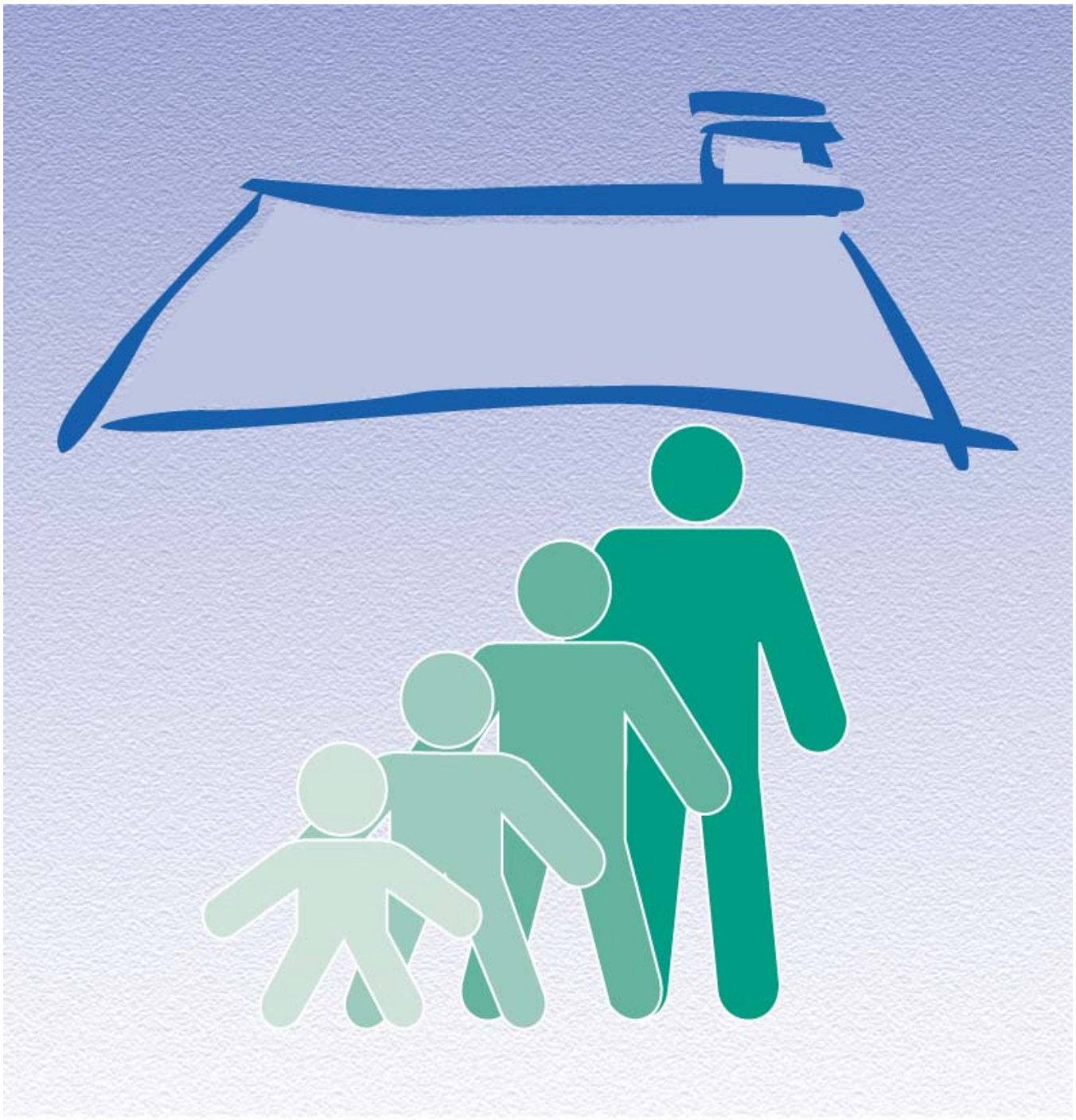
Canadian Community Health Survey (CCHS) – Annual component

User guide
2007-2008 Microdata files

June 2009



Statistics Canada Statistique Canada

Canada

TABLE OF CONTENTS

1. INTRODUCTION

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

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

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

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

The purpose of this document is to facilitate the manipulation of the CCHS microdata files and to describe the methodology used. The CCHS produces three types of microdata files: master files, share files and public use microdata files (PUMF). The characteristics of each of these files are presented in this guide. The PUMF will be released in the summer of 2009 and will include the data collected for the years 2007 and 2008.

