CANADIAN COMMUNITY HEALTH SURVEY /

HEALTH SERVICES ACCESS SURVEY
(2003)

MASTER FILE DOCUMENTATION

STATISTICS CANADA

December 2004

TABLE OF CONTENTS

1. INTRODUCTION 1

2. BACKGROUND AND OBJECTIVES 2

3. SURVEY CONTENT 4

4. SAMPLING DESIGN 5

4.1 Target population 5
4.2 Sampling frame 5
4.3 Sample size and allocation 5
4.4 Sample allocation at the sub-provincial level 7
4.5 Sample selection of Part A 7
4.6 Sample selection of Part B 7
4.7 Sample allocation over the collection period 7

5. DATA COLLECTION 9

5.1 Questionnaire design and data collection method 9
5.2 Supervision and control 9
5.3 Interviewing 9

6. DATA PROCESSING 10

6.1 Editing 10
6.2 Coding 10
6.3 Creation of derived and grouped variables 10
6.4 Weighting 10

7. WEIGHTING 11

7.1 Weighting procedures for the CCHS 11
7.2 Weighting procedures for the HSAS 12
7.2.1 Weighting steps 13

8. DATA QUALITY 21

8.1 Response rate 21
8.2 Survey errors 23
8.2.1 Non-sampling errors 23
8.2.2 Sampling errors 24
8.2.2.1 Bootstrap Method for Variance Estimation 24
8.2.2.2 Bootvar Program for Variance Estimation 25

9. GUIDELINES FOR TABULATION, ANALYSIS AND RELEASE 26

9.1 Rounding guidelines 26
9.2 Sample weighting guidelines for tabulation 26
9.2.1 Definitions: categorical estimates, quantitative estimates 27
9.2.2 Tabulation of categorical estimates 27
9.2.3 Tabulation of quantitative estimates 28
9.3 Guidelines for statistical analysis 28
9.4 Release guidelines 28

10. FILE USAGE 30

10.1 Use of weights 30
10.2 Variable naming convention 30
10.2.1 Variable name component structure in HSAS 30
10.2.2 Positions 1-3: variable / questionnaire section name 31
10.2.3 Position 4: cycle 31
10.2.4 Position 5: variable type 31
10.2.5 Positions 6-8: variable name 32
10.3 Access to master file data 32

1. Introduction

The Health Services Access Survey (HSAS) was released for the first time in 2002 as a supplement to the Canadian Community Health Survey (CCHS) - 2000-2001 (Cycle 1.1), as a one-time survey. For the cycle 2.1 of the CCHS it was decided to incorporate the HSAS into its content, but to release it as a separate file.

This file contains data collected in the second year of collection for the HSAS. Information was collected between January 2003 and December 2003, for 126 health regions, covering all 10 provinces. The HSAS 2003 collects responses from persons aged 15 or older, living in private occupied dwellings. Excluded from the sampling frame were individuals living on Indian Reserves and on Crown Lands, institutional residents, full-time members of the Canadian Armed Forces, and residents of certain remote regions.

This document has been produced to facilitate the manipulation of the HSAS (2003) master file, which is described in detail in the following text.

Any questions about the data sets or their use should be directed to:

Data Access Unit, Health Statistics Division: 1 (613) 951-1653
Email: cchs-escc@statcan.ca

For custom tabulations/general data support:
Client Custom Services, Health Statistics Division: 1 (613) 951-1746
Email: hd-ds@statcan.ca

2. Background and objectives

Following the September 2000 First Ministers' Health Accord, it was agreed that federal, provincial and territorial governments would report to their constituents on the results and performance of the health care systems starting in September 2002. These 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 (HSAS) was designed to collect this information including patient experiences, acceptance and perceptions of waiting for care.

Information on 24/7 access to first contact services includes:

- 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.; and
- experience of respondents in getting health care services in general.

Respondents were asked about their use of first contact services at different times of the day, difficulties accessing services, and where services were obtained.

Information on access to specialized services includes:

- experience of respondents requiring care from a medical specialist such as a cardiologist, allergist, etc. to obtain a diagnostic for a new illness or condition;
- experience of respondents requiring non-emergency surgery such as cardiac surgery, joint surgery, etc.; and
- experience of respondents requiring selected diagnostic tests: MRIs, CAT scans and angiographies.

Respondents were asked about their experiences accessing these services including waiting times, acceptability of the waiting time and impact of the wait on the respondent.

The Health Services Access Survey was released for the first time in 2001 as a supplement to the Canadian Community Health Survey cycle 1.1, as a one-time survey. For the cycle 2.1 of the CCHS it was decided to incorporate the HSAS into its content and is now part of this statistical program, but was released as a separate file. For the next cycle of the CCHS, to be released in 2006, this survey will be completely amalgamated and there will no longer be a separate file and a separate questionnaire.

The HSAS 2003 consisted of two questionnaire modules: Access to Health Care Services and Waiting Times. Data collection for HSAS was integrated with CCHS in the following ways:

Part A: A total of 18 981 respondents were interviewed for HSAS at the same time as their CCHS interview. All of these respondents are also members of sub-sample 3. (Waiting Times is not included on the CCHS data file.) These interviews were conducted between September and December 2003 (plus 45 interviews in January 2004).

Part B: A total of 13 024 respondents were re-contacted after having been interviewed previously for CCHS. All of these respondents **were not** part of sub-sample 3 during their first interview. These respondents were asked the Access to Health Services and Waiting Times modules during their second interview. Respondents to Part B of HSAS are not included in Sub-sample 3.

3. Survey content

The HSAS was designed to collect information about access to first contact services and specialized services. First contact services include routine care, health information or advice, and immediate care for a minor health problem provided by a family or general physician, nurse or other health care provider, not including medical specialists. Specialized services are services including specialist visits for a new illness or condition, non-emergency surgery other than dental surgery, and selected diagnostic tests (non-emergency MRIs, CT scans, and angiographies).

The HSAS 2003 consisted of two questionnaire modules: Access to Health Care Services and Waiting Times. In the first module, questions focused on difficulties accessing first contact services such as routine care, health information or advice, and immediate care for a minor health problem for individuals or members of their families. Those who reported difficulties were asked about the type of barriers they faced. Individuals were also asked whether they had a regular family physician.

The second module of the HSAS focused on access to specialized services, such as specialist visits for a new illness or condition, non-emergency surgery and selected diagnostic tests. Waiting times clearly remain an issue for access to specialized services. The HSAS provides comparable self reported waiting time data at the national and provincial levels. To ensure comparability, standard definitions for waiting times for specialist visits, nonemergency surgery and diagnostic tests were used. Respondents were asked whether they felt their waiting time was acceptable and whether waiting for specialized services affected their lives. The HSAS also provides information on other difficulties and barriers faced by those accessing specialized services.

