Canadian Community Health Survey (CCHS)
Annual component

User guide
2011 Microdata files

June 2012

Statistics Canada Statistique Canada

Canada

TABLE OF CONTENTS

WHAT'S NEW IN THE 2011 CANADIAN COMMUNITY HEALTH SURVEY? 1

1. INTRODUCTION 3

2. BACKGROUND 4

3. CCHS REDESIGN IN 2007 5

4. CONTENT STRUCTURE OF THE CCHS 6

4.1 Common content 6
4.2 Optional content 7
4.3 Rapid response content 7
4.4 Content included in data files 7

5. SAMPLE DESIGN 10

5.1 Target population 10
5.2 Health regions 10
5.3 Sample size and allocation 10
5.4 Frames, household sampling strategies 12
5.4.1 Sampling of households from the area frame 12
5.4.2 Sampling of households from the list frame of telephone numbers 14
5.4.3 Sampling of households from the Random Digit Dialing frame of telephone numbers 14
5.5 Sample allocation over the collection periods 15
5.6 Sampling of interviewees 15
5.7 Supplementary buy-in sample in Ontario 16

6. DATA COLLECTION 18

6.1 Computer-assisted interviewing 18
6.2 CCHS application development 18
6.3 Interviewer training 19
6.4 The interview 20
6.5 Field operations 22
6.6 Quality control and collection management 23

7. DATA PROCESSING 24

7.1 Editing 24
7.2 Coding 24
7.3 Creation of derived variables 24
7.4 Weighting 24
7.5 Income imputation 25

8. WEIGHTING 26

8.1 Overview 26
8.2 Weighting of the area frame sample 27
8.3 Weighting of the telephone frame sample 28
8.4 Integration of the telephone and area frames (I1) 30
8.5 Post-integration weighting steps 31
8.6 Particular aspects of the weighting in the three territories 32
8.7 Creation of a share weight 33
8.8 Weighting for a two-year file 33

9. DATA QUALITY 34

9.1 Response rates 34
9.2 Survey Errors 35

9.2.1 Non-sampling Errors *35*
9.2.2 Sampling Errors *36*

10. GUIDELINES FOR TABULATION, ANALYSIS AND RELEASE 37

10.1 ROUNDING GUIDELINES 37
10.2 SAMPLE WEIGHTING GUIDELINES FOR TABULATION 38
10.2.1 Definitions: categorical estimates, quantitative estimates *38*
10.2.2 Tabulation of categorical estimates *39*
10.2.3 Tabulation of quantitative estimates *39*
10.3 GUIDELINES FOR STATISTICAL ANALYSIS 39
10.4 RELEASE GUIDELINES 40

11. APPROXIMATE SAMPLING VARIABILITY TABLES 42

11.1 HOW TO USE THE CV TABLES FOR CATEGORICAL ESTIMATES 42
11.2 EXAMPLES OF USING THE CV TABLES FOR CATEGORICAL ESTIMATES 44
11.3 HOW TO USE THE CV TABLES TO OBTAIN CONFIDENCE LIMITS 47
11.4 EXAMPLE OF USING THE CV TABLES TO OBTAIN CONFIDENCE LIMITS 48
11.5 HOW TO USE THE CV TABLES TO DO A Z-TEST 48
11.6 EXAMPLE OF USING THE CV TABLES TO DO A Z-TEST 48
11.7 EXACT VARIANCES/COEFFICIENTS OF VARIATION 49
11.8 RELEASE CUT-OFFS FOR THE CCHS 50

12. MICRODATA FILES: DESCRIPTION, ACCESS AND USE 51

12.1 MASTER FILES 51
12.1.1 Research Data Centre *51*
12.1.2 Custom tabulations *51*
12.1.3 Remote access *51*
12.2 SHARE FILES 51
12.3 PUBLIC USE MICRODATA FILES 52
12.4 HOW TO USE THE CCHS DATA FILES: ANNUAL DATA FILE OR TWO-YEAR DATA FILE? 53
12.5 USE OF WEIGHT VARIABLE 53
12.6 VARIABLE NAMING CONVENTION BEGINNING IN 2007 53
12.7 VARIABLE NAMING CONVENTION BEFORE 2007 55
12.8 GUIDELINES FOR THE USE OF SUB-SAMPLE VARIABLES 56
12.8.1 Variable naming convention for sub-sample modules *56*
12.8.2 Separate file *56*
12.8.3 Analysis combining sub-sample and optional content *57*
12.9 DATA DICTIONARIES 57
12.10 DIFFERENCES IN CALCULATION OF COMMON CONTENT VARIABLES USING DIFFERENT FILES 57
APPENDIX A – CANADIAN COMMUNITY HEALTH SURVEY CONTENT OVERVIEW (2011-2012) 58
APPENDIX B – OPTIONAL CONTENT SELECTION BY HEALTH REGIONS (GROUPED BY PROVINCE) (2011) 61
APPENDIX C - AVAILABLE GEOGRAPHY IN THE MASTER AND SHARE FILES AND THEIR CORRESPONDING CODES: CANADA, PROVINCES/TERRITORIES, HEALTH REGIONS AND PEER GROUPS 65
APPENDIX D (2011) SAMPLE ALLOCATION BY HEALTH REGION AND FRAME, BY LOCAL HEALTH INTEGRATED NETWORK (LHIN) AND FRAMES IN THE CCHS IN ONTARIO, AND BY PROVINCE AND FRAME FOR THE HEALTH SERVICES ACCESS SURVEY (HSAS) SUBSAMPLE 70
APPENDIX E (2011) - RESPONSE RATES BY HEALTH REGION AND FRAME, RESPONSE RATES BY LOCAL HEALTH INTEGRATED NETWORK (LHIN) AND FRAME IN THE CCHS IN ONTARIO AND RESPONSE RATES BY PROVINCE AND FRAMES FOR THE HEALTH SERVICES ACCESS SURVEY (HSAS) SUB-SAMPLE 75

WHAT'S NEW IN THE 2011 CANADIAN COMMUNITY HEALTH SURVEY?

Content

There have been a few changes to existing modules in the Canadian Community Health Survey (CCHS) content in 2011. Also, new modules were introduced for one year as part of the 2011 two year common content and optional content.

Changes

- **Restriction of Activities (RAC)**: This module has been a core module throughout the years, with the exception of 2011 when it was only offered in the territories.
- **Consultations about mental health (CMH):** This module was previously offered as an optional module. In 2011 – 2012, it is part of the 2 year common content as mental well being.

New modules

- **Positive Mental Health (PMH)**: This new theme module was introduced in 2011 as part of the 2 year common content on mental well being.
- **Physical activities – Stages of Changes (SCP)**: This new optional module was added in 2011.
- **Social Provisions (SPS)**: This new optional module was added in 2011. It includes questions on five social provisions sub-scales - Attachment, Social integration, Reassurance of worth, Reliable alliance, and Guidance.

Data Processing

- Beginning with the 2011 reference year, the household income variable will be imputed. Missing values due to either respondent refusal or respondent's lack of knowledge of household income will be completed using statistical techniques. The main variable of interest is INC_3: 'Total household income - best estimate' but all variables that are derived based on income will also be affected. The income variables along with an imputation flag (INCFIMP) indicating which values were imputed will be provided on the data file. For more information on the imputation process, please refer to the document 'Income Imputation for the Canadian Community Health Survey' available under the 'Documentation' section of the Definitions, data sources and methods page on the Statistics Canada website.

Geography

- Nova Scotia has made changes in health region names and codes. Zones will be renamed to District Health Authorities.

- The province of New Brunswick has made minor name changes to its health regions. 'Regions' are now referred to as 'Zones' and include a descriptive name.

- In Alberta, Regional Health Authorities have been replaced with five new Zones.

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 intra-provincial 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 115 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. The PUMF is released every two years and contains two years of data. The next PUMF file will be released in September 2013 and will include the data collected for the years 2011and 2012.
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: 613-951-1746
E-mail: hd-ds@statcan.gc.ca

For remote access support: 613-951-1746
E-mail: cchs-escc@statcan.gc.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 program 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. Henceforth, the x.1 cycles of the CCHS are designated as "the annual component" of the CCHS. The full title is "The Canadian Community Health Survey – Annual component, 2011" and the short title is simply "CCHS –2011". 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 into three: the annual common content (previously referred to as core content), the one year and two-year common content (previously referred to as theme content). The one year common content is asked for one year and re-introduced every two or four years. The two year common content is asked for two years and re-introduced every four years. The two year and one year common 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 annual common 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. Since 2008, CCHS data are released annually. Every two years, a file combining the two years' sample (130,000 respondents) is also 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 annual common content, the two year and one year common content, the optional content component, and the rapid response component. Appendix A lists the modules included in the 2011questionnaire 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 • Annual • One and two-year	**30 minutes** (20 minutes) (10 minutes)
Optional content	**10 minutes**
Rapid response content (As requested on a cost recovery basis)	**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 three components: the annual common content, one-year and two year common content.

The annual common 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 up to 2014, unless a major concern is raised about quality.

The one year and two-year common content (previously called theme content) comprises questions related to a specific topic. Combined, the two-year and one year common content take about 10 minutes of the interview time. Modules comprising this content type 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 one year common content 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 2005 (cycle 3.1), 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 the current year and two year combined 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, excluding the Territories.

4.4 Content included in data files

The survey produces different data files:

- one year reference period
- combined two years reference periods and
- one year sub-sample data files.

Table 4.2 provides clarification about the data files available for the 2011 and 2012 CCHS.

