# HEALTH SERVICES ACCESS SURVEY

# USER GUIDE

May 2002

Statistics Canada Statistique Canada

Canada

## Table of Contents

# *1.0 Introduction*

The Health Services Access Survey was conducted by Statistics Canada in November 2001 with the cooperation and support of Health Canada and the provinces of Prince Edward Island, Alberta and British Colombia. 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

Claudia Sanmartin
Senior Analyst
Health Analysis and Measurement Group
Statistics Canada
Tunney's Pasture
Ottawa (Ontario)
K1A 0T6
Tel.: (613) 951-6059
Fax: (613) 951-3959
e-mail: claudia.sanmartin@statcan.ca

Christian Houle
Chief, Analysis Section
Health Analysis and Measurement Group
Statistics Canada
Tunney's Pasture
Ottawa (Ontario)
K1A 0T6
Tel: (613) 951-3767
Fax: (613) 951-3959
e-mail: houlchr@statcan.ca

Jean-Marie Berthelot
Manager,
Health Analysis and Measurement Group
Statistics Canada
Tunney's Pasture
Ottawa (Ontario)
K1A 0T6
Tel: (613) 951-3760
Fax: (613) 951-3959
e-mail: berthel@statcan.ca

## *2.0 Background*

Following from the September, 2000 First Ministers' Health Accord, it was agreed that federal, provincial and territorial governments would report to their constituents starting in September 2002. Those reports would contain measures for 14 health indicator areas. Within the quality of service, there were two measures for which data were not available: a) waiting time for key diagnostic and treatment service and b) access to 24/7 first contact health services. The Health Services Access Survey was designed to collect this information including patient experiences, acceptance and perceptions of waiting for care.

A number of frame options were evaluated such as a supplement to the Labour Force Survey, a Random Digit Dialling survey and recontacting respondents from the Canadian Community Health Survey. It is the last option that was selected as it provided more background health information on the respondents and the survey could be conducted within the time frame available.

The content of the survey, while not expected to address perfectly the need for information on these two indicator areas, should be able to shed considerable light on Canadians' experiences and perceptions with regard to the general question of access to health services.

## *3.0 Objectives*

The objective of the survey was to provide information on the experiences of respondents in using some selected health care services. The survey focussed on two main topics: waiting for specialized services for a new illness or condition and access to basic health care.

The topics in the area of waiting times included:

- experience of respondents requiring care from a medical specialist such as a cardiologist, allergist, etc.;
- experience of respondents requiring non-emergency surgery such as cardiac surgery, joint surgery, etc.;
- experience of respondents requiring diagnostic tests: MRIs, CAT scans and angiographies .

Questions on their experiences covered information such as waiting time to access the service, acceptability of the waiting time and impact of the wait on the respondent.

The topics covered in the access to basic health care services included:

- experience of respondents in getting health information or advice;
- experience of respondents in getting health care services for routine or on-going care;
- experience of respondents in getting immediate care for minor heath problems such as fever, headache; sprained ankle, vomiting or unexplained rash, etc.
- experience of respondents in getting health care services in general.

Questions on their experiences covered information on access to services at different times of the day, difficulties encountered in getting services, type of services required, and where services were obtained.

# 4.0 Concepts and Definitions

This chapter outlines concepts and definitions of interest to the users. The concepts and definitions used in the Canadian Communities Health Survey are described in section 4.1 while those specific to the Health Services Access Survey are given in section 4.2. Users are referred to Chapter 12 of this document for a copy of the actual survey forms used.

## 4.1 Canadian Communities Health Survey Concepts and Definitions

Information from the Canadian Community Health Survey can be found in the documentation provided by the CCHS survey team with the data released.

## 4.2 Health Services Access Survey Concepts and Definitions

24/7

24 hours a day, 7 days a week

Access

The process when you are seeking and/or receiving health care services or health information.

Affected

Any effects, good or bad, experienced due to waiting for health services such as: stress and anxiety, pain, loss of income.

Diagnostic Test

An MRI, CT scan or angiography requested by your physician to determine or confirm a diagnosis. Does NOT include x-rays, blood tests, etc.

Difficulties (accessing health care services or information)

Any problems getting health services or information such as: getting a referral or appointment, waiting too long, service not available in the area, getting a family doctor, etc.

Family Member

A family member living in the household for whom the respondent is responsible for the health care needs (e.g. child, parent, spouse)

Health Care Information or Advice

Information regarding a new or existing health condition or disease obtained from a health care professional or telephone based service (e.g. tele-triage). Information may be about appropriate treatments or care or about who to contact for care (e.g. doctor, emergency room, hospital clinic). Does NOT include information from a resource book or the internet. Need a live interaction.

Medical Specialist

Any medical doctor with a specialization e.g. cardiologist, oncologist, radiologist, psychiatrist, orthopedist, gynaecologist, etc.; does not include optometrists or family or general practitioners.

