. When producing simple estimates including the production of ordinary statistical tables, users must apply the proper survey weights.

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 CSE-PHC data can be tabulated and analyzed, 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 CSE-PHC.

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 would definitely or probably recommend their primary care provider to a friend or relative or the proportion of people who have been an overnight patient in a hospital, nursing home or convalescent home, for at least one night, in the past 12 months 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 general, would you say YOUR health is:
R: Excellent / Very good / Good / Fair / Poor

Q: In the past 12 months, did you require any routine or on-going care?
R: Yes / No

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 nights spent as a patient in a hospital, nursing home or convalescent home in the past 12 months by respondents who spent at least one night in such a facility. The numerator ($\hat{X}$) is an estimate of the total number of nights spent in institutions in the past 12 months and its denominator ($\hat{Y}$) is the number of persons who reported having spent at least one night in such a facility.

Examples of Quantitative Questions:

Q: For how many nights in the past 12 months?
R: |_|_|_| nights

Q: Including yourself, how many persons usually live here?
R: |_|_| persons

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 ($\hat{X}$),
b) summing the final weights of records having the characteristic of interest for the denominator ($\hat{Y}$), then
c) dividing estimate a) by estimate b) ($\hat{X} / \hat{Y}$).

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 average number of times women saw or talked to a family doctor or general practitioner about their physical, emotional or mental health in the past 12 months, multiply the value reported in question D02 (number of times women saw or talked to a family doctor or general practitioner) by the final weight for the record, then sum this value over all records with D_SEX = 2 (women).

To obtain a weighted average of the form $\hat{X} / \hat{Y}$, the numerator ($\hat{X}$) is calculated as for a quantitative estimate and the denominator ($\hat{Y}$) is calculated as for a categorical estimate. For example, to estimate the average number of times women saw or talked to a family doctor or general practitioner about their physical, emotional or mental health in the past 12 months,

a) estimate the total number of times ($\hat{X}$) as described above,
b) estimate the number of women ($\hat{Y}$) in this category by summing the final weights of all records with D_SEX = 2, then
c) divide estimate a) by estimate b) ($\hat{X} / \hat{Y}$).

9.4 Guidelines for Statistical Analysis

The CSE-PHC 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 may 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 D_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 D_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 estimates from the CSE-PHC 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 rounded weighted 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 E (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 F (or some similar identifier) and the following warning should accompany the estimates: “Please be warned that these estimates [flagged with the letter F] 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 Survey of Experiences with Primary Health Care

The following table provides an indication of the precision of population estimates as it shows the release cut-offs associated with each of the three quality levels presented in the previous section. These cut-offs are derived from the coefficient of variation (CV) tables discussed in Chapter 10.0.

For example, the table shows that the quality of a weighted estimate of 145,000 women possessing a given characteristic in the 65 and over age group is marginal.

Note that these cut-offs apply to estimates of population totals only. To estimate ratios, users should not use the numerator value (nor the denominator) in order to find the corresponding quality level. Rule 4 in Section 10.1 and Example 4 in Section 10.1.1 explain the correct procedure to be used for ratios.

Age Group	**Sex**	**Acceptable CV 0.0% to 16.5%**	**Marginal CV 16.6% to 33.3%**	**Unacceptable CV > 33.3%**
Less than 65 years	Men	690,000 & over	180,000 to < 690,000	under 180,000
	Women	605,000 & over	155,000 to < 605,000	under 155,000
	All	665,000 & over	167,000 to < 665,000	under 167,000
65 and over	Men	381,000 & over	111,000 to < 381,000	under 111,000
	Women	330,000 & over	92,000 to < 330,000	under 92,000
	All	382,000 & over	101,000 to < 382,000	under 101,000
All	Men	654,000 & over	168,000 to < 654,000	under 168,000
	Women	560,000 & over	142,000 to < 560,000	under 142,000
	All	**616,000 & over**	**155,000 to < 616,000**	**under 155,000**

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 in this case, 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 age group and sex, which were used to produce the Approximate Sampling Variability Tables for The Canadian Survey of Experiences with Primary Health Care (CSE-PHC).