One year data files

The survey produces data files every year. In June 2012, two annual files based on the 2011 reference period have been released. The main file includes respondents from the 2011 data collection and variables from the common annual content, common two year content as well as optional content.

The one year common content was released in a separate file as it is based on a sub-samle of respondents.

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

Two year data files

Every two years, a file combining the most recent two years is released. A combined file also to be released in June 2013 will contain data from 2011 and 2012. The following two year file is scheduled to be released in 2015, and will include both the 2013 and 2014 reference years.

The two-year data file includes all respondents and the questions that were in the survey over the two year reference period. Unless otherwise specified, it is the question component from the common annual and two-year content and selected optional content over the two year period. The one-year common content and optional content selected for one year only are not available in the two-year data file.

Sub-sample data files

Any modules collected from a sub-sample of the population will continue to be disseminated in separate files. These files include the annual and one year common content collected from a sub-sample of respondents. Sub-sample files have been released as follow:

Year	**Modules**
2000	Waiting times and Access to health care services
2003	Dental visits, Driving and safety, Health utility index, Medication use, Oral health 2
2005	Waiting times, Access to health care services, Patient satisfaction, Health Utility Index, Measured height and weight, Fruit and vegetable consumption, Labour force – long form
2007	Waiting times, Access to health care services and Patient satisfaction
2008	Measured height and weight
2009	Waiting times, Access to health care services
2011	Waiting times, Access to health care services

Table 4.2 Content components for the 2011 and 2012 data files

Files		Annual common content	2011 one year common content[1]	2012 one year common content	2011-2012 two-year common content	Optional content[2]
2011	**Main**	Yes	No	N/A	Yes	Yes
	Sub-sample (2 modules)	Yes	Yes	N/A	No	No
2012	**Main**	Yes	N/A	Yes	Yes	Yes
2011-2012	**Main**	Yes	No	No	Yes	Yes

1 The 2011 annual common content was comprised of two modules (Access to health care services and Waiting times) which were all asked to a sub-sample of respondents.

2 Optional content will be included in the 2011-2012 data file if it is asked of respondents in a jurisdiction during the two year period. Otherwise, it will only be included in the file of the year in which it was collected. Note that if an annual common content module from one year is selected for the optional content of a jurisdiction during the second year, the module will be included in the two-year data file and will be processed as optional content.

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 2011, data was collected in 112 HRs in the ten provinces, as well as to one HR per territory, totalling 115 HRs (Appendix C).

In 2011, the definition of HRs was modified between the time of sampling and the creation of the data files for two provinces. In Alberta, there are now 5 HRs, which are simple aggregations of the 9 HRs that were defined at the time of sampling[4]. In Prince Edward Island, the 3 HRs defined at the time of sampling (1101, 1102 and 1103) have been aggregated into a single HR (1100). The current chapter on sample design, as well as the figures on sample sizes provided in Appendix D, refer to the HRs as they were defined at the time of sampling.

5.3 Sample size and allocation

To provide reliable estimates for each HR 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. In this first step, 60,350 units were allocated 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. This sample allocation strategy was used for CCHS 2005 and the sample sizes have remained mainly the

[4] The correspondence between the 5 new HRs and the 9 HRs defined at the time of sampling is as follows: HR 4831 (4821 and 4822), HR 4832 (4823), HR 4833 (4824 and 4825), HR 4834 (4826) and HR 4835 (4827, 4828 and 4829).

same since then. The sample was then divided evenly between the 2 collection years. Table 5.1 gives the targeted sample sizes for 2011.

Table 5.1 Number of health regions and targeted sample sizes by province/territory, 2011

Province	Number of HRs	Targeted sample size
Newfoundland and Labrador	4	2,005
Prince Edward Island	1	1,001
Nova Scotia	6	2,521
New Brunswick	7	2,575
Quebec	16	12,145
Ontario[1]	36	22,198
Manitoba	10	3,750
Saskatchewan	11	3,860
Alberta	5	6,100
British Columbia	16	8,045
Yukon	1	600
Northwest Territories	1	600
Nunavut	1	350
Canada	**121**	**65,750**

[1] The sample size for Ontario includes the buy-in extra sample by LHIN. The initial annual sample size for Ontario before the buy-in was 20,880 units (refer to section 5.7 for further details).

In the third step, the 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 proportionality of the allocation by province.

Note that the three territories were not part of the above allocation strategy as they were dealt with separately. Each year, 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 for the territories is done proportionally to the population sizes of the strata. The strata used were 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 was then divided between the area frame and the list frame[5], as described in the next section. It should be mentioned that the number of units taken from each frame was increased in order to account for the anticipated out-of-scope and non-response rates based on the rates obtained in previous CCHS cycles. These adjusted sample sizes are called the *raw sample sizes*, and correspond to the number of units that have to be selected in order to obtain the targeted sample sizes. The sample sizes by HR and frame are provided in Appendix D.

[5] Except for two 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

The CCHS used three sampling frames to select the sample of households: 40.5% of the sample of households came from an area frame, 58.5% 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[6]. 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. This change is dealt with during weighting.

6 Statistics Canada (2008). *Methodology of the Canadian Labour Force Survey.* Statistics Canada. Cat. No. 71-526-XIE.

Table 5.2 Major first-stage units, sizes and yields

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 about 26,000 respondents for a given year of CCHS, about 40,000 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 required 40,000 dwellings distributed across the HRs, which have different geographic boundaries from those of the LFS economic regions. Overall, the CCHS required a lower number of dwellings than those generated by the LFS selection mechanism, which corresponds to an average *adjustment factor* of 0.66 (40,000/60,000). . However, since the adjustment factors varied 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[7].

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 *stabilization*.

Sampling of households from the area frame in the three territories

For operational reasons, the LFS area frame sample design for the three territories was different. For each territory, the larger communities each have their own stratum while smaller 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

[7] 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.

five design strata in the Yukon, ten in the Northwest Territories and seven in Nunavut. For strata consisting of a group of communities, 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, the second stage consisted of selecting households using the same sampling strategy as 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, another community was selected for the CCHS. For larger communities with their own stratum, only one stage design was necessary where households were selected directly using the same sampling strategy described above.

It is worth mentioning that the frame for the CCHS covered 90% of the private households in the Yukon, 97% in the Northwest Territories and 71% in Nunavut[8].

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 postal code conversion files to map each telephone number to a stratum. 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.

5.4.3 Sampling of households from the Random Digit Dialing frame of telephone numbers

In four HRs, a Random Digit Dialling (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[9]. 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

[8] In Nunavut, 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.

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

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 periods

In order to balance interviewer workload and to minimize possible seasonal effects on estimates, the initial sample size for each frame is allocated equally over the six 2-month collection periods. Exceptionally in 2011, this was slightly modified because of the additional workload related to the collection of data for the 2011 Census, which started in May. A portion of the CCHS sample allocated to the May-June and July-August collection periods was transferred to the other four collection periods.

In the area frame, each PSU 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 in scope Canadian population for each two month period.

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 needs, cost, design efficiency, response burden and operational constraints. One person is selected 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 formed of two adults of age 45 to 64 and one 15-year-old, the teenager would have a 7/9 chance of being selected (i.e., 70/(70+10+10)) while each of the adults would have a 1/9 chance of being selected. 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+**
Factors	70	50	20	10	10

5.7 Supplementary buy-in sample 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. The CCHS sample was increased in order to obtain a minimum size of 2,000 respondents 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 respondents 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 targeted sample sizes by LHIN for 2011.

Table 5.4 Targeted sample sizes by Local Health Integrated Network (LHIN), CCHS 2011

LHIN	**Targeted sample size**
01-Erie St. Clair	1,550
02-South West	2,561
03-Waterloo Wellington	1,243
04-Hamilton Niagara Haldimand Brant	2,597
05-Central West	1,049
06-Mississauga Halton	1,115
07-Toronto Central	1,123
08-Central	1,411
09-Central East	2,108
10-South East	1,313
11-Champlain	2,057
12-North Simcoe Muskoka	1,041
13-North East	1,990
14-North West	1,040
Ontario	**22,198**

The total sample size of the HR-LHIN overlapping areas was then allocated between the two frames, with 59% going to the list frame and the 41% to 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 sub-sample of 31,502 respondents ensures that reliable estimates can be produced at the provincial level. Table 5.5 provides the targeted sample sizes by province for the HSAS 2011 survey.

Table 5.5: Targeted sample sizes for CCHS 2011 and HSAS 2011 by province.

Province	Targeted sample size	
	CCHS 2011	HSAS 2011
Newfoundland and Labrador	2,005	2,005
Prince Edward Island	1,001	1,001
Nova Scotia	2,521	2,521
New Brunswick	2,575	2,575
Quebec	12,145	4,600
Ontario	22,198	4,800
Manitoba	3,750	3,200
Saskatchewan	3,860	3,200
Alberta	6,100	3,600
British Columbia	8,045	4,000
CANADA	**64,200**	**31,502**

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 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 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) as well as for the proxy interviews, which are not accepted in HSAS. The final sample sizes and the expected number of respondents by province and frame are given in Appendix D.

Sample selection was performed independently for each collection period, among the units selected for the CCHS.

6. DATA COLLECTION

6.1 Computer-assisted interviewing

Between January and December 2011, over 60,000 valid interviews were conducted using computer assisted interviewing (CAI). Approximately 40% of the interviews were conducted in person using computer assisted personal interviewing (CAPI) and the other 60% 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 application's functionality to the type of interview being conducted. Each application consisted of entry, health content, 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 health content 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 health content and exit components of the applications. It also ensures that variables affecting skip patterns and logic flows are correctly passed between modules within the health content. 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, health content and exit components.