Minor Health Problem

Fever, vomiting, major headache, ankle sprained, minor burns, cuts, skin irritation, unexplained rash, etc. Non life threatening health problems or injuries due to a minor accident.

New illness or condition

An illness or condition that was just newly diagnosed; does not include specialist visits for on-going care for a previously diagnosed situation.

Non-emergency Surgery or Diagnostic Test

A booked or planned surgery or diagnostic test provided on an outpatient or inpatient basis. Does not refer to surgery provided through an admission to the hospital emergency room as a result of, for example, an accident or life-threatening situation.

Over Night Hospital Stay

Patient is admitted to a hospital and occupies a hospital bed for at least one night. Does not include overnight stay in an emergency room.

Patient in a hospital

Patient is admitted to a hospital and occupies a bed.

Routine or On-going Care

Health care provided by a family or general practitioner including an annual check-up, blood tests or routine care for an on-going illness (e.g. prescription refill)

# *5.0 Survey Methodology*

The Health Services Access Survey was administered in November and December 2001 to a sub-sample of the respondents to the Canadian Communities Health Survey (CCHS) sample. Sections 6.1 and 6.2 give a short overview of the CCHS sample design. Section 5.3 shows the modifications to this design that took place to accommodate the HSAS. Section 5.4 indicates the HSAS sample size by province.

## *5.1 Population Coverage*

The CCHS targets persons aged 12 years or older who are living in private dwellings in the ten provinces and the three territories. Persons living on Indian Reserves or Crown lands, clientele of institutions, full-time members of the Canadian Armed Forces and residents of certain remote regions are excluded from this survey. The CCHS covered approximately 98% of the Canadian population aged 12 or older.

## *5.2 Sample Design*

The CCHS uses a combination of two sampling frames to select their sample. The first is the same frame used by the Canadian Labour Force Survey (LFS). This survey design is a stratified, multi-stage design employing probability sampling at all stages. In regions where the LFS frame could not supply a sufficient sample size for the CCHS, a Random Digit Dialling (RDD) method was used to supplement the area frame sample.

Within each selected household, either one or two people aged 12 and over were selected. The number depended upon the household composition and was intended to increase the representation of two age groups of interest - youths (aged 12-19) and seniors (age 65+).

For more details on the CCHS sample design, see Béland (2002).[1]

## *5.3 Modifications to the CCHS design for the HSAS Supplement*

A sub-sample of CCHS respondents were selected to receive the HSAS. The sample size varied from one province to the next, depending upon the size of the province and whether the provincial ministry of health decided to supplement the collection of extra sample. In order for a CCHS respondent to be eligible to be selected for the HSAS he/she had to meet the following criteria
i) The individual had to be at least fifteen years old as of November 1, 2001
ii) The individual had to have agreed to share his/her CCHS data with the provincial partners during the CCHS interview
iii) The individual had to have provided Statistics Canada with a telephone number at which he/she could be contacted
iv) No sample was selected from the three territories

In addition to the above requirements, there were other restrictions. Only one person per household could be selected (in some cases CCHS had selected two). Households that were selected for

[1]Béland (2002). Canadian Community Health Survey - Methodological overview. Health Reports, Vol. 13, No. 3,
9-14

September or October 2000 CCHS collection were ineligible to receive the HSAS (October respondents in PEI were the exception to this rule). Finally, people who were chosen for the CCHS via the Random Digit Dialling frame were not eligible to be selected, with the exception of those households chosen from five health regions where no CCHS sample had been selected from the LFS frame.

Once the sampling frame and sample size were finalized, the HSAS sample was selected by using a systematic sampling method from the frame sorted by a geographic identifier. This allowed all regions of a province to be covered by the HSAS.

## *5.4 Sample size by Province for the HSAS Supplement*

The following table shows the number of responding individuals from the CCHS who were selected to receive the HSAS supplement.

| PROVINCE | SAMPLE SIZE |
|---|---|
| Newfoundland and Labrador | 1000 |
| Prince Edward Island | 1259 |
| Nova Scotia | 1000 |
| New Brunswick | 1000 |
| Quebec | 1250 |
| Ontario | 1400 |
| Manitoba | 1000 |
| Saskatchewan | 1000 |
| Alberta | 3868 |
| British Columbia | 4839 |
| CANADA | 17616 |

# 6.0 Data Collection

Data collection for the HSAS was carried out between November 19 and December 28, 2001.

## 6.1 Interviewing for the HSAS

Statistics Canada interviewers are part-time employees hired and trained specifically to conduct Computer Assisted Personal Interviews for the Labour Force Survey and other major surveys such as the CCHS.

The interviews for the Health Services Access Survey were conducted by telephone with the selected respondent. Proxy interviews were not allowed for the HSAS.

