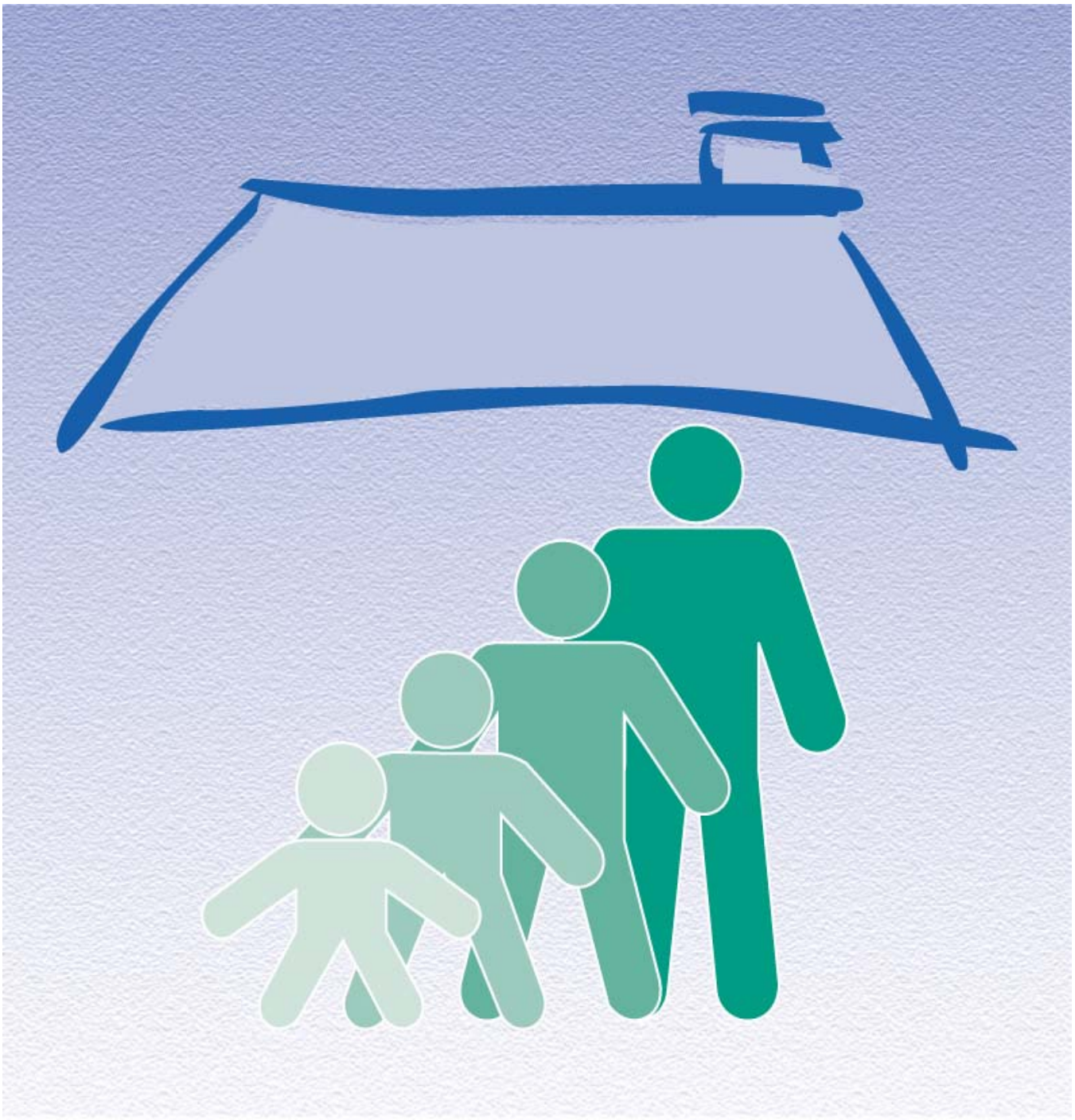
Canadian Community Health Survey (CCHS)

2007 Microdata files
User guide

June 2008



Statistics Canada Statistique Canada

Canada

TABLE OF CONTENTS

1. INTRODUCTION 1

2. BACKGROUND 2

3. CCHS REDESIGN IN 2007 3

4. CONTENT STRUCTURE OF THE CCHS 4

4.1 Common content 4
4.2 Optional content 5
4.3 Rapid response content 5
4.4 Content included in data files 5

5. SAMPLE DESIGN 7

5.1 Target population 7
5.2 Health regions 7
5.3 Sample size and allocation 7
5.4 Frames, household sampling strategies 9
5.4.1 Sampling of households from the area frame 9
5.4.2 Sampling of households from the list frame of telephone numbers 11
5.4.3 Sampling of households from the RDD frame of telephone numbers 12
5.5 Sample allocation over the collection period 12
5.6 Sampling of interviewees 12
5.7 Supplementary buy-in sample in three health regions in Ontario 13
5.8 Sub-sample for the Health Services Access Survey (HSAS) 14

6. DATA COLLECTION 16

6.1 Computer-assisted interviewing 16
6.2 CCHS application development 16
6.3 Interviewer training 17
6.4 The interview 18
6.5 Field operations 20
6.6 Quality control and collection management 21

7. DATA PROCESSING 23

7.1 Editing 23
7.2 Coding 23
7.3 Creation of derived variables 23
7.4 Weighting 24

8. WEIGHTING 25

8.1 Overview 25
8.2 Weighting of the area frame sample 26
8.3 Weighting of the telephone frame sample 27
8.4 Integration of the telephone and area frames (I1) 29
8.5 Post-integration weighting steps 30
8.6 Particular aspects of the weighting in the three territories 31

9. DATA QUALITY 32

9.1 Response rates 32
9.2 Survey Errors 33
9.2.1 Non-sampling Errors 33
9.2.2 Sampling Errors 34

10. GUIDELINES FOR TABULATION, ANALYSIS AND RELEASE 35

10.1 ROUNDING GUIDELINES 35
10.2 SAMPLE WEIGHTING GUIDELINES FOR TABULATION 36
10.2.1 DEFINITIONS: CATEGORICAL ESTIMATES, QUANTITATIVE ESTIMATES 36
10.2.2 TABULATION OF CATEGORICAL ESTIMATES 37
10.2.3 TABULATION OF QUANTITATIVE ESTIMATES 37
10.3 GUIDELINES FOR STATISTICAL ANALYSIS 38
10.4 RELEASE GUIDELINES 38

11. APPROXIMATE SAMPLING VARIABILITY TABLES 40

11.1 HOW TO USE THE CV TABLES FOR CATEGORICAL ESTIMATES 40
11.2 EXAMPLES OF USING THE CV TABLES FOR CATEGORICAL ESTIMATES 42
11.3 HOW TO USE THE CV TABLES TO OBTAIN CONFIDENCE LIMITS 45
11.4 EXAMPLE OF USING THE CV TABLES TO OBTAIN CONFIDENCE LIMITS 46
11.5 HOW TO USE THE CV TABLES TO DO A Z-TEST 46
11.6 EXAMPLE OF USING THE CV TABLES TO DO A Z-TEST 47
11.7 EXACT VARIANCES/COEFFICIENTS OF VARIATION 47
11.8 RELEASE CUT-OFFS FOR THE CCHS 48

12. MICRODATA FILES: DESCRIPTION, ACCESS AND USE 49

12.1 MASTER FILES 49
12.1.1 RESEARCH DATA CENTRE 49
12.1.2 CUSTOM TABULATIONS 49
12.1.3 REMOTE ACCESS 49
12.2 SHARE FILES 50
12.3 PUBLIC USE MICRODATA FILES 50
12.4 USE OF WEIGHT VARIABLE 50
12.5 VARIABLE NAMING CONVENTION BEGINNING IN 2007 51
12.6 VARIABLE NAMING CONVENTION BEFORE 2007 53
12.7 GUIDELINES FOR THE USE OF SUB-SAMPLE VARIABLES 53
12.7.1 VARIABLE NAMING CONVENTION FOR SUB-SAMPLE MODULES 53
12.7.2 SEPARATE FILE 54
12.7.3 ANALYSIS COMBINING SUB-SAMPLE AND OPTIONAL CONTENT 55
12.8 DATA DICTIONARIES 56
12.9 DIFFERENCES IN CALCULATION OF COMMON CONTENT VARIABLES USING DIFFERENT FILES 56

APPENDIX A CANADIAN COMMUNITY HEALTH SURVEY (2007-2008) 57
APPENDIX B SELECTION OF OPTIONAL CONTENT BY PROVINCE OR TERRITORY (2007) 61
APPENDIX C CCHS GEOGRAPHY 2007 67
APPENDIX D SAMPLE ALLOCATION BY HEALTH REGION AND FRAME 73
APPENDIX E RESPONSE RATE BY HEALTH REGION AND FRAMES 79

1. INTRODUCTION

The Canadian Community Health Survey (CCHS) is a cross-sectional survey that collects information related to health status, health care utilization and health determinants for the Canadian population. It surveys a large sample of respondents and is designed to provide reliable estimates at the health region level. In 2007, major changes were made to the CCHS design. Data is now collected on an ongoing basis with annual releases, rather than every two years as was the case prior to 2007. The survey's objectives were also revised and are as follows:

- support health surveillance programs by providing health data at the national, provincial and infraprovincial levels;
- provide a single data source for health research on small populations and rare characteristics;
- timely release of information easily accessible to a diverse community of users; and
- create a flexible survey instrument that includes a rapid response option to address emerging issues related to the health of the population.

Details of the other redesign changes are provided in section 3.

The CCHS data is always collected from persons aged 12 and over living in private dwellings in the 121 health regions covering all provinces and territories. Excluded from the sampling frame are individuals living on Indian Reserves and on Crown Lands, institutional residents, full-time members of the Canadian Forces, and residents of certain remote regions. The CCHS covers approximately 98% of the Canadian population aged 12 and over.

The purpose of this document is to facilitate the manipulation of the CCHS microdata files and to describe the methodology used. The CCHS produces three types of microdata files: master files, share files and public use microdata files (PUMF). The characteristics of each of these files are presented in this guide. It is important, however, to mention that a PUMF has not been released for the 2007 CCHS. This will take place in summer 2009 when the 2007 and 2008 data will be available.

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

Electronic Products Help Line:	1-800-949-9491
For custom tabulations or general data support:	
Client Custom Services, Health Statistics Division:	613-951-1653
E-mail:	hd-ds@statcan.ca
For remote access support:	613-951-1653
E-mail:	cchs-escc@statcan.ca
Fax:	613-951-0792

2. BACKGROUND

In 1991, the National Task Force on Health Information cited a number of issues and problems with the health information system. The members felt that data was fragmented; incomplete, could not be easily shared, was not being analysed to the fullest extent, and the results of research were not consistently reaching Canadians.[1]

In responding to these issues, the Canadian Institute for Health Information (CIHI), Statistics Canada and Health Canada joined forces to create a Health Information Roadmap. From this mandate, the Canadian Community Health Survey (CCHS) was conceived. The format, content and objectives of the CCHS evolved through extensive consultation with key experts and federal, provincial and community health region stakeholders to determine their data requirements.[2]

To meet many data requirements, the CCHS had a two-year data collection cycle. Until the redesign in 2007, the first year of the survey cycle, designated by ".1", was a general population health survey, designed to provide reliable estimates at the health region level. The second year of the survey cycle, designated by ".2", had a smaller sample and was designed to provide provincial level results on specific health topics.

New designations for Cycles .1 and .2

As of 2007, the regional component of the CCHS began being collected on an ongoing basis. To avoid confusion with the health focused surveys, the two components stopped using the ".1" and ".2" designations to distinguish them. Now, the x.1 cycles of the CCHS are designated simply by the name of the survey followed by the collection period (e.g., "Canada Community Health Survey, 2007" or "CCHS – 2007"). The focused content component of the survey remains unchanged. It will continue to examine in greater detail more specific topics or populations. It will be designated by the name of the survey followed by the topic of the themes covered by each survey (e.g., "Canadian Community Health Survey on Healthy Aging" or "CCHS – Healthy Aging").

[1] 1999. Health Information Roadmap: Responding to Needs, Health Canada, Statistics Canada. p. 3.
[2] 1999. Health Information Roadmap: Beginning the Journey. Canadian Institute for Health Information/Statistics Canada. ISBN 1-895581-70-2. p. 19.

3. CCHS REDESIGN IN 2007

Until 2005, the CCHS data were collected every two years over a one year period and released every two years, about six months after the end of the collection period. There were two main objectives for the 2007 CCHS redesign: to address the needs of partners to increase the survey's content and the frequency of data releases, and to ensure better use of operational resources. For these reasons, the proposed changes to the CCHS design focused on improving the survey's efficiency and flexibility through ongoing data collection.

Extensive consultations were held across Canada with key experts and federal, provincial and health region stakeholders to gather input on the proposed changes and detailed information on the data requirements and products of the various partners.

Below are the main changes arising from the CCHS redesign:

- In the past, the CCHS data were collected from 130,000 respondents over a 12-month period. Now, data collection takes place on an ongoing basis. The sample, which retains the same size, is divided into 12 two-month collection periods. Each collection period is representative of the population living in the ten Canadian provinces during the two months. For operational reasons, the sample in the territories is representative of their population after 12 months.

- The common content component is divided in two: the core content and the theme content. The theme content was created to take advantage of the continuous collection approach. The data collection time for this component can be adjusted based on the prevalence of the desired estimates and their geographic level. The core content will remain relatively stable over time. At the discretion of the provinces and regions, the optional content can also be adjusted on an annual basis, rather than every two years.

- Content and collection changes inevitably impact the dissemination strategy. Previously, data were released every two years. As of 2008, CCHS data will be released annually. Every two years, a file combining the two years' sample (130,000 respondents) will also be released. In addition to these regular files, other special files will be made available when additional content has been collected during collection periods that do not correspond to the standard annual periods, which is January to December.

- The annual data collection is divided into six two-month periods. Unlike the previous collection strategy, these periods no longer overlap, which provides more efficient oversight of collection and offers the possibility of changing the collection interface every two months, if necessary.

4. CONTENT STRUCTURE OF THE CCHS

In addition to socio-demographic and administrative data, the content of the CCHS includes three components, each of which addresses a different need: the common content component comprising the core content and theme content, the optional content component, and the rapid response component. Appendix A lists the modules included in the 2007 questionnaire by component.

The average length of a CCHS interview is estimated at 40 to 45 minutes.

Table 4.1 Length of survey by component

CCHS component	Average interview time
Common content • Core content • Theme content	**30 minutes** (20 minutes) (10 minutes)
Optional content	**10 minutes**
Rapid response content (optional)	**2 minutes**

4.1 Common content

The CCHS common content component includes questions asked of respondents in all provinces and territories (unless otherwise specified). It is divided into two components: the core content and the theme content.

The core content consists of questions asked of all survey respondents. These questions will remain relatively stable in the questionnaire for a period of about six years, unless a major concern is raised about quality.

Theme content comprises questions related to a specific topic. It covers a two-year theme and an annual theme and takes about 10 minutes of the interview time. Some themes could be reintroduced in the survey every two, four or six years, if required. This component enables CCHS to better plan its content in the medium term.

Some of the modules in the theme component may be asked of a sub-sample of respondents if the objective of these questions is to provide reliable data at the national or provincial level, rather than at the health region level. This approach is used to minimize the related response burden and costs.

4.2 Optional content

The optional content component gives health regions the opportunity to select content that addresses their provincial or regional public health priorities. The optional content is selected from a long list of modules available for inclusion in the CCHS. The content modules selected by a region are asked only of residents in the regions that selected these modules. In reality, since cycle 3.1 of the CCHS, the regions and provinces have opted to coordinate the optional content selected in order to ensure a uniform selection of optional modules provincially. The optional content may vary annually depending on needs and must be reviewed every two years.

It should be noted that, unlike the modules included in the common content, the resulting data from the optional content modules is not easily generalized across Canada.[3].

Appendix B presents the selection results of the optional content for 2007 by province of residence.

4.3 Rapid response content

The rapid response component is offered on a cost-recovery basis to organizations interested in obtaining national estimates on an emerging or specific topic related to the health of the population. The rapid response content takes a maximum of two minutes of interview time. The questions appear in the questionnaire for a single collection period (two months) and are asked of all CCHS respondents during that period.

4.4 Content included in data files

In June 2009, two main files will be made available: a main data file based on the 2008 collection period, which will be similar to the main 2007 data file, as well as a main data file based on the 2007-2008 collection period. The 2007-2008 file will be similar in size to files from the previous cycles (approximately 130,000 respondents). It will include core, optional and the theme content collected over the two-year period. One-year themes will not be available on the two-year data file. Also, theme modules collected from sub-sample of respondents will continue to be disseminated in separate files. These files include core content and sub-sample theme modules only. See table 4.2 for a clarification of what will be available with the 2007 and 2008 releases.