6.3 Interviewer training

Project managers, senior interviewers and interviewers from regional collection offices were sent self study training packages before the start of collection. These packages were prepared by the CCHS project team and were used by existing experienced CCHS interviewers to reinforce their previous training. Project managers and senior interviewers also conducted customised training sessions for new CCHS interviewing staff as needed. There were also specific training sessions to deal with various topics related to CCHS collection on a monthly basis. The focus of the training sessions were to get interviewers comfortable using the CCHS 2010 applications, and familiarise interviewers with survey content and to introduce interviewers to interviewing procedures specific to the CCHS. The training 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, roughly three quarters 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 health content 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.

e) Youth interviews

In 2011, interviewers needed to obtain verbal permission from parents/guardians to interview youths between the ages of 12 to 15 who were selected for interviews. This information was collected in the Parental/Guardian Consent (PGC) block. 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.

The Person Most Knowledgeable (PMK) block collected household level information found at the end of the survey (

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.

For final response rates, refer to Appendix E.

6.6 Quality control and collection management

During the collection year, several methods are 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.

In 2011, an additional quality control system was introduced for CAPI interviews. Upon obtaining consent from the respondents, specific sections of the CAPI interviews were recorded. These recordings were transmitted back to the regional offices, and then randomly chosen for analysis. 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.

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 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 12.6.

7.4 Weighting

The principle behind estimation in a probability sample such as CCHS 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.

7.5 Income imputation

Beginning with the 2011 reference year, the household income variable will be imputed. Missing values due to either respondent refusal or respondent's lack of knowledge of household income will be completed using statistical techniques. The main variable of interest is INC_3: 'Total household income - best estimate' but all variables that are derived based on income will also be affected. The income variables along with an imputation flag (INCFIMP) indicating which values were imputed will be provided on the data file. For more information on the imputation process, please refer to the document 'Income Imputation for the Canadian Community Health Survey' available under the 'Documentation' section of the Definitions, data sources and methods page on the Statistics Canada website or contact Client Services (613-951-1746; fax: 613-951-0792; hd-ds@statcan.gc.ca).

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 made up of telephone numbers used to complement 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.

Depending on the need, one or two frames are used for the selection of the sample within a given health region (HR). When two frames are used, the weighting strategy treats both the area and telephone frames independently to come up with separate household-level weights for each of the frames used. These household-level weights are then combined into a single set of household weights through a step called "*integration*". After applying person-level selection weights and some further adjustments, this integrated weight becomes the final person-level weight.

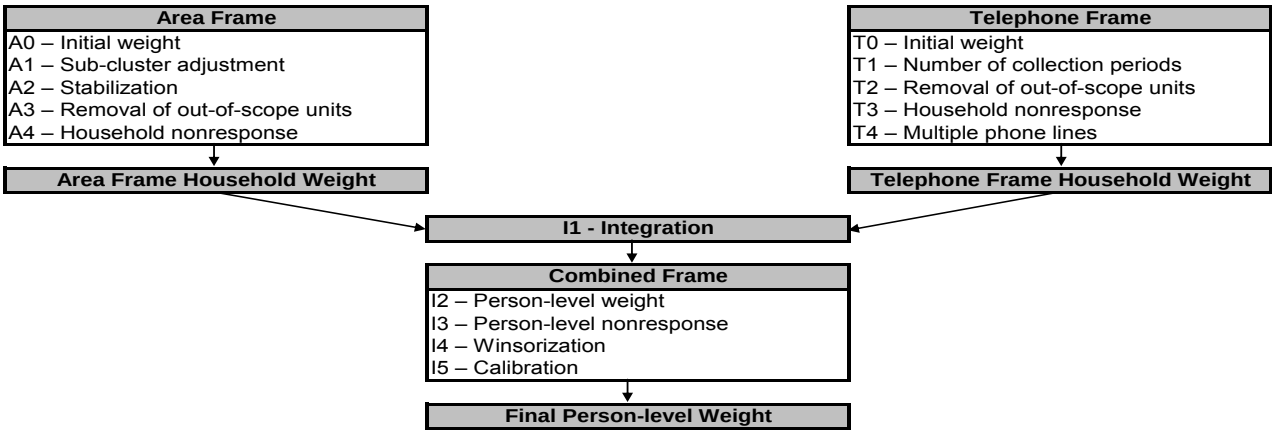
8.1 Overview

As mentioned earlier, 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 steps including calibration, where the weights are adjusted to control for seasonality and to match known population totals. These steps are explained in sub-section 8.5.

Although the two frames are used to cover the three territories, the sampling methods used are slightly different from those used in the provinces. 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 Labour Force Survey (LFS). This weight is based on the LFS design since the CCHS area frame sample design is based on the LFS. The LFS design consists of a sample of dwellings within clusters selected from LFS strata. In the initial adjustment A0, the LFS weight is adjusted to take into consideration the fact that the CCHS selects a sample to be representative of the Health Region. To do so, 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 weight A0. For more details about the selection mechanism, as well as a more complete definition of LFS strata and clusters, refer to Statistics Canada (1998)[10].

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)).

A2 – Stabilization

In some HRs, the increase of the sample size as described in section 5, results in a larger sample than necessary. Stabilization is used to bring the sample size back down to the desired level. The

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

stabilization process consists of randomly sub-sampling dwellings at the HR level from the dwellings originally selected 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 is representative of, in-scope dwellings or households. These remaining in-scope dwellings maintain the same weight as in the previous step, which is now called weight 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 homogeneity 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 or process data, which includes the number of contact attempts, the time/day of attempt, and whether the household was called on a weekend or weekday. Starting in 2008, RHGs were formed within province to better control for provincial totals. An adjustment factor is calculated within each response group 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 do not always conform to the HR geography. Consequently, some units may have been sampled from one HR but the information collected at the time of the interview places them in a neighbouring HR. This is 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 adjustments (T3 and T4) are 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 design 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[11] 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 randomly 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 of the population 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 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.

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

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

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 response homogeneity group is obtained as follows:

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

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 because these households have a higher probability of being selected. 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[12]. Those units on the area frame that are not on the telephone frame do not have their weights adjusted. For all others units, an adjustment factor α between 0 and 1 is applied to the weights. 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. Starting in 2008, a fixed α of 0.4 has been used for those units on both frames to ensure greater comparability of estimates across years. 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.

[12] 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.

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 5.6 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, the selected person is interviewed. 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 person respondents to account for this nonresponse. Using the same methodology that is used in the treatment of household nonresponse, the adjustment is applied within response homogeneity groups. In this process, the scoring method is used to define a response probability based on characteristics available for both respondents and non-respondents. All characteristics collected when creating the roster of household members are available for the estimation of the response probabilities as well as geographic information and some paradata. The probabilities are grouped into response homogeneity groups and the following adjustment factor is calculated within each group:

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

Weight I2 for responding persons is multiplied by the above adjustment factor to produce weight I3. Nonresponding persons are 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 to other units of the same domain of interest. These units could represent a large proportion of their HR or have a large impact on the variance. In order to prevent this, the weight of these outlier units is adjusted downward using a "winsorization" trimming approach.

I5 – Calibration

The last step necessary to obtain the final CCHS weight is calibration (I5). Calibration is done using CALMAR[13] to ensure that the sum of the final weights corresponds to the population estimates

[13] 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.

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. Starting in 2009, additional controls at sub-HR levels were introduced for the applicable HRs. These controls included grouped CLSCs in health regions 2403 (National Capital Region, Quebec) and 2415 (Laurentides, Quebec) as well as DHAs across Nova Scotia. 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 each collection period (two-month period) is equally represented within the sample. Note that the calibration is done using the most up to date geography and may not match the geography used in sampling.

The population estimates are based on the 2006 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 is adjusted using CALMAR to obtain the final weight I5. Weight I5 corresponds to the ***final CCHS person-level weight*** and can be found on the data file with the variable name WTS_M for master or PUMF users. Prior to the 2010 and 2009-2010 reference period, 2001 Census population counts were used. Evaluation studies have confirmed that the impact of this change on CCHS estimates should be minimal.

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 its use is exclusive to the capitals of the Yukon and the Northwest Territories. 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 three 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.

Finally, starting with the 2008 and 2007-2008 reference year products, controls have been put in place to ensure that the proportion of aboriginals and the proportion of individuals in the capital regions are controlled in the Northwest Territories and Yukon. A similar control based on Inuit status was introduced for Nunavut. Starting in 2009, the proportion of individuals in the capital regions is controlled in Nunavut. These controls ensure that the proportion of the estimates represented by these different groups is consistent with proportions indicated by the 2006 Census.

8.7 Creation of a share weight

Along with the master file and PUMF which contain all CCHS respondents, a share file is created which contains only a portion (>90%) of the original CCHS respondents. The individuals on this share file have agreed to share their data with certain partners. To compensate for the loss of some respondents from the file, the weights of these "sharers" must be adjusted by the factor:

$$\frac{\text{Sum of weight I5 for all respondents}}{\text{Sum of weight I5 for all respondents agreeing to share their data}}$$

Similar to the nonresponse adjustments, this factor is calculated within homogeneity groups, where in this case, individuals with similar estimated propensity to share will be grouped together. The final weight after this adjustment is called WTS_S.