## 6.2 Non-Response to the HSAS

In total, 17,616 individuals were eligible for the Health Services Access Survey. A total of 14,210 of these individuals either responded or were found to no longer be in our target population (dead or institutionalized) for a response rate of 81%. The overall response rate to the CCHS had been 84.68%.

# 7.0 Data Processing

The main output of the HSAS is a "clean" microdata file. This chapter presents a brief summary of the processing steps involved in producing this file.

## 7.1 Data Capture

Capture of survey data was accomplished using minicomputers. During this process any document containing at least one interviewer-completed item was captured and an unedited version of the computer record was electronically transmitted to Ottawa for further processing. In total 17,616 documents were captured and transmitted for the survey.

## 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

A few data items on the questionnaire were recorded by interviewers in an open-ended format. A total of nine partially open-ended questions were included in the survey. These questions related to type of surgery and health conditions. A number of possible answers were provided for each question but giving the possibility to enter a write-in answer if the answer was not included in the list. In some cases, the answer given was recoded to an existing code while in other cases new answer categories were created.

## 7.4 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. All questions relating to waiting time were recoded to “days” as the respondents could provide an answer in using days, weeks or months. In order to simplify analysis the information was recoded into days. This process involved nine questions.

## 7.5 Weighting

The principle behind estimation in a probability sample such as the HSAS 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 have had contact with a specialist during the past 12 months is to be estimated, it is 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.

# 8.0 Data Quality

## 8.1 Response Rates

The following table summarizes the response rates to the CCHS and to the HSAS.

| | Number of CCHS Respondents | CCHS Overall Response Rate* | HSAS Sample Size | Number of HSAS Completed Cases** | HSAS Percentage of Completed Cases | Number of Records on HSAS File*** | Percentage of Cases on HSAS File |
|---|---|---|---|---|---|---|---|
| Newfoundland and Labrador | 3,870 | 86.82% | 1,000 | 851 | 85.10% | 773 | 77.30% |
| Prince Edward Island | 3,651 | 84.73% | 1,259 | 1,063 | 84.43% | 1,012 | 80.38% |
| Nova Scotia | 5,319 | 88.83% | 1,000 | 870 | 87.00% | 854 | 85.40% |
| New Brunswick | 4,996 | 88.50% | 1,000 | 835 | 83.50% | 800 | 80.00% |
| Québec | 22,667 | 85.60% | 1,250 | 1,096 | 87.68% | 1,030 | 82.40% |
| Ontario | 39,278 | 81.98% | 1,400 | 1,171 | 83.64% | 1,099 | 78.50% |
| Manitoba | 8,470 | 89.46% | 1,000 | 843 | 84.30% | 799 | 79.90% |
| Saskatchewan | 8,009 | 86.78% | 1,000 | 873 | 87.30% | 833 | 83.30% |
| Alberta | 14,456 | 85.09% | 3,868 | 3,237 | 83.69% | 3,087 | 79.81% |
| British Columbia | 18,302 | 84.65% | 4,839 | 4,074 | 84.19% | 3,923 | 81.07% |
| CANADA | 131,535 | 84.68% | 17,616 | 14,913 | 84.66% | 14,210 | 80.67% |

* The CCHS response rate accounts for non-response at both the household and person level.
** The number of HSAS completed cases includes selected individuals who were found to be dead or living in an institution at the time of the HSAS interview.
*** Completed cases that were dropped from the HSAS file include those who people were found to be dead or institutionalized at the time of the HSAS interview, who were later found not to have passed all of the CCHS minimum edit criteria or who did not agree to share their HSAS data with the survey partners

## 8.2 Survey Errors

The estimates derived from this survey are based on a sample of households. Somewhat different figures might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. as those actually used. The difference between the estimates obtained from the sample and the results 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 included 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 HSAS was a supplement to the CCHS, any impacts that the frame had on the CCHS apply to the HSAS as well. The CCHS used both the LFS frame and an RDD frame. The LFS frame excludes about 2% of all households in the 10 provinces of Canada. This includes people living on Indian reserves and in extremely remote regions. The RDD frame excludes all households that do not have a telephone number (slightly more than 1% of the population).

Any non-response to the CCHS had an impact on the HSAS frame. The CCHS weighting procedures were derived in order to minimize this impact. In addition, certain individuals were excluded from HSAS based upon the CCHS available. There were approximately 4% of the people for whom a complete date of birth was not available. However for most of these people, a year and month of birth was available so that the age could be easily estimated. For the rest, an approximate age was included, so an age as of November 2001 could be derived for all individuals. Approximately 5% of the CCHS respondents did not provide Statistics Canada with a phone number, which was a necessity to be eligible to be selected for the HSAS. Likewise, about 4% of the respondents had not agreed to share their CCHS data with the provincial partners and therefore were also ineligible to be selected for the HSAS. Both of these groups were accounted for in the weighting steps. If this sub-group of ineligible people are different from the overall population, then it could impact upon the quality of the results.