4. Sampling design

4.1 Target population

The HSAS targets persons aged 15 or older living in private dwellings in the ten provinces. Residents of the three territories, persons living on Indian Reserves or Crown lands, in institutions, full-time members of the Canadian Forces and residents of certain remote regions are excluded from this survey. The HSAS covered approximately 98% of the population aged 15 or older in the 10 provinces.

4.2 Sampling frame

Given that the HSAS is a sub-sample of the CCHS 2.1, the HSAS uses the same sample frames as the CCHS, which uses a multiple frames design. First, it uses the area frame designed for the Canadian Labour Force Survey (LFS). The sampling plan of the LFS is a multistage stratified cluster design in which the dwelling is the final sampling unit. The CCHS uses two types of telephone frames: a list frame of telephone numbers and a Random Digit Dialling (RDD) sampling frame of telephone numbers. For more detailed information in regards to these sampling frames, please refer to the CCHS 2.1 users' guide[1].

4.3 Sample size and allocation

For the purpose of producing reliable estimates for the ten provinces, namely in regards to the estimations of the waiting times, the following sample sizes (table 4.1) were established.

Table 4.1 Targeted sample size by province for HSAS

Province	Targeted total sample size
Newfoundland and Labrador	2,400
Prince Edward Island	1,200
Nova Scotia	3,000
New Brunswick	3,000
Quebec	4,000
Ontario	4,800
Manitoba	3,200
Saskatchewan	3,200
Alberta	3,600
British Columbia	4,000
Canada	**32,400**

In order to reach these targeted sample sizes, a sub-sample of the CCHS respondents was selected. However, due to the timeframe for the implementation of the HSAS (summer-fall 2003), the sub-sample selection of CCHS respondents had to be divided in two parts. First, a first part of the HSAS sample (named the part A) comes directly from the CCHS. From September 2003, the waiting times module was added to the CCHS questionnaire and this module along with the other HSAS modules were asked as part of the CCHS interview for a

[1] Statistics Canada (2004). *CCHS 2.1 Public Use Microdata File.* Catalogue No. 82M0013XCB.

sub-sample of the CCHS sample cases between the months of September and December 2003. However, in certain provinces, the CCHS sample size was insufficient to reach the HSAS targeted sample size in some provinces even when selecting 100% of the CCHS respondents for that period. In order to meet the sample size requirements, it was decided to re-contact (by telephone) a sub-sample of the CCHS respondents interviewed between January and September 2003. This part of the HSAS sample is named the part B. It is important to mention that, for consistency purposes, the part A/part B split was forced to be approximately the same in every province although it was not necessary in some provinces. The table 4.2 presents the sample allocation for the two HSAS parts.

Since the CCHS respondents were selected from two sampling frames (area and telephone), the HSAS sample allocation was also controlled at the sampling frame level. For part A, an equal allocation was made between the two types of frames (approximately 50% of the CCHS sample came from the area frame and the other 50% from the telephone frame). For the part B, a larger proportion of the sample was allocated to the CCHS respondents selected from the telephone frame because the re-contacts were going to be conducted over the phone. But because in three provinces all (eligible) CCHS respondents had to be contacted to reach targeted sample sizes, sample was allocated to both types of frames everywhere for consistency purposes. Table 4.2 also gives the samples by sampling design.

Table 4.2 Targeted sample sizes by HSAS part (A & B)

Province	Part A (targeted sizes)			Part B (targeted sizes)		
	Total	Area Frame	Phone Frame	Total	Area Frame	Phone Frame
Newfoundland and Labrador	880	440	440	1,520	720	800
Prince Edward Island	560	280	280	640	310	330
Nova Scotia	1,200	600	600	1,800	850	950
New Brunswick	1,160	580	580	1,840	740	1,100
Quebec	3,000	1,500	1,500	1,000	250	750
Ontario	3,800	1,900	1,900	1,000	250	750
Manitoba	1,640	820	820	1,560	520	1,040
Saskatchewan	1,810	905	905	1,390	410	980
Alberta	3,000	1,500	1,500	600	200	400
British Columbia	3,200	1,600	1,600	800	200	600
Canada	**20,250**	**10,125**	**10,125**	**12,150**	**4,450**	**7,700**

4.4 Sample allocation at the sub-provincial level

The CCHS aims to produce reliable estimates at the health region level whereas the HSAS aims to produce reliable estimates at the provincial level. For this reason, the CCHS sample allocation by health region is not optimal for the HSAS. Therefore and in the provinces where it was possible, the HSAS sample was allocated within each province in such a way that the sample size by health region is as close as possible to the sample size that would be obtained with a sample allocation proportional to the health region sizes (the optimal allocation). For example, the Montréal-Centre region represents 25% of the population in Québec, but in CCHS, only 13% of the Québec total sample was allocated to that region. For HSAS, the sample was allocated in such a way that the Montréal-Centre region gets approximately 25% of the Québec total sample.

4.5 Sample selection of Part A

Once the sample was allocated to the two parts of HSAS, the sample of part A was selected in each health region by frame (area and telephone). This selection consisted in randomly choosing a sub-sample of dwellings already selected in the CCHS sample (between September and December 2003) according to the targeted sample sizes in each health region by sampling frame. In some provinces, the part A sample consisted in all dwellings selected in the CCHS during that period. Note that the targeted sample sizes for the part A were previously inflated to take into account the anticipated non-response and the vacant dwellings.

4.6 Sample selection of Part B

The HSAS part B consisted in re-contacting (over the phone) a sub-sample of CCHS respondents. However, not all the CCHS respondents were eligible to be re-contacted for the HSAS. The eligibility criteria were the following:

i) The respondent was interviewed for CCHS between January and August 2003;
ii) The respondent was aged 15 or older at the time of CCHS interview;
iii) The respondent was not part of the sub-sample 3 of CCHS (i.e.,had not already completed the “Access to health care services” module);
iv) Was not a proxy respondent in CCHS;
v) A valid and unique telephone number had to be provided by the respondent (this criterion applied only to those selected from the area frame).

Once the eligible respondents were identified, the sample selection of the part B consisted simply in choosing randomly a sub-sample of eligible respondents according to the targeted sample sizes in each health region by sampling frame. Note that the targeted sample sizes for the part B were previously inflated to take into account the anticipated non-response. A response rate of 80% was assumed for the part B.

4.7 Sample allocation over the collection period

The data collection of the HSAS part A sample cases was conducted between September and December 2003 (a few cases were completed at the beginning of January 2004). The part A sample was divided into three collection periods: September, October and November. In general, the sample size for September and October were identical whereas the sample size for November was half of the size of the other two months.