[3] Unless all health regions in Canada select an optional module in the same collection period, which has never happened to date.

Table 4.2 Content components included in 2007 and 2008 data files

Files		Core content	2007 Theme[1]	2008 Theme[2]	2007-2008 Theme	Optional content[3]
2007 Files	**Main**	Yes	N/A	N/A	Yes	Yes
	Sub-sample (3 modules)	Yes	Yes	N/A	No	No
2008 File	**Main**	Yes	N/A	Yes (except the sub-sample theme module)	Yes	Yes
	Sub-sample (1 module)	Yes	N/A	(Measured heath and weight)	No	No
2007-2008 File	**Main**	Yes	No	No	Yes	Yes

1 The 2007 theme was comprised of three modules (Patient satisfaction, Access to health care services and Waiting times) which were all asked to a sub-sameple of respondents.

2 The 2008 theme is formed of a group of modules related to chronic disease screening and a module on measured height and weight. This last module is asked of a sub-sample of respondents.

3 This assumes that optional content remains the same for the two years. If not, it will only be included in the file of the year in which it was collected.

5. SAMPLE DESIGN

5.1 Target population

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

5.2 Health regions

For administrative purposes, each province is divided into health regions (HR) and each territory is designated as a single HR. Statistics Canada is sometimes asked to make minor changes to the boundaries of some of the HRs to correspond to the geography of the Census, or to better account for the health data needs determined by the new geographic boundaries. For CCHS 2007, data was collected in 118 HRs in the ten provinces, as well as to one HR per territory, totalling 121 HRs (Appendix C).

5.3 Sample size and allocation

To provide reliable estimates for each HR, and given the budget allocated to the CCHS component, it was determined that the survey should consist of a sample of nearly 130,000 respondents over a period of 2 years. Although producing reliable estimates for each HR was a primary objective, the quality of the estimates for certain key characteristics at the provincial level was also deemed important. Therefore, the sample allocation strategy, consisting of three steps, gave relatively equal importance to the HRs and the provinces. In the first step, a minimum size of 500 respondents per HR was imposed. This is considered the minimum for obtaining a reasonable level of data quality. However, due to response burden, a maximum sampling fraction of 1 out of 20 dwellings was imposed to avoid sampling too many dwellings in smaller regions also targeted by other surveys. Note that very few HRs have a size lower than 500 due to limit of the sampling fraction. This first step allocated 60,350 units in total. The second step involves allocating the rest of the available sample by using an allocation proportional to the population size by province. The total sample size by province is therefore the sum of the sizes established by the two first steps. The sample is then divided evenly between the 2 collection years. Table 5.1 gives the annual sample size for 2007.

Table 5.1 Number of health regions and targeted sample sizes by province/territory, 2007

Province	Number of HRs	Total sample size (targeted)
Prince Edward Island[1]	3	1,480
Newfoundland and Labrador	4	2,005
Nova Scotia	6	2,520
New Brunswick	7	2,580
Quebec	16	12,145
Ontario[2]	36	22,230
Manitoba	10	3,755
Saskatchewan	11	3,860
Alberta	9	6,105
British Columbia	16	8,050
Yukon	1	600
Northwest Territories	1	600
Nunavut	1	350
Canada	**121**	**66,280**

[1] The sample size for Prince Edward Island was increase to meet the needs of the Health Services Access Survey for 2007. The initial sample size was 1,000 units (refer to section 5.8 for further details).
[2] The sample size for Ontario includes the buy-in extra sample by LHIN. The initial sample size for Ontario before the buy-in was 20,900 units (refer to section 5.7 for further details).

In the third step, provincial sample was allocated among its HRs proportionally to the square root of the estimated population in each HR. This three-step approach gives sufficient sample for each HR with minimal disturbance to the provincial allocation of the sample.

Note that the three territories were not part of the above allocation strategy as they were dealt with separately. In total, for 2007, 600 sample units were allocated to the Yukon, 600 to the Northwest Territories and 350 to Nunavut. These sizes are determined according to the available budget. The sample allocation is done proportionally to the population sizes in the strata. The strata used are the same as those defined by the Labour Force Survey (LFS), which group together communities (For more details, see section 5.4.1).

The sample is then equally divided in 2 in order to obtain the same sample sizes between the area frame sample and the list frame sample for each HR[4], as described in the next section. We should finally mention that the size of the samples taken from each frame was increased before data collection in order to account for the out-of-scope and for the foreseen non-response rates based on the rates obtained in previous CCHS cycles. The sample sizes by HR and frame are provided in Appendix D.

[4] Except for 2 regions which use a random digit dialing frame (RDD) only (section 5.4.3) and the three territories which use only area frame and random digit dialing frame (RDD) (sections 5.4.1 and 5.4.3) .

5.4 Frames, household sampling strategies

CCHS 2007 used three sampling frames to select the sample of households: 49% of the sample of households came from an area frame, 50% came from a list frame of telephone numbers and the remaining 1% came from a Random Digit Dialling (RDD) sampling frame.

5.4.1 Sampling of households from the area frame

The CCHS used the area frame designed for the Canadian Labour Force Survey (LFS) as a sampling frame. The sampling plan of the LFS is a multistage stratified cluster design in which the dwelling is the final sampling unit[5]. In the first stage, homogeneous strata are formed and independent samples of clusters are drawn from each stratum. In the second stage, dwelling lists are prepared for each cluster and dwellings, or households, are selected from these lists.

For the purpose of the LFS plan, each province is divided into three types of regions: major urban centres, cities, and rural regions. Geographic or socio-economic strata are created within each major urban centre. Within the strata, between 150 and 250 dwellings are grouped together to create clusters. Some urban centres have separate strata for apartments or for census Dissemination Areas (DA) to pinpoint households with high income, immigrants and aboriginals. In each stratum, six clusters or residential buildings (sometimes 12 or 18 apartments) are chosen by a random sampling method with a probability proportional to size (PPS), the size of which corresponds to the number of households. The number six is used throughout the sample design to allow for one sixth of the LFS sample to be rotated each month.

The other cities and rural regions of each province are stratified first on a geographical basis, then according to socio-economic characteristics. In the majority of strata, six clusters (usually census DAs) are selected using the PPS method. Some geographically isolated urban centres are covered by a three-stage sampling design. This type of sampling plan is used for Quebec, Ontario, Alberta and British Columbia.

Once the new clusters are listed, the sample is obtained using a systematic sampling of dwellings. The sample size for each systematic sample is called the "yield". Table 5.2 gives an overview of the types of PSUs used in the LFS sample and the yield predicted by systematic sample. As the sampling rates are determined in advance, there is frequently a difference between the expected sample size and the numbers that are obtained. The yield of the sample, for example, is sometimes excessive. This can particularly happen in sectors where there is an increase in the number of dwellings due to new construction. To reduce the cost of collection, an excessive output is corrected by eliminating, from the beginning, a part of the units selected and by modifying the weight of the sample design.

[5] Statistics Canada (1998). *Methodology of the Canadian Labour Force Survey.* Statistics Canada. Cat. No. 71-526-XPB.

Table 5.2 Major first-stage units, sizes and yields, 2007

Area	Primary Sampling Unit (PSU)	Size (households per PSU)	Yield (sampled households)
Toronto, Montreal, Vancouver	Cluster	150-250	6
Other cities	Cluster	150-250	8
Most rural areas / small urban centres	Cluster	100-250	10

Due to the specific of the CCHS, some modifications had to be incorporated in this sampling strategy. To obtain an annual sample of 33,000 respondents for CCHS 2007, more than 47,000 dwellings had to be selected from the area frame to account for vacant dwellings and non-responding households. Each month, the LFS design provides approximately 60,000 dwellings distributed across the various economic regions in the ten provinces, whereas the CCHS 2007 required 47,000 dwellings distributed across the HRs, which have different geographic boundaries from those of the LFS economic regions. Overall, the CCHS 2007 required a lower number of dwellings than those generated by the LFS selection mechanism, which corresponds to an *adjustment factor* of 0.78 (47,000/60,000). However, since the adjustment factors varied from 0.3 to 3.0 at the HR level, certain adjustments were required.

The changes made to the selection mechanism in the regions varied depending on the size of the adjustment factors. For HRs that had a factor smaller than or equal to 1, the number of PSUs selected was reduced if necessary. For example, if the factor was 0.5 then only 3 PSUs were selected in each stratum instead of the usual number of 6 PSUs. For those HRs with a factor greater than 1 but smaller than or equal to 2, the sampling process of dwellings within a PSU was repeated for a subset of the selected PSUs that were part of the same HR. For example, if the factor was 1.6 then the selection of dwellings within a PSU was repeated for 4 of the 6 PSUs in all strata of that HR. When it was necessary to have a repeated selection of dwellings within a PSU and there were no more dwellings available in that PSU, then another PSU was selected. When the factor was greater than 2, the sampling process of dwellings was repeated among other PSUs that were part of the same HR[6].

Finally, when the number of dwellings available in the selected PSUs was greater than the requested number of dwellings for a given HR, a sub-sample of dwellings was selected. This process is called ‘stabilisation’.

[6] To reduce listing costs, the sampling process of dwellings was repeated up to 3 times within PSUs already selected in urban areas only. These cases were exceptions, however.

Sampling of households from the area frame in the three territories

For operational reasons, the LFS area frame sample design for the three territories is different. For each territory, in-scope communities are grouped into strata based on various characteristics (population, geographical information, proportion of Inuit and/or Aboriginal persons, and median household income). The LFS defined five design strata in the Yukon, ten in the Northwest Territories and six in Nunavut. The first stage of selection consisted of randomly selecting one community with a probability proportional to population size within each design stratum. Then, within the selected community, a household sampling strategy was put in place identically to the one described above. The CCHS selected its sample from the same communities sampled by the LFS, while ensuring that different dwellings were selected. If too many or too few dwellings were available for a community within a stratum, the LFS chose another community for the CCHS.

It is worth mentioning that the frame for the CCHS 2007 covered 90% of the private households in the Yukon, 97% in the Northwest Territories and 71% in Nunavut[7].

5.4.2 Sampling of households from the list frame of telephone numbers

With the exception of 5 HRs (the two RDD only HRs and the three territories), the list frame of telephone numbers was used in all HRs to complement the area frame. The list frame consists of the Canada Phone directory which is an external administrative database of names, addresses and telephone numbers from telephone directories in Canada updated every six months. It was linked to administrative conversion files to obtain postal codes, and these were mapped to HRs to create list frame strata. There was one list frame stratum per HR. Within each stratum, the required number of telephone numbers was selected using a simple random sampling process from the list. As for the RDD frame, additional telephone numbers were selected to account for the numbers not in service or out-of-scope.

It is important to mention that the undercoverage of the list frame is higher than the one for the RDD as unlisted numbers do not have a chance of being selected. Nevertheless, as the list frame is always used as a complement to the area frame, the impact of the undercoverage of the list frame is minimal and is dealt with during weighting.

[7] In Nunavut, because of operational difficulties inherent to remote locales, only the 10 largest communities are covered by the survey: Iqaluit, Cambridge Bay, Baker Lake, Arviat, Rankin Inlet, Kugluktuk, Pond Inlet, Cape Dorset, Pangnirtung and Igloolik.

5.4.3 Sampling of households from the RDD frame of telephone numbers

In four HRs, a Random Digit Dialing (RDD) sampling frame of telephone numbers was used to select a sample of households. The sampling of households from the RDD frame used the Elimination of Non-Working Banks (ENWB) method, a procedure adopted by the General Social Survey[8]. A bank of one hundred telephone numbers (the first eight digits of a ten-digit telephone number) is considered to be non-working if it does not contain any residential telephone numbers. At first, the frame consists of a list of all possible banks and, as non-working banks are identified, they are eliminated from the frame. It should be noted that these banks are eliminated only when there is evidence from various sources that they are non-working. When there is no information about a bank it is left on the frame. The Canada Phone Directory and telephone companies' billing address files were used in conjunction with various internal administrative files to eliminate non-working banks.

Using available geographic information (postal codes), the banks on the frame were regrouped to create RDD strata to encompass, as closely as possible, the HR areas. Within each RDD stratum, a bank was randomly chosen and a number between 00 and 99 was generated at random to create a complete, ten-digit telephone number. This procedure was repeated until the required number of telephone numbers within the RDD stratum was reached. Frequently, the number generated is not in service or is out-of-scope, and therefore, many additional numbers must be generated to reach the targeted sample size. This success rate varies from region to region. Within the CCHS, the success rates ranged from 25% to 50% among the four HRs which required the use of the RDD frame.

5.5 Sample allocation over the collection period

In order to balance interviewer workload and to minimize possible seasonal effects on estimates of certain key characteristics such as physical activity, the initial sample of dwellings / telephone numbers was allocated at random, within each HR, over a two-month data collection period.

In the area frame, each start selected within each HR was randomly assigned to a collection period accounting for a number of constraints related to field operations or weighting, while maintaining a uniform size for each period. For example, a sample that is representative of the Canadian population is ensured every six months by ensuring that the dwelling sample covers all LFS strata during this period.

For the lists of telephone numbers, independent samples were selected in each collection period. This strategy ensures that each sample is representative of the Canadian population that is within the scope of the survey in each two months.

5.6 Sampling of interviewees

As was done for the previous cycles, the selection of individual respondents was designed to ensure over-representation of youths (12 to 19). The selection strategy that was adopted accounted for user

[8] Norris, D.A. and Paton, D.G. (1991). Canada's General Social Survey: Five Years of Experience, *Survey Methodology*, 17, 227-240.

needs, cost, design efficiency, response burden and operational constraints. As in Cycle 3.1, CCHS 2007 once again selected one person per household using varying probabilities taking into account the age and the household composition. The selection probabilities resulted from simulations using various parameters in order to determine the optimal approach without causing extreme sampling weights.

Table 5.3 gives the selection weight multiplicative factors used to determine the probabilities of selection of individuals in sampled households by age group. For example, for a three-person household (two adults of age 45 to 64 and one 15-year-old), the teenager would have 5 times more chance of being selected compared to the adults. To avoid extreme sampling weights, there is one exception to this rule: if the size of the household is greater than or equal to 5 or if the number of 12-19 year olds is greater than or equal to 3 then the selection weight multiplicative factor equals 1 for each individual in the household. Consequently, all people in that household have the same probability of being selected.

Table 5.3 Selection weight multiplicative factors for the person-level sampling strategy by age

	Selection Weight Multiplicative Factors				
Age	**12-19**	**20-29**	**30-44**	**45-64**	**65+**
Factor	5	2	2	1	1

5.7 Supplementary buy-in sample in three health regions in Ontario

The province of Ontario requested a sample increase in order to produce estimates at the Local Health Integrated Network (LHIN) geography level. Ontario contains 14 LHIN (see Appendix Y). The CCHS sample was increased in order to obtain a minimum size of 2,000 per LHIN over a period of 2 years. As the HR and LHIN boundaries intersect each other, the stratification level used was the HR-LHIN overlap. The preliminary sample sizes allotted by HR are therefore preserved. In cases where the HR allocation prevented the sample from reaching sizes of 2,000 per LHIN, the sample was then increased, and was allocated proportionally to the size of the population within the HR-LHIN overlap. Table 5.4 provides the sample sizes of targeted respondents by LHIN for 2007.

Table 5.4 Targeted respondents by Local Health Integrated Network (LHIN), CCHS 2007.

LHIN	Targeted respondents
01-Erie St. Clair	1 552
02-South West	2 568
03-Waterloo Wellington	1 230
04-Hamilton Niagara Haldimand Brant	2 620
05-Central West	1 062
06-Mississauga Halton	1 136
07-Toronto Central	1 084
08-Central	1 404
09-Central East	2 076
10-South East	1 344
11-Champlain	2 058
12-North Simcoe Muskoka	1 058
13-North East	1 992
14-North West	1 044
Ontario	**22 228**

The total sample size of the HR-LHIN overlapping areas was then allocated equally between the list frame and the area frame. The usual sample selection procedures within each frame were then applied to the total sample. The additional sample was included as part of the full CCHS sample. Sample sizes by Local Health Integrated Network and frame are given in Appendix D.

5.8 Sub-sample for the Health Services Access Survey (HSAS)

A sub-sample of the CCHS was taken to obtain additional information on the services and access to health care. The survey covers the same population as the CCHS, except for the territories and persons less than 15 years of age.