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

Electronic Products Help Line: 1-800-949-9491

For custom tabulations or general data support:
Client Custom Services, Health Statistics Division: 613-951-1653
E-mail: hd-ds@statcan.gc.ca

For remote access support: 613-951-1653
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, 2008" and the short title is simply "CCHS – 2008". The focused content component of the survey remains unchanged. It will continue to examine in greater detail more specific topics or populations. It will be designated by the name of the survey followed by the topic of the themes covered by each survey (e.g., "Canadian Community Health Survey on Healthy Aging" or "CCHS – Healthy Aging").

[1] 1999. Health Information Roadmap: Responding to Needs, Health Canada, Statistics Canada. p. 3.

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

3. CCHS REDESIGN IN 2007

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

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

Below are the main changes arising from the CCHS redesign:

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

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

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

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

4. CONTENT STRUCTURE OF THE CCHS

In addition to socio-demographic and administrative data, the content of the CCHS includes three components, each of which addresses a different need: the common content component comprising the core content and theme content, the optional content component, and the rapid response component. Appendix A lists the modules included in the 2008 questionnaire by component.

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

Table 4.1 Length of survey by component

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

4.1 Common content

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

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

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

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

4.2 Optional content

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

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

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

4.3 Rapid response content

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

4.4 Content included in data files

The survey produces annual files. Also, a file combining the most recent two years is released every two years. In June 2009, two microdata files were released: an annual file based on the 2008 reference period, which resembles the 2007 data file, and a data file based on the 2007-2008 reference period. The main 2007-2008 data file is similar in size to files from previous cycles (2001, 2003 and 2005), i.e. approximately 130,000 respondents.

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

Also, theme modules collected from a sub-sample of the population will continue to be disseminated in separate files. These include the core content and the theme module content collected from a sub-sample of respondents. Table 4.2 provides clarification about the content available in the 2007, 2008 and combined 2007-2008 data files.

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

Table 4.2 Content components included in 2007 and 2008 data files

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

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

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

3 Optional content is included in the 2007-2008 data file if it was asked of respondents in a province 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 a theme module from one year is selected for the optional content of a jurisdiction during the second year, the module is included in the two-year data file and is 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 2007-2008, data was collected in 118 HRs in the ten provinces, as well as to one HR per territory, totalling 121 HRs (Appendix C).

5.3 Sample size and allocation

To provide reliable estimates for each HR 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 3.1 and the sample sizes have remained mainly the same since then. The sample was then divided evenly between the 2 collection years. Table 5.1 gives the total sample size for the years 2007 and 2008.

Table 5.1 Number of health regions and targeted sample sizes by province/territory, 2007-2008

Province	Number of HRs	Total sample size (targeted)
Newfoundland and Labrador	4	4,010
Prince Edward Island	3	2,480
Nova Scotia	6	5,040
New Brunswick	7	5,160
Quebec	16	24,290
Ontario[1]	36	44,460
Manitoba	10	7,510
Saskatchewan	11	7,720
Alberta	9	12,210
British Columbia	16	16,100
Yukon	1	1,200
Northwest Territories	1	1,200
Nunavut	1	700
Canada	**121**	**132,080**

[1] The sample size for Ontario includes the buy-in extra sample by LHIN. The initial sample size for Ontario before the buy-in was 20,900 units (refer to section 5.7 for further details).