### *8.2.2 Data Collection*

Interviewer training consisted of reading the Health Services Access Interviewers' Manual, practicing with the training cases on the laptop computer, and discussing any questions with senior interviewers before the start of the survey. 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 collection period ran from November 19th to December 28th, 2001.

### *8.2.3 Data Processing*

During processing of the data, a small number of record were dropped from the file because of data quality issues. As well, item non-response is separated in three categories i.e. “Don’t know” where the respondent was unable to provide the information, “refused” where the respondent did not want to respond to the question and “Not stated” for missing

information because of a “Don’t know” or “refused “ in a previous question. The number of partial response are minimal for this survey usually less than five by item.

### *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 22.5% of Canadians aged 15 and over visited a medical specialist for a new illness or condition in the past 12 months, and this estimate is found to have standard error of 0.0079. Then the coefficient of variation of the estimate is calculated as :

$$\left( \frac{0.0079}{0.225} \right) \; x \; 100\% \; = \; 3.5\%$$

There is more information on the calculation of CV in Chapter 11.

# *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 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 HSAS was not self-weighting. When producing simple estimates, including the production of ordinary statistical tables, users must apply the proper sampling weight.

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

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

### *9.2.1 Definitions of types of estimates: Categorical vs. Quantitative*

Before discussing how the HSAS 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 HSAS.

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 or proportion of people visiting a specialist in the past twelve 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 the past 12 months, did you require a visit to a medical specialist for a diagnosis or a consultation for a new illness or condition?
R: Yes / No

Q: Were you referred by ...?
R: A family doctor / Another specialist / Another health care provider

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 X/Y where X is an estimate of surveyed population quantity total and 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 days that a person had to wait to see a specialist. The numerator is an estimate of the total number of days that all people who saw a specialist had to wait for their appointment and its denominator is the number of persons who saw a specialist.

Example of Quantitative Questions :

Q: How long did you have to wait to see the specialist?
R: |_|_|_|_| Days

### *9.2.2 Tabulation of Categorical Estimates*

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

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

### *9.2.3 Tabulation of Quantitative Estimates*

Estimates of quantities can be obtained from the microdata file by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest. For example, to obtain an estimate of the total number of days that all people who had a visit to a specialist in the last twelve months waited for that visit, multiply the value reported in WT_D08 (number of days that the person had to wait) by the final weight for the record, then sum this value over all records with WT_Q02=1 (required a visit to a medical specialist in the last twelve months).

To obtain a weighted average of the form X/Y, the numerator (X) is calculated as for a quantitative estimate and the denominator (Y) is calculated as for a categorical estimate. For example, to estimate the average number of days that people who visited a specialist felt was an acceptable waiting time,

(a) estimate the total number of days by multiplying the answer to WT_D11 by the final weight for the record for all people in this category (WT_Q02=1),
(b) estimate the number of people in this category by summing the final weights of all records with WT_Q02=1, then
(c) divide estimate (a) by estimate (b).

## *9.3 Guidelines for Statistical Analysis*

The HSAS 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. Variances for simple estimates

such as totals, proportions and ratios (for qualitative variables) are provided in 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:

- select all respondents from the file who reported SEX=male
- 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=male
- for each of these respondents, calculate a RESCALED weight equal to the original person weight divided by the AVERAGE weight
- 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 truly meaningful 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.

## *9.4 CV Release Guidelines*

Before releasing and/or publishing any estimate from HSAS, 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 9. 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 9 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:<br>a sample size of 30 or more, and<br>low coefficients of variation in the range 0.0% - 16.5%<br><br>No warning is required. |
| 2. Marginal | Estimates have:<br>a sample size of 30 or more, and<br>high coefficients of variation in the range 16.6% - 33.3%.<br><br>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:<br>a sample size of less than 30, or<br>very high coefficients of variation in excess of 33.3%.<br><br>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:<br><br>"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." |

## *9.5 Release cut-off's for the HSAS*

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 introduced in the previous sub-section. These cut-offs are derived from the CV tables discussed in Section 11.

For example, the table shows that the quality of a weighted estimate of 20,000 people possessing a given characteristic in Newfoundland and Labrador is marginal.

Note that these cut-offs are appropriate for estimates of population totals only. In the case of estimate of ratios, users should not use the numerator value (nor the denominator) in order to find the corresponding quality level. Rule 4 and Example 4 in Section 11.1 explain the right procedure to follow in the case of a ratio.