The budget allocated to this sub-sample was similar to the previous survey, at nearly 34,000 respondents, which ensured that reliable estimates could be produced at the provincial level. Furthermore, the same sample sizes were targeted at the provincial level. On the other hand, in the Maritime Provinces, the available sample size from the CCHS was lower than the 2005 HSAS sample size. In fact, CCHS 2007 only covered half of the sample used in previous cycles given its continuous collection strategy. Except for Prince Edward Island (PEI), the other Maritime Provinces used the CCHS 2007 available sizes only. In PEI, the CCHS sample was increased from 1,000 to 1,480 to meet its needs. In addition to that, Ontario decided to use the entire CCHS sample for HSAS. Here are the final sizes for the HSAS 2007 survey and the comparison with the 2005 HSAS sizes.

Table 5.5: Sample sizes of CCHS 2007 and the HSAS 2007 in comparison with 2005 by province.

	Sample Size		
Province	**CCHS 2007**	**HSAS 2007**	**HSAS 2005**
Newfoundland and Labrador	2,005	2,005	2,700
Prince Edward Island	1,480	1,480	1,850
Nova Scotia	2,520	2,520	3,000
New Brunswick	2,580	2,580	3,000
Quebec	12,145	4,600	4,600
Ontario	22,230	22,230	4,800
Manitoba	3,755	3,200	3,200
Saskatchewan	3,860	3,200	3,200
Alberta	6,105	3,600	3,600
British Columbia	8,050	4,000	4,000
CANADA	**66,280**	**49,415**	**33,950**

Once the size is defined by province, the sample was allocated by HR proportionally to the HR population size, which thus ensured a better sample allocation by province while accounting for the stratification of CCHS by HR. In provinces where the sample size by HR was insufficient, a power allocation with a power less than 1 had to be used. A power of 0.9 was used in Quebec, Alberta and British Columbia, whereas a power of 0.55 had to be used in Manitoba and Saskatchewan, which made the design less optimal. For the other provinces, no allocation by HR was necessary since the entire CCHS sample was used.

Finally, the sample was allocated evenly between the list frame and the area frame. The size was also increased to account for out-of-scope units, and for the predicted non-response rate. Where possible, the size was once again inflated to account for the population not covered by HSAS (12-14 year-olds) and proxy interviews that were not accepted in HSAS. Sample sizes by province and frame for the HSAS sub-sample are given in Appendix D.

Sample selection was performed independently in each collection period based on the CCHS samples. A equally sub-sample of dwellings or telephone numbers was selected randomly in each HR every collection period.

6. DATA COLLECTION

6.1 Computer-assisted interviewing

Between January and December 2007, a total of 65,946 valid interviews were conducted using computer assisted interviewing (CAI). Approximately half the interviews were conducted in person using computer assisted personal interviewing (CAPI) and the other half were conducted over the phone using computer assisted telephone interviewing (CATI).

CAI offers two main advantages over other collection methods. First, CAI offers a case management system and data transmission functionality. This case management system automatically records important management information for each attempt on a case and provides reports for the management of the collection process. CAI also provides an automated call scheduler, i.e. a central system to optimise the timing of call-backs and the scheduling of appointments used to support CATI collection.

The case management system routes the questionnaire applications and sample files from Statistics Canada's main office to regional collection offices (in the case of CATI) and from the regional offices to the interviewers laptops (for CAPI). Data returning to the main office takes the reverse route. To ensure confidentiality, the data is encrypted before transmission. The data are then unencrypted when they are on a separate secure computer with no remote access.

Second, CAI allows for custom interviews for every respondent based on their individual characteristics and survey responses. This includes:

- questions that are not applicable to the respondent are skipped automatically
- 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.
- 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.

6.2 CCHS application development

The CCHS uses two separate CAI applications to collect data, one for telephone interviews (CATI) and one for personal interviews (CAPI). This was done in order to customise each applications' functionality to the type of interview being conducted. Each application consisted of entry, health content (known as the C2), and exit components.

Entry and exit components contain standard sets of questions designed to guide the interviewer through contact initiation, collection of important sample information, respondent selection and determination of cases status. The C2 consists of the health modules themselves and made up the bulk of the applications. This includes common modules asked of all respondents and optional modules which differed by health region. Each application underwent three stages of testing: block, integrated and end to end.

Block level testing consists of independently testing each content module or "block" to ensure skip patterns, logic flows and text, in both official languages, are specified correctly. Skip patterns or logic flows across modules are not tested at this stage as each module is treated as a stand alone questionnaire. Once all blocks are verified by several testers they are added together along with entry and exit components into integrated applications. These newly integrated applications are then ready for the next stage of testing.

Integrated testing occurs when all of the tested modules are added together, along with the entry and exit components, into an integrated application. This second stage of testing ensures that key information such as age and gender are passed from the entry to the C2 and exit components of the applications. It also ensures that variables affecting skip patterns and logic flows are correctly passed between modules within the C2. Since, at this stage the applications essentially function as they will in the field, all possible scenarios faced by interviewers are simulated to ensure proper functionality. These scenarios test various aspects of the entry and exit components including, establishing contact, collecting contact information, determining whether a case is in scope, rostering households, creating appointments and selecting respondents. The applications are also tested to ensure that during an interview, correct modules are triggered reflecting health region optional content selections.

End to end testing occurs when the fully integrated applications are placed in simulated collection environment. The applications are loaded onto computers that are connected to a test server. Data is then collected, transmitted and extracted in real time, exactly as it would be done in the field. This last stage of testing allows for the testing of all technical aspects of data input, transmission and extraction for each of the CCHS applications. It also provided a final chance of finding errors within the entry, C2 and exit components.

6.3 Interviewer training

Project managers from regional collection offices attended, before the start of collection, training sessions at Statistics Canada. These sessions were conducted by the CCHS project team and were used to outline the interviewer training courses to be used in the regions. Project managers then returned to the regions and conducted customised training sessions for their interviewing staff. Members of the survey team attended these training sessions to offer additional support and clarify any questions or concerns that may have arisen.

The focus of these sessions were to get interviewers comfortable using the CCHS 2007 applications and familiarise interviewers with survey content. The training sessions focused on:

- goals and objectives of the survey including a focus on the survey redesign
- survey methodology
- application functionality
- review of the questionnaire content and exercises with an emphasis on significant content changes
- interviewer techniques for maintaining response - complete exercises to minimise non-response
- use of mock interviews to simulate difficult situations and practise potential non-response situations
- survey management
- transmission procedures

One of the key aspects of the training was a focus on minimizing non-response. Exercises to minimise non-response were prepared for interviewers. The purpose of these exercises was to have the interviewers practice convincing reluctant respondents to participate in the survey. There was also a series of refusal avoidance workshops given to the senior interviewers responsible for refusal conversion in each regional collection office.

6.4 The interview

Sample units selected from the telephone list and RDD (Random Digit Dialling) frames were interviewed from centralised call centres using CATI. The CATI interviewers were supervised by a senior interviewer located in the same call centre. Units selected from the area frame were interviewed by decentralised field interviewers using CAPI. While in some situations field interviewers were permitted to complete some or part of an interview by telephone, almost three-quarters (72.1%) of these interviews were conducted exclusively in person. CAPI interviewers worked independently from their homes using laptop computers and were supervised from a distance by senior interviewers. The variable SAM_TYP on the microdata files indicates whether a case was selected from the area frame (CAPI) or from the telephone or RDD frame (CATI).

In all selected dwellings, a knowledgeable household member was asked to supply basic demographic information on all residents of the dwelling. One member of the household was then selected for a more in-depth interview, which is referred to as the C2 Interview.

CAPI interviewers were trained to make an initial personal contact with each sampled dwelling. In cases where this initial visit resulted in non-response, telephone follow-ups were permitted. The variable ADM_N09 on the microdata files indicates whether the interview was completed face-to-face, by telephone or using a combination of the two techniques.

To ensure the quality of the data collected, interviewers were instructed to make every effort to conduct the interview with the selected respondent in privacy. In situations where this was unavoidable, the respondent was interviewed with another person present. Flags on the microdata files indicate whether somebody other than the respondent was present during the interview (ADM_N10) and whether the interviewer felt that the respondent's answers were influenced by the presence of the other person (ADM_N11).

To ensure the best possible response rate attainable, many practices were used to minimise non-response, including:

a) Introductory letters

Before the start of each collection period introductory letters explaining the purpose of the survey were sent to the sampled households. These explained the importance of the survey and provided examples of how CCHS data would be used.

b) Initiating contact

Interviewers were instructed to make all reasonable attempts to obtain interviews. When the timing of the interviewer's call (or visit) was inconvenient, an appointment was made to call back at a more convenient time. If requests for appointments were unsuccessful over the telephone, interviewers were instructed to follow-up with a personal visit. If no one was home on first visit, a brochure with information about the survey and intention to make contact was left at the door. Numerous call-backs were made at different times on different days.

c) Refusal conversion

For individuals who at first refused to participate in the survey, a letter was sent from the nearest Statistics Canada Regional Office to the respondent, stressing the importance of the survey and the household's collaboration. This was followed by a second call (or visit) from a senior interviewer, a project supervisor or another interviewer to try to convince respondent of the importance of participating in the survey.

d) Language barriers

To remove language as a barrier to conducting interviews, each of the Statistics Canada Regional Offices recruited interviewers with a wide range of language competencies. When necessary, cases were transferred to an interviewer with the language competency needed to complete an interview. In addition, the survey questions were translated into the following languages: Chinese, Punjabi and Inuktitut. Chinese and Punjabi were the most common language barriers identified by the regional offices. The Inuktitut translation was used to facilitate collection in Nunavut.

e) Youth interviews

Interviewers were obliged to obtain verbal permission from parents/guardians to interview youths between the ages of 12 to 15 who were selected for interviews. Several procedures were followed by interviewers to alleviate potential parental concerns and to ensure a completed interview. Interviewers carried with them a card entitled "Note to parents / guardians about interviewing youths for the Canadian Community Health Survey". This card explained the purpose of collecting information from youth, lists the subjects to be covered in the survey, asks for permission to share and link the obtained information and explains the need to respect a child's right to privacy and confidentiality.

If a parent/guardian asked to see the actual questions; interviewers were instructed to either show the survey questions, or if the interviewer was being conducted by phone, to immediately have the regional office send a copy of the questionnaire.

If privacy could not be obtained to interview the selected youth either in person or over the phone (another person listening in) the interview was coded a refusal. However, for CAPI interviews, if privacy could not be obtained to interview the selected youth, the interviewer was able to propose to the parent/guardian that the interviewer read the questions out loud and the youth enter their answers directly on the computer.

During all interviews conducted with youths, survey questions regarding income and food security were answered by the parent/guardian. These questions were asked at the end of the survey questionnaire, so that when they came up, the parent/guardian could complete the interview.

f) Proxy interviews

In cases where the selected respondent was, for reasons of physical or mental health, incapable of completing an interview, another knowledgeable member of the household supplied information about the selected respondent. This is known as a proxy interview. While proxy interviewees were able to provide accurate answers to most of the survey questions, the more sensitive or personal questions were beyond the scope of knowledge of a proxy respondent. This resulted in some questions from the proxy interview being unanswered. Every effort was taken to keep proxy interviews to a minimum. The variable ADM_PRX indicates whether a case was completed by proxy.

6.5 Field operations

The majority of the 2007 sample was divided into six non-overlapping two-month collection periods. Regional collection offices were instructed to use the first 4 weeks of each collection period to resolve the majority of the sample, with next 4 weeks being used finalise the remaining sample and to follow up on outstanding non-response cases. All cases were to have been attempted by the second week of each collection period.

Sample files were sent approximately two weeks before the start of each collection period to centralised collection offices. A series of dummy cases were included with each CAPI sample. These cases were completed by senior interviewers for the purposes of ensuring that all data transmission procedures were working through the collection cycle. Once, the samples were received, project supervisors were responsible for planning CAPI interviewer assignments. Wherever possible, assignments were generally no larger than 15 cases per interviewer.

Transmission of cases from each of the CATI offices to head office was the responsibility of the regional office project supervisor, senior interviewer and the technical support team. These transmissions were performed nightly and sent all completed cases to Statistics Canada's head office. Completed CAPI interviews were transmitted daily from the interviewer's home directly to Statistics Canada's head office using a secure telephone transmission.

At the end of data collection, a national response rate of 78% was achieved. Complete details regarding the response rates can be found in Appendix D

6.6 Quality control and collection management

During the 2007 collection year, several methods were used to ensure data quality and to optimize collection. These included using internal measures to verify interviewer performance and the use of a series of ongoing reports to monitor various collection targets and data quality.

A system of validation was used for CAPI cases whereby interviewers had their work validated on a regular basis by the Regional Office. Each collection period, randomly selected cases were flagged in the sample. Regional office managers and supervisors created lists of cases to be validated. These cases were handed to the validation team who then contacted households to verify that a legitimate interview took place. Validation procedures generally occurred during the first few weeks of a collection period to ensure that any issues were detected promptly. Interviewers were provided feedback by their supervisors on a regular basis.

CATI interviewers were also randomly chosen for validation. Validation in the CATI collection offices consisted of senior interviewers monitoring interviews to ensure proper techniques and procedures (reading the questions as worded in the applications, not prompting respondents for answers, etc.) were followed by the interviewer.

A series of reports were produced to effectively track and manage collection targets and to assist in identifying other collection issues.

Cumulative reports were generated at the end of each collection period, showing response, link, share and proxy rates for both the CATI and CAPI samples by individual health region. The reports were useful in identifying health regions that were below collection target levels, allowing the regional offices to focus efforts in these regions.

Using information obtained from the CAI applications, further analysis was done in head office in order to identify interviews that were completed below acceptable time frames. These short interviews were flagged, removed from the microdata and treated as non-response.

Customised reports were also created and used to examine specific data quality issues that arose during collection. For example, the CCHS protocol allows interviews by proxy when the selected respondent is mentally or physically incapable of undertaking the survey. Proxy interviews have limited value since many modules are skipped during these interviews for data quality reasons. Therefore, it is important to determine the rates of proxy interviews and the rationales behind them. A proxy report was created to keep track of proxy interviews and their justification. Through these reports it was possible to identify interviewers who seemed to be unclear as to the circumstances in which a proxy interview would be acceptable. These interviewers then received additional/refresher training reviewing the procedures related to proxy interviews.

7. DATA PROCESSING

7.1 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".

7.2 Coding

Pre-coded answer categories were supplied for all suitable variables. Interviewers were trained to assign the respondent's answers to the appropriate category.

In the event that a respondent's answer could not be easily assigned to an existing category, several questions also allowed the interviewer to enter a long-answer text in the "Other-specify" category. All such questions were closely examined in head office processing. 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. For all questions, the 'Other-specify' responses are taken into account when refining the answer categories for future cycles.

7.3 Creation of derived variables

To facilitate data analysis and to minimize the risk of error, a number of variables on the file have been derived using items found on the CCHS 2007 questionnaire. Derived variables generally have a "D", "G" or "F" in the fourth 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. The *Derived Variables Documentation (DV)* provides details on how these more complex variables were derived. For more information on the naming convention, please go to Section 12.5.

7.4 Weighting

The principle behind estimation in a probability sample such as the CCHS Cycle 3.1 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 PUMF, 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 8.

8. WEIGHTING

In order for estimates produced from survey data to be representative of the covered population, and not just the sample itself, users must incorporate the survey weights in their calculations. A survey weight is given to each person included in the final sample, that is, the sample of persons having responded to the survey. This weight corresponds to the number of persons in the entire population that are represented by the respondent.

As described in Section 5, the CCHS has recourse to three sampling frames for its sample selection: an area frame acting as the primary frame and two frames formed of telephone numbers complementing the area frame. Since only minor differences differentiate the two telephone frames in terms of weighting, they are treated together as one and referred to as being part of the telephone frame.

The weighting strategy was developed by treating both the area and telephone frames independently. Household-level weights resulting from these two frames are then combined into a single set of household weights through a step called "*integration*". After applying person-level weights and some further adjustments, this integrated weight becomes the final weight. Note that depending on the need, one or two frames were used for the selection of the sample within a given health region (HR). The weighting strategy deals with this aspect at the integration step.