Age Group	Sex	Design Effect	Sample Size	Population
Less than 65 years	Men	1.45	757	10,539,613
	Women	1.51	913	10,558,375
	All	1.48	1,670	21,097,988
65 and over	Men	1.40	196	1,822,869
	Women	1.53	328	2,254,479
	All	1.47	524	4,077,348
All	**Men**	**1.45**	**953**	**12,362,482**
	Women	**1.54**	**1,241**	**12,812,854**
	All	**1.51**	**2,194**	**25,175,336**

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* estimate 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 is most likely unacceptable and Statistics Canada recommends not to release such an estimate, regardless of the value of the coefficient of variation.

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 age/sex groups, 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 people taking prescription medication regularly who experienced side effects in the past 12 months is more reliable than the estimated number of people taking prescription medication regularly who experienced side effects in the past 12 months.

When the proportion or percentage is based upon the total population 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 suffering from a chronic disease), 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 people who needed routine or on-going care for the past 12 months and the numerator is the number of people who, over the past 12 months, had difficulty accessing the services they needed.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of people who needed routine or on-going care for the past 12 months as compared to the number of people who needed immediate care for a minor health problem for the same period, the standard error 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 CVs 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 CSE-PHC 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 8,914,814 persons needed routine or on-going care in the past 12 months. How does the user determine the coefficient of variation of this estimate?

1) Refer to the coefficient of variation table for CANADA - All ages.

2) The estimated aggregate 8,914,814 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 9,000,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 3.4%.

4) So the approximate coefficient of variation of the estimate is 3.4%. The finding that 8,914,814 (to be rounded according to the rounding guidelines in Section 9.1)

persons needed routine or on-going care in the past 12 months is publishable with no qualifications.

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

Suppose that the user estimates 2,362,719 / 8,914,814 = 26.5% of persons who needed routine or on-going care in the past 12 months reported experiencing difficulties getting the services they needed. How does the user determine the coefficient of variation of this estimate?

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

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., those who needed routine or on-going care over the past 12 months), it is necessary to use both the percentage (26.5%) and the numerator portion of the percentage (2,362,719) in determining the coefficient of variation.

3) The numerator, 2,362,719, 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 2,000,000. 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 8.1% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 8.1%. The finding that 26.5% of persons who needed routine or on-going care in the past 12 months and reported experiencing difficulties getting the services they needed can be published with no qualifications.

Example 3: Estimates of Differences Between Aggregates or Percentages

Suppose that a user estimates the proportion of persons who needed routine or on-going care in the past 12 months and reported experiencing difficulties getting the services they needed was 1,942,741 / 7,991,443 = 24.3% for persons who had a regular doctor, and 419,978 / 923,371 = 45.5% for persons who didn't have a regular doctor. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the CANADA – All ages coefficient of variation table in the same manner as described in Example 2 gives the CV of the estimate for persons who had a regular doctor as 8.1%, and the CV of the estimate for persons who didn't have a regular doctor as 14.7%.

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 (persons who had a regular doctor), $\hat{X}_2$ is estimate 2 (persons who didn't have a regular doctor) 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 $\hat{d} = 0.243 – 0.455 = -0.212$ is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(0.243)(0.081)]^2 + [(0.455)(0.147)]^2} \\ &= \sqrt{(0.000387) + (0.004474)} \\ &= 0.070\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d} = 0.070 / 0.212 = 0.330$

4) So the approximate coefficient of variation of the difference between the estimates is 33.0%. The difference between the estimates is considered marginal and Statistics Canada recommends this estimate not be released. However, should the user choose to do so, the estimate should be flagged with the letter E (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 in the past 12 months 8,914,814 persons needed routine or on-going care while 7,414,864 persons needed immediate care for minor health problems. The user is interested in comparing the two estimates 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 persons who needed routine or on-going care. The denominator of the estimate ($\hat{X}_2$) is the number of persons who needed immediate care for a minor health problem.

2) Refer to the coefficient of variation table for CANADA – All ages.

3) The numerator of this ratio estimate is 8,914,814. The figure closest to it is 9,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 3.4%.

4) The denominator of this ratio estimate is 7,414,864. The figure closest to it is 7,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 4.2%.

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:

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0.034)^2 + (0.042)^2} \\ &= \sqrt{0.001156 + 0.001764} \\ &= 0.054\end{aligned}$$

6) The obtained ratio of the number of persons who needed routine or on-going care versus those who needed immediate care for a minor health problem was 8,914,814 / 7,414,864 which is 1.20 (to be rounded according to the rounding guidelines in Section 9.1). The coefficient of variation of this estimate is 5.4%, which makes the estimate releasable with no qualifications.