8.8 Weighting for a two-year file

When two years of data are combined to create a two-year file, new weights are calculated straightforwardly by halving the annual weights. This ensures that the sum of the final weights is equal to the average population size over the two years. For more information on combining multiple years, please refer to the article “Combining cycles of the Canadian Community Health Survey” published in the Statistics Canada Health Reports publication (82-003) at the following link: http://www.statcan.gc.ca/pub/82-003-x/2009001/article/10795-eng.pdf

9. DATA QUALITY

9.1 Response rates

In total, 91,039 of the selected units in the CCHS 2011 were in-scope for the survey[14]. Out of these, 72,386 households accepted to participate in the survey resulting in an overall household-level response rate of 79.5%. Among these responding households, 72,386 individuals (one per household) were selected to participate to the survey, out of which a response was obtained for 63, 542 individuals, resulting in an overall person-level response rate of 87.8%. At the Canada level, this yields a combined response rate of 69.8% for the CCHS 2011. Table 9.1 provides combined response rates as well as relevant information for their calculation by health region or group of health regions. Table 9.2 provides the same data by Local Health Integrated Network (LHIN) level. Table 9.3 provides response rates by province for the Health Services Access Survey (HSAS) sub-sample.

Table 9.1 : 2011 response rates by health region and frames

(see Appendix E)

Table 9.2 : 2011 response rates by Local Health Integrated Network (LHIN) and frames in Ontario

(see Appendix E)

Table 9.3 : 2011 response rates by province and frames for the Health Services Access Survey (HSAS) sub-sample

(see Appendix E)

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

Household-level response rate

HHRR =	$\frac{\textit{Number of responding households in both frames}}{\textit{All in-scope households in both frames}}$

[14] 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.

Person-level response rate

PPRR =	*Number of responding persons in both frames* / *All selected persons in both frames*

Combined response rate = HHRR x PPRR

Below is an example on how to calculate the combined response rate for Canada using the information found in Table 9.1. The same method applies to rates computed for smaller regions such as province or health region.

HHRR =	$\frac{29{,}965 + 42{,}421}{36{,}633 + 54{,}406}$	=	$\frac{72{,}386}{91{,}039}$	= 0.795
PPRR =	$\frac{27{,}011 + 36{,}531}{29{,}965 + 42{,}421}$	=	$\frac{63{,}542}{72{,}386}$	= 0.878

Combined response rate = 0.795 x 0.878

= 0.698

= **69.8%**

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 2010. 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 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 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 in the numerator and 20 in the denominator. 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	$16.6 \leq CV \leq 33.3$	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 Approximate Sampling Variability Tables, along with the design effects, the sample sizes and the population counts that were used to produce them, are provided in the document *Approximate Sampling Variability Tables*, which is available to the share file and PUMF users. 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, 1.7%.

4) So the approximate coefficient of variation of the estimate is 1.7%. 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 1.0% is the coefficient of variation (expressed as a percentage) to be used.

5) 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 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 2.4% (expressed as a percentage), and the CV for estimate 2 as 2.4% (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 :

$$\sigma_{\hat{d}} = \sqrt{[(0.194)(0.024)]^2 + [(0.162)(0.024)]^2}$$
$$= 0.0061$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.0061/0.032 = 0.190.

4) So the approximate coefficient of variation of the difference between the estimates is 19.0% (expressed as a percentage). According to the Sampling Variability Guidelines presented in Section 10.4, this estimate can be published but a warning has to be issued.

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, 3.7%.

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. 7%.

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,

$$\alpha_{\hat{R}} = \sqrt{(.037)^2 + (.017)^2}$$

$$= 0.041$$

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 4.1% (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

$\alpha_{\hat{X}}$ = 0.01 is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.777 - (2) (0.777) (0.01) , 0.777 + (2) (0.777) (0.01)}

$CI_{\hat{x}}$ = {0.761, 0.793}

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 = 0.00461. Hence,

$$z = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.194 - 0.162}{0.0061} = \frac{0.032}{0.0061} = 5.25$$

Since $z = 5.25$ 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

The document *Approximate Sampling Variability Table*, which is available to the share file and PUMF users, 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). Table 12.1 includes the list of all available 2011 data files.

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 613-951-1746 or by e-mail at hd-ds@statcan.gc.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.gc.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.

Variables most likely to lead to identification of an individual are deleted from the data file or are collapsed to broader categories.

The PUMF contains the data collected over two years. It includes questions that were asked over two years. Unless otherwise specified, these questions are usually those included in the annual common content and in the two-year common content as well as the optional content selected for two years by the provinces and territories.

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 at 613-951-1746 or by e-mail at hd-ds@statcan.gc.ca.

Table 12.1 2011 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 and Waiting Times 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 and Waiting Times module.	Records of all respondents selected for the sub-sample who agreed to share their data

12.4 How to use the CCHS data files: annual data file or two-year data file?

Since the 2008 and 2007-2008 data were released, users that have access to share files or master files have had the choice of using one-year or two-year data files. Decisions about which period to use in a given data analysis should be guided by the level of detail and the quality required. With a one-year file, estimates will not always be available because of the quality associated with limited sample sizes.

Before interpreting and using a CCHS estimate, it is recommended to make sure that the estimates meets the following rules:

- Coefficient of Variation 33.3% or less
- a minimum of 10 respondents in the domain with the characteristic and
- total domain of interest includes at least 20 respondents.

This will not be possible for rare characteristics and detailed domains with one-year files. Instead, users will have to rely on two-year files or multi-year files.

Where the use of either a one-year or two-year file is viable, the user should consider the trade-off between accuracy and currency. If it is important to reflect the current characteristics of a population as closely as possible, the one-year file would be preferable. However, with the increased sample size, more detailed estimates and analyses can be carried out with a two-year file.

12.5 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 weight variable **WTS_M** must be used for regional analyses.

The **Food Security** module, included in certain reference period data files, 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.

12.6 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 = 2000 (cycle 1.1), C = 2003 (cycle 2.1) and E = 2005 (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 12.2 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.

12.7 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 2000 to 2005 data (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	**2000 (**Cycle 1.1) : Canadian Community Health Survey
B	**2002 (**Cycle 1.2) : Canadian Community Health Survey - Mental Health and Well-Being
C	**2003 (**Cycle 2.1) : Canadian Community Health Survey

D **2004 (**Cycle 2.2) : Canadian Community Health Survey - Nutrition

E **2005 (**Cycle 3.1) : Canadian Community Health Survey

12.8 Guidelines for the use of sub-sample variables

In 2011, a sub-sample of respondents was selected at random to answer questions from two modules about the performance of the Canadian health care system: Access to Health Care Services (ACC) and Waiting Times (WTM).[15]

12.8.1 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 British Columbia health regions chose the Access to Health Care Services module. There are two sets of variables for British Columbia (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 12.3 illustrates the possibilities.

Table 12.3 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

12.8.2 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

[15] 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.

(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. Table 12.1 describes the two data files released in 2011.

12.8.3 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, one of the two sub-sample modules was 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.

12.9 Data dictionaries

Separate data dictionary reports, including universe statements and frequencies, are provided for the main master and share files 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 DOWST variable indicates if the Work stress module applies to a given respondent.

12.10 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, when a one year and a two-year data file is available. 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 content overview (2011-2012)

Appendix A – Canadian community health survey content overview (2011 - 2012)

Annual common content (all health regions)[i]

- Age of respondent (ANC)
- Alcohol use (ALC)
- Chronic conditions (CCC)
- Contact with health professionals (CHP)[iii]
- Exposure to second-hand smoke (ETS)
- Fruit and vegetable consumption (FVC)
- Flu shots (FLU)
- General health (GEN)
- Health care utilization (HCU)
- Height and weight – Self –reported (HWT)
- Pain and discomfort (HUP)
- Physical activities (PAC)
- Smoking (SMK)

Administration and Socio-demographics

- Administration information (ADM)
- Education (EDU)
- Income (INC)
- Labour force (LBS)
- Person most knowledgeable about the household situation (PMK)
- Socio-demographic characteristics (SDC)

Two-year biennial common content (all health regions)

2011-2012

1) *Healthy Living*

- Sedentary activities (SAC)
- Food security (FSC)
- Maternal experiences - breastfeeding (MEX)

2) Mental well-being

- Consultation about mental health (CMH)
- Positive mental health (PMH)

One-year biennial/quadrennial common content (all health regions)

2011

1) Health Services Access Survey [ii]

- Access to health care services (ACC)
- Waiting times (WTM)

2) Neurological Conditions (NEU)

2012

1) *Chronic Disease Screening*

- PAP smear test (PAP)
- Mammography (MAM)
- Colorectal cancer screening (CCS)
- Spirometry (SPI)
- Blood test (BLT)
- Physical check-up (PCU)

Rapid Response (national estimates only)

2011

- Neighbourhood environment (NBE) (July – August)

2012

- Canada's food guide (CFG) (May – June)

i RAC has been a core module throughout the years, with the exception of 2011 when it was asked only in the territories.
ii Asked of a sub-sample of respondents. These theme modules were not asked of respondents in the territories.
iii In 2011,CHP changed from being a common content module to an optional module. In 2012, CHP returned to being a common content module but was divided into two modules (CHP and CP2), CP2 is an optional module.