**Table of Release Cut-offs**

| Province | Acceptable | Marginal | Unacceptable |
|---|---|---|---|
| Newfoundland and Labrador | 30,000 & over | 8,000 to < 30,000 | under 8,000 |
| Prince Edward Island | 6,000 & over | 2,000 to <6,000 | under 2,000 |
| Nova Scotia | 53,000 & over | 14,000 to < 53,000 | under 14,000 |
| New Brunswick | 37,000 & over | 10,000 to < 37,000 | under 10,000 |
| Quebec | 391,000 & over | 101,000 to < 391,000 | under 101,000 |
| Ontario | 611,000 & over | 158,000 to < 611,000 | under 158,000 |
| Manitoba | 81,000 & over | 21,000 to < 81,000 | under 21,000 |
| Saskatchewan | 57,000 & over | 15,000 to < 57,000 | under 15,000 |
| Alberta | 53,000 & over | 14,000 to <53,000 | under 14,000 |
| British Columbia | 49,000 & over | 49,000 to < 12,000 | under 12,000 |
| CANADA | 339,000 & over | 84,000 to <339,000 | under 84,000 |

# *10.0 Approximate Sampling Variability Tables*

In order to supply coefficients of variation which would be applicable to a wide variety of categorical estimates produced from this microdata file and which could be readily accessed by the user, a set of Approximate Sampling Variability Tables has been produced. These 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 (CV) 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 (the 75$^{th}$ percentile) to be used in the look-up 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.

| PROVINCE | DESIGN EFFECT | NUMBER OF RESPONDENTS | POPULATION |
|---|---|---|---|
| Newfoundland and Labrador | 1.52 | 773 | 439155 |
| Prince Edward Island | 1.6 | 1012 | 111095 |
| Nova Scotia | 1.75 | 854 | 753068 |
| New Brunswick | 1.4 | 800 | 606417 |
| Quebec | 1.95 | 1030 | 6003232 |
| Ontario | 2.05 | 1099 | 9528227 |
| Manitoba | 2.23 | 799 | 863421 |
| Saskatchewan | 1.84 | 833 | 760508 |
| Alberta | 1.91 | 3087 | 2389364 |
| British Columbia | 1.61 | 3923 | 3298846 |
| Canada | 5.36 | 14210 | 24753333 |

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 an other. 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 CV tables for Categorical Estimates*

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

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

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

**Rule 2: Estimates of Proportions or Percentages Possessing a Characteristic**

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the 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 who had a non-emergency surgery in the reference period is more reliable than the estimated number of people who had a non-emergency surgery in the reference period. (Note that in the tables the cv's decline in value reading from left to right).

When the proportion or percentage is based upon the total population of the geographic area covered by the table, 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 ($d = X_1 - X_2$) is:

$$\sigma_d = \sqrt{(X_1\alpha_1)^2 + (X_2\alpha_2)^2}$$

where $X_1$ is estimate 1, $X_2$ is estimate 2, and $\alpha_1$ and $\alpha_2$ are the coefficients of variation of $X_1$ and $X_2$ respectively. The coefficient of variation of đ is given by $\sigma_d/d$. 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 had a non-emergency surgery and the numerator is the number of “people who had a cancer related non-emergency surgery”.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of "people who required health information during regular office hours in the reference period" as compared to the number of "people who required health information during evenings and weekends in the reference period", the standard deviation of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by R. That is, the standard error of a ratio ($R = X_1 / X_2$) is:

$$\sigma_R = R\sqrt{\alpha_1^2 + \alpha_2^2}$$

where $\alpha_1$ and $\alpha_2$ are the coefficients of variation of $X_1$ and $X_2$ respectively. The coefficient of variation of R is given by $\sigma_R/R$. The formula will tend to overstate the error if $X_1$ and $X_2$ are positively correlated and understate the error if $X_1$ and $X_2$ 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 CV tables for Categorical Estimates*

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

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

Suppose that a user estimates that 1,388,106 Canadians had a non-emergency surgery in the reference period. How does the user determine the coefficient of variation of this estimate?

(1) Refer to the cv table for CANADA.

(2) The estimated aggregate (1,388,106) 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 1,500,000.

(3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 7.5%.

(4) So the approximate coefficient of variation of the estimate is 7.5%.
The finding that there were 1,388,106 Canadians who had a non-emergency surgery in the reference period is publishable with no qualifications.

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

Suppose that the user estimates that 147,918/1,388,106=10.67% of the Canadians who had a non-emergency surgery in the reference period reported that it was a hip or knee replacement surgery. How does the user determine the coefficient of variation of this estimate?

(1) Refer to the table for CANADA.

(2) Because the estimate is a percentage which is based on a subset of the total population (i.e., people who had a non-emergency surgery), it is necessary to use both the percentage (10.67%) and the numerator portion of the percentage (147,918) in determining the coefficient of variation.

(3) The numerator, 147,918, 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 150,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the figure closest to it, 10.0%.

(4) The figure at the intersection of the row and column used, namely 23.7% is the coefficient of variation to be used.