There was only one collection period for the part B sample cases. The collection started in October 2003 and ended in December 2003.

5. Data collection

5.1 Questionnaire design and data collection method

The HSAS 2003 questionnaire was administered using computer-assisted interviewing (CAI). Sample units selected from the area frame were interviewed using the Computer-Assisted Personal Interviewing (CAPI) method while units selected from the Random Digit Dialling (RDD) and telephone list frames were interviewed using the Computer-Assisted Telephone Interviewing (CATI) method.

CAI offers a number of data quality advantages over other collection methods. First, question text, including reference periods and pronouns, is customised automatically based on factors such as the age and sex of the respondent, the date of the interview and answers to previous questions. Second, edits to check for inconsistent answers or out-of-range responses are applied automatically and on-screen prompts are shown when an invalid entry is recorded. Immediate feedback is given to the respondent and the interviewer is able to correct any inconsistencies. Third, questions that are not applicable to the respondent are skipped automatically.

5.2 Supervision and control

CAPI interviewers worked independently from their homes using laptop computers and were supervised from a distance by senior interviewers. Completed interviews were transmitted daily to Statistics Canada's head office using a secure telephone transmission directly from the interviewer's home.

CATI interviewers worked in centralised offices and were supervised by a senior interviewer located in the same office. Transmission of cases from each of 5 CATI offices to head office was the responsibility of the regional office project supervisor, senior interviewer and the technical support team.

An automated call scheduler, ie. a central system to optimise the timing of call-backs and the scheduling of appointments, was <u>not</u> available to support CATI collection. Instead, at the start of each collection period a batch of cases was assigned to each personal computer in each CATI office. The caseload on each PC was then managed manually. Because the number of CATI cases was relatively small, this approach was reasonably efficient and the absence of a call scheduler is not thought to have had an adverse effect on data quality.

5.3 Interviewing

For Part A of HSAS, CAPI interviewers were trained to make an initial personal contact with each sampled individual. In cases where this initial visit resulted in non-response, telephone follow-ups were permitted.

All Part B interviews were conducted by telephone, either by CAPI interviewers working from their homes or by CATI interviewers working in centralised call centres.

6. Data processing

6.1 Editing

Most editing of the data was performed at the time of the interview by the computer-assisted interviewing (CAI) application. It was not possible for interviewers to enter out-of-range values and flow errors were controlled through programmed skip patterns. For example, CAI ensured that questions that did not apply to the respondent were not asked.

In response to some types of inconsistent or unusual reporting, warning messages were invoked but no corrective action was taken at the time of the interview. Where appropriate, edits were instead developed to be performed after data collection at Head Office. Inconsistencies were usually corrected by setting one or both of the variables in question to "not stated".

6.2 Coding

Pre-coded answer categories were supplied for all suitable variables. Several questions in the HSAS 2003 questionnaire allow write-in responses. For some of these questions, write-in responses were coded into one of the existing listed categories if the write-in information duplicated a listed category.

6.3 Creation of derived and grouped variables

To facilitate data analysis, a number of variables on the file have been derived using items found on the HSAS 2003 questionnaire. Derived variables generally have a "D" or "G" in the fifth character of the variable name. In some cases, the derived variables are straightforward, involving collapsing of response categories. In other cases, several variables have been combined to create a new variable. Please consult the Derived Variables Documentation for more information.

6.4 Weighting

The principle behind estimation in a probability sample such as the HSAS 2003 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. In the terminology used here, it can be said that each person has a weight of 50.

The weighting phase is a step that calculates, for each person, his or her associated sampling weight. 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 smoke daily is to be estimated, it is done by selecting the records referring to those individuals in the sample having that characteristic and summing the weights entered on those records. Details of the method used to calculate sampling weights are presented in Section 7.

7. Weighting

The sampling weight of a respondent can be seen as the number of persons in the population represented by the respondent. When computing a point estimate, survey weights must be used with survey data in order to get estimates representative of the covered population.

As described in the section 5, the HSAS is a sub-sample of the CCHS 2.1. The derivation of weights for the HSAS is strongly tied to the weighting procedures used for the CCHS. In fact, the starting weight used for the HSAS weighting procedures is one of the intermediate subweights of CCHS (after a certain step). The CCHS weighting procedures are briefly described below in section 7.1 and it is followed, in section 7.2, by a detailed explanation of the weighting procedures for the HSAS.

7.1 Weighting procedures for the CCHS

The CCHS was designed to provide reliable estimates for 126 health regions in the 10 provinces. Consequently, the sampling frames used for the CCHS were stratified into those 126 health regions. The CCHS used two types of frames; an area frame and a telephone frame. The targeted total sample size of 130,000 responding persons was first allocated to the provinces and then to the health regions within each province. Finally, within each health region, the sample was again allocated to the two sampling frames (about half of the sample in each frame). In very few health regions, only one sampling frame was used in the CCHS. The weighting of the CCHS was done independently for each of the two frames. An integration of the two frames was then performed before adjusting the weights to minimize seasonal effects and before performing the poststratification adjustment. The weighting procedures for the CCHS are summarized in the diagram A.

Diagram A Weighting Strategy Overview for the CCHS

Area Frame	Telephone Frame
A0 – Initial weight	T0 – Initial weight
A1 – Sample increase	T1 – Number of months
A2 – Stabilization	T2 – Removal of out-of-scope numbers
A3 – Removal of out-of-scope units	T3 – Coverage of the list frames
A4 – Household non-response	T4 – Combination of the list frames
A5 – Creation of person level weight	T5 – Household non-response
A6 – Person non-response	T6 – Households without telephone
Final area frame weight	T7 – Creation of person level weight
↘	T8 – Person non-response
	T9 – Multiple lines
	Final telephone frame weight
	↙
I1 – Integration	
I2 – Seasonal effect	
I3 - Post-stratification	
Final CCHS (Cycle 2.1) weight	

For more details on the CCHS weighting procedures, the reader should refer to the CCHS 2.1 users' guide.

7.2 Weighting procedures for the HSAS

As described in the section 4, the HSAS sample was divided into two parts: part A and part B. The CCHS weights after integration were used as the starting point for the weighting of the two parts of the HSAS. The weighting of each part was done separately and, at the end, the weights of the two parts were integrated. A poststratification was performed on the HSAS integrated weights. The weighting procedures for the HSAS are summarized in the diagram B below.