Appendix B – Optional content selection by health regions (grouped by province) (2011)

Appendix B – Optional content selection by health regions (grouped by province) (2011)

DescriptionE	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.	Nvt
ACC Access to health care services	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐
ADL Activities of Daily Living	☑	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☑
ALW Alcohol use during the past week	☑	☐	☐	☐	☑	☑	☑	☑	☐	☐	☐	☐	☐
BPC Blood pressure check	☑	☑	☐	☐	☐	☐	☑	☐	☐	☐	☐	☑	☐
BLT Blood test	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑
BRX Breast examinations	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐	☑	☑
BSX Breast self-examinations	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑
CIH Changes made to improve health	☑	☑	☑	☐	☑	☐	☑	☐	☑	☐	☑	☑	☐
CCS Colorectal cancer screening	☑	☑	☐	☐	☐	☑	☐	☐	☑	☐	☐	☐	☑
CHP Contacts with Health Professionals	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐
DEN Dental visits	☑	☐	☐	☐	☐	☑	☑	☐	☑	☐	☐	☑	☑
DEP Depression	☑	☑	☑	☑	☐	☐	☐	☑	☐	☐	☐	☐	☑
DIA Diabetes care	☑	☐	☑	☑	☐	☑	☐	☐	☐	☐	☐	☐	☐
DSU Dietary supplement use - Vitamins and minerals	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☑
DIS Distress	☐	☐	☐	☐	☐	☑	☐	☑	☐	☐	☐	☐	☐
DRV Driving and safety	☐	☑	☐	☑	☐	☐	☑	☐	☑	☐	☑	☑	☐
EYX Eye examinations	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑
FDC Food choices	☐	☑	☑	☐	☐	☐	☐	☑	☑	☑	☑	☑	☑
HCS Health care system satisfaction	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☑	☐	☑
HMC Home care services	☑	☐	☐	☑	☐	☑	☐	☑	☑	☐	☐	☐	☐
HMS Home safety	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑

DescriptionE	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.	Nvt
HRT Hormone replacement therapy	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
IDG Illicit drugs use	☐	☐	☐	☐	☐	☑	☐	☑	☐	☐	☐	☐	☑
INJ Injuries	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☑	☐
INS Insurance coverage	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☑
LOP Loss of Productivity	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐
MAM Mammography	☑	☐	☐	☐	☐	☑	☐	☐	☑	☐	☐	☐	☑
MAS Mastery	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐
MXA Maternal experiences - Alcohol use during pregnancy	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑
MXS Maternal experiences - Smoking during pregnancy	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☑
MED Medication use	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐
OH1 Oral health 1	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐
OH2 Oral health 2	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐
PAP PAP smear test	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☑
PSC Patient satisfaction - Community-based care	☐	☑	☐	☐	☐	☑	☐	☐	☐	☑	☑	☐	☐
PAS Patient satisfaction - Health care services	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐
SCP Physical activity - Stages of change	☑	☐	☑	☐	☑	☐	☐	☐	☐	☑	☑	☑	☐
PCU Physical check-up	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐
PSA Prostate cancer screening	☐	☐	☑	☐	☐	☑	☐	☐	☐	☐	☐	☑	☑
REP Repetitive strain - Sub Block	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☑	☐
RAC Restriction of activities	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑	☑
SWL Satisfaction with life	☐	☐	☐	☐	☐	☐	☐	☐	☑	☑	☐	☑	☐
SFE Self-esteem	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐
SXB Sexual behaviours	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☑	☑

DescriptionE	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.	Nvt
SLP Sleep	☐	☐	☑	☐	☑	☐	☑	☐	☑	☐	☑	☑	☐
TAL Smoking - Other tobacco products	☐	☐	☐	☐	☐	☑	☐	☐	☑	☐	☐	☐	☐
SPC Smoking - Physician counselling	☐	☑	☐	☐	☐	☐	☐	☐	☐	☑	☑	☐	☑
SCH Smoking - Stages of change	☐	☐	☐	☐	☑	☐	☐	☐	☐	☑	☑	☐	☐
YSM Smoking - Youth smoking	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐
SCA Smoking cessation methods	☐	☐	☑	☐	☐	☐	☐	☐	☐	☑	☐	☐	☑
SPS Social Provisions	☐	☐	☐	☐	☑	☐	☐	☐	☐	☑	☑	☑	☑
SSA Social support - Availability	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐
SPR Spiritual values	☐	☐	☐	☑	☐	☐	☑	☐	☐	☐	☐	☐	☐
SPI Spirometry	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
CST Stress - Childhood and adult stressors	☐	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐
RLE Stress - Recent life events	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐
STS Stress - Sources	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐	☑	☐	☐
SUI Suicidal thoughts and attempts	☑	☐	☐	☐	☐	☐	☐	☑	☑	☐	☐	☐	☑
SSB Sun safety behaviours	☐	☐	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐
UCN Unmet health care needs	☐	☑	☑	☐	☐	☑	☐	☐	☐	☑	☐	☑	☑
UPE Use of protective equipment	☑	☑	☐	☐	☐	☑	☐	☐	☑	☐	☑	☐	☑
ORG Voluntary organizations - Participation	☐	☑	☐	☐	☑	☐	☐	☐	☐	☐	☐	☑	☐
INW Workplace Injury - Sub Block	☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☑	☐

Appendix C - Available geography in the master and share files and their corresponding codes: Canada, provinces/territories, health regions and peer groups

Appendix C - Available geography in the master and share files and their corresponding codes: Canada, provinces/territories, health regions and peer groups

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
1100-A	Prince Edward Island
12	Nova Scotia
1210-C	South Shore -- South West Nova
1223-C	Annapolis Valley District Health Authority
1230--C	Colchester East Hants -- Cumberland
1240-C	Pictou County -- Guysborough Antigonish Strait
1258-I	Cape Breton District Health Authority
1269-A	Capital District Health Authority
13	New Brunswick
1301-C	Zone 1 (Moncton area)
1302-C	Zone 2 (Saint John area)
1303-C	Zone 3 (Fredericton area)
1304-C	Zone 4 (Edmunston area)
1305-I	Zone 5 (Campbellton area)
1306-I	Zone 6 (Bathurst area)
1307-I	Zone 7 (Miramichi area)
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-A	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-D	Région de la Chaudière-Appalaches
2413-A	Région de Laval
2414-A	Région de Lanaudière
2415-A	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-A	Elgin-St. Thomas Health Unit
3533-D	Grey Bruce Health Unit
3534-A	Haldimand-Norfolk Health Unit
3535-A	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-D	Huron County Health Unit
3540-C	Chatham-Kent Health Unit
3541-A	Kingston, Frontenac and Lennox and Addington Health Unit
3542-A	Lambton Health Unit
3543-A	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-A	Oxford County Health Unit
3553-J	Peel Regional Health Unit
3554-D	Perth District Health Unit
3555-A	Peterborough County-City Health Unit
3556-C	Porcupine Health Unit
3557-D	Renfrew County and District Health Unit
3558-A	Eastern Ontario Health Unit
3560-B	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-A	Windsor-Essex County Health Unit
3570-J	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-D		North Eastman Regional Health Authority
4625-E		South Eastman Regional Health Authority
4630-D		Interlake Regional Health Authority
4640-D		Central Regional Health Authority
4645-D		Assiniboine Regional Health Authority
4660-H		Parkland Regional Health Authority
4670-H		NOR-MAN 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-H		Prince Albert Parkland Regional Health Authority
4710-H		Prairie North Regional Health Authority
4714-F		Mamawetan/Keewatin/Athabasca
48	Alberta	
4831-A		South Zone
4832-B		Calgary Zone
4833-E		Central Zone
4834-B		Edmonton Zone
4835-E		North Zone
59	British Columbia	
5911-D		East Kootenay Health Service Delivery Area
5912-D		Kootenay-Boundary Health Service Delivery Area
5913-A		Okanagan Health Service Delivery Area
5914-A		Thompson/Cariboo Health Service Delivery Area
5921-A		Fraser East Health Service Delivery Area
5922-J		Fraser North Health Service Delivery Area
5923-J		Fraser South Health Service Delivery Area
5931-J		Richmond Health Service Delivery Area
5932-G		Vancouver Health Service Delivery Area
5933-J		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-A		North Vancouver Island Health Service Delivery Area
5951-H		Northwest Health Service Delivery Area
5952-H		Northern Interior Health Service Delivery Area
5953-E		Northeast Health Service Delivery Area
60	Yukon	
6001-E		Yukon

61	Northwest Territories
6101-E	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
J	Peer group J

Appendix D (2011) Sample allocation by health region and frame, by Local Health Integrated Network (LHIN) and frames in the CCHS in Ontario, and by province and frame for the Health Services Access Survey (HSAS) subsample

Appendix D (2011) Sample allocation by health region and frame[16]

Geography		Area Frame		Phone frames		Combined	
Prov./Terr.	Health Region	expected # of respondents	raw sample size	expected # of respondents	raw sample size	expected # of respondents	raw sample size
Canada	**Total**	27,466	43,899	38,284	68,455	65,750	112,354
N.L.	**Total**	824	1,300	1,181	1,928	2,005	3,228
	1011	333	559	477	793	810	1,352
	1012	193	311	277	442	470	753
	1013	175	248	250	404	425	652
	1014	123	182	177	289	300	471
P.E.I.	**Total**	411	716	590	1,196	1,001	1,912
	1101	73	130	105	244	178	374
	1102	189	320	271	536	460	856
	1103	149	266	214	416	363	682
N.S.	**Total**	1,037	1,699	1,484	2,399	2,521	4,098
	1210	162	258	234	382	396	640
	1223	132	212	188	302	320	514
	1230	148	242	212	355	360	597
	1240	144	231	206	330	350	561
	1258	173	276	247	422	420	698
	1269	278	480	397	608	675	1,088
N.B.	**Total**	1,059	1,808	1,516	2,351	2,575	4,159
	1301	206	340	294	446	500	786
	1302	199	360	286	466	485	826
	1303	193	360	277	413	470	773
	1304	111	179	159	238	270	417
	1305	103	185	147	216	250	401
	1306	142	196	203	322	345	518
	1307	105	188	150	250	255	438
Que.	**Total**	4,831	7,379	7,314	13,410	12,145	20,789
	2401	247	345	353	594	600	939
	2402	258	382	370	560	628	942
	2403	382	598	544	883	926	1,481
	2404	330	472	473	743	803	1,215
	2405	254	454	364	582	618	1,036
	2406	639	1,046	914	1,732	1,553	2,778
	2407	265	431	380	625	645	1,056
	2408	247	360	353	549	600	909
	2409	247	348	353	756	600	1,104
	2410	0	0	400	1,628	400	1,628
	2411	247	349	353	684	600	1,033
	2412	297	396	426	728	723	1,124
	2413	275	437	395	666	670	1,103
	2414	295	435	423	693	718	1,128

[16] As mentioned in section 5.2, the figures for Alberta and Prince Edward Island are based on the definitions of health regions that were used at the time of sampling.