(5) So the approximate coefficient of variation of the estimate is 23.7%. The finding that 10.67% of the Canadians who had a non-emergency surgery reported that it was a hip or knee replacement surgery can be published, but should be flagged with the letter M (or some similar identifier). It should be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimate.

**Example 3 : Estimates of Differences Between Aggregates or Percentages**

Suppose that a user estimates that 608,351/5,559,472=10.94% of Canadians who required a visit to a medical specialist for a diagnosis or a consultation for a new condition reported the reason for the consultation as heart condition or stroke, while 420,518/5,559,472=7.56% reported the reason as arthritis or other joint conditions. How does the user determine the coefficient of variation of the difference between these two estimates?

(1) Using the CANADA cv table for in the same manner as described in example 2 gives the cv of the estimate for heart condition as 13.0%, and the cv of the estimate for arthritis as 14.5%.

(2) Using Rule 3, the standard error of a difference ($d = X_1 - X_2$) is:

$$\sigma_d = \sqrt{(X_1\alpha_1)^2 + (X_2\alpha_2)^2}$$

where $X_1$ is estimate 1, $X_2$ is estimate 2, and $\alpha_1$ and $\alpha_2$ are the coefficients of variation of $X_1$ and $X_2$ respectively.

That is, the standard error of the difference d = (0.1094-0.0756) = 0.0338 is:

$$\sigma_d = \sqrt{[(0.1094)(0.130)]^2 + [(0.0756)(0.145)]^2}$$

$$= \sqrt{(0.000202) + (0.000120)}$$

$$= 0.0179$$

(3) The coefficient of variation of đ is given by $\sigma_d/d = 0.0179/0.0338 = 0.5296$.

(4) So the approximate coefficient of variation of the difference between the estimates is 52.96%.Statistics Canada recommends not to release this estimate. However, if the user chooses to do so then it 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."

**Example 4 : Estimates of Ratios**

Suppose that the user estimates that 1,388,106 Canadians had a non-emergency surgery in the reference period, while 1,855,213 Canadians had a diagnostic test during the same period. The user is interested in comparing these 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 (= $X_1$ ) is the number of Canadians who reported having a non-emergency surgery in the reference period. The denominator of the estimate (= $X_2$ ) is the number of Canadians who reported having a diagnostic test in the reference period.

(2) Refer to the table for CANADA.

(3) The numerator of this ratio estimate is 1,388,106. The figure closest to it is 1,500,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 7.5%.

(4) The denominator of this ratio estimate is 1,855,213. The figure closest to it is 2,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 6.5%.

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

$$\alpha_R = \sqrt{\alpha_1^2 + \alpha_2^2}$$

where $\alpha_1$ and $\alpha_2$ are the coefficients of variation of $X_1$ and $X_2$ respectively.

That is ,

The obtained ratio of Canadians who had a non-emergency surgery versus those

$$\alpha_R = \sqrt{(0.065)^2 + (0.075)^2}$$

$$= 0.0992$$

who had a diagnostic test is 1,388,106/1,855,213 which is 0.7482:1. The coefficient of variation of this estimate is 9.92%, which is releasable with no qualifications.

**Example 5: Estimates of Differences of Ratios**

Suppose that the user estimates that the ratio of people who had required health information to the people who had required health care services is 1.291:1 for Alberta while it is 1.659:1 for British Columbia. The user is interested in comparing the ratios between the two provinces. How does the user determine the coefficient of variation of the difference?

(1) First calculate the approximate CV for the Alberta ratio ($R_1$) and British Columbia ratio ($R_2$) using the same formula as in example 4. In this case, the provincial CV tables are used. The estimates for the number of people in Alberta who required health information and health care services are 1030545 and 798336 with CVs of 2.7% and 3.6% respectively. Similar estimates for British Columbia are 1683117 and 1014265 with CVs of 2.1% and 3.0%. Using the formula from example 4, the approximate CV for the Alberta ratio is 4.50% and 3.66% for British Columbia.

(2) Using rule 3, the standard error of a difference ($d = R_1 - R_2$) is:

$$\sigma_d = \sqrt{(R_1\alpha_1)^2 + (R_2\alpha_2)^2}$$

where $\alpha_1$ and $\alpha_2$ are the coefficients of variation of $R_1$ and $R_2$ respectively.

That is, the standard error of the difference d = (1.291.-1.659) =- 0.368 is:

$$\sigma_d = \sqrt{[(1.291)(0.0450)]^2 + [(1.659)(0.0366)]^2}$$

$$= \sqrt{(0.0034) + (0.0037)}$$

$$= 0.0843$$

(3) The coefficient of variation of d is given by $\sigma_d/d = 0.0843/(0.368) = 0.2291$.