Diagram B HSAS weighting steps

0	Initial weight (CCHS weight)
1	Removal of out-of-scope persons
2	Part A and Part B split

3	**Part A**
3.1	Uncovered sampling frames
3.2	Sub-sampling of part A
3.3	Non-response

4	**Part B**
4.1	Uncovered health regions
4.2	Uncovered sampling frames
4.3	Exclusion of CCHS sub-sample 3 respondents
4.4	Exclusion of respondents from multiple dwellings
4.5	Exclusion of proxy respondents
4.6	Exclusion of respondents with invalid telephone numbers
4.7	Sub-sampling of part B
4.8	Non-response

5	Integration of part A and part B
6	Poststratification → Master weight
7	Share adjustment
8	Poststratification → Share weight

7.2.1 Weighting steps

Adjustment 0 – Initial weight

The CCHS weights after integration were used as the initial weights for the HSAS. These weights will be referred as weight 0 in HSAS.

Adjustment 1 – Removal of out-of-scope persons

As opposed to the CCHS, the following groups were not part of the target population of the HSAS: people who were less than 15 years old, lived in the three territories, or resided in the James Bay health region (Québec). The CCHS respondents who belong to one of these groups were considered out-of-scope and removed from the sample. The in-scope respondents kept the same weight as in the previous step, which is now called weight 1.

Adjustment 2 – Part A and part B split

The HSAS sample was selected in two parts: part A and part B. The part A consisted in a sub-sample of the CCHS dwellings between September and November 2003. The part B consisted in a sub-sample of the CCHS respondents in the collection periods of January to August 2003. Since, the collection strategies for these two parts were different and independent, it was decided to separately adjust the weights of the CCHS respondents associated to each part in such a way that each part represents the population covered by the survey (as if it was two independent surveys).

The selection of the sub-sample for part A and part B was done by controlling for the health region by sampling frame level. Therefore, all weight adjustments up to the adjustment for the selection of the sub-samples, were performed at the health region and sampling frame levels. The adjustment factor of this step was calculated as follows:

Part A

For the CCHS respondents in the collection periods of September to November 2003 (eligible for the HSAS part A), the adjustment factor was as follows:

$$\frac{\textit{Sum of weight 1 for all CCHS respondents}}{\textit{Sum of weight 1 for all CCHS respondents in the collection periods of Sept. to Nov. 2003}}$$

The weight 1 of the CCHS respondents in the collection periods of Sept. to Nov. 2003 was multiplied by this adjustment factor to produce weight 2.

Part B

For the CCHS respondents in the collection periods of January to August 2003 (eligible for the HSAS part B), the adjustment factor was as follows:

$$\frac{\textit{Sum of weight 1 for all CCHS respondents}}{\textit{Sum of weight 1 for all CCHS respondents in the collection periods of Jan. to Aug. 2003}}$$

The weight 1 of the CCHS respondents in the collection periods of Jan. to Aug. 2003 was multiplied by this adjustment factor to produce weight 2.

Adjustment 3 – Weighting steps for part A

For the weight adjustments that applied only to the HSAS part A, only the CCHS respondents in the collection periods of Sept. to Nov. 2003 were kept. For those, the weight is now called the weight 3. The adjustments 3.1 to 3.3 below apply to the HSAS part A only.

Adjustment 3.1 – Uncovered sampling frames

Originally, all CCHS respondents in the collection periods of Sept. to Nov. 2003 were supposed to have a chance of being selected in the HSAS part A sub-sample. However and due to some changes in the selection process of the CCHS sample for these collection periods, the telephone frame respondents of two health regions (Muskoka-Parry Sound and Northwestern in Ontario) did not have a chance to be selected in the HSAS part A sub-sample. The weights of those CCHS respondents who did not have a chance of being selected were re-distributed to the CCHS respondents from the area frame in those two health regions. In each of the two health regions, the adjustment factor was calculated as follows:

$$\frac{\textit{Sum of weight 3 for all CCHS respondents}}{\textit{Sum of weight 3 for all CCHS respondents from the area frame}}$$

For the CCHS respondents in the other health regions, the adjustment factor was set to 1. The weight 3 was multiplied by this adjustment factor to produce weight 3.1. The CCHS respondents from the telephone frame in the two problematic health regions were dropped out of the weighting process at this point.

Adjustment 3.2 – Sub-sampling of part A

At this point, all the CCHS respondents were eligible to be selected in the HSAS part A sub-sample. Therefore, a sub-sample was selected and an adjustment was done to account for this sample selection. Since the sub-sample selection was done at the health region and sampling frame levels the adjustment factor was calculated accordingly. The adjustment factor, which represents the inverse of the probability of selection, was calculated as the following:

$$\frac{\textit{Sum of weight 3.1 for all CCHS respondents}}{\textit{Sum of weight 3.1 for all selected persons in HSAS part A}}$$

The weight 3.1 of the selected persons in the HSAS part A sample was multiplied by this adjustment factor to produce weight 3.2. The CCHS respondents not selected in the HSAS part A sample were dropped out of the process at this point.

Adjustment 3.3 – Non-response

Although the remaining units are all CCHS respondents and that the HSAS interview was done at the same time as the CCHS interview, some of them could not be considered respondents for the HSAS. To be considered respondent for HSAS, the person had to have answered to all the HSAS modules. However, it was not the case for some of the CCHS respondents such as the proxy respondents (the HSAS modules were not asked to proxy respondents) and for some

partial CCHS respondents. For this reason, CCHS respondents who did not meet the criteria above were considered nonrespondents for HSAS. Consequently, the weight of those considered HSAS nonrespondents had to be distributed to the HSAS respondents using response propensity classes.

At this point in the weighting process, the adjustment classes could be defined at the provincial level. Because not all health regions are covered in the HSAS part B (see adjustment 4.1) and that there will be an integration of the part A and part B weights the adjustment classes was first divided into two groups within each province. The first group contained all the health regions covered in both part A and part B, and the second group contained the other health regions. For each province and for each group, response propensity classes were created. The CHAID (Chi-Square Automatic Interaction Detector) algorithm available in Knowledge Seeker[2] was used to identify which characteristics best split the sample into groups that were dissimilar with respect to response/non-response within each province by group. Depending on the group and province, the following characteristics were used to form the adjustment classes: restriction of activities, difficulty with social situations, hearing problems, smoking type, self-rated mental health, sex, age, education, and household size. As a result, in each class, an adjustment factor was calculated as follows:

$$\frac{\textit{Sum of weight 3.2 for all selected persons in HSAS part A}}{\textit{Sum of weight 3.2 for all responding persons in HSAS part A}}$$

The weight 3.2 of the responding persons in the HSAS part A sample was multiplied by this adjustment factor to produce weight 3.3. The nonresponding persons were dropped out of the weighting process at this point.

These are the final weights for the HSAS part A sample and they will be integrated with the HSAS part B final weights.

Adjustment 4 – Weighting steps for part B

For the weight adjustments that applied only to the HSAS part B, only the CCHS respondents in the collection periods of Jan. to Aug. 2003 were kept. For those, the weight 2 is now called the weight 4. The adjustments 4.1 to 4.8 below apply to the HSAS part B only.