Geography		Area Frame		Phone frames		Combined	
Prov./Terr.	Health Region	expected # of respondents	raw sample size	expected # of respondents	raw sample size	expected # of respondents	raw sample size
	2415	313	520	448	784	761	1,304
	2416	535	806	765	1,203	1,300	2,009
Ont.	**Total**	9,125	14,460	13,073	23,279	22,198	37,739
	3526	175	246	250	438	425	684
	3527	167	270	238	398	405	668
	3530	335	507	480	831	815	1,338
	3531	140	226	200	324	340	550
	3533	206	300	295	582	501	882
	3534	159	261	227	424	386	685
	3535	195	396	280	542	475	938
	3536	290	427	415	700	705	1,127
	3537	339	570	486	820	825	1,390
	3538	193	291	277	497	470	788
	3539	121	169	174	304	295	473
	3540	164	219	236	410	400	629
	3541	208	335	297	533	505	868
	3542	179	271	256	408	435	679
	3543	195	342	280	497	475	839
	3544	308	492	442	777	750	1,269
	3546	315	477	450	804	765	1,281
	3547	164	274	236	506	400	780
	3549	162	329	232	468	394	797
	3551	421	724	604	989	1,025	1,713
	3552	154	203	221	358	375	561
	3553	544	808	781	1,340	1,325	2,148
	3554	134	176	191	314	325	490
	3555	175	302	250	461	425	763
	3556	154	246	221	408	375	654
	3557	154	238	221	402	375	640
	3558	214	330	306	528	520	858
	3560	453	798	650	1,254	1,103	2,052
	3561	222	380	318	535	540	915
	3562	265	444	381	716	646	1,160
	3563	103	190	147	266	250	456
	3565	315	477	450	742	765	1,219
	3566	239	319	344	572	583	891
	3568	294	439	421	752	715	1,191
	3570	388	576	555	964	943	1,540
	3595	881	1,408	1,261	2,415	2,142	3,823
Man.	**Total**	1,542	2,451	2,208	3,783	3,750	6,234
	4610	434	696	621	957	1,055	1,653
	4615	115	191	165	273	280	464
	4620	103	157	147	278	250	435
	4625	123	198	177	298	300	496
	4630	142	256	203	368	345	624
	4640	164	224	236	422	400	646

Geography		Area Frame		Phone frames		Combined	
Prov./Terr.	Health Region	expected # of respondents	raw sample size	expected # of respondents	raw sample size	expected # of respondents	raw sample size
	4645	146	206	209	349	355	555
	4660	109	191	156	262	265	453
	4670	103	158	147	282	250	440
	4685	103	174	147	294	250	468
Sask.	**Total**	1,483	2,196	2,377	4,557	3,860	6,753
	4701	123	167	177	287	300	454
	4702	123	177	177	286	300	463
	4703	109	170	156	250	265	420
	4704	255	351	365	596	620	947
	4705	127	197	183	323	310	520
	4706	271	418	389	632	660	1,050
	4707	111	184	159	256	270	440
	4708	107	157	153	225	260	382
	4709	134	188	191	342	325	530
	4710	123	187	177	326	300	513
	4714	0	0	250	1,034	250	1,034
Alta.	**Total**	2,510	4,190	3,590	6,283	6,100	10,473
	4821	210	334	300	488	510	822
	4822	171	247	244	438	415	685
	4823	574	917	821	1,434	1,395	2,351
	4824	288	474	412	714	700	1,188
	4825	183	307	262	462	445	769
	4826	539	986	771	1,339	1,310	2,325
	4827	222	358	318	568	540	926
	4828	191	289	274	470	465	759
	4829	132	278	188	370	320	648
B.C.	**Total**	3,309	5,453	4,736	8,653	8,045	14,106
	5911	125	195	180	304	305	499
	5912	127	216	183	316	310	532
	5913	243	340	347	593	590	933
	5914	206	308	294	532	500	840
	5921	214	327	306	566	520	893
	5922	312	472	448	818	760	1,290
	5923	329	513	471	842	800	1,355
	5931	175	261	250	496	425	757
	5932	329	578	471	986	800	1,564
	5933	224	496	321	620	545	1,116
	5941	278	454	397	726	675	1,180
	5942	216	330	309	511	525	841
	5943	109	185	156	240	265	425
	5951	134	224	191	360	325	584
	5952	175	324	250	441	425	765
	5953	113	230	162	302	275	532
Y.T.	**6001**	475	589	125	318	600	907
N.W.T.	**6101**	510	914	90	298	600	1,212
Nvt.	**6201**	350	744	0	0	350	744

Appendix D (2011) - Sample allocation by Local Health Integrated Network (LHIN) and frames in the CCHS in Ontario

Geography		Area Frame		Phone frames		Combined	
Province	LHIN	expected # of respondents	raw sample size	expected # of respondents	raw sample size	expected # of respondents	raw sample size
Ont.	Total	9,125	14,460	13,073	23,279	22,198	37,739
	3501	637	929	913	1,570	1,550	2,499
	3502	1,053	1,542	1,508	2,625	2,561	4,167
	3503	511	744	732	1,192	1,243	1,936
	3504	1,069	1,706	1,528	2,667	2,597	4,373
	3505	431	663	618	1,114	1,049	1,777
	3506	458	658	657	1,114	1,115	1,772
	3507	461	744	662	1,244	1,123	1,988
	3508	580	882	831	1,482	1,411	2,364
	3509	867	1,453	1,241	2,274	2,108	3,727
	3510	540	882	773	1,400	1,313	2,282
	3511	845	1,378	1,212	2,046	2,057	3,424
	3512	428	770	613	1,214	1,041	1,984
	3513	818	1,336	1,172	2,153	1,990	3,489
	3514	427	773	613	1,184	1,040	1,957

Appendix D (2011) - Sample allocation by province and frames for the Health Services Access Survey (HSAS) subsample[17]

Geography	Area Frame		Phone frames		Combined	
Province	expected # of respondents	raw sample size	expected # of respondents	raw sample size	expected # of respondents	raw sample size
Canada	12,690	21,734	18,236	33,718	30,925	55,452
N.L.	761	1,300	1,091	1,928	1,852	3,228
P.E.I.	374	716	537	1,196	912	1,912
N.S.	966	1,699	1,382	2,399	2,348	4,098
N.B.	993	1,808	1,422	2,351	2,415	4,159
Que.	1,886	3,107	2,713	4,909	4,600	8,016
Ont.	1,968	3,305	2,832	5,385	4,800	8,690
Man.	1,312	2,219	1,888	3,419	3,200	5,638
Sask.	1,312	2,075	1,888	3,531	3,200	5,607
Alta.	1,476	2,641	2,124	3,945	3,600	6,586
B.C.	1,640	2,863	2,359	4,655	4,000	7,518

[17] The CCHS respondents aged 12 to 14 years old or that responded by proxy are out-of-scope for HSAS. This explains the discrepancies observed for some of the provinces between the figures found in this table and the ones found in section 5.8. For the provinces where all CCHS respondents were sampled for HSAS, the expected number of respondents was adjusted in the current table to account for the expected number of CCHS units that would be out-of-scope for HSAS. For other provinces, it was possible to account for the presence of such units at the time of sampling by selecting a larger fraction of CCHS respondents, to obtain the desired expected number of respondents for the HSAS.