(4) So the approximate coefficient of variation of the difference between the estimates is 22.91%. It should be flagged with the letter M (or some similar identifier). It should 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 CV tables to obtain Confidence Limits*

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

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

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

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

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

$$CI_X = [X - tX\alpha_X \, , \, X + tX\alpha_X]$$

where αX is the determined coefficient of variation of 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.

### *10.2.1 Example of using the CV tables to obtain confidence limits*

A 95% confidence interval for the estimated proportion of Canadians who had a non-emergency surgery in the reference period and reported that it was a hip or knee replacement surgery (from Example 2, section 11.1.1) would be calculated as follows.

X = 10.67% (or expressed as a proportion = 0.1067)

t = 2

$\alpha_X$ = 23.7% (0.237 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_X$ = {0.1067 - (2) (0.1067) (0.237), 0.1067 + (2) (0.1067) (0.237)}

$CI_X$ = {0.1067 - 0.0506, 0.1067 + 0.0506}

$CI_X$ = {0.0561, 0.1573}

With 95% confidence it can be said that between 5.61% and 15.73% of the Canadians who had non-emergency surgeries in the reference period reported a hip or knee replacement surgery.

## *10.3 How to use the CV tables to do a t-test*

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

Let $\mathbf{X_1}$ and $\mathbf{X_2}$ be sample estimates for 2 characteristics of interest. Let the standard error on the difference $X_1 - X_2$ be $\sigma_d$ .

If $$t = \frac{X_1 - X_2}{\sigma_d}$$

is between -2 and 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. That is to say that the characteristics are significantly different

### *10.3.1 Example of using the CV tables to do a t-test*

Let us suppose we wish to test, at 5% level of significance, the hypothesis that there is no difference between the ratio of people who required health information to those who required health care in Alberta and in British Columbia in the reference period. From example 5, section 11.1.1, the standard error of the difference between these two estimates was found to be = 0.0843. Hence ,

$$t = \frac{R_1 - R_2}{\sigma_d} = \frac{1.291 - 1.659}{0.0843} = \frac{-0.368}{0.0843} = -4.37.$$

Since t = -4.37 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 Health Services Access 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 length of time people had to wait for surgery would be greater than the coefficient of variation of the proportion of people who required surgery. 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 CV Tables*

The CV tables have been produced at the provincial and national levels and can be found in a file called HSAS_CVtable_can.pdf provides Cvs at the Canada level and HSAS_CVtable_prov.pdf provides the CV at the provincial levels.

# *11.0 Weighting*

Since the HSAS used a sub-sample of the CCHS sample, the derivation of weights for the HSAS survey records is clearly tied to the weighting procedure used for the CCHS. The CCHS weighting procedure is briefly described below followed by a more detailed explanation of the adjustments to these weights required for the HSAS.

**12.1 Weighting Procedures for the CCHS**
Since the CCHS used two frames from which to select their sample, weighting was done independently for the records from each of these two frames until the end of the weighting procedures, at which time they were integrated together. However the steps leading up to this integration were similar.

An initial household-level weight was derived by taking into account the probability of the household being selected, the under-coverage of the frame, and the removal of out of scope cases. This weight was then adjusted for household level non-response. Further adjustments then took place to account for the selection of one or two people within the household, followed by an adjustment for the non-responding selected people. This produced the final area frame and telephone frame weights. These were used as the basis for the HSAS weighting. The CCHS then integrated these two weights together and calibrated them to ensure that they added up to Census population projections.

For more details on the CCHS weighting procedures, see Béland (2002).[2]

## *11.1 Weighting Procedures for the HSAS*

The principles behind the calculation of the weights for the HSAS are identical to those for the CCHS. However, further adjustments are made to the CCHS weights in order to derive a final weight for the individual records on the HSAS microdata file.

The starting point for these weight adjustments are the CCHS final area frame weights and the final telephone frame weight. These weights will be referred to as W0. A total of eight adjustments were made

(1) **An adjustment to account for the months of CCHS collection that were not included in the HSAS frame:** Since the CCHS respondents from September and October 2000 were not eligible to receive the HSAS (with the exception of the PEI October respondents), an adjustment to the CCHS weights was necessary. This new weight (W1) applied to only those CCHS respondents in the in scope months and was equal to

$$W1 = W0 \times \frac{\Sigma\ W0 \text{ of all households of size } i}{\Sigma\ W0 \text{ of all households of size } i \text{ in the in scope months}}$$

[2]Béland (2002). Canadian Community Health Survey - Methodological overview. Health Reports, Vol. 13, No. 3, 9-14

This adjustment was made at the CCHS health region level. The household size was defined as the number of people aged twelve and over during the CCHS. All households of size of six or more people were grouped together. Any household size group that contained fewer than 25 households was collapsed with the next smaller household size group.