Example 5: Estimates of Differences of Ratios

Suppose that the user estimates that in the past 12 months the ratio of persons who needed routine or on-going care, to those who needed immediate care for a minor health problem was 1.06 for men and 1.34 for women. The user is interested in comparing the two ratios to see if there is a statistical difference between them. How does the user determine the coefficient of variation of the difference?

1) First calculate the approximate coefficient of variation for the ratio for men ($\hat{R}_1$) and the ratio for women ($\hat{R}_2$) as in Example 4. Refer to the coefficient of variation tables for CANADA – Men, All ages and CANADA – Women, All ages. The approximate CV for the ratio for men is 7.8% and 6.7% for the women.

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

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

where α_1 and α_2 are the coefficients of variation of $\hat{R}_1$ and $\hat{R}_2$ respectively. That is, the standard error of the difference $\hat{d}$ = 1.06 – 1.34 = -0.28 is:

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{\left[(1.06)(0.078)\right]^2 + \left[(1.34)(0.067)\right]^2} \\ &= \sqrt{(0.006836) + (0.008060)} \\ &= 0.122\end{aligned}$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.122 / (-0.28) = -0.436.

4) So the approximate coefficient of variation of the difference between the estimates is 43.6%. The difference between the estimates is considered unacceptable and Statistics Canada recommends this estimate not be released. However, should the user choose to do so, the estimate should be flagged with the letter F (or some similar identifier) and be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimate.

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 difference 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 persons who needed routine or on-going care in the past 12 months and reported experiencing difficulties getting the services they needed (from Example 2, Section 10.1.1) would be calculated as follows:

$\hat{X}$ = 26.5% (or expressed as a proportion 0.265)

t = 2

$\alpha_{\hat{x}}$ = 8.1% (0.081 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$$CI_{\hat{x}} = \{0.265 – (2)\ (0.265)\ (0.081),\ 0.265 + (2)\ (0.265)\ (0.081)\}$$

$$CI_{\hat{x}} = \{0.265 – 0.043,\ 0.265 + 0.043\}$$

$$CI_{\hat{x}} = \{0.222,\ 0.308\}$$

With 95% confidence it can be said that between 22.2% and 30.8% of persons who needed routine or on-going care in the past 12 months experienced difficulty getting the services they needed.

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 $t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ 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 that the user wishes to test, at 5% level of significance, the hypothesis that for persons who needed routine or on-going care in the past 12 months and reported experiencing difficulties getting the services they needed, there is no difference between the proportion of persons who had a regular doctor and persons who didn’t have a regular doctor. From Example 3, Section 10.1.1, the standard error of the difference between these two estimates was found to be 0.070. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.243 - 0.455}{0.070} = \frac{-0.212}{0.070} = -3.03$$

Since t = -3.03 is less 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 CSE-PHC 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 times people have personally used a hospital emergency department in the past 12 months would be greater than the coefficient of variation of the corresponding proportion of people who have used these services. Hence, if the coefficient of variation of the proportion is unacceptable (making the proportion not releasable), then the coefficient of variation of the corresponding quantitative estimate will also be unacceptable (making the quantitative estimate not 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

Refer to CSE-PHC2006-2007_CVTabsE.pdf for the coefficient of variation tables.

11.0 Weighting

Since The Canadian Survey of Experiences with Primary Health Care (CSE-PHC) used a sub-sample of the Canadian Community Health Survey Cycle 3.1 (CCHS) sample, the derivation of weights for the survey records is clearly tied to the weighting procedure used for the CCHS. The CCHS weighting procedure is briefly described below.

11.1 Weighting Procedures for the Canadian Community Health Survey

Both an area frame and a telephone frame were used for the CCHS Cycle 3.1. As noted in Section 5.4, only respondents from the area frame were eligible for the CSE-PHC. In the CCHS, the respondents from each of the two frames are weighted separately before the two frames are combined. Hence, the initial CSE-PHC weight is the weight of the selected CCHS respondents, as calculated before the frames are combined. That weight is supposed to make it possible for the area frame sample to properly represent all of the survey's target population. The weighting strategy for units from the CCHS area frame is described in detail in the Public Use Microdata File User Guide for the CCHS Cycle 3.1. The CCHS Cycle 3.1 area frame final weight takes into account the selection probability for each household, household-level non-response, household person selection and person-level non-response.