Adjustment 4.1 – Uncovered health regions

Following the sample allocation of the HSAS part B, there were 21 health regions where no HSAS part B sample was allocated to; those regions were therefore not covered for this part. For that reason, the CCHS respondents in the collection periods of Jan. to Aug. 2003 from those regions were dropped out of the process at this point. No adjustment was done to the weight 4 of the CCHS respondents in the other health regions. Weight 4 is now called weight 4.1. From now, the HSAS part B covered the population of only 109 health regions (130 - 21).

[2] ANGOSS Software (1995). Knowledge Seeker IV for Windows – User's Guide. ANGOSS Software International Limited.

Adjustment 4.2 – Uncovered sampling frames

For the HSAS part B, there were also 31 other health regions (from the 109 health regions covered) where no HSAS sample was allocated to the CCHS respondents from the area frame. Therefore, within each of these health regions, the weights of the CCHS respondents from the area frame were distributed to the CCHS respondents from the telephone frame. In each of these 31 health regions, an adjustment factor was calculated as follows:

$$\frac{\textit{Sum of weight 4.1 for all CCHS respondents}}{\textit{Sum of weight 4.1 for all CCHS respondents from the telephone frame}}$$

For the CCHS respondents in the other health regions, the adjustment factor was set to 1. The weight 4.1 of the respondents in the covered frames was multiplied by its adjustment factor to produce weight 4.2. The CCHS respondents from the area frame in the 31 health regions were dropped out of the weighting process at this point.

Adjustment 4.3 – Exclusion of CCHS sub-sample 3 respondents

The CCHS sub-sample 3 consists in a subset of the CCHS respondents starting from the collection period April 2003 (there was no sub-sample 3 before April). The sample units selected for the sub-sample 3 were asked two additional questionnaire modules: “Patient satisfaction” and “Access to health care services”. As the HSAS part B consists mainly in asking the “Access to health care services” module and the module on waiting times, it was decided to exclude the CCHS respondents who had already answered to the “Access to health care services” module (i.e., in the CCHS sub-sample 3) from the HSAS part B. Within each health region/sampling frame combination, the weights of the respondents in the CCHS sub-sample 3 were distributed to the respondents that were not in the CCHS sub-sample 3 (between April and August 2003). The adjustment factor was calculated as follows:

$$\frac{\textit{Sum of weight 4.2 for all CCHS respondents}}{\textit{Sum of weight 4.2 for all respondents not in the CCHS sub - sample 3}}$$

For CCHS respondents before April 2003, the adjustment factor was set to 1. The new weight 4.3 was equal to weight 4.2 times its adjustment factor. The respondents in the CCHS sub-sample 3 were dropped out of the weighting process at this point.

Adjustment 4.4 – Exclusion of respondents from multiple dwellings

This adjustment only applies to the CCHS respondents selected from the area frame.

The CCHS respondents coming from multiple dwellings selected from the area frame did not have a chance of being selected for the HSAS for technical reasons; those respondents were excluded from the HSAS part B selection process. Their sampling weights were therefore distributed to the other respondents from the area frame within the same health region. The adjustment factor was calculated as follows:

$$\frac{\textit{Sum of weight 4.3 for all CCHS respondents}}{\textit{Sum of weight 4.3 for all respondents from non - multiple dwellings}}$$

The weight 4.3 of respondents from non-multiple dwellings was multiplied by this adjustment factor to produce weight 4.4. Respondents from multiple dwellings were dropped out of the weighting process at this step.

Adjustment 4.5 – Exclusion of proxy respondents

As the HSAS questionnaire modules cannot be completed by proxy, it was decided to not re-contact the CCHS proxy respondents in the HSAS part B sample. Therefore, the weights of the proxy respondents were distributed to the non-proxy respondents using classes. As for the non-response adjustment 3.3, the CHAID algorithm was used to identify which characteristics best split the sample into groups that were dissimilar with respect to proxy/non-proxy within each health region by sampling frame. Characteristics such as the household size, restriction of activities, difficulty with social situations, language(s) in with the respondent can converse, immigration flag, income, education, age, sex, and other characteristics were used to define classes. An adjustment factor was calculated as follows within each class defined at the health region and sampling frame levels:

$$\frac{\textit{Sum of weight 4.4 for all CCHS respondents}}{\textit{Sum of weight 4.4 for all CCHS non-proxy respondents}}$$

The weight 4.4 of non-proxy respondents was multiplied by this adjustment factor to produce weight 4.5. The CCHS proxy respondents were dropped out of the weighting process.

Adjustment 4.6 – Exclusion of invalid phone numbers

This adjustment only applies to the CCHS respondents selected from the area frame.

The HSAS part B was conducted by re-contacting the CCHS respondents using the telephone numbers provided during the CCHS interview. Respondents who refused to provide their telephone number or those who provided an incorrect phone number were excluded from the HSAS part B selection. The CHAID algorithm was again used to identify classes within each health region to determine which characteristics were most relevant in separating who were most likely to refuse giving their phone number or to provide an invalid phone number. For this adjustment, characteristics that were most often used were household income, tenure of the dwelling and household size. Within each class defined at the health region level, the adjustment factor was then calculated as follows:

$$\frac{\textit{Sum of weight 4.5 for all CCHS respondents}}{\textit{Sum of weight 4.5 for all CCHS respondents with valid phone numbers}}$$

The weight 4.5 of CCHS respondents with valid phone numbers was multiplied by this adjustment factor to produce weight 4.6. The CCHS respondents with invalid phone numbers were dropped out of the weighting process.

Adjustment 4.7 – Sub-sampling of part B

At this point, all the remaining CCHS respondents were eligible to be selected in the HSAS part B sample. Within each health region and controlling for the sampling frame, a sub-sample of those CCHS respondents was selected. An adjustment factor was calculated within each health region and sampling frame combination to reflect the probabilities of selection.

$$\frac{\textit{Sum of weight 4.6 for all CCHS respondents}}{\textit{Sum of weight 4.6 for all selected persons in the HSAS part B}}$$

The weight 4.6 of the selected persons in the HSAS part B sample was multiplied by this adjustment factor to produce weight 4.7. The CCHS respondents not selected in the HSAS part B sample were dropped out of the weighting process at this point.