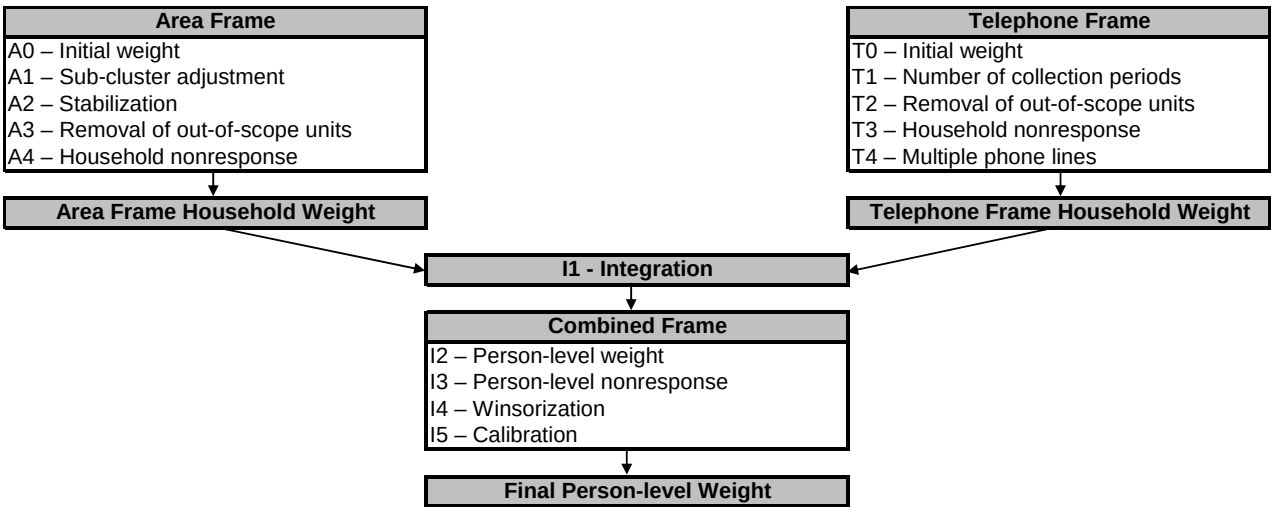
8.1 Overview

As mentioned previously, units from both the area and telephone frames are treated separately up to the integration step. The following sections describe the weighting process for the provinces. Sub-section 8.2 provides details on the weighting strategy for the area frame, while sub-section 8.3 deals with the strategy for the telephone frame. The integration of the two frames is discussed in 8.4. This is followed by the last weighting step - calibration, where the weights are adjusted to control for seasonality and to match known population totals. This step is explained in sub-section 8.5.

Although the two frames are used to cover the three territories, the sampling methods used are slightly modified. These modifications affect the weighting of these three regions substantially, and they are reported in sub-section 8.6.

Diagram A presents an overview of the different adjustments that are part of the weighting strategy. A numbering system is used to identify each adjustment and will be used throughout the section. Letters *A* and *T* are used as prefixes to refer to adjustments applied to the units on the *A*rea and *T*elephone frames respectively, while prefix *I* identifies adjustments applied from the *I*ntegration step onwards.

Diagram A Weighting strategy overview



8.2 Weighting of the area frame sample

A0 – Initial weight

The weighting on the area frame sample begins with a weight provided by the LFS. This weight is based on the LFS design since the CCHS sample is derived from the LFS. The LFS design consists of a sample of dwellings within the selected clusters of the LFS strata. In the initial adjustment, the LFS weight is adjusted to take into consideration the fact that the CCHS selects a sample to be representative of the Health Region. As well, the CCHS selects a different number of clusters than the LFS and can repeat the sampling of dwellings within the selected clusters. The resulting weight is called A0. For more details about the selection mechanism, as well as a more complete definition of strata and clusters, refer to Statistics Canada (1998)[9].

A1 – Sub-cluster adjustment

In clusters that experience significant growth, a sub-sampling methodology is used to ensure that the workload of the interviewers is kept at a reasonable level. This can consist of sub-sampling from the selected dwellings, dividing the cluster into sub-clusters, or reclassifying the cluster as a stratum and creating new clusters within the stratum. In all these cases, a sub-sample adjustment is calculated and applied to the CCHS weight. This adjustment is applied to weight A0 to produce weight A1. Again, more information can be found in the LFS documentation (Statistics Canada (1998)).

[9] Statistics Canada. 1998. *Methodology of the Canadian Labour Force Survey.* Statistics Canada. Cat. No. 71-526-XPB.

A2 – Stabilization

In some HRs, the increase of the sample size as described in section 5, results in a significantly larger sample than necessary. Stabilization is used to bring the sample size back down to the desired level. The stabilization process consists of randomly sub-sampling dwellings at the HR level from the dwellings originally obtained within each cluster. An adjustment factor representing the effect of this stabilization is calculated in order to adjust the probability of selection appropriately. This factor, multiplied by weight A1, produces weight A2.

A3 – Removal of out-of-scope units

Among all dwellings sampled, a certain proportion is identified during collection as being out-of-scope. Dwellings that are demolished or under construction, vacant, seasonal or secondary, and institutions are examples of out-of-scope cases for the CCHS. These dwellings and their associated weight are simply removed from the sample. This leaves a sample that consists of, and representative of, in-scope dwellings. These remaining dwellings maintain the same weight as in the previous step, which is now called A3.

A4 – Household nonresponse

During collection, a certain proportion of sampled households inevitably result in nonresponse. This usually occurs when a household refuses to participate in the survey, provides unusable data, or cannot be reached for an interview. Weights of the nonresponding households are redistributed to responding households within response homogeneity groups (RHGs). In order to create the response groups, a scoring method based on logistic regression models is used to determine the propensity to respond and these response probabilities are used to divide the sample into groups with similar response properties. The information available for nonrespondents is limited so the regression model uses characteristics such as the collection period and geographic information, as well as paradata, which includes the number of contact attempts, the time/day of attempt, and whether the household was called on a weekend or weekday. An adjustment factor is calculated within each class as follows:

$$\frac{\textit{Sum of weight A3 for all households}}{\textit{Sum of weight A3 for all responding households}}$$

Weight A3 is multiplied by this factor to produce weight A4 for the responding households. Non-responding households are dropped from the process at this point.

8.3 Weighting of the telephone frame sample

As mentioned earlier, the telephone frame is composed of two frames: a Random Digit Dialling (RDD) frame and a list frame. Only one of the frames can be used within an HR. When the list frame is used, it is always used as a complement to the area frame within the HR. When the RDD frame is used, it is always used as the only frame within the HR. For the purposes of

weighting, units coming from the two telephone frames are treated together and therefore are subject to the same adjustments.

The geographical boundaries used to select the sample from the telephone frame did not perfectly replicate the HR geography. Consequently, some units were samlped from one HR but the information collected at the time of the interview placed them in a neighbouring HR. This particularity was handled in the weighting by applying the first 3 telephone adjustments (T0, T1 and T2) relative to the HR assigned at the time of sample selection. The remaining 2 adjusments (T3 and T4) were applied to the HR based on information collected from the respondent to ensure that all units belong to their correct HR.

T0 –Initial weight

The initial weight is defined as the inverse of the probability of selection and is computed separately for the RDD and list frame samples since the method of selection differs between these two frames. For the RDD frame, the selection of telephone numbers is done within each RDD stratum. An RDD stratum is an aggregation of area code prefixes (ACP: the first six digits of a 10-digit telephone number), with each ACP containing valid banks of one hundred numbers (see Norris and Paton[10] for more details). Therefore, the probability of selection is the ratio between the number of sampled units and one hundred times the number of banks within the RDD stratum.

For the list frame, telephone numbers are selected among those assigned to the specific HR. The probability of selection corresponds to the ratio of the number of sampled units to the number of telephone numbers on the list within the HR. The ratio is based on the frame available and the number of units selected for the particular two-month collection period. The probability of selection can therefore change depending on sample allocation and frame updates. The inverse of these probabilities represents the initial weight T0.

T1 – Number of collection periods

On the area frame, the entire sample is selected at the beginning of the year. This is in contrast to the telephone frame, where samples are drawn every two months. Each of these samples comes with an initial weight that allows each sample to be representative at the HR level. To ensure that the total sample represents the population only once, an adjustment factor is applied to reduce the weights of each two-month sample. The adjustment factor applied to each two-month sample is equal to the the inverse of the number of samples being combined (i.e. the number of collection periods). Following this adjustment, the entire list frame sample corresponds to the average over the entire combined collection period. The initial weights are multiplied by this adjustment factor to produce weight T1.

[10] Norris, D.A. and Paton, D.G. 1991. Canada's General Social Survey: Five Years of Experience. *Survey Methodology*. 17, 227-240.

T2 - Removal of out-of-scope numbers

Telephone numbers associated with businesses, institutions or other out-of-scope dwellings, as well as numbers not in service or any other non-working numbers are all examples of out-of-scope cases for the telephone frame. Similar to the methods used on the area frame, these cases are simply removed from the process, leaving only in-scope dwellings in the sample. These in-scope dwellings keep the same weight as in the previous step, now called weight T2.

T3 – Household nonresponse

The adjustment applied here to compensate for the effect of household nonresponse is identical to the one applied for the area frame (adjustment A4) although the paradata used does differ because of the differences in collection applications for personal and telephone interviews. The adjustment factor calculated within each class was obtained as follows:

$$\frac{\textit{Sum of weights T2 for all households}}{\textit{Sum of weights T2 for all responding households}}$$

The weight T2 of responding households is multiplied by this factor to produce the weight T3. Nonresponding households are removed from the process at this point.

T4 - Multiple phone lines

Some households can possess more than one residential telephone line. This has an impact on the weighting as these households have a higher probability of being selected so the weights for these households need to be adjusted for the number of residential telephone lines within the household. The adjustment factor represents the inverse of the number of lines in the household. The weight T4 is obtained by multiplying this factor by the weight T3.

8.4 Integration of the telephone and area frames (I1)

This step consists of integrating the weights for households common to the area and telephone frames into a single weight by applying a method of integration[11]. Those units on the area frame that are not on the telephone frame do not have their weights adjusted. For all others, an adjustment factor between 0 and 1 is determined at the health region level in such a way that it represented the relative importance of each sample in the total sample. This relative importance is measured in terms of sample size. The higher the proportion that the sample represents of the total sample, the higher its relative importance. The integration factor can be calculated as follows:

$$\alpha = \; n_A / \left(n_A + n_T \right)$$

[11] Skinner, C.J. and Rao, J.N.K. 1996. Estimation in Dual Frame Surveys with Complex Designs. *Journal of the American Statistical Association*. 91, 433, 349-356.

where n_A and n_T represent the area and telephone frames sample sizes respectively. The weight of the area frame units is multiplied by this factor α, while the weight of the telephone frame units is multiplied by 1- α. Note that in the case where an HR is covered by only one frame, the adjustment factor is equal to 1. The product between the factor derived here and the final household weight calculated earlier (A4 or T4, depending on which frame the unit belongs to), gives the integrated household weight I1.

8.5 Post-integration weighting steps

I2 – Creation of person level weight

Since persons are the desired sampling units, the household level weights computed to this point need to be converted to the person level. This weight is obtained by multiplying the weight I1 by the inverse of the probability of selection of the person selected in the household. This gives the weight I2. As mentioned earlier, the probability of selection for an individual changes depending on the number of people in the household and the ages of those individuals (see Section 5.5 for more details).

I3 – Person nonresponse

A CCHS interview can be seen as a two-part process. First, the interviewer gets the complete roster of the people within the household. Second, (s)he interviews the selected person. In some cases, interviewers can only get through the first part, either because they cannot get in touch with the selected person, or because that selected person refuses to be interviewed. Such individuals are defined as person nonrespondents and an adjustment factor must be applied to the weights of respondents to account for this nonresponse. Using the same methodology that was used in the treatment of household nonresponse, the adjustment was applied within response homogeneity groups based on characteristics available for both respondents and non-respondents. All characteristics collected when creating the roster of household members were in fact available for the creation of the classes as well as geographic information and some paradata. The scoring method was used to define the classes. In the end, the following adjustment factor was calculated:

$$\frac{\textit{Sum of weight I2 for all selected persons}}{\textit{Sum of weight I2 for all responding selected persons}}$$

Weight I2 for responding persons was multiplied by the above adjustment factor to produce weight I3. Nonresponding persons were dropped from the weighting process from this point onward.

I4 – Winsorization

Following the series of adjustments applied to the respondents, some units may come out with extreme weights compared with other units of the same domain of interest. Some respondents could represent a large proportion of their HR or have a large impact on the variance. In order to prevent this, the weight of the outlier units that represent a large proportion of their HR-age-sex group is adjusted downward using a “winsorization” trimming approach.

I5 – Calibration

The final step necessary to obtain the final CCHS weight is calibration (I5). Calibration is done using CALMAR[12] to ensure that the sum of the final weights corresponds to the population estimates defined at the HR level, for all 10 age-sex groups of interest. The five age groups are 12-19, 20-29, 30-44, 45-64, 65+, for both males and females. A minimum domain size of 20 respondents is required to calibrate at the HR by age by sex level. For domains that have less 20 respondents, some collapsing is done within province and / or within gender. At the same time, weights are adjusted to ensure that the each collection period (two-month period) is equally represented within the sample. Note that the calibration is done using the most up to data geography and may not match the geography used in sampling.

The population estimates are based on the most recent Census counts and counts of birth, death, immigration and emigration since that time. The average of these monthly estimates for each of the HR-age-sex post-strata by collection period is used to calibrate. The weight I4 was therefore adjusted using CALMAR to obtain the final weight I5. Weight I5 corresponds to the ***final CCHS weight*** that can be found on the data file with the variable name WTS_M.

8.6 Particular aspects of the weighting in the three territories

As described in Section 5, the sampling frame used in the three territories is somewhat different from the one used in the provinces. Therefore, the weighting strategy is adapted to comply with these differences. This section summarises the changes applied to the steps described in sub-sections 8.1 to 8.5.

For the area frame, as mentioned in sub-section 5.4.1, an additional stage of selection is added in the territories where each territory is stratified into groupings of communities and one community is selected within each group. The capital of each territory forms a stratum on its own and is selected automatically at the first stage. This has an effect in the computation of the probability of selection, and therefore in the value of the initial weight (A0). Once the initial weight is calculated, the same series of adjustments (A1 to A4) is applied to the area frame units. Household-level and person-level nonresponse adjustment classes are built in the same way as for the provinces, using the same set of variables.

For the weighting of the telephone frame units, it should be noted that only the RDD frame is used and exclusively in the Yukon and Northwest Territories capitals. All of the telephone frame adjustments are applied to derive a final weight for the telephone units. The two sets of weights (area and telephone) are subsequently integrated and post-stratified in a similar way to what is done for the provinces, with two exceptions. First, the integration is applied only to units located in the Yukon and Northwest Territories capitals since the other communities are covered only by the area frame. Second, the population counts used for calibration for Nunavut represent 70% of the entire population because of the under-coverage of the area frame that was described in section 5.4.1.

[12] Sautory O. CALMAR 2: A New Version of the CALMAR Calibration Adjustment Program. *Proceedings of Statistics Canada Symposium* (Statistics Canada, Catalogue no. 11-522-XCB), 2003.

9. DATA QUALITY

9.1 Response rates

In total, 84,973 of the selected units in the CCHS 2007 were in-scope for the survey[13]. Out of these, 71,922 households accepted to participate in the survey resulting in an overall household-level response rate of 84.6%. Among these responding households, 71,922 individuals (one per household) were selected to participate in the CCHS Cycle 3.1, out of which a response was obtained for 65,946 individuals, resulting in an overall person-level response rate of 91.7%. At the Canada level, this yields a combined response rate of 77.6% for the CCHS 2007. Appendix E provides combined response rates as well as relevant information for their calculation by health region or combined health region, as well as by Local Health Integrated Network (LHIN) level in Ontario and by province for the Health Services Access Survey (HSAS) sub-sample.

Next, we describe how the various components of the equation should be handled to correctly compute combined response rates.

Household-level response rate

HHRR =		*# of responding households in both frames*
		all in-scope households in both frames

Person-level response rate

PPRR =		*# of responding persons in both frames*
		all selected persons in both frames

[13] Among the units selected, some are not in-scope for the survey. They are, for examples, vacant, demolished or non-residential dwellings or invalid phone numbers such as phone numbers without service or non-residential lines. These units are identified during the data collection, otherwise, they would have been excluded before the sample selection. These units are not considered in the calculation of response rates.

Combined response rate = HHRR x PPRR

Next is an example on how to calculate the combined response rate for Canada using the information found in Table 9.1.

HHRR =	35,748 + 36,174	=	71,922	= 0.846
	40,668 + 44,310		84,978	

PPRR =	33,473 + 32,473	=	65,946	= 0.917
	35,748 + 36,174		71,922	

Combined response rate = 0.846 x 0.917

= 0.776

= **77.6%**

9.2 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. than 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*.

9.2.1 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 CCHS 2007. 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 the CCHS 2007 was minimal; once the questionnaire was started, it tended to be completed with very little non-response. Total non-response occurred either because a person refused to participate in the survey or because the interviewer was unable to contact the selected person. 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 on the weight adjustment for non-response.

9.2.2 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 it is estimated that 25% of Canadians aged 12 and over are regular smokers and that this estimate is found to have a standard deviation of 0.003. Then the CV of the estimate is calculated as:

$$(0.003/0.25) \times 100\% = 1.20\%$$

Statistics Canada commonly uses CV results when analyzing data and urges users producing estimates from the CCHS Cycle 3.1 data files to also do so. For details on how to determine CVs, see Section 11. For guidelines on how to interpret CV results, see the table at the end of Sub-section 10.4.