11.2 Weighting Procedures for The Canadian Survey of Experiences with Primary Health Care

The initial weight for the CSE-PHC is the CCHS Cycle 3.1 area frame final weight. It is adjusted to compensate for the selection of a small sample of CCHS respondents and for the CSE-PHC non-response. The weights are also adjusted to control for the presence of outlier weights and to ensure that the estimates for the CSE-PHC match the population projections for certain population subgroups. All of the adjustments are explained in this section.

Selecting the sample
The initial weight taken from the CCHS Cycle 3.1 provides an adequate representation of the target population as long as all respondents are included. For the CSE-PHC, a sample of 3,800 respondents was selected at random from all eligible area-frame respondents. The sample was chosen independently in each Health Region (HR) by means of systematic random sampling. Thus, each CCHS area frame respondent aged 18 and over in a given HR had the same probability of being selected. The CSE-PHC selection weight was combined with the initial weight provided by the CCHS in such a way as to ensure that the sum of the weights of all respondents in each HR remained unchanged.

The CSE-PHC adjusted selection weight is given by:

$$CSE-PHC\ weight = \left(\frac{CCHS\ Cycle\ 3.1\ weight\ * \sum weights\ of\ all\ CCHS\ Cycle\ 3.1\ area\ frame\ respondents\ 18\ years\ and\ over\ in\ the\ HR}{\sum weights\ of\ all\ respondents\ selected\ for\ the\ CSE-PHC\ in\ the\ HR} \right)$$

Non-response adjustment
For various reasons, some people could not be interviewed for the CSE-PHC. In some cases, current contact information was unavailable. In others, the collection period ended before the respondent could be contacted. Other people refused to participate in the survey. Thus, part of the CSE-PHC initial sample was "lost", and an adjustment factor had to be applied to the weights of responding persons to compensate for that non-response.

The weight of non-respondents is redistributed to respondents within response groups. This is done using logistic regression. A model to predict the probability of responding to the survey was built using the variables available for all persons selected for the CSE-PHC. Because so much information was available from the CCHS, there was a wide range of options for building the response model. Using the model, respondents were divided into nine groups on the basis of their probability of responding to the survey. Groups of equal size were created, except for the group with the smallest response probabilities, which was twice the size of the eight others. Each non-respondent was then added to the group that matched his/her own response probability. In each group, the weight of the respondents was then increased by a factor equal to the sum of the weights of all units in the response group divided by the weight of all respondents in the group.

The possibility of making separate adjustments for different types of non-response (no contact information, etc.) was considered, but since that approach did not appear to improve the adjustment, only one adjustment was made.

Controlling for outlier weights
Because respondent weights undergo a number of successive adjustments, first by the CCHS and then by the CSE-PHC, some units may end up with weights that are substantially different from the weights of the other respondents in the same population group, or even weights that are outliers. In other words, some respondents may represent an abnormally large proportion of their group and strongly influence the estimates for those groups. To prevent that, the weight of respondents who make an outlier contribution to their population group is adjusted downward by a method known as “winsorization”. Because of the small sample size, that adjustment had to be confined to rather large groups. The groups used were composed of:

- six regions: the Atlantic Provinces, Quebec, Ontario, Alberta, British Columbia and the Yukon, and the remaining provinces and territories;
- four age groups: 18 to 34, 35 to 49, 50 to 64 and 65 and over; and
- gender.

Very few units had their weights “winsorized”.

Post-stratification
The last step in determining the final weight for the CSE-PHC is post-stratification. That technique is used to ensure that the sum of the final weights matches the population estimates for each of the above-mentioned 48 groups (six regions, four age groups and gender). The population estimates for January 20, 2007, were used for post-stratification. The CSE-PHC final weight is given by:

$$Final\ \ weight = \left(\frac{Winsoried\ \ weight\ \ *\ the\ \ population\ \ projections\ \ for\ \ the\ \ group\ \ to\ \ which\ \ the\ \ respondents\ \ belong}{\sum weights\ \ of\ \ all\ \ respondents\ \ in\ \ the\ \ group} \right)$$

The resulting weight WTPS is the final weight that appears in the CSE-PHC Share microdata file.

12.0 Questionnaires

Refer to CSE-PHC2006-2007_QuestE.pdf for the English questionnaire used to collect the data.

13.0 Record Layout with Univariate Frequencies

See CSE-PHC2006-2007_CdBk.pdf for the record layout with univariate counts.