Adjustment 4.8 – Non-response

When the interviewer could not reach the selected person, or the selected person refused to participate in the HSAS, this person was considered as a nonrespondent for the HSAS and they were removed from the survey. Within each province, the weights of those HSAS nonrespondents were distributed to the HSAS respondents. The CHAID algorithm was used to determine the best split of classes within each province. The following characteristics, obtained from the CCHS interview, were mainly used to define the classes: self-rated general health, self-rated mental health, restriction of activities, two-week disability, smoking type, marital status, age, sex and income. An adjustment factor was calculated within each province and class as follows:

$$\frac{\textit{Sum of weight 4.7 for all selected persons in HSAS part B}}{\textit{Sum of weight 4.7 for all HSAS part B respondents}}$$

The weight 4.7 of the responding persons to the HSAS was multiplied by this adjustment factor to produce weight 4.8. The non-responding persons were dropped out of the weighting process from this point onward.

These are the final weights for the HSAS part B sample and they will be integrated with the HSAS part A final weights.

Adjustment 5 – Integration of part A and part B

At this point, two sets of HSAS weights were created. The final weights of part A represent the entire Canadian population (covered by the survey), whereas the final weights of part B represents the Canadian population minus the population of 21 health regions. The objective of this adjustment is to integrate the weights of the respondents of the two parts into one weight in such a way that the integrated weights represent the entire Canadian population. Here is how the integration was done.

First, the weights of the part A respondents in the 21 health regions not covered in the part B did not need to be adjusted. The integrated weight (weight 5) is equal to weight 3.3 for those respondents. The sampling weights of only the respondents from the health regions covered in both parts were adjusted. For those respondents, an adjustment factor between 0 and 1 was determined in such a way that it represented the relative importance of each sample in the total sample by province. This relative importance was measured as a function of sample sizes. The larger the proportion a sample represented, the higher was its relative importance in the total sample. To obtain the integration adjustment factor, in each province, a factor α was first calculated as follows:

$$\alpha = \frac{n_A}{n_A + n_B}$$

where n_A and n_B represent the part A and part B sample sizes respectively within each province. The weight of the part A respondents was multiplied by this factor α, while the weight of the part B respondents was multiplied by 1- α. The multiplication of the factor derived here with the final weight calculated earlier (3.3 or 4.8 depending on which part the respondent belongs to) gave the integrated weight 5.

Note that following the series of adjustments applied to the respondents, some units may come out with outlier weights compared with other units of the same province. Some respondents could represent a large proportion of their province and hence strongly influence estimates for their province. In order to prevent that, the weight of the outlier units that represent a large proportion of their province is "winsorized" downward.

Adjustment 6 – Poststratification

The last adjustment factor consisted in a poststratification. It is performed to ensure that the sum of the final weights equals to the 2003 population projections in every province-age-sex group and also to those in every census metropolitan area (CMA). The following population counts and control groups were used for this step:

1. For both males and females, four age groups (15-24, 25-44, 45-64, 65+) were defined within each province. The population totals for these control groups were the average of the Census monthly projections for January to December 2003.
2. Also, CMAs (25 in Canada) and non-CMA areas in each province were used as control groups. The population totals of the CMA and non-CMA areas were the average of the Census monthly projections for January to December 2003.

The Generalized regression (GREG) estimation method was used to modify the integrated weight (weight 5) so that the sum of the poststratified weights (weight 6) for each control group above is consistent with the projected population counts.

HSAS master weight

After applying all these adjustments, the HSAS master weight was created and it can be found on the data file with the variable name WTSC_HSM.

Adjustment 7 – Share adjustment

For the purpose of sharing health-related data with the provincial Ministries of Health and Health Canada, a file containing only those respondents who agreed to share their health-related data with those organizations was created. The respondents who refused to share their data were dropped out at this step and the weights of those respondents were distributed to those who agreed to share their data using classes. Classes were defined by using the CHAID algorithm within each province. The personal income question was used as the first split in every province since it was significantly related to the answer of the share question. Other characteristics such as education, age, sex, immigration flag and self-rated general health were also used in the creation of classes. Within each class, an adjustment factor was calculated as follows:

$$\frac{\textit{Sum of weight 6 for all HSAS respondents}}{\textit{Sum of weight 6 for all HSAS sharers}}$$

The weight 6 of sharing respondents was multiplied by this adjustment factor to produce weight 7. The non-sharing respondents were removed from the file at this point.

Adjustment 8 – Poststratification, after share adjustment

Following the share adjustment, the sum of weight 7 in each province-age-sex group and census metropolitan area (CMA) did not match exactly the population counts. Therefore, a poststratification was again performed so that the sum of the final share weights is consistent with the 2003 population projections. As in adjustment 6, the method of generalized regression (GREG) estimation was used to produce the final share weight (weight 8).

HSAS share weight

After applying these additional adjustments, the HSAS share weight was created and it can be found on the data file with the variable name WTSC_HSS.

8. Data quality

8.1 Response rate

Given that the collection for HSAS was divided in two parts, it is not appropriate to report an overall response rate for the survey; independent response rates (part A and B) only can be reported.

For part A and having removed the out-of-scope units, 26,394 households were selected to participate in the CCHS/HSAS. From those, 22,972 accepted to participate to the survey, which resulted in a household response rate of 87.0%. Among these responding households, 21,680 persons were selected for the survey and were part of the target population of the HSAS (the 12 to 14 years old were excluded from HSAS). From those, 18,981 were considered respondents for the HSAS, which resulted in a response rate of 87.6% at the person level. A person was considered a respondent for the HSAS if that person answered all HSAS questionnaire modules (note that a person could be considered a respondent for CCHS but not for HSAS). Therefore, at the national level, a combined response rate of **76.2%** was observed for the part A of the HSAS. The combined response rate is obtained simply by multiplying the response rates at the household and person level. Response rates at the provincial level for part A are presented in the table 8.1.

For part B, 15,051 persons were initially selected (from the CCHS respondents) to participate in the HSAS. From those, 13,024 accepted to participate in the survey, which resulted in a response rate of **86.5%** at the national level for part B. Note that all the selected persons for part B (15,051) were considered in-scope in the calculation of the response rate although a few persons were impossible to reach (e.g. wrong telephone numbers). Response rates at the provincial level for part B are also presented in table 8.1.

Next we describe how the various components of the equation should be handled to correctly compute combined response rates for the HSAS part A.