10. GUIDELINES FOR TABULATION, ANALYSIS AND RELEASE

This section of the documentation outlines the guidelines to be used by users in tabulating, analyzing, publishing or otherwise releasing any data derived from the survey 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.

10.1 Rounding guidelines

In order that estimates for publication or other release derived from the data files (Master, Share or PUMF) 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.

10.2 Sample weighting guidelines for tabulation

The sample design used for this survey 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 data file 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.

10.2.1 Definitions: categorical estimates, quantitative estimates

Before discussing how the survey 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 data files.

Categorical estimates:

Categorical estimates are estimates of the number or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of individuals who smoke daily 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:

*At the present do/does ... smoke cigarettes daily, occasionally or not at all? (**SMK_202**)*
__ Daily
__ Occasionally
__ Not at all

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 cigarettes smoked per day by individuals who smoke daily. The numerator is an estimate of the total number of cigarettes smoked per day by individuals who smoke daily, and its denominator is an estimate of the number of individuals who smoke daily.

Example of quantitative question:

*How many cigarettes do/does you/he/she smoke each day now? **(SMK_204)***
|_|_| Number of cigarettes

10.2.2 Tabulation of categorical estimates

Estimates of the number of people with a certain characteristic can be obtained from the data file 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.

10.2.3 Tabulation of quantitative estimates

Estimates of sums or averages for quantitative variables can be obtained using the following three steps (only step a) is necessary to obtain the estimate of a sum):

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.

For example, to obtain the estimate of the average number of cigarettes smoked each day by individuals who smoke daily, first compute the numerator ($\hat{X}$) by summing the product between the value of variable **SMK_204** and the weight **WTS_M.** Next, sum this value over those records with a value of "daily" to the variable **SMK_202**. The denominator ($\hat{Y}$) is obtained by summing the final weight of those records with a value of "daily" to the variable **SMK_202**. Divide ($\hat{X}$) by ($\hat{Y}$) to obtain the average number of cigarettes smoked each day by daily smokers.

10.3 Guidelines for statistical analysis

The CCHS 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.

10.4 Release guidelines

Before releasing and/or publishing any estimate from the data files, users must first determine the number of sampled respondents having the characteristic of interest (for example, the number of respondents who smoke when interested in the proportion of smokers for a given population) in order to ensure that enough observations are available to calculate a quality estimate. For users of the PUMF, if this number is less than 30, the unweighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. For users of the master or share files, it is recommended to have at least 10 observations. For weighted estimates, based on sample sizes of 10 or more (30 for the PUMF), users should determine the coefficient of variation of the estimate and follow the guidelines below.

Table 10.1 Sampling variability guidelines

Type of Estimate	CV (in %)	Guidelines
Acceptable	$0.0 \leq CV \leq 16.5$	Estimates can be considered for general unrestricted release. Requires no special notation.
Marginal	$16.6 < CV \leq 33.3$	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."

11. APPROXIMATE SAMPLING VARIABILITY TABLES

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

The coefficients of variation (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, for each table produced, a conservative value among all design effects relative to that table. The value chosen was then used to generate a table that applies to the entire set of characteristics.

The design effects, sample sizes and population counts used to produce the Approximate Sampling Variability Tables as well as the tables are presented in Appendix E, which is included on the PUMF CD. All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Options concerning the computation of exact coefficients of variation are discussed in sub-section 11.7.

Remember: As indicated in Sampling Variability Guidelines in Section 10.4, if the number of observations on which an estimate is based is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation. Coefficients of variation based on small sample sizes are too unpredictable to be adequately represented in the tables.

11.1 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 appropriate Approximate Coefficients of Variations Table, 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. Since not all the possible values for the estimate are available, the smallest value which is the closest must be taken (as an example, if the estimate is equal to 1,700 and the two closest available values are 1,000 and 2,000, the first has to be chosen). This figure is the approximate coefficient of variation.

Rule 2: Estimates of proportions or percentages of people possessing a characteristic

The coefficient of variation of an estimated proportion (or percentage) depends on both the size of the proportion and the size of the numerator upon which the proportion is based. Estimated proportions are relatively more reliable than the corresponding estimates of the numerator of the proportion when the proportion is based upon a sub-group of the population. This is due to the fact that the coefficients of variation of the latter type of estimates are based on the largest entry in a row of a particular table, whereas the coefficients of variation of the former type of estimators are based on some entry (not necessarily the largest) in that same row. (Note that in the tables the CVs decline in value reading across a row from left to right). For example, the estimated proportion of individuals who smoke daily out of those who smoke at all is more reliable than the estimated number who smoke daily.

When the proportion (or percentage) is based upon the total population covered by each specific table, the CV of the proportion is the same as the CV of the numerator of the proportion. In this case, this is equivalent to applying Rule 1.

When the proportion (or percentage) is based upon a subset of the total population (e.g., those who smoke at all), reference should be made to the proportion (across the top of the table) and to the numerator of the proportion (down the left side of the table). Since not all the possible values for the proportion are available, the smallest value which is the closest must be taken (for example, if the proportion is 23% and the two closest values available in the column are 20% and 25%, 20% must be chosen). The intersection of the appropriate row and column gives the coefficient of variation.

Rule 3: Estimates of differences between aggregates or percentages

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

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

where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$. This formula is accurate for the difference between independent populations or subgroups, but is only approximate otherwise. It will tend to overstate the error, if $\hat{X}_1$ and $\hat{X}_2$ are positively correlated and understate the error if $\hat{X}_1$ and $\hat{X}_2$ are negatively correlated.

Rule 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 individuals who smoke at all and the numerator is the number of individuals who smoke daily out of those who smoke at all.

Consider the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of individuals who smoke daily or occasionally as compared to the number of individuals who do not smoke at all. The standard deviation of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by $\hat{R}$, where $\hat{R}$ is the ratio of the estimates ($\hat{R} = \hat{X}_1 / \hat{X}_2$). That is, the standard error of a ratio is:

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

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

The coefficient of variation of $\hat{R}$ is given by $\sigma_{\hat{R}} / \hat{R} = \sqrt{\alpha_1^{\,2} + \alpha_2^{\,2}}$. The formula will tend to overstate the error, if $\hat{X}_1$ and $\hat{X}_2$ are positively correlated and understate the error if $\hat{X}_1$ and $\hat{X}_2$ are negatively correlated.

Rule 5: Estimates of differences of ratios

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

11.2 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 4,722,617 individuals smoke daily in Canada. How does the user determine the coefficient of variation of this estimate?

1) Refer to the CANADA level CV table.

2) The estimated aggregate (4,722,617) does not appear in the left-hand column (the "Numerator of Percentage" column), so it is necessary to use the smallest figure closest to it, namely 4,000,000.

3) The coefficient of variation for an estimated aggregate (expressed as a percentage) is found by referring to the first non-asterisk entry on that row, namely, 1.0%.

4) So the approximate coefficient of variation of the estimate is 1.0%. According to the Sampling Variability Guidelines presented in Section 10.4, the finding that there were 4,722,617 individuals who smoke daily is publishable with no qualifications.

Example 2 : Estimates of proportions or percentages possessing a characteristic

Suppose that the user estimates that 4,722,617/6,081,453=77.7% of individuals in Canada who smoke at all smoke daily. How does the user determine the coefficient of variation of this estimate?

1) Refer to the CANADA level CV table.

2) Because the estimate is a percentage which is based on a subset of the total population (i.e., individuals who smoke at all, that is to say, daily or occasionally), it is necessary to use both the percentage (77.7%) and the numerator portion of the percentage (4,722,617) in determining the coefficient of variation.

3) The numerator (4,722,617) does not appear in the left-hand column (the "Numerator of Percentage" column) so it is necessary to use the smallest figure closest to it, namely 4,000,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the figure closest to it, 70.0%.

4) The figure at the intersection of the row and column used, namely 0.6% is the coefficient of variation (expressed as a percentage) to be used.

5) So the approximate coefficient of variation of the estimate is 0.6%. According to the Sampling Variability Guidelines presented in Section 10.4, the finding that 77.7% of individuals who smoke at all smoke daily can be published with no qualifications.

Example 3 : Estimates of differences between aggregates or percentages

Suppose that a user estimates that, among men, 2,535,367/13,078,499 = 19.4% smoke daily (estimate 1), while for women, this percentage is estimated at 2,187,250 / 13,476,931 = 16.2% (estimate 2). How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the CANADA level CV table in the same manner as described in example 2 gives the CV for estimate 1 as 1.5% (expressed as a percentage), and the CV for estimate 2 as 1.5% (expressed as a percentage).

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

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

Where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The standard error of the difference $\hat{d}$ = (0.194 - 0.162) = 0.032 is :

$$\sigma_{\hat{d}} = \sqrt{[0(.194)(0.015)]^2 + [(0.162)(0.015)]^2}$$
$$= 0.004$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.004/0.032 = 0.125.

4) So the approximate coefficient of variation of the difference between the estimates is 12.5% (expressed as a percentage). According to the Sampling Variability Guidelines presented in Section 10.4, this estimate can be published with no qualifications.

Example 4 : Estimates of ratios

Suppose that the user estimates that 4,722,617 individuals smoke daily, while 1,358,836 individuals smoke occasionally. The user is interested in comparing the estimate of daily to occasional smokers in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

1) First of all, this estimate is a ratio estimate, where the numerator of the estimate (= $\hat{X}_1$) is the number of individuals who smoke occasionally. The denominator of the estimate (= $\hat{X}_2$) is the number of individuals who smoke daily.

2) Refer to the CANADA level CV table.

3) The numerator of this ratio estimate is 1,358,836. The smallest figure closest to it is 1,000,000. The coefficient of variation for this estimate (expressed as a percentage) is found by referring to the first non-asterisk entry on that row, namely, 2.3%.

4) The denominator of this ratio estimate is 4,722,617. The figure closest to it is 4,000,000. The coefficient of variation for this estimate (expressed as a percentage) is found by referring to the first non-asterisk entry on that row, namely, 1.0%.

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

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

That is,

$$\alpha_{\hat{R}} = \sqrt{(.023)^2 + (.01)^2}$$

$$= 0.025$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The obtained ratio of occasional to daily smokers is 1,358,836/4,722,617 which is 0.29:1. The coefficient of variation of this estimate is 2.5% (expressed as a percentage), which is releasable with no qualifications, according to the Sampling Variability Guidelines presented in Section 10.4.

11.3 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 of 100 that the differences would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

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

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

$$CI_X = [\ \hat{X} - z\ \hat{X}\ \alpha_{\hat{X}}\ ,\ \hat{X} + z\ \hat{X}\ \alpha_{\hat{X}}\ ]$$

Where $\alpha_{\hat{X}}$ is determined coefficient of variation for $\hat{X}$, and

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

Note: Release guidelines presented in section 10.4 which apply to the estimate also apply to the confidence interval. For example, if the estimate is not releasable, then the confidence interval is not releasable either.

11.4 Example of using the CV tables to obtain confidence limits

A 95% confidence interval for the estimated proportion of individuals who smoke daily from those who smoke at all (from example 2, sub-section 11.2) would be calculated as follows:

$\hat{X}$ = 0.777

z = 2

$\alpha_{\hat{X}}$ = 0.006 is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.777 - (2) (0.777) (0.006) , 0.777 + (2) (0.777) (0.006)}

$CI_{\hat{x}}$ = {0.768 , 0.786}

11.5 How to use the CV tables to do a Z-test

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

Let $\hat{X}_1$ and $\hat{X}_2$ be sample estimates for 2 characteristics of interest. Let the standard error on the difference $\hat{X}_1 - \hat{X}_2$ be $\sigma_{\hat{d}}$. If the ratio of $\hat{X}_1 - \hat{X}_2$ over $\sigma_{\hat{d}}$ is between -2 and 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level.

11.6 Example of using the CV tables to do a Z-test

Let us suppose we wish to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of men who smoke daily AND the proportion of women who smoke daily. From example 3, sub-section 11.2, the standard error of the difference between these two estimates was found to be = 0.004. Hence,

$$z = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.194 - 0.162}{0.004} = \frac{0.032}{0.004} = 8$$

Since $z = 8$ is greater than 2, it must be concluded that there is a significant difference between the two estimates at the 0.05 level of significance. Note that the two sub-groups compared are considered as being independent, so the test is correct.

11.7 Exact variances/coefficients of variation

All coefficients of variation in the Approximate Sampling Variability Tables (CV Tables) are indeed approximate and, therefore, unofficial.

The computation of exact coefficients of variation is not a straightforward task since there is no simple mathematical formula that would account for all CCHS sampling frame and weighting aspects. Therefore, other methods such as resampling methods must be used in order to estimate measures of precision. Among these methods, the bootstrap method is the one recommended for analysis of CCHS data.

The computation of coefficients of variation (or any other measure of precision) with the use of the bootstrap method requires access to information that is considered confidential and not available on the PUMF. This computation must be done using the Master file. Access to the Master file is discussed in section 12.3.

For the computation of coefficients of variation, the bootstrap method is advised. A macro program, called “Bootvar”, was developed in order to give users easy access to the bootstrap method. The Bootvar program is available in SAS and SPSS formats, and is made up of macros that calculate the variances of totals, ratios, differences between ratios, and linear and logistic regressions.

There are a number of reasons why a user may require an exact variance. A few are given below.

Firstly, if a user desires estimates at a geographic level other than those available in the tables (for example, at the rural/urban level), then the CV tables provided are not adequate. Coefficients of variation of these estimates may be obtained using "domain" estimation techniques through the exact variance program.

Secondly, should a user require more sophisticated analyses such as estimates of parameters from linear regressions or logistic regressions, the CV tables will not provide correct associated coefficients of variation. Although some standard statistical packages allow sampling weights to be incorporated in the analyses, the variances that are produced often do not take into account the stratified and clustered nature of the design properly, whereas the exact variance program would do so.

Thirdly, for estimates of quantitative variables, separate tables are required to determine their sampling error. Since most of the variables for the CCHS are primarily categorical in nature, this has not been done. Thus, users wishing to obtain coefficients of variation for quantitative variables can do so through the exact variance program. 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 estimate of the total number of cigarettes smoked each day by individuals who smoke daily would be greater than the coefficient of variation of the corresponding estimate of the number of individuals who smoke daily. Hence if the coefficient of variation of the latter is not releasable, then the coefficient of variation of the corresponding quantitative estimate will also not be releasable.

Lastly, should users find themselves in a position where they can use the CV tables, but this renders a coefficient of variation in the "marginal" range (16.6% - 33.3%), the user should release the associated estimate with a warning cautioning users of the high sampling variability associated with the estimate. This would be a good opportunity to recalculate the coefficient of variation through the exact variance program to find out if it is releasable without a qualifying note. The reason for this is that the coefficients of variation produced by the tables are based on a wide range of variables and are therefore considered crude, whereas the exact variance program would give an exact coefficient of variation associated with the variable in question.

11.8 Release cut-offs for the CCHS

Appendix E presents tables giving the minimum cut-offs for estimates of totals at the Canada, provincial, health region and CLSC levels and those for various age groups at the Canada level. Estimates smaller than the value given in the "Marginal" column may not be released under any circumstances.

12. MICRODATA FILES: DESCRIPTION, ACCESS AND USE

The CCHS produces three types of microdata files: master files, share files and public use microdata files (PUMF).

12.1 Master files

The master files contain all variables and all records from the survey collected during a collection period. These files are accessible at Statistics Canada for internal use and in Statistics Canada's Research Data Centres (RDC), and are also subject to custom tabulation requests.

12.1.1 Research Data Centre

The RDC Program enables researchers to use the survey data in the master files in a secure environment in several universities across Canada. Researchers must submit research proposals that, once approved, give them access to the RDC. For more information, please consult the following web page: http://www.statcan.ca/english/rdc/index.htm

12.1.2 Custom tabulations

Another way to access the master files is to offer all users the option of having staff in Client Services of the Health Statistics Division prepare custom tabulations. This service is offered on a cost-recovery basis. It 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. For more information, please contact Client Services at 613-951-1653 or by e-mail at hd-ds@statcan.ca.

12.1.3 Remote access

Finally, the remote access service to the survey master files is another way to have access to these data if, for some reason, the user cannot access a Research Data Centre. Each purchaser of the microdata product can be supplied with a synthetic or 'dummy' master file and a corresponding record layout. With these tools, the researcher can develop his own set of analytical computer programs. The code for the custom tabulations is then sent via e-mail to cchs-escc@statcan.ca. The code will then be transferred into Statistics Canada's internal secured network and processed using the appropriate master file of CCHS data. Estimates generated will be released to the user, subject to meeting the guidelines for analysis and release outlined in Section 10 of this document. Results are screened for confidentiality and reliability concerns and then the output is returned to the client. There is no charge for this service.