(2) **An adjustment to account for the households that did not provided the CCHS with a telephone number:** Since the HSAS was a telephone survey, an adjustment was needed to account for the CCHS respondents who did not provide Statistics Canada with a phone number. This new weight (W2) was applied only to people who had given Statistics Canada a phone number and was equal to

$$W2 = W1 \times \frac{\Sigma\ W1 \text{ of all households of size } i}{\Sigma\ W1 \text{ of all households of size } i \text{ who provided a phone number}}$$

This adjustment was made at the CCHS health region level. The household size was defined as the number of people aged twelve and over during the CCHS. All households of size of six or more people were grouped together. Any household size group that contained fewer than 25 households was collapsed with the next smaller household size group.

(3) **An adjustment to account for the CCHS respondents who did not agree to share their data:** Since CCHS respondents who did not agree to share their data were ineligible to be selected for the HSAS, it was necessary to adjust the weights for the people who did not agree to share. This new weight (W3) was applied only to CCHS respondents who had agreed to share their data and was equal to

$$W3 = W2 \times \frac{\Sigma\ W2 \text{ of all CCHS respondents}}{\Sigma\ W2 \text{ of all CCHS respondents who agreed to share their data}}$$

This adjustment was made at the household level

(4) **An adjustment to account for the selection of one individual from a household in cases where there were two eligible respondents from the CCHS:** In households with two HSAS-eligible respondents, one was selected at random to remain in the sampling frame. The new weight for this person (W4) was equal to W3 times two. The other person was dropped from the frame.

(5) **An adjustment to account for the sub-sampling of the CCHS respondents that took place for the HSAS**: At this point, the sampling frame for the HSAS was completed and the HSAS sub-sample was selected. An adjustment to account for this sub-sampling was made. The new weight (W5) was applied only to the people selected for the HSAS and was equal to

$$W5 = W4 \times \frac{\Sigma\ W4 \text{ of all people left on the HSAS sample frame}}{\Sigma\ W4 \text{ of all people selected for the HSAS}}$$

This adjustment was done at the health region level

(6) **An adjustment to account for records selected to receive the HSAS that were later re-coded to be CCHS non-respondents**: Following the selection of the HSAS sample, some people who were initially considered to be CCHS repondents were later determined to have not met the CCHS minimum edit criteria and therefore were re-coded to CCHS non-respondents. Since complete CCHS data was a requirement to be in the HSAS sample, these people had to be considered as HSAS non-respondents, even if they provided the HSAS with data. The new weight (W6) applied only to the HSAS selected sample that had valid CCHS data and was equal to

$$W6 = W5 \times \frac{\Sigma\ W5 \text{ of all people selected for HSAS}}{\Sigma\ W5 \text{ of all people selected for HSAS with valid CCHS data}}$$

This adjustment also took place at the health region level.

(7) **An adjustment to account for the non-response of people selected to be in the HSAS or the people who did not agree to share their data with the survey sponsors:** There was some non-response to the HSAS. An adjustment was necessary to account for the people from HSAS data was not received or who did not agree to share it with the survey sponsors. The new weight (W7) applied only to HSAS respondents who agreed to share their data and was equal to

$$W7 = W6 \times \frac{\Sigma\ W6 \text{ of all people in non-response group } i}{\Sigma\ W6 \text{ of all respondents who shared their data in non-response group } i}$$

The groupings were more complex for this adjustment. First, the data was broken into groups by province and four rural/urban codes - i) Census Metropolitan Areas (CMAs), ii) Census Agglomerations (CAs), iii) non-CMA/CA urban, iv) non-CMA/CA rural. Within each of these groups, the software package Knowledge Seeker was used to help identify the further non-response groupings. A number of CCHS and socio-demographic variables were input into this software and it determined which ones best predicted non-response. These variables were then used to create the non-response groupings. A minimum of thirty people were required within a grouping before it was accepted.

(8) **An adjustment to adjust the weights so that they sum to Census projections for independent province-sex-age groups and census metropolitan area (CMA) counts*:*** An adjustment was made to W7 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:
i) Population totals by province, sex and the following age groups: 15-24, 25-44, 45-64 and 65+. These totals were the average of the projections for November and December 2001.

ii) Total age 15+ populations for each of the 25 CMAs and the non-CMA areas in each province. These totals were the average of the projections for November and December 2001.

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 resulting weight (FINWT) is the final weight which appears on the HSAS microdata file.

# *12.0 Questionnaires*

## *12.1 The Supplementary Survey Questionnaire*

The Health Service Access Survey questionnaire was used in November and December 2001 to collect the information for the survey. Note that the data was collected using a Computer Assisted Telephone Interviewing system and that some of the edits used in the application are not reflected in the paper version of the questionnaire. The file HSASQuestionnaire_E.pdf contains the English questionnaire.

## *13.0 File Layout with Univariate Frequencies*

The record layout and univariate counts for the Health Service Access Survey can be found in a file called HSASCodeBook_E.pdf.