Response rate at the household level

$$\text{HHRR} = \frac{\text{\# of responding households}}{\text{\# of in-scope households}}$$

Response rate at the person level

$$\text{PPRR} = \frac{\text{\# of respondents}}{\text{\# of selected in-scope persons}}$$

Combined response rate (part A) *= HHRR x PPRR*

Next is an example on how to calculate the combined response rate for the part A of HSAS using the information found in Table 8.1.

$$\text{HHRR} = \frac{22{,}972}{26{,}394} = 0.870$$

$$\text{PPRR} = \frac{18{,}981}{21{,}680} = 0.876$$

Combined response rate (part A) = 0.870 x 0.876
= 0.762
= **76.2%**

Table 8.1 Response rate for HSAS part A and B

Province	Part A							Part B		
	Household			Person			Comb. Resp. Rate	# pers. selected	# pers. Resp.	Resp. Rate
	# HH In-scope	# HH Resp.	HH Resp. Rate	# resp. selected in-scope	# pers. resp.	Person Resp. Rate				
Newfoundland and Labrador	1,065	991	93.1	934	826	88.4	82.3	1,934	1,694	87.6
Prince Edward Island	744	653	87.8	622	559	89.9	78.9	801	691	86.3
Nova Scotia	1,397	1,245	89.1	1,174	1,055	89.9	80.1	2,160	1,875	86.8
New Brunswick	1,327	1,224	92.2	1,149	1,025	89.2	82.2	2,281	1,870	82.0
Quebec	4,038	3,373	83.5	3,204	2,828	88.3	73.7	1,216	1,048	86.2
Ontario	5,363	4,442	82.8	4,192	3,505	83.6	69.2	1,225	1,001	81.7
Manitoba	2,063	1,877	91.0	1,767	1,591	90.0	81.9	1,933	1,734	89.7
Saskatchewan	2,213	1,992	90.0	1,876	1,661	88.5	79.7	1,745	1,575	90.3
Alberta	3,776	3,357	88.9	3,146	2,700	85.8	76.3	776	652	84.0
British Columbia	4,408	3,818	86.6	3,616	3,231	89.4	77.4	980	884	90.2
Canada	**26,394**	**22,972**	**87.0**	**21,680**	**18,981**	**87.6**	**76.2**	**15,051**	**13,024**	**86.5**

The table 8.2 shows the total number of respondents for the HSAS (part A and part B).

Table 8.2 Total number of respondents for the HSAS by province

Province	Number of respondents
Newfoundland and Labrador	2,520
Prince Edward Island	1,250
Nova Scotia	2,930
New Brunswick	2,895
Quebec	3,876
Ontario	4,506
Manitoba	3,325
Saskatchewan	3,236
Alberta	3,352
British Columbia	4,115
Canada	32,005

8.2 Survey errors

The estimates derived from this survey are based on a sample of individuals. 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 under similar conditions is called the sampling error of the estimate.

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

8.2.1 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 HSAS. Quality assurance measures were implemented at each step of data collection and processing to monitor the quality of the data. These measures included the use of highly skilled interviewers, extensive training with respect to the survey procedures and questionnaire, and the observation of interviewers to detect problems. Testing of the CAI application and field tests were also essential procedures to ensure that data collection errors were minimized.

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. Partial non-response to HSAS was minimal; once the questionnaire was started, it tended to be completed with very little non-response. Total non-response occurred either because a respondent refused to participate in the survey, or because

the interviewer was unable to contact the selected respondent. Total non-response was handled by adjusting the weight of persons who responded to the survey to compensate for those who did not respond. See section 8 for details of the weight adjustment for non-response.

8.2.2 Sampling errors

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error. The basis for measuring the potential size of sampling errors is the standard deviation of the estimates derived from survey results. However, because of the large variety of estimates that can be produced from a survey, the standard deviation 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 deviation of the estimate by the estimate itself and is expressed as a percentage of the estimate.

For example, suppose hypothetically that one estimates that 25% of Canadians aged 15 and over needed to consult a specialist doctor in the last 12 months and that this estimate is found to have a standard deviation of 0.012. Then the CV of the estimate is calculated as:

$$(0.012/0.25) \times 100\ \% = 4.80\ \%$$

Statistics Canada commonly uses CV results when analyzing data, and urges users producing estimates from HSAS data files to also do so.

8.2.2.1 Bootstrap Method for Variance Estimation

In order to determine the quality of the estimate and to calculate the CV, the standard deviation must be calculated. Confidence intervals also require the standard deviation of the estimate.

The HSAS uses a multi-stage survey design, which means that there is no simple formula that can be used to calculate variance estimates. Therefore, an approximate method was needed. The bootstrap method is used because the sample design information needs to be taken into account when calculating variance estimates. The bootstrap method does this, and with the use of the Bootvar program, discussed in the next subsection, remains a method that is fairly easy for users to use.

The bootstrap re-sampling method used in the HSAS involves the selection of simple random samples known as replicates, and the calculation of the variation in the estimates from replicate to replicate. In each stratum, a simple random sample of (n-1) of the n clusters is selected with replacement to form a replicate. Note that since the selection is with replacement, a cluster may be chosen more than once. In each replicate, the survey weight for each record in the (n-1) selected clusters is recalculated. These weights are then post-stratified according to demographic information in the same way as the sampling design weights in order to obtain the final bootstrap weights.

The entire process (selecting simple random samples, recalculating and post-stratifying weights for each stratum) is repeated B times, where B is large. The HSAS typically uses B=500, to produce 500 bootstrap weights. To obtain the bootstrap variance estimator, the point estimate for each of the B samples must be calculated. The standard deviation of these estimates is the

bootstrap variance estimator. Statistics Canada has developed a program that can perform all of these calculations for the user: the Bootvar program.

8.2.2.2 Bootvar Program for Variance Estimation

The Bootvar program is available in both SAS and SPSS formats. It is made up of macros that compute variances for totals, ratios, differences between ratios and for linear and logistic regression. The Bootvar program is provided with bootstrap weights and a document explaining how to modify and use the program to suit user's needs.

9. Guidelines for tabulation, analysis and release

This section 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 figures that are in close agreement with 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 proceeding 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 that 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. That is to say, the sampling weights are not identical for all individuals in the sample. 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 might 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: categorical estimates, quantitative estimates

Before discussing how HSAS data can be tabulated and analyzed, it is useful to describe the two main types of point estimates of population characteristics that can be generated from the microdata file.

Categorical estimates:

Categorical estimates are estimates of the number or percentage of the surveyed population possessing certain characteristics or falling into some defined category. What kind of doctor referring a patient is an example of such an estimate. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Example of categorical question:

WTMC_03 Were you referred by:

…a family doctor?
…another specialist?
…another health care provider?
Did not require a referral

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.

An example of a quantitative estimate is the average number of days waited before seeing a specialist. The numerator is an estimate of the total number of days waited by individuals, and its denominator is an estimate of the number of individuals who waited.

Example of quantitative question:

WTM_Q07A How long did you have to wait between when [you and your doctor decided that you should see a specialist/you and your health care provider decided that you should see a specialist/the appointment was initially scheduled] and when you actually visited the specialist?
|_|_|_| (number of units)

9.2.2 Tabulation of categorical estimates

Estimates of the number of people with a certain characteristic can be obtained from the microdata files by summing the final weights of all records possessing the characteristic 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 the numerator estimate by the denominator estimate.