12.2 Share files

The share files contain all variables and all records of CCHS respondents who agreed to share their data with Statistic Canada's partners, which are the provincial and territorial health departments, Health Canada and the Public Health Agency of Canada. Statistics Canada also asks respondents living in Quebec for their permission to share their data with the Institut de la statistique du Québec. The share file is released only to these organizations. Personal identifiers are removed from the share files to respect respondent confidentiality. Users of these files must first certify that they will not disclose, at any time, any information that might identify a survey respondent.

12.3 Public use microdata files

The public use microdata files (PUMF) are developed from the master files using a technique that balances the need to ensure respondent confidentiality with the need to produce the most useful data possible at the health region level. The PUMF must meet stringent security and confidentiality standards required by the *Statistics Act* before they are released for public access. To ensure that these standards have been achieved, each PUMF goes through a formal review and approval process by an executive committee of Statistics Canada.

There is no charge to access the PUMF in a post-secondary educational institution that is part of the Data Liberation Initiative. They are also free of charge from Client Services on request.

12.4 Use of weight variable

The weight variable **WTS_M** represents the sampling weight for key survey files. For a given respondent, the sampling weight can be interpreted as the number of people the respondent represents in the Canadian population. This weight must always be used when computing statistical estimates in order to make inference at the population level possible. The production of unweighted estimates is not recommended. The sample allocation, as well as the survey design specifics can cause such results to not correctly represent the population. Refer to section 8 on weighting for a more detailed explanation on the creation of this weight.

The **Food Security** module measures concepts that apply not only to the respondent's situation, but also to that of the respondent's entire household. Depending on the level of analysis, the analysis of the variables may require use of a weight calculated to represent the number of Canadian households, rather than the number of persons. This weight variable **WTS_HH** is found in a separate file (HS_HHWT.txt). It can be used in place of the variable **WTS_M** for household analyses at the national and provincial levels. The weight variable **WTS_M** must be used for regional analyses.

12.5 Variable naming convention beginning in 2007

The variable naming convention adopted allows data users to easily use and identify the data based on the module and variable type. The CCHS variable naming convention fulfils two requirements: to restrict variable names to a maximum of eight characters for ease of use by analytical software products and to identify easily conceptually identical variables from one survey collection period to the next. Questions to which changes are made between two collection periods, and where the changes alter the concept measured by the question, are entirely renamed to avoid any confusion in the analysis.

The CCHS variable naming convention was changed beginning with the data from the 2007 collection period. The letter corresponding to the survey version (e.g., A = cycle 1.1, C = cycle 2.1 and E = 3.1) is no longer used in the variable names. A new variable (REFPER, format = YYYYMM-YYYYMM) was added to the microdata files in order to identify the beginning and the end of the reference during which data included in the file were collected. This variable will be useful, notably for users wanting to use data from several collection periods at a time. Therefore, variable names for identical modules or questions from one collection year to the next (e.g., 2007 and 2008) will be the same.

The naming convention used for variables beginning with the 2007 CCHS use up to eight characters. The variable names are structured as follows:

Positions 1 to 3:	Module/questionnaire section name
Position 4:	Variable type (underscore, C, D, F or G)
Positions 5 to 8:	Question number and answer option for multiple response questions

Example 1 shows that the structure of the variable name for question 202, Smoking Module, is SMK_202 :

Positions 1 to 3:	SMK	Smoking module
Position 4 :	_	(underscore = collected data)
Position 5 to 8:	202	Question number

Example 2 shows the structure of the variable name for question 2 of the Health Care Utilization Module (HCU_02A), which is a multi-response question:

Positions 1 to 3:	**HCU**	Health care utilization module
Position 4 :	_	(underscore = collected data)
Position 5 to 8:	02AA	Corresponding question number and answer option

Positions 1 to 3 contain the acronyms for each of the modules. These acronyms appear beside the module names given in the table in Appendix A.

Position 4 designates the variable type based on whether it is a variable collected directly from a questionnaire question ("_"), from a coded ("C"), derived ("D"), grouped ("G"), or flag ("F") variable.

In general, the last four positions (5 to 8) 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 or three digit format. For example, question Q01A in the questionnaire becomes simply 01A, and question Q15 becomes simply 15.

Table 12.1 Designation of codes used in the 4th position of the CCHS variable names

_	Collected variable	A variable that appears 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)

For questions that 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.

12.6 Variable naming convention before 2007

As mentioned earlier, the variable naming convention was changed in 2007. The flag for the cycle in which the variables were collected was removed. This flag was found in the 4th position for cycles 1.1 to 3.1.

Here is the list of letters used in the CCHS microdata files between cycles 1.1 and 3.1 and their corresponding cycle.

Letter	Cycle and cycle name
A	Cycle 1.1: Canadian Community Health Survey
B	Cycle 1.2: Canadian Community Health Survey - Mental Health and Well-Being
C	Cycle 2.1: Canadian Community Health Survey
D	Cycle 2.2: Canadian Community Health Survey - Nutrition
E	Cycle 3.1: Canadian Community Health Survey

12.7 Guidelines for the use of sub-sample variables

In 2007, a sub-sample of respondents was selected at random to answer questions from three modules about the performance of the Canadian health care system: Access to Health Care Services (ACC), Waiting Times (WTM) and Patient Satisfaction (PAS).[14]

12.7.1 Variable naming convention for sub-sample modules

Variables from sub-sample content modules are designated with a "Z" in position 4 of the variable name (e.g., ACC**Z**_01).

Each of the sub-sample modules was also made available to health regions as optional content. The Ontario and New Brunswick health regions chose the Access to Health Care Services module, while the Patient Satisfaction module was selected by Ontario, Saskatchewan, the Yukon and the Northwest Territories. There are two sets of variables for these provinces (e.g., ACC_11 and ACC**Z**_11). These variables are applicable or not applicable for a given respondent according to whether or not the respondent is a member of the relevant sub-sample **and** whether or not the respondent lives in a health region where the module was chosen as optional content. Table 1 illustrates the possibilities.

[14] In 2000-2001, this sub-sample was part of a follow-up survey to the CCHS known as the Health Services and Access Survey (HSAS). This survey was incorporated in the CCHS in 2003.

Table 12.2 Possible not-applicable values for a sub-sample variable

Member of sub-sample	Respondent living in a health region where the Access to Health Care Services (ACC) module was part of the optional content	
	YES	**NO**
YES	ACC_11 and ACCZ_11 apply	ACC_11 does not apply and ACCZ_11 applies
NO	ACC_11 applies and ACCZ_11 does not apply	ACC_11 and ACCZ_11 do not apply

12.7.2 Separate file

To encourage appropriate use of CCHS data, sub-sample content is provided on a physically separate file. Each file has a corresponding sample weight and set of bootstrap weights that must be used to produce valid estimates for all variables on the file. The sampling weight for sub-sample 1 (in which all modules apply to respondents aged 15 and older) is calibrated so that it represents the total Canadian population aged 15 and older (xxxx). Table 3 describes the two data files released in 2007.

Table 12.3 Names and content of 2007 CCHS data files

Files	File name	Sampling weight	Bootstrap weights file	Variables included	Records included
Main master file	HS.txt	WTS_M	b5.txt	All common and all optional modules.	All respondent records
Sub-sample 1 master file	HSS1.txt	WTS_S1M	b5_s1.txt	All common modules, plus the "**Z**" set of variables for the Access to Health Care Services, Waiting Times and Patient Satisfaction module.	Records of all respondents selected for the sub-sample
Share file	HS.txt	WTS_S	b5.txt	All common and all optional modules.	Records of all respondents who agreed to share their data
Sub-sample 1 share file 1	HSS1.txt	WTS_S1S	b5_s1.txt	All common modules, plus the "**Z**" set of variables for the Access to Health Care Services, Waiting Times and Patient Satisfaction module.	Records of all respondents selected for the sub-sample who agreed to share their data

12.7.3 Analysis combining sub-sample and optional content

The aim of the CCHS sub-sample modules is to permit the calculation of estimates at the provincial and national levels. The territories are excluded from the sub-sample. The sub-sample file and the associated weights are not intended to support health region level estimates.

As mentioned earlier, two of the three sub-sample modules were also selected by all health regions of at least one province as optional content. In these instances, it is possible to calculate health region and provincial estimates of the variables in that module using the **main master file** and the **master sampling weight**. This offers the advantage of a large sample size for the calculation and therefore smaller sampling error. The provincial estimate can then be compared to the estimate for other provinces calculated using the appropriate sub-sample file and sub-sample sampling weight.

12.8 Data dictionaries

Separate data dictionary reports, including universe statements and frequencies, are provided for the main master file and each of the sub-sample files.

In the master file data dictionary reports, optional content modules are treated in the same way as previous CCHS cycles. For each module, a flag indicates whether a given respondent lives in a health region where the module was selected as optional content. When the flag is equal to 2 (No), all variables in the module have "not applicable" values. For example, the variable PASFOPT indicates whether the module Patient Satisfaction is applicable for a given respondent.

12.9 Differences in calculation of common content variables using different files

Variables from common content modules can be estimated using either of the two data files provided. Depending on which file is used, very small differences will be observed.

All official Statistics Canada estimates of variables from common modules are based on the main master file sampling weight.

APPENDIX A

CANADIAN COMMUNITY HEALTH SURVEY (2007-2008)

<table>
<tr><th colspan="5">Core content (all regions)</th></tr>
<tr><td colspan="2">• Age of respondent (ANC)
• Alcohol consumption (ALC)
• Chronic health problem (CCC)
• Exposure to second-hand smoke (ETS)
• Flu vaccination (FLU)
• General health status (GEN)
• Health care utilization (HCU)
• Pain and illness (HUP)</td><td colspan="2">• Height and weight – self-reported (HWT)
• Motherhood– Breastfeeding (MEX)
• Fruit and vegetable consumption (FVC)
• Physical activities (PAC)
• Limitation of activities (RAC)
• Smoking (SMK)</td><td>Administration and socio-demographic information
• Administrative information (ADM)
• Dwelling information (DWL)
• Education (EDU)
• Income (INC)
• Labour force (LBS)
• Socio-demographic information (DEM)</td></tr>
<tr><th colspan="5">Theme content (all regions)</th></tr>
<tr><td colspan="2">Theme for 2007-2008: Healthy living</td><td colspan="2">Theme for 2007: Health Service and Access Surveyi</td><td>Theme for 2008: Prevention of Chronic Illnesses</td></tr>
<tr><td>• Changes to improve health (CIH)
• Food safety (FSC)
• Oral health 1 (OH1)</td><td>• Physical activities - Workplace facilities (PAF)
• Sedentary activities (SAC)</td><td colspan="2">• Access to Health Care Services (ACCZ)
• Patient Satisfaction with Health Care Services (PASZ)
• Waiting Times (WTMZ)</td><td>• Blood test (BLT)
• PAP smear test (PAP)
• Mammogram (MAM)</td><td>• Spirometry (SPI)
• General physical (PCU)
• Colorectal cancer exams (CCS)</td></tr>
</table>

Optional content (certain regions)			
• Access to health care services (ACC)[ii] • Alcohol consumption - Dependency (ALD) • Alcohol consumption – Former drinkers (ALN) • Drank alcohol in past week (ALW) • Blood pressure check (BPC) • Breast examination (BRX) • Breast self-examination (BSX) • Stress – Childhood and adult stressors (CST) • Colorectal cancer exam (CCS) • Mental health consultations (CMH) • Dental visits (DEN) • Depression (DEP) • Diabetes care (DIA) • Dietary supplement use – Vitamins and minerals (DSU) • Distress (DIS) • Driving and safety (DRV)	• Eye examinations (EYX) • Food choices (FDC) • Health Status Index (HUI) • Health status (SFR) • Home care (HMC) • Home safety (HMS) • Use of illegal drugs (DRG) • Injuries (INJ) • Insurance coverage (INS) • Mammogram (MAM) • Self-control (MAS) • Motherhood – Drank alcohol during pregnancy (MXA) • Motherhood – Smoked during pregnancy (MXS) • Smoking – Nicotine addiction (NDE)	• Oral health 2 (OH2) • PAP smear test (PAP) • Patient satisfaction with health care services (PAS)[ii] • Patient satisfaction with community health care services (PSC) • Canadian Problem Gambling Index (CPG) • Prostate cancer screening (CCS) • Psychological well-being (PWB) • Stress – Recent events (RLE) • Satisfaction with life (SWL) • Self-esteem (SFE) • Sexual behaviour (SXB) • Sleep (SLP) • Smoking – Smoking cessation methods (SCA) • Smoking – Physician counselling (SPC)	• Smoking - Coping (SCH) • Social support - Availability (SSA) • Social support - Utilization (SSU) • Spiritual values (SPR) • Stress - Coping (STC) • Stress – Sources (STS) • Suicidal thoughts and attempts (SUI) • Sun safety (SSB) • Smoking – Other tobacco products (TAL) • Use of protective equipment (UPE) • Voluntary organizations - Participation (ORG) • Work stress (WST) • Youth smoking (YSM)

i Asked of a sub-sample of respondents. The 2007 theme (Health Service and Access Survey) was not asked of territorial respondents.
ii These 2007 theme content modules were also selected as optional content by certain regions.

APPENDIX B

SELECTION OF OPTIONAL CONTENT BY PROVINCE OR TERRITORY (2007)

Optional modules	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.	Nvt.
Access to health care services				*		*							
Alcohol - Dependency								*		*			
Alcohol - Former drinkers	*									*			
Drank alcohol in past week	*		*			*				*			*
Blood pressure check								*	*			*	
Breast examinations				*				*				*	
Breast self-examinations	*			*				*					
Stress - Childhood and adult stressors												*	
Colorectal cancer exams	*	*				*		*				*	*
Mental health consultations		*	*	*		*			*			*	*
Dental visits	*								*			*	
Depression			*	*	*				*		*		*
Diabetes care	*	*	*	*		*				*	*	*	
Dietary supplement use - Vitamins and minerals											*	*	
Distress			*		*								
Driving and safety							*		*			*	
Eye examinations						*							
Food choices		*					*		*	*		*	
Health care system satisfaction						*					*	*	
Health Status Index					*								
Home care						*							
Home safety												*	
Use of illegal drugs			*							*			
Injuries			*							*			
Insurance coverage				*									

Optional modules	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.	Nvt.
Mammogram	*		*	*		*		*				*	
Self-control							*						
Motherhood– Drank alcohol during pregnancy						*		*		*		*	
Motherhood– Smoked during pregnancy						*		*		*		*	*
Smoking – Nicotine addiction									*			*	*
Oral health 2	*												
PAP smear test				*		*		*				*	
Health care services patient satisfaction						*		*			*	*	
Community health care services patient satisfaction				*				*			*	*	
Canadian Problem Gambling Index					*	*		*					*
Prostate cancer screening	*	*						*				*	
Psychological well-being				*									
Stress – Recent events											*		
Satisfaction with life					*				*				*
Self-esteem													*
Sexual behaviour			*	*		*		*				*	*
Health status (SF-36)	*	*					*						
Sleep			*					*			*		
Smoking - Physician counselling									*				
Stress - Coping						*							
Smoking – Smoking cessation methods						*							
Social support - Availability			*		*					*	*		*
Social support - Utilization										*	*		*

Optional modules	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.	Nvt.
Spiritual values								*					*
Stress - Coping											*		*
Stress - Sources											*		*
Suicidal thoughts and attempts						*			*	*		*	
Sun safety		*											
Smoking - Other tobacco products						*		*	*				
Use of protective equipment			*					*	*				
Voluntary organizations - Participation			*									*	
Work stress									*				
Youth smoking										*			

APPENDIX C

CCHS GEOGRAPHY 2007

0	Canada
10	Newfoundland and Labrador
1011-C	Eastern Regional Integrated Health Authority
1012-I	Central Regional Integrated Health Authority
1013-I	Western Regional Integrated Health Authority
1014-H	Labrador-Grenfell Regional Integrated Health Authority
11	Prince Edward Island
1101-D	Kings County
1102-A	Queens County
1103-C	Prince County
12	Nova Scotia
1201-C	Zone 1
1202-C	Zone 2
1203-C	Zone 3
1204-C	Zone 4
1205-I	Zone 5
1206-A	Zone 6
13	New Brunswick
1301-C	Region 1
1302-C	Region 2
1303-C	Region 3
1304-C	Region 4
1305-I	Region 5
1306-I	Region 6
1307-I	Region 7
24	Quebec
2401-C	Région du Bas-Saint-Laurent
2402-C	Région du Saguenay - Lac-Saint-Jean
2403-A	Région de la Capitale-Nationale
2404-C	Région de la Mauricie et du Centre-du-Québec
2405-C	Région de l'Estrie
2406-G	Région de Montréal
2407-A	Région de l'Outaouais
2408-C	Région de l'Abitibi-Témiscamingue
2409-H	Région de la Côte-Nord
2410-H	Région du Nord-du-Québec
2411-I	Région de la Gaspésie - Îles-de-la-Madeleine
2412-E	Région de la Chaudière-Appalaches
2413-A	Région de Laval
2414-E	Région de Lanaudière
2415-E	Région des Laurentides
2416-A	Région de la Montérégie
35	Ontario by Local Health Integration Network
3501	Erie St. Clair Health Integration Network
3502	South West Health Integration Network
3503	Waterloo Wellington Health Integration Network
3504	Hamilton Niagara Haldimand Brant Health Integration Network