In the third step, 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. In total, for 2007-2008, 1,200 sample units were allocated to the Yukon, 1,200 to the Northwest Territories and 700 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 equally divided in 2 in order to obtain the same sample sizes between the area frame sample and the list frame sample for each HR[4], as described in the next section. We should finally mention that the size of the samples taken from each frame was increased before data collection in order to account for the anticipated out-of-scope and non-response rates based on the rates obtained in previous CCHS cycles. The sample sizes by HR and frame are provided in Appendix D.

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

5.4 Frames, household sampling strategies

CCHS 2008 used three sampling frames to select the sample of households: 49% of the sample of households came from an area frame, 50% came from a list frame of telephone numbers and the remaining 1% came from a Random Digit Dialling (RDD) sampling frame.

5.4.1 Sampling of households from the area frame

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

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

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

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

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

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

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

Due to the specific of the CCHS, some modifications had to be incorporated in this sampling strategy. To obtain an annual sample of 33,000 respondents for CCHS, close to 47,000 dwellings had to be selected from the area frame to account for vacant dwellings and non-responding households. Each month, the LFS design provides approximately 60,000 dwellings distributed across the various economic regions in the ten provinces, whereas the CCHS required 47,000 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 *adjustment factor* of 0.78 (47,000/60,000). However, since the adjustment factors varied from 0.3 to 3.0 at the HR level, certain adjustments were required.

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

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

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

Sampling of households from the area frame in the three territories

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

It is worth mentioning that the frame for the CCHS 2007-2008 covered 90% of the private households in the Yukon, 97% in the Northwest Territories and 71% in Nunavut[7].

5.4.2 Sampling of households from the list frame of telephone numbers

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

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

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

5.4.3 Sampling of households from the RDD frame of telephone numbers

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

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

5.5 Sample allocation over the collection period

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

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

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

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

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

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

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

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

The province of Ontario requested a sample increase in order to produce estimates at the Local Health Integrated Network (LHIN) geography level. Ontario contains 14 LHIN (see Appendix Y). The CCHS sample was increased in order to obtain a minimum size of 2,000 per LHIN over a period of 2 years. As the HR and LHIN boundaries intersect each other, the stratification level used was the HR-LHIN overlap. The preliminary sample sizes allotted by HR are therefore preserved. In cases where the HR allocation prevented the sample from reaching sizes of 2,000 per LHIN, the sample was then increased, and was allocated proportionally to the size of the population within the HR-LHIN overlap. Table 5.4 provides the sample sizes of targeted respondents by LHIN for 2007-2008.

Table 5.4 Targeted respondents by Local Health Integrated Network (LHIN), CCHS 2007-2008.

LHIN	Targeted respondents
01-Erie St. Clair	3,104
02-South West	5,136
03-Waterloo Wellington	2,460
04-Hamilton Niagara Haldimand Brant	5,248
05-Central West	2,124
06-Mississauga Halton	2,272
07-Toronto Central	2,170
08-Central	2,808
09-Central East	4,152
10-South East	2,688
11-Champlain	4,116
12-North Simcoe Muskoka	2,116
13-North East	3,984
14-North West	2,082
Ontario	**44,460**

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

6. DATA COLLECTION

6.1 Computer-assisted interviewing

Between January 2007 and December 2008, a total of 131,959 valid interviews were conducted using computer assisted interviewing (CAI). Approximately half the interviews were conducted in person using computer assisted personal interviewing (CAPI) and the other half were conducted over the phone using computer assisted telephone interviewing (CATI).

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

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

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

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

6.2 CCHS application development

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

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

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

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

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

6.3 Interviewer training

Project managers, 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 2007 and 2008 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, almost three-quarters (73.9%) of these interviews were conducted exclusively in person. CAPI interviewers worked independently from their homes using laptop computers and were supervised from a distance by senior interviewers. The variable SAM_TYP on the microdata files indicates whether a case was selected from the area frame (CAPI) or from the telephone or RDD frame (CATI).

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

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

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

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

a) Introductory letters

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

b) Initiating contact

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

c) Refusal conversion

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

d) Language barriers

To remove