9.2.3 Tabulation of quantitative estimates

Estimates of quantities can be obtained from the microdata files by:

a) multiplying the value of the variable of interest by the final weight and summing this quantity over all records of interest to obtain the numerator ($\hat{X}$);

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

c) dividing the numerator estimate by the denominator estimate.

9.3 Guidelines for statistical analysis

The HSAS 2003 is based upon a complex design, with stratification and multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used.

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures can differ from what is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are almost meaningless.

For many analysis techniques (for example linear regression, logistic regression, analysis of variance), a method exists that can make the application of standard packages more meaningful. If the weights on the records are rescaled so that the average weight is one (1), then the results produced by the standard packages will be more reasonable; they still will not take into account the stratification and clustering of the sample's design, but they will take into account the unequal probabilities of selection. The rescaling can be accomplished by using in the analysis a weight equal to the original weight divided by the average of the original weights for the sampled units (people) contributing to the estimator in question.

9.4 Release guidelines

Before releasing and/or publishing any estimate from these microdata files, users should first determine the number of sampled respondents who contribute to the calculation of the estimate. If this number is less than 10, the weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. For weighted estimates based on sample sizes of 10 or more, users should determine the coefficient of variation of the rounded estimate and follow the guidelines below.

Table 9.1 Sampling variability guidelines

Type of estimate	CV (in %)	Guidelines
Acceptable	0.0 ≤ CV ≤ 16.6	Estimates can be considered for general unrestricted release. Requires no special notation.
Marginal	16.6 < CV ≤ 33.3	Estimates can be considered for general unrestricted release but should be accompanied by a warning cautioning subsequent users of the high sampling variability associated with the estimates. Such estimates should be identified by the letter E (or in some other similar fashion).
Unacceptable	CV > 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 in some other fashion) and the following warning should accompany the estimates: "The user is advised that . . .(specify the data) . . . do not meet Statistics Canada's quality standards for this statistical program. Conclusions based on these data will be unreliable and most likely invalid. These data and any consequent findings should not be published. If the user chooses to publish these data or findings, then this disclaimer must be published with the data."

10. File usage

10.1 Use of weights

Only one weight, WTSC_HSM, appears on the file. This weight is applicable to all age groups, regions, provinces and territories. VARIABLES ON THE FILE SHOULD BE ANALYZED USING THIS WEIGHT. For a more detailed explanation on the creation of this weight, see Section 5 of the documentation on weighting.

10.2 Variable naming convention

The CCHS adopted a variable naming convention that allows data users to easily use and identify the data based on module and cycle. The HSAS has the same variable naming convention as the CCHS. The variable naming convention includes the following mandatory requirements: restrict variable names to a maximum of 8 characters for ease of use by analytical software products; identify the survey occasion (Cycle 1.1, 1.2 …) in the name; and allow conceptually identical variables to be easily identifiable over survey occasions. The variable names for these identical modules and questions should only differ in the cycle position identifying the particular survey occasion in which they were collected.

10.2.1 Variable name component structure in HSAS

Each of the eight characters in a variable name contains information about the type of data contained in the variable.

Positions 1-3:	Module/Questionnaire section name
Position 4:	Survey cycle
Position 5:	Variable type
Positions 6-8:	Question number

For example: the variable from question 6, Waiting Times Module, HSAS Cycle 2.1 (WTMC_06A) is:

Position 1-3:	**WTM**	Waiting Times module
Position 4:	**C**	Cycle 2.1
Position 5:	_	underscore (_ = collected data)
Position 6-8:	**06A**	question number & answer option

10.2.2 Positions 1-3: variable / questionnaire section name

The following values are used for the section name component of the variable name:

ACC	Access to health care services	INC	Income
ADM	Administration	LBF	Labour force
CCC	Chronic conditions	RAC	Restriction of activities
DHH	Demographic and household variables	SAM	Sample identifiers
EDU	Education	SDC	Socio-demographic characteristics
GEN	General Health	WTM	Waiting times
GEO	Geographic identifiers	WTS	Weights

10.2.3 Position 4: cycle

Cycle **Description**

A Cycle 1.1: Canadian Community Health Survey
- Regional level survey, stratified by health region;
- Common content and optional content selected by health region;
- Estimates for health regions, provinces, territories and Canada.

B Cycle 1.2: Canadian Community Health Survey, Mental Health and Well-Being
- Provincial level survey;
- Focus content with additional, general content;
- Estimates for the provinces, territories and Canada.

C Cycle 2.1: Canadian Community Health Survey / Health Services Access Survey
- Regional level survey, stratified by health region;
- Common content and optional content selected by health region;
- Estimates for health regions, provinces, territories and Canada.

10.2.4 Position 5: variable type

_	Collected variable	A variable that appeared directly on the questionnaire
C	Coded variable	A variable coded from one or more collected variables (e.g., SIC, Standard Industrial Classification code)
D	Derived variable	A variable calculated from one or more collected or coded variables, usually calculated during head office processing (e.g., Health Utility Index)
F	Flag variable	A variable calculated from one or more collected variables (like a derived variable), but usually calculated by the data collection computer application for later use during the interview (e.g., work flag)
G	Grouped variable	Collected, coded, suppressed or derived variables collapsed into groups (e.g., age groups)

10.2.5 Positions 6-8: variable name

In general, the last three positions follow the variable numbering used on the questionnaire. The letter "Q" used to represent the word "question" is removed, and all question numbers are presented in a two- digit format. For example, question Q01A in the questionnaire becomes simply 01A. Question fifteen becomes simply 15.

For questions which have more than one response option, the final position in the variable naming sequence is represented by a letter. For this type of question, new variables were created to differentiate between a “yes” or “no” answer for each response option. For example, if Q2 had 4 response options, the new questions would be named Q2A for option 1, Q2B for option 2, Q2C for option 3, etc. If only options 2 and 3 were selected, then Q2A = No, Q2B = Yes, Q2C = Yes and Q2D = No.

10.3 Access to master file data

A first approach for any client is the production of custom tabulations done by the Client Custom Services staff in Health Statistics Division. This service allows users who do not possess knowledge of tabulation software products to get custom results. The results are screened for confidentiality and reliability concerns before release. There is a charge for this service. You can contact this service by writing to: hd-ds@statcan.ca.

As a second approach, the Research Data Centres Program allows researchers to submit to Statistics Canada a research project that uses data from the Master Files. These projects are accepted based on a set of specific rules. When the project is accepted, the researcher is designated as a "deemed employee" of Statistics Canada for the duration of the research, and given access to the Master Files data from designated Statistics Canada sites. Please consult the following web page for more information: http://www.statcan.ca/english/rdc/index.htm.