Appendix E (2011) - Response rates by health region and frame, Response rates by Local Health Integrated Network (LHIN) and frame in the CCHS in Ontario and Response rates by province and frames for the Health Services Access Survey (HSAS) sub-sample

Appendix E (2011) - Table 9.1 response rates by health region and frame

Geography		Area frame							Phone frames							Combined
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**	36,633	29,965	81.8	29,965	27,011	90.1	73.7	54,406	42,421	78.0	42,421	36,531	86.1	67.1	69.8
N.L.	**Total**	1,041	903	86.7	903	822	91.0	79.0	1,578	1,225	77.6	1,225	1,038	84.7	65.8	71.0
	1011	467	389	83.3	389	347	89.2	74.3	650	501	77.1	501	426	85.0	65.5	69.2
	1012	234	211	90.2	211	187	88.6	79.9	356	282	79.2	282	237	84.0	66.6	71.9
	1013	194	171	88.1	171	161	94.2	83.0	336	259	77.1	259	221	85.3	65.8	72.1
	1014	146	132	90.4	132	127	96.2	87.0	236	183	77.5	183	154	84.2	65.3	73.6
P.E.I.	**Total**	515	438	85.0	438	381	87.0	74.0	786	612	77.9	612	510	83.3	64.9	68.5
	1100	515	438	85.0	438	381	87.0	74.0	786	612	77.9	612	510	83.3	64.9	68.5
N.S.	**Total**	1,382	1,152	83.4	1,152	1,019	88.5	73.7	1,878	1,568	83.5	1,568	1,376	87.8	73.3	73.5
	1210	197	182	92.4	182	165	90.7	83.8	298	242	81.2	242	214	88.4	71.8	76.6
	1223	174	153	87.9	153	138	90.2	79.3	230	198	86.1	198	175	88.4	76.1	77.5
	1230	194	160	82.5	160	144	90.0	74.2	258	217	84.1	217	199	91.7	77.1	75.9
	1240	176	146	83.0	146	126	86.3	71.6	261	216	82.8	216	188	87.0	72.0	71.9
	1258	232	199	85.8	199	174	87.4	75.0	318	261	82.1	261	226	86.6	71.1	72.7
	1269	409	312	76.3	312	272	87.2	66.5	513	434	84.6	434	374	86.2	72.9	70.1
N.B.	**Total**	1,442	1,180	81.8	1,180	1,043	88.4	72.3	1,869	1,500	80.3	1,500	1,286	85.7	68.8	70.3
	1301	284	210	73.9	210	183	87.1	64.4	354	275	77.7	275	240	87.3	67.8	66.3
	1302	286	219	76.6	219	184	84.0	64.3	382	305	79.8	305	274	89.8	71.7	68.6
	1303	263	229	87.1	229	209	91.3	79.5	342	284	83.0	284	248	87.3	72.5	75.5
	1304	151	130	86.1	130	123	94.6	81.5	183	148	80.9	148	124	83.8	67.8	74.0
	1305	143	116	81.1	116	103	88.8	72.0	158	133	84.2	133	113	85.0	71.5	71.8
	1306	165	152	92.1	152	140	92.1	84.8	251	189	75.3	189	144	76.2	57.4	68.3
	1307	150	124	82.7	124	101	81.5	67.3	199	166	83.4	166	143	86.1	71.9	69.9
Que.	**Total**	6,344	5,203	82.0	5,203	4,822	92.7	76.0	10,623	8,300	78.1	8,300	7,185	86.6	67.6	70.8
	2401	264	231	87.5	231	226	97.8	85.6	501	399	79.6	399	345	86.5	68.9	74.6
	2402	341	302	88.6	302	287	95.0	84.2	476	391	82.1	391	352	90.0	73.9	78.2

Geography		Area frame							Phone frames							Combined
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
	2403	540	423	78.3	423	392	92.7	72.6	785	581	74.0	581	507	87.3	64.6	67.8
	2404	389	327	84.1	327	301	92.0	77.4	651	543	83.4	543	479	88.2	73.6	75.0
	2405	382	304	79.6	304	290	95.4	75.9	477	396	83.0	396	352	88.9	73.8	74.7
	2406	931	692	74.3	692	634	91.6	68.1	1,521	1,132	74.4	1,132	953	84.2	62.7	64.7
	2407	346	294	85.0	294	267	90.8	77.2	501	395	78.8	395	352	89.1	70.3	73.1
	2408	289	253	87.5	253	231	91.3	79.9	477	403	84.5	403	357	88.6	74.8	76.8
	2409	273	237	86.8	237	224	94.5	82.1	596	439	73.7	439	367	83.6	61.6	68.0
	2410	.	.	.	.	.	.	.	463	372	80.3	372	320	86.0	69.1	69.1
	2411	293	271	92.5	271	253	93.4	86.3	571	424	74.3	424	355	83.7	62.2	70.4
	2412	330	272	82.4	272	254	93.4	77.0	673	544	80.8	544	465	85.5	69.1	71.7
	2413	417	325	77.9	325	287	88.3	68.8	585	450	76.9	450	382	84.9	65.3	66.8
	2414	378	316	83.6	316	293	92.7	77.5	587	462	78.7	462	400	86.6	68.1	71.8
	2415	449	361	80.4	361	330	91.4	73.5	648	490	75.6	490	415	84.7	64.0	67.9
	2416	722	595	82.4	595	553	92.9	76.6	1,111	879	79.1	879	784	89.2	70.6	72.9
Ont.	**Total**	12,256	9,781	79.8	9,781	8,721	89.2	71.2	18,970	15,067	79.4	15,067	12,937	85.9	68.2	69.4
	3526	209	177	84.7	177	168	94.9	80.4	328	279	85.1	279	253	90.7	77.1	78.4
	3527	236	179	75.8	179	149	83.2	63.1	307	245	79.8	245	210	85.7	68.4	66.1
	3530	452	371	82.1	371	332	89.5	73.5	730	591	81.0	591	495	83.8	67.8	70.0
	3531	193	155	80.3	155	143	92.3	74.1	267	212	79.4	212	189	89.2	70.8	72.2
	3533	241	200	83.0	200	182	91.0	75.5	436	358	82.1	358	320	89.4	73.4	74.2
	3534	214	162	75.7	162	138	85.2	64.5	326	262	80.4	262	222	84.7	68.1	66.7
	3535	255	186	72.9	186	162	87.1	63.5	379	301	79.4	301	263	87.4	69.4	67.0
	3536	394	317	80.5	317	287	90.5	72.8	616	502	81.5	502	438	87.3	71.1	71.8
	3537	487	368	75.6	368	319	86.7	65.5	719	557	77.5	557	477	85.6	66.3	66.0
	3538	241	204	84.6	204	171	83.8	71.0	386	314	81.3	314	280	89.2	72.5	71.9
	3539	135	122	90.4	122	114	93.4	84.4	229	187	81.7	187	163	87.2	71.2	76.1
	3540	172	153	89.0	153	145	94.8	84.3	337	272	80.7	272	245	90.1	72.7	76.6
	3541	270	201	74.4	201	171	85.1	63.3	422	349	82.7	349	304	87.1	72.0	68.6

Geography		Area frame							Phone frames							Combined
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
	3542	236	200	84.7	200	181	90.5	76.7	295	226	76.6	226	201	88.9	68.1	71.9
	3543	286	250	87.4	250	220	88.0	76.9	406	348	85.7	348	307	88.2	75.6	76.2
	3544	410	331	80.7	331	310	93.7	75.6	684	541	79.1	541	463	85.6	67.7	70.7
	3546	414	355	85.7	355	325	91.5	78.5	680	538	79.1	538	467	86.8	68.7	72.4
	3547	177	147	83.1	147	128	87.1	72.3	362	300	82.9	300	269	89.7	74.3	73.7
	3549	221	172	77.8	172	145	84.3	65.6	291	242	83.2	242	218	90.1	74.9	70.9
	3551	660	489	74.1	489	431	88.1	65.3	855	715	83.6	715	621	86.9	72.6	69.4
	3552	175	156	89.1	156	150	96.2	85.7	303	262	86.5	262	229	87.4	75.6	79.3
	3553	746	649	87.0	649	565	87.1	75.7	1,154	861	74.6	861	692	80.4	60.0	66.2
	3554	156	117	75.0	117	111	94.9	71.2	268	231	86.2	231	210	90.9	78.4	75.7
	3555	257	206	80.2	206	186	90.3	72.4	347	281	81.0	281	247	87.9	71.2	71.7
	3556	224	156	69.6	156	129	82.7	57.6	328	268	81.7	268	236	88.1	72.0	66.1
	3557	184	167	90.8	167	144	86.2	78.3	316	268	84.8	268	244	91.0	77.2	77.6
	3558	286	249	87.1	249	224	90.0	78.3	441	383	86.8	383	338	88.3	76.6	77.3
	3560	636	459	72.2	459	409	89.1	64.3	963	803	83.4	803	695	86.6	72.2	69.0
	3561	298	237	79.5	237	208	87.8	69.8	404	328	81.2	328	283	86.3	70.0	69.9
	3562	355	269	75.8	269	245	91.1	69.0	536	429	80.0	429	392	91.4	73.1	71.5
	3563	122	88	72.1	88	82	93.2	67.2	217	182	83.9	182	162	89.0	74.7	72.0
	3565	425	317	74.6	317	285	89.9	67.1	664	536	80.7	536	448	83.6	67.5	67.3
	3566	287	247	86.1	247	224	90.7	78.0	492	408	82.9	408	360	88.2	73.2	75.0
	3568	381	305	80.1	305	282	92.5	74.0	640	508	79.4	508	423	83.3	66.1	69.0
	3570	536	426	79.5	426	385	90.4	71.8	847	616	72.7	616	497	80.7	58.7	63.8
	3595	1,285	994	77.4	994	871	87.6	67.8	1,995	1,364	68.4	1,364	1,076	78.9	53.9	59.4
Man.	**Total**	2,064	1,708	82.8	1,708	1,531	89.6	74.2	2,822	2,197	77.9	2,197	1,945	88.5	68.9	71.1
	4610	637	495	77.7	495	433	87.5	68.0	830	661	79.6	661	576	87.1	69.4	68.8
	4615	174	148	85.1	148	136	91.9	78.2	212	167	78.8	167	150	89.8	70.8	74.1
	4620	127	107	84.3	107	98	91.6	77.2	143	110	76.9	110	102	92.7	71.3	74.1
	4625	175	157	89.7	157	142	90.4	81.1	239	193	80.8	193	169	87.6	70.7	75.1