3505	Central West Health Integration Network
3506	Mississauga Halton Health Integration Network
3507	Toronto Central Health Integration Network
3508	Central Health Integration Network
3509	Central East Health Integration Network
3510	South East Health Integration Network
3511	Champlain Health Integration Network
3512	North Simcoe Muskoka Health Integration Network
3513	North East Health Integration Network
3514	North West Health Integration Network
35	Ontario by Health Unit
3526-C	District of Algoma Health Unit
3527-A	Brant County Health Unit
3530-B	Durham Regional Health Unit
3531-E	Elgin-St. Thomas Health Unit
3533-E	Grey Bruce Health Unit
3534-E	Haldimand-Norfolk Health Unit
3535-E	Haliburton, Kawartha, Pine Ridge District Health Unit
3536-B	Halton Regional Health Unit
3537-A	City of Hamilton Health Unit
3538-A	Hastings and Prince Edward Counties Health Unit
3539-E	Huron County Health Unit
3540-A	Chatham-Kent Health Unit
3541-A	Kingston, Frontenac and Lennox and Addington Health Unit
3542-A	Lambton Health Unit
3543-E	Leeds, Grenville and Lanark District Health Unit
3544-A	Middlesex-London Health Unit
3546-A	Niagara Regional Area Health Unit
3547-C	North Bay Parry Sound District Health Unit
3549-H	Northwestern Health Unit
3551-B	City of Ottawa Health Unit
3552-E	Oxford County Health Unit
3553-B	Peel Regional Health Unit
3554-E	Perth District Health Unit
3555-A	Peterborough County-City Health Unit
3556-H	Porcupine Health Unit
3557-E	Renfrew County and District Health Unit
3558-E	Eastern Ontario Health Unit
3560-E	Simcoe Muskoka District Health Unit
3561-C	Sudbury and District Health Unit
3562-C	Thunder Bay District Health Unit
3563-C	Timiskaming Health Unit
3565-B	Waterloo Health Unit
3566-B	Wellington-Dufferin-Guelph Health Unit
3568-B	Windsor-Essex County Health Unit
3570-B	York Regional Health Unit
3595-G	City of Toronto Health Unit

46	Manitoba	
4610-A		Winnipeg Regional Health Authority
4615-A		Brandon Regional Health Authority
4620-E		North Eastman Regional Health Authority
4625-E		South Eastman Regional Health Authority
4630-E		Interlake Regional Health Authority
4640-D		Central Regional Health Authority
4645-D		Assiniboine Regional Health Authority
4660-D		Parkland Regional Health Authority
4670-H		Norman Regional Health Authority
4685-F		Burntwood/Churchill
47	Saskatchewan	
4701-D		Sun Country Regional Health Authority
4702-D		Five Hills Regional Health Authority
4703-D		Cypress Regional Health Authority
4704-A		Regina Qu'Appelle Regional Health Authority
4705-D		Sunrise Regional Health Authority
4706-A		Saskatoon Regional Health Authority
4707-D		Heartland Regional Health Authority
4708-D		Kelsey Trail Regional Health Authority
4709-C		Prince Albert Parkland Regional Health Authority
4710-H		Prairie North Regional Health Authority
4714-F		Mamawetan/Keewatin/Athabasca
48	Alberta	
4821-E		Chinook Regional Health Authority
4822-E		Palliser Health Region
4823-B		Calgary Health Region
4824-E		David Thompson Regional Health Authority
4825-E		East Central Health
4826-E		Capital Health
4827-E		Aspen Regional Health Authority
4828-E		Peace Country Health
4829-H		Northern Lights Health Region
59	British Columbia	
5911-E		East Kootenay Health Service Delivery Area
5912-C		Kootenay-Boundary Health Service Delivery Area
5913-A		Okanagan Health Service Delivery Area
5914-C		Thompson/Cariboo Health Service Delivery Area
5921-A		Fraser East Health Service Delivery Area
5922-B		Fraser North Health Service Delivery Area
5923-B		Fraser South Health Service Delivery Area
5931-B		Richmond Health Service Delivery Area
5932-G		Vancouver Health Service Delivery Area
5933-B		North Shore/Coast Garibaldi Health Service Delivery Area
5941-A		South Vancouver Island Health Service Delivery Area
5942-A		Central Vancouver Island Health Service Delivery Area
5943-C		North Vancouver Island Health Service Delivery Area
5951-H		Northwest Health Service Delivery Area

5952-H	Northern Interior Health Service Delivery Area
5953-H	Northeast Health Service Delivery Area
60	Yukon Territory
6001-H	Yukon Territory
61	Northwest Territories
6101-H	Northwest Territories
62	Nunavut – 10 largest communities
6201-F	Nunavut – 10 largest communities
A	Peer group A
B	Peer group B
C	Peer group C
D	Peer group D
E	Peer group E
F	Peer group F
G	Peer group G
H	Peer group H
I	Peer group I

APPENDIX D

SAMPLE ALLOCATION BY HEALTH REGION AND FRAME

		Area Frame		Phone frames		Combined	
Prov./Terr.	**Health Region**	**# expected respondents**	**raw sample size**	**# expected respondents**	**raw sample size**	**# expected respondents**	**raw sample size**
Canada	Total	33375	47120	32903	53978	66278	101098
N.L.	Total	1003	1397	1003	1353	2006	2750
	1011	405	556	405	548	810	1104
	1012	235	340	235	324	470	664
	1013	213	271	213	282	426	553
	1014	150	230	150	199	300	429
P.E.I.	Total	741	1110	739	1134	1480	2244
	1101	88	150	88	138	176	288
	1102	410	618	418	648	828	1266
	1103	243	342	233	348	476	690
N.S.	Total	1261	1864	1261	1748	2522	3612
	1201	198	300	198	282	396	582
	1202	160	219	160	212	320	431
	1203	180	239	180	261	360	500
	1204	175	300	175	257	350	557
	1205	210	297	210	286	420	583
	1206	338	509	338	450	676	959
N.B.	Total	1289	1811	1289	2115	2578	3926
	1301	250	358	250	413	500	771
	1302	243	361	243	403	486	764
	1303	235	340	235	391	470	731
	1304	135	183	135	232	270	415
	1305	125	172	125	209	250	381
	1306	173	232	173	251	346	483
	1307	128	165	128	216	256	381
Que.	Total	5872	8161	6272	10753	12144	18914
	2401	300	395	300	449	600	844
	2402	314	402	314	453	628	855
	2403	464	608	464	711	928	1319
	2404	401	523	401	601	802	1124
	2405	309	470	309	471	618	941
	2406	776	1110	776	1357	1552	2467
	2407	323	485	323	504	646	989
	2408	300	436	300	429	600	865
	2409	300	423	300	457	600	880
	2410	0	0	400	1584	400	1584
	2411	300	418	300	452	600	870
	2412	361	480	361	556	722	1036
	2413	335	466	335	540	670	1006
	2414	359	472	359	570	718	1042
	2415	380	604	380	633	760	1237
	2416	650	869	650	986	1300	1855
Ont.	Total	11114	15657	11114	18290	22228	33947
	3526	213	342	213	338	426	680

		Area Frame		Phone frames		Combined	
Prov./Terr.	Health Region	# expected respondents	raw sample size	# expected respondents	raw sample size	# expected respondents	raw sample size
	3527	203	284	203	315	406	599
	3530	408	582	408	629	816	1211
	3531	170	222	170	275	340	497
	3533	237	340	237	401	474	741
	3534	193	283	193	331	386	614
	3535	238	391	238	434	476	825
	3536	353	445	353	544	706	989
	3537	413	592	413	663	826	1255
	3538	235	336	235	389	470	725
	3539	148	186	148	231	296	417
	3540	200	243	200	311	400	554
	3541	253	390	253	410	506	800
	3542	218	304	218	360	436	664
	3543	238	337	238	380	476	717
	3544	375	582	375	597	750	1179
	3546	383	497	383	623	766	1120
	3547	200	281	200	325	400	606
	3549	164	258	164	336	328	594
	3551	513	746	513	774	1026	1520
	3552	188	258	188	256	376	514
	3553	657	896	657	1130	1314	2026
	3554	163	221	163	241	326	462
	3555	213	327	213	380	426	707
	3556	188	264	188	297	376	561
	3557	188	271	188	326	376	597
	3558	260	347	260	387	520	734
	3560	571	844	571	947	1142	1791
	3561	270	359	270	437	540	796
	3562	358	486	358	589	716	1075
	3563	125	186	125	196	250	382
	3565	383	515	383	560	766	1075
	3566	282	367	282	431	564	798
	3568	358	481	358	566	716	1047
	3570	471	604	471	814	942	1418
	3595	1084	1590	1084	2067	2168	3657
Man.	Total	1877	2552	1877	2829	3754	5381
	4610	528	705	528	747	1056	1452
	4615	140	185	140	217	280	402
	4620	125	180	125	240	250	420
	4625	150	199	150	221	300	420
	4630	173	280	173	263	346	543
	4640	200	251	200	272	400	523
	4645	178	240	178	252	356	492
	4660	133	189	133	193	266	382
	4670	125	169	125	224	250	393
	4685	125	154	125	200	250	354

		Area Frame		Phone frames		Combined	
Prov./Terr.	**Health Region**	**# expected respondents**	**raw sample size**	**# expected respondents**	**raw sample size**	**# expected respondents**	**raw sample size**
Sask.	Total	1806	2510	2056	3890	3862	6400
	4701	150	193	150	220	300	413
	4702	150	213	150	228	300	441
	4703	133	187	133	198	266	385
	4704	310	427	310	436	620	863
	4705	155	219	155	229	310	448
	4706	330	443	330	474	660	917
	4707	135	227	135	188	270	415
	4708	130	191	130	180	260	371
	4709	163	217	163	259	326	476
	4710	150	193	150	227	300	420
	4714	0	0	250	1251	250	1251
Alta.	Total	3052	4244	3052	4642	6104	8886
	4820	255	338	255	381	510	719
	4821	208	255	208	300	416	555
	4822	698	1019	698	1036	1396	2055
	4823	350	503	350	529	700	1032
	4824	223	297	223	333	446	630
	4825	655	904	655	963	1310	1867
	4826	270	415	270	445	540	860
	4827	233	312	233	376	466	688
	4828	160	201	160	279	320	480
B.C.	Total	4025	5707	4025	6625	8050	12332
	5911	152	205	152	245	304	450
	5912	155	211	155	244	310	455
	5913	295	356	295	479	590	835
	5914	250	380	250	408	500	788
	5921	260	346	260	412	520	758
	5922	380	471	380	640	760	1111
	5923	400	568	400	657	800	1225
	5931	213	293	213	367	426	660
	5932	400	646	400	750	800	1396
	5933	273	411	273	447	546	858
	5941	338	468	338	534	676	1002
	5942	263	365	263	399	526	764
	5943	132	182	132	200	264	382
	5951	163	275	163	293	326	568
	5952	213	330	213	323	426	653
	5903	138	200	138	227	276	427
Terr.	60	475	733	125	334	600	1067
	61	510	815	90	265	600	1080
	62	350	559	0	0	350	559

Sample allocation by Local Health Integrated Network and frames in the CCHS 2007 in Ontario.

		Area Frame		Phone frames		Combined	
Prov./Terr.	LHIN	# expected respondents	raw sample size	# expected respondents	raw sample size	# expected respondents	raw sample size
Ont.	Total	11114	15657	11114	18290	22228	33947
	3501	776	1028	776	1237	1552	2265
	3502	1284	1815	1284	2008	2568	3823
	3503	615	818	615	916	1230	1734
	3504	1310	1800	1310	2110	2620	3910
	3505	531	729	531	921	1062	1650
	3506	568	756	568	945	1136	1701
	3507	542	796	542	1034	1084	1830
	3508	702	948	702	1247	1404	2195
	3509	1038	1560	1038	1786	2076	3346
	3510	672	990	672	1096	1344	2086
	3511	1029	1460	1029	1595	2058	3055
	3512	529	781	529	877	1058	1658
	3513	996	1432	996	1593	1992	3025
	3514	522	744	522	925	1044	1669

Sample allocation by province and frames for the Health Services Access Survey (HSAS) sub-sample 2007

	Area Frame		Phone frames		Combined	
Prov./Terr.	# expected respondents[1]	raw sample size	# expected respondents[1]	raw sample size	# expected respondents[1]	raw sample size
Canada	23,690	35,789	23,690	40,844	47,380	76,633
N.L.	937	1,397	937	1,353	1,874	2,750
P.E.I.	693	1,110	693	1,134	1,386	2,244
N.S.	1,178	1,864	1,178	1,748	2,356	3,612
N.B.	1,204	1,811	1,204	2,115	2,408	3,926
Que.	2,300	3,449	2,300	4,003	4,600	7,452
Ont.	10,378	15,657	10,378	18,290	20,756	33,947
Man.	1,600	2,363	1,600	2,598	3,200	4,961
Sask.	1,600	2,325	1,600	3,071	3,200	5,396
Alta.	1,800	2,752	1,800	2,944	3,600	5,696
B.C.	2,000	3,061	2,000	3,588	4,000	6,649

APPENDIX E

RESPONSE RATE BY HEALTH REGION AND FRAMES

		Area frame							Phone frames							
Prov. Terr.	Health Region	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	Combined resp. rates
Canada	**Total**	40663	35748	87.9	35748	33473	93.6	82.3	44310	36174	81.6	36174	32473	89.8	73.3	77.6
N.L.	**Total**	1192	1111	93.2	1111	1036	93.2	86.9	1247	1065	85.4	1065	958	90	76.8	81.8
	1011	491	453	92.3	453	426	94	86.8	488	415	85	415	370	89.2	75.8	81.3
	1012	269	257	95.5	257	240	93.4	89.2	323	277	85.8	277	252	91	78	83.1
	1013	227	217	95.6	217	210	96.8	92.5	252	215	85.3	215	194	90.2	77	84.3
	1014	205	184	89.8	184	160	87	78	184	158	85.9	158	142	89.9	77.2	77.6
P.E.I.	**Total**	889	790	88.9	790	744	94.2	83.7	888	733	82.5	733	659	89.9	74.2	79
	1101	82	70	85.4	70	65	92.9	79.3	61	50	82	50	44	88	72.1	76.2
	1102	553	481	87	481	451	93.8	81.6	567	466	82.2	466	429	92.1	75.7	78.6
	1103	254	239	94.1	239	228	95.4	89.8	260	217	83.5	217	186	85.7	71.5	80.5
N.S.	**Total**	1543	1368	88.7	1368	1270	92.8	82.3	1665	1415	85	1415	1304	92.2	78.3	80.2
	1201	219	208	95	208	199	95.7	90.9	261	225	86.2	225	210	93.3	80.5	85.2
	1202	191	174	91.1	174	159	91.4	83.2	188	160	85.1	160	148	92.5	78.7	81
	1203	215	196	91.2	196	190	96.9	88.4	232	194	83.6	194	184	94.8	79.3	83.7
	1204	214	182	85	182	163	89.6	76.2	227	189	83.3	189	171	90.5	75.3	75.7
	1205	253	235	92.9	235	222	94.5	87.7	274	233	85	233	213	91.4	77.7	82.5
	1206	451	373	82.7	373	337	90.3	74.7	483	414	85.7	414	378	91.3	78.3	76.6
N.B.	**Total**	1518	1358	89.5	1358	1277	94	84.1	1792	1540	85.9	1540	1427	92.7	79.6	81.7
	1301	305	266	87.2	266	242	91	79.3	350	294	84	294	275	93.5	78.6	78.9
	1302	278	244	87.8	244	232	95.1	83.5	338	286	84.6	286	265	92.7	78.4	80.7
	1303	271	239	88.2	239	226	94.6	83.4	318	271	85.2	271	250	92.3	78.6	80.8
	1304	169	141	83.4	141	136	96.5	80.5	183	162	88.5	162	153	94.4	83.6	82.1
	1305	150	138	92	138	132	95.7	88	186	166	89.2	166	148	89.2	79.6	83.3
	1306	199	192	96.5	192	178	92.7	89.4	231	204	88.3	204	192	94.1	83.1	86