Geography		Area frame							Phone frames							Combined
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
	4630	188	170	90.4	170	153	90.0	81.4	257	195	75.9	195	173	88.7	67.3	73.3
	4640	200	173	86.5	173	157	90.8	78.5	336	269	80.1	269	242	90.0	72.0	74.4
	4645	167	138	82.6	138	123	89.1	73.7	269	214	79.6	214	198	92.5	73.6	73.6
	4660	142	117	82.4	117	107	91.5	75.4	213	157	73.7	157	137	87.3	64.3	68.7
	4670	119	93	78.2	93	84	90.3	70.6	187	136	72.7	136	120	88.2	64.2	66.7
	4685	135	110	81.5	110	98	89.1	72.6	136	95	69.9	95	78	82.1	57.4	64.9
Sask.	**Total**	1,759	1,538	87.4	1,538	1,441	93.7	81.9	3,165	2,413	76.2	2,413	2,128	88.2	67.2	72.5
	4701	126	116	92.1	116	113	97.4	89.7	236	175	74.2	175	156	89.1	66.1	74.3
	4702	149	138	92.6	138	135	97.8	90.6	231	160	69.3	160	143	89.4	61.9	73.2
	4703	131	126	96.2	126	121	96.0	92.4	211	165	78.2	165	144	87.3	68.2	77.5
	4704	294	263	89.5	263	250	95.1	85.0	485	367	75.7	367	327	89.1	67.4	74.1
	4705	163	148	90.8	148	139	93.9	85.3	267	208	77.9	208	181	87.0	67.8	74.4
	4706	329	259	78.7	259	229	88.4	69.6	542	437	80.6	437	382	87.4	70.5	70.1
	4707	137	112	81.8	112	106	94.6	77.4	206	167	81.1	167	150	89.8	72.8	74.6
	4708	130	115	88.5	115	108	93.9	83.1	186	143	76.9	143	127	88.8	68.3	74.4
	4709	158	139	88.0	139	130	93.5	82.3	256	192	75.0	192	167	87.0	65.2	71.7
	4710	142	122	85.9	122	110	90.2	77.5	239	188	78.7	188	161	85.6	67.4	71.1
	4714	.	.	.	.	.	.	.	306	211	69.0	211	190	90.0	62.1	62.1
Alta.	**Total**	3,498	2,831	80.9	2,831	2,540	89.7	72.6	5,224	3,826	73.2	3,826	3,261	85.2	62.4	66.5
	4831	485	415	85.6	415	373	89.9	76.9	778	582	74.8	582	515	88.5	66.2	70.3
	4832	798	657	82.3	657	607	92.4	76.1	1,223	870	71.1	870	728	83.7	59.5	66.1
	4833	624	472	75.6	472	415	87.9	66.5	974	719	73.8	719	625	86.9	64.2	65.1
	4834	878	706	80.4	706	630	89.2	71.8	1,136	851	74.9	851	719	84.5	63.3	67.0
	4835	713	581	81.5	581	515	88.6	72.2	1,113	804	72.2	804	674	83.8	60.6	65.1
B.C.	**Total**	4,516	3,672	81.3	3,672	3,276	89.2	72.5	7,154	5,442	76.1	5,442	4,620	84.9	64.6	67.7
	5911	158	145	91.8	145	127	87.6	80.4	254	197	77.6	197	169	85.8	66.5	71.8
	5912	173	153	88.4	153	135	88.2	78.0	248	192	77.4	192	171	89.1	69.0	72.7
	5913	303	274	90.4	274	261	95.3	86.1	512	405	79.1	405	364	89.9	71.1	76.7

Geography		Area frame							Phone frames							Combined
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
	5914	259	213	82.2	213	197	92.5	76.1	438	332	75.8	332	284	85.5	64.8	69.0
	5921	294	229	77.9	229	192	83.8	65.3	460	375	81.5	375	311	82.9	67.6	66.7
	5922	426	367	86.2	367	340	92.6	79.8	687	527	76.7	527	425	80.6	61.9	68.7
	5923	447	360	80.5	360	324	90.0	72.5	707	534	75.5	534	456	85.4	64.5	67.6
	5931	220	187	85.0	187	175	93.6	79.5	392	268	68.4	268	218	81.3	55.6	64.2
	5932	489	363	74.2	363	333	91.7	68.1	807	545	67.5	545	440	80.7	54.5	59.6
	5933	331	266	80.4	266	207	77.8	62.5	506	370	73.1	370	303	81.9	59.9	60.9
	5941	378	299	79.1	299	280	93.6	74.1	627	495	78.9	495	429	86.7	68.4	70.5
	5942	284	241	84.9	241	215	89.2	75.7	441	354	80.3	354	311	87.9	70.5	72.6
	5943	150	103	68.7	103	99	96.1	66.0	208	171	82.2	171	154	90.1	74.0	70.7
	5951	158	123	77.8	123	101	82.1	63.9	261	192	73.6	192	164	85.4	62.8	63.2
	5952	259	207	79.9	207	177	85.5	68.3	354	283	79.9	283	255	90.1	72.0	70.5
	5953	187	142	75.9	142	113	79.6	60.4	252	202	80.2	202	166	82.2	65.9	63.6
Y.T.	**6001**	639	549	85.9	549	512	93.3	80.1	196	160	81.6	160	146	91.3	74.5	78.8
N.W.T.	**6101**	714	613	85.9	613	541	88.3	75.8	141	111	78.7	111	99	89.2	70.2	74.9
Nvt.	**6201**	463	397	85.7	397	362	91.2	78.2	.	.	.	.	.	.	.	78.2

Appendix E (2011) - Table 9.2 Response rate by Local Health Integrated Network (LHIN) and frame in the CCHS in Ontario

Geography		Area frame							Phone frames							Combined
Province	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**	12,256	9,781	79.8	9,781	8,721	89.2	71.2	18,970	15,067	79.4	15,067	12,937	85.9	68.2	69.4
	3501	789	658	83.4	658	608	92.4	77.1	1,272	1,006	79.1	1,006	869	86.4	68.3	71.7
	3502	1,289	1,061	82.3	1,061	992	93.5	77.0	2,153	1,763	81.9	1,763	1,543	87.5	71.7	73.6
	3503	670	531	79.3	531	476	89.6	71.0	1,044	857	82.1	857	736	85.9	70.5	70.7
	3504	1,481	1,165	78.7	1,165	1,022	87.7	69.0	2,242	1,777	79.3	1,777	1,533	86.3	68.4	68.6
	3505	596	502	84.2	502	420	83.7	70.5	948	705	74.4	705	566	80.3	59.7	63.9
	3506	620	526	84.8	526	479	91.1	77.3	976	751	76.9	751	629	83.8	64.4	69.4
	3507	665	514	77.3	514	449	87.4	67.5	1,003	683	68.1	683	554	81.1	55.2	60.1
	3508	827	661	79.9	661	602	91.1	72.8	1,275	917	71.9	917	722	78.7	56.6	63.0
	3509	1,196	937	78.3	937	832	88.8	69.6	1,831	1,427	77.9	1,427	1,199	84.0	65.5	67.1
	3510	719	587	81.6	587	505	86.0	70.2	1,117	928	83.1	928	815	87.8	73.0	71.9
	3511	1,208	973	80.5	973	856	88.0	70.9	1,719	1,459	84.9	1,459	1,289	88.3	75.0	73.3
	3512	590	420	71.2	420	375	89.3	63.6	924	766	82.9	766	669	87.3	72.4	69.0
	3513	1,030	805	78.2	805	715	88.8	69.4	1,639	1,357	82.8	1,357	1,203	88.7	73.4	71.9
	3514	576	441	76.6	441	390	88.4	67.7	827	671	81.1	671	610	90.9	73.8	71.3

Appendix E (2011) - Table 9.3 Response rate by province and frames for the Health Services Access Survey (HSAS) sub-sample

Geography	Area frame							Phone frames							Combined
Province	# 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	18,151	14,909	82.1	14,182	12,790	90.2	74.1	27,330	21,107	77.2	19,661	16,856	85.7	66.2	69.4
N.L.	1,041	903	86.7	845	770	91.1	79.0	1,578	1,225	77.6	1,125	953	84.7	65.8	71.1
P.E.I.	515	438	85.0	413	360	87.2	74.1	786	612	77.9	554	468	84.5	65.8	69.1
N.S.	1,382	1,152	83.4	1,100	977	88.8	74.0	1,878	1,568	83.5	1,457	1,277	87.6	73.2	73.5
N.B.	1,442	1,180	81.8	1,121	991	88.4	72.3	1,869	1,500	80.3	1,411	1,204	85.3	68.5	70.2
Que.	2,715	2,171	80.0	2,088	1,930	92.4	73.9	4,261	3,323	78.0	3,113	2,691	86.4	67.4	70.0
Ont.	2,873	2,281	79.4	2,155	1,913	88.8	70.5	4,504	3,461	76.8	3,218	2,706	84.1	64.6	66.9
Man.	1,877	1,555	82.8	1,480	1,326	89.6	74.2	2,584	2,012	77.9	1,869	1,657	88.7	69.0	71.2
Sask.	1,655	1,446	87.4	1,378	1,291	93.7	81.9	2,706	2,079	76.8	1,955	1,727	88.3	67.9	73.2
Alta.	2,250	1,829	81.3	1,736	1,557	89.7	72.9	3,305	2,411	73.0	2,251	1,904	84.6	61.7	66.3
B.C.	2,401	1,954	81.4	1,866	1,675	89.8	73.1	3,859	2,916	75.6	2,708	2,269	83.8	63.3	67.1