		Area frame							Phone frames							
Prov. Terr.	Health Region	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	Combined resp. rates
	1307	146	138	94.5	138	131	94.9	89.7	186	157	84.4	157	144	91.7	77.4	82.8
Que.	**Total**	7202	6338	88	6338	6017	94.9	83.5	8506	6960	81.8	6960	6233	89.6	73.3	78
	2401	344	320	93	320	314	98.1	91.3	364	321	88.2	321	294	91.6	80.8	85.9
	2402	353	321	90.9	321	300	93.5	85	382	323	84.6	323	291	90.1	76.2	80.4
	2403	606	531	87.6	531	493	92.8	81.4	651	534	82	534	483	90.4	74.2	77.6
	2404	479	429	89.6	429	411	95.8	85.8	555	465	83.8	465	422	90.8	76	80.6
	2405	404	336	83.2	336	320	95.2	79.2	370	324	87.6	324	300	92.6	81.1	80.1
	2406	969	820	84.6	820	772	94.1	79.7	1248	904	72.4	904	777	86	62.3	69.9
	2407	416	341	82	341	323	94.7	77.6	431	352	81.7	352	320	90.9	74.2	75.9
	2408	382	359	94	359	347	96.7	90.8	346	311	89.9	311	291	93.6	84.1	87.6
	2409	373	330	88.5	330	315	95.5	84.5	383	327	85.4	327	292	89.3	76.2	80.3
	2410	.	.	.	.	.	.	.	541	433	80	433	386	89.1	71.3	71.3
	2411	342	322	94.2	322	311	96.6	90.9	374	327	87.4	327	283	86.5	75.7	83
	2412	404	376	93.1	376	364	96.8	90.1	457	390	85.3	390	360	92.3	78.8	84.1
	2413	429	368	85.8	368	333	90.5	77.6	539	424	78.7	424	362	85.4	67.2	71.8
	2414	444	383	86.3	383	369	96.3	83.1	476	387	81.3	387	352	91	73.9	78.4
	2415	445	401	90.1	401	375	93.5	84.3	509	414	81.3	414	374	90.3	73.5	78.5
	2416	812	701	86.3	701	670	95.6	82.5	880	724	82.3	724	646	89.2	73.4	77.8
Ont.	**Total**	13841	12031	86.9	12031	11183	93	80.8	15419	12251	79.5	12251	10784	88	69.9	75.1
	3526	298	249	83.6	249	229	92	76.8	278	238	85.6	238	217	91.2	78.1	77.4
	3527	256	231	90.2	231	205	88.7	80.1	269	219	81.4	219	201	91.8	74.7	77.3
	3530	521	454	87.1	454	418	92.1	80.2	559	448	80.1	448	393	87.7	70.3	75.1
	3531	193	169	87.6	169	157	92.9	81.3	217	176	81.1	176	161	91.5	74.2	77.6
	3533	269	242	90	242	233	96.3	86.6	308	248	80.5	248	227	91.5	73.7	79.7
	3534	238	208	87.4	208	188	90.4	79	275	218	79.3	218	194	89	70.5	74.5
	3535	344	299	86.9	299	293	98	85.2	326	269	82.5	269	241	89.6	73.9	79.7
	3536	425	362	85.2	362	337	93.1	79.3	476	388	81.5	388	340	87.6	71.4	75.1

		Area frame							Phone frames							
Prov. Terr.	Health Region	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	Combined resp. rates
	3537	528	441	83.5	441	395	89.6	74.8	557	437	78.5	437	388	88.8	69.7	72.2
	3538	298	266	89.3	266	249	93.6	83.6	304	249	81.9	249	213	85.5	70.1	76.7
	3539	170	163	95.9	163	159	97.5	93.5	210	167	79.5	167	148	88.6	70.5	80.8
	3540	219	211	96.3	211	200	94.8	91.3	286	224	78.3	224	195	87.1	68.2	78.2
	3541	344	288	83.7	288	270	93.8	78.5	316	259	82	259	231	89.2	73.1	75.9
	3542	229	214	93.4	214	198	92.5	86.5	274	220	80.3	220	198	90	72.3	78.7
	3543	276	250	90.6	250	236	94.4	85.5	305	262	85.9	262	238	90.8	78	81.6
	3544	516	403	78.1	403	386	95.8	74.8	530	426	80.4	426	369	86.6	69.6	72.2
	3546	433	395	91.2	395	377	95.4	87.1	513	403	78.6	403	365	90.6	71.2	78.4
	3547	239	207	86.6	207	182	87.9	76.2	269	224	83.3	224	204	91.1	75.8	76
	3549	194	178	91.8	178	155	87.1	79.9	212	169	79.7	169	140	82.8	66	72.7
	3551	674	559	82.9	559	512	91.6	76	670	548	81.8	548	486	88.7	72.5	74.3
	3552	240	222	92.5	222	205	92.3	85.4	226	195	86.3	195	173	88.7	76.5	81.1
	3553	854	745	87.2	745	675	90.6	79	971	736	75.8	736	616	83.7	63.4	70.7
	3554	206	186	90.3	186	169	90.9	82	215	185	86	185	164	88.6	76.3	79.1
	3555	249	212	85.1	212	201	94.8	80.7	272	227	83.5	227	208	91.6	76.5	78.5
	3556	237	205	86.5	205	183	89.3	77.2	237	206	86.9	206	194	94.2	81.9	79.5
	3557	216	199	92.1	199	182	91.5	84.3	242	197	81.4	197	185	93.9	76.4	80.1
	3558	321	277	86.3	277	264	95.3	82.2	348	292	83.9	292	259	88.7	74.4	78.2
	3560	746	642	86.1	642	597	93	80	828	684	82.6	684	600	87.7	72.5	76
	3561	311	273	87.8	273	259	94.9	83.3	349	289	82.8	289	264	91.3	75.6	79.2
	3562	410	360	87.8	360	340	94.4	82.9	491	407	82.9	407	363	89.2	73.9	78
	3563	125	118	94.4	118	116	98.3	92.8	166	132	79.5	132	116	87.9	69.9	79.7
	3565	472	392	83.1	392	359	91.6	76.1	489	407	83.2	407	371	91.2	75.9	76
	3566	349	317	90.8	317	301	95	86.2	390	321	82.3	321	293	91.3	75.1	80.4
	3568	416	366	88	366	347	94.8	83.4	507	412	81.3	412	346	84	68.2	75.1

		Area frame							Phone frames							
Prov. Terr.	Health Region	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	Combined resp. rates
	3570	566	508	89.8	508	477	93.9	84.3	732	533	72.8	533	449	84.2	61.3	71.3
	3595	1459	1220	83.6	1220	1129	92.5	77.4	1802	1236	68.6	1236	1034	83.7	57.4	66.3
Man.	**Total**	2202	1996	90.6	1996	1895	94.9	86.1	2261	1968	87	1968	1831	93	81	83.5
	4610	652	568	87.1	568	541	95.2	83	655	550	84	550	512	93.1	78.2	80.6
	4615	165	148	89.7	148	137	92.6	83	172	142	82.6	142	130	91.5	75.6	79.2
	4620	158	143	90.5	143	135	94.4	85.4	175	157	89.7	157	151	96.2	86.3	85.9
	4625	167	158	94.6	158	147	93	88	187	167	89.3	167	149	89.2	79.7	83.6
	4630	203	186	91.6	186	180	96.8	88.7	205	184	89.8	184	167	90.8	81.5	85
	4640	228	212	93	212	200	94.3	87.7	223	201	90.1	201	191	95	85.7	86.7
	4645	205	190	92.7	190	180	94.7	87.8	203	184	90.6	184	174	94.6	85.7	86.8
	4660	145	133	91.7	133	132	99.2	91	161	139	86.3	139	131	94.2	81.4	85.9
	4670	134	128	95.5	128	117	91.4	87.3	149	127	85.2	127	121	95.3	81.2	84.1
	4685	145	130	89.7	130	126	96.9	86.9	131	117	89.3	117	105	89.7	80.2	83.7
Sask.	**Total**	2068	1910	92.4	1910	1810	94.8	87.5	2579	2191	85	2191	2008	91.6	77.9	82.2
	4701	174	167	96	167	160	95.8	92	165	148	89.7	148	137	92.6	83	87.6
	4702	177	166	93.8	166	156	94	88.1	189	154	81.5	154	140	90.9	74.1	80.9
	4703	159	143	89.9	143	133	93	83.6	171	131	76.6	131	121	92.4	70.8	77
	4704	389	352	90.5	352	333	94.6	85.6	435	366	84.1	366	332	90.7	76.3	80.7
	4705	154	148	96.1	148	143	96.6	92.9	168	141	83.9	141	136	96.5	81	86.6
	4706	387	351	90.7	351	330	94	85.3	448	375	83.7	375	340	90.7	75.9	80.2
	4707	154	145	94.2	145	139	95.9	90.3	162	135	83.3	135	120	88.9	74.1	82
	4708	148	141	95.3	141	134	95	90.5	130	111	85.4	111	101	91	77.7	84.5
	4709	151	131	86.8	131	123	93.9	81.5	248	215	86.7	215	197	91.6	79.4	80.2
	4710	175	166	94.9	166	159	95.8	90.9	166	145	87.3	145	138	95.2	83.1	87.1
	4714	.	.	.	.	.	.	.	297	270	90.9	270	246	91.1	82.8	82.8
Alta.	**Total**	3669	3212	87.5	3212	2961	92.2	80.7	4097	3363	82.1	3363	3032	90.2	74	77.2
	4821	297	270	90.9	270	244	90.4	82.2	319	257	80.6	257	242	94.2	75.9	78.9

		Area frame							Phone frames							
Prov. Terr.	Health Region	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	Combined resp. rates
	4822	213	198	93	198	190	96	89.2	268	232	86.6	232	210	90.5	78.4	83.2
	4823	904	764	84.5	764	704	92.1	77.9	917	760	82.9	760	688	90.5	75	76.4
	4824	454	395	87	395	363	91.9	80	466	396	85	396	356	89.9	76.4	78.2
	4825	260	246	94.6	246	231	93.9	88.8	301	256	85	256	238	93	79.1	83.6
	4826	796	692	86.9	692	639	92.3	80.3	953	750	78.7	750	661	88.1	69.4	74.3
	4827	309	281	90.9	281	264	94	85.4	337	277	82.2	277	246	88.8	73	78.9
	4828	274	239	87.2	239	213	89.1	77.7	316	259	82	259	233	90	73.7	75.6
	4829	162	127	78.4	127	113	89	69.8	220	176	80	176	158	89.8	71.8	70.9
B.C.	**Total**	4857	4130	85	4130	3863	93.5	79.5	5591	4451	79.6	4451	4026	90.5	72	75.5
	5911	167	161	96.4	161	155	96.3	92.8	221	173	78.3	173	160	92.5	72.4	81.2
	5912	160	137	85.6	137	130	94.9	81.3	187	146	78.1	146	137	93.8	73.3	76.9
	5913	356	308	86.5	308	296	96.1	83.1	447	359	80.3	359	326	90.8	72.9	77.5
	5914	289	266	92	266	255	95.9	88.2	307	257	83.7	257	229	89.1	74.6	81.2
	5921	299	274	91.6	274	261	95.3	87.3	339	291	85.8	291	270	92.8	79.6	83.2
	5922	433	355	82	355	342	96.3	79	538	429	79.7	429	394	91.8	73.2	75.8
	5923	497	422	84.9	422	382	90.5	76.9	578	456	78.9	456	395	86.6	68.3	72.3
	5931	256	226	88.3	226	218	96.5	85.2	312	243	77.9	243	220	90.5	70.5	77.1
	5932	556	426	76.6	426	407	95.5	73.2	627	462	73.7	462	415	89.8	66.2	69.5
	5933	292	253	86.6	253	224	88.5	76.7	373	285	76.4	285	259	90.9	69.4	72.6
	5941	426	360	84.5	360	331	91.9	77.7	426	348	81.7	348	324	93.1	76.1	76.9
	5942	305	263	86.2	263	246	93.5	80.7	341	284	83.3	284	264	93	77.4	78.9
	5943	175	163	93.1	163	155	95.1	88.6	179	152	84.9	152	137	90.1	76.5	82.5
	5951	214	163	76.2	163	149	91.4	69.6	226	170	75.2	170	143	84.1	63.3	66.4
	5952	317	277	87.4	277	246	88.8	77.6	306	243	79.4	243	215	88.5	70.3	74
	5953	115	76	66.1	76	66	86.8	57.4	184	153	83.2	153	138	90.2	75	68.2
Y.T.	**6001**	615	531	86.3	531	500	94.2	81.3	160	146	91.3	146	129	88.4	80.6	81.2

		Area frame							Phone frames							
Prov. Terr.	Health Region	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	Combined resp. rates
N.W.T.	6101	632	583	92.2	583	543	93.1	85.9	105	91	86.7	91	82	90.1	78.1	84.8
Nvt.	6201	435	390	89.7	390	374	95.9	86	.	.	.	.	.	.	.	86

Reponse rate by Local Health Integrated Network (LHIN) and frames in the CCHS 2007 in Ontario

		Area frame / Base aréolaire							Phone frames / Bases téléphoniques							
Prov. Terr.	LHIN	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	# in scope HH	# resp. HH	HH resp. rates	# pers. select.	# resp.	Pers. resp. rates	Resp. rates	Combined resp. rates
Ont.	Total	13841	12031	86.9	12031	11183	93	80.8	15419	12251	79.5	12251	10784	88	69.9	75.1
	3501	864	791	91.6	791	745	94.2	86.2	1067	856	80.2	856	739	86.3	69.3	76.9
	3502	1592	1387	87.1	1387	1310	94.4	82.3	1699	1390	81.8	1390	1238	89.1	72.9	77.4
	3503	767	662	86.3	662	614	92.7	80.1	812	668	82.3	668	612	91.6	75.4	77.6
	3504	1599	1394	87.2	1394	1279	91.8	80	1774	1406	79.3	1406	1263	89.8	71.2	75.4
	3505	682	583	85.5	583	536	91.9	78.6	794	611	77	611	501	82	63.1	70.3
	3506	723	639	88.4	639	582	91.1	80.5	824	646	78.4	646	550	85.1	66.7	73.2
	3507	705	562	79.7	562	514	91.5	72.9	865	587	67.9	587	509	86.7	58.8	65.2
	3508	895	795	88.8	795	747	94	83.5	1157	811	70.1	811	682	84.1	58.9	69.6
	3509	1390	1209	87	1209	1141	94.4	82.1	1475	1167	79.1	1167	1021	87.5	69.2	75.5
	3510	829	721	87	721	678	94	81.8	830	693	83.5	693	609	87.9	73.4	77.6
	3511	1300	1118	86	1118	1035	92.6	79.6	1355	1114	82.2	1114	1003	90	74	76.8
	3512	681	580	85.2	580	538	92.8	79	765	637	83.3	637	559	87.8	73.1	75.9
	3513	1210	1052	86.9	1052	969	92.1	80.1	1299	1089	83.8	1089	995	91.4	76.6	78.3
	3514	604	538	89.1	538	495	92	82	703	576	81.9	576	503	87.3	71.6	76.4

Response rate by province and frames for the Health Services Access Survey (HSAS) sub-sample 2007.

	Area frame							Phone frames							
Prov. Terr.	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**Combined resp. rates**
Canada	31046	27295	87.9	25968	24171	93.1	81.8	34553	28149	81.5	26136	22486	86	70.1	75.7
N.L.	1192	1111	93.2	1058	988	93.4	87	1247	1065	85.4	981	862	87.9	75	80.9
P.E.I.	889	790	88.9	761	716	94.1	83.6	888	733	82.5	669	592	88.5	73	78.4
N.S.	1543	1368	88.7	1317	1220	92.6	82.1	1665	1415	85	1302	1189	91.3	77.6	79.8
N.B.	1518	1358	89.5	1288	1209	93.9	84	1792	1540	85.9	1409	1284	91.1	78.3	80.9
Que.	3076	2699	87.7	2569	2427	94.5	82.9	3519	2832	80.5	2656	2258	85	68.4	75.2
Ont.	13841	12031	86.9	11444	10592	92.6	80.5	15419	12251	79.5	11470	9634	84	66.7	73.3
Man.	2039	1842	90.3	1741	1641	94.3	85.1	2085	1808	86.7	1675	1479	88.3	76.6	80.8
Sask.	1923	1773	92.2	1673	1574	94.1	86.7	2253	1911	84.8	1774	1567	88.3	74.9	80.4
Alta.	2383	2063	86.6	1958	1783	91.1	78.8	2637	2156	81.8	1997	1733	86.8	71	74.7
B.C.	2642	2260	85.5	2159	2021	93.6	80.1	3048	2438	80	2203	1888	85.7	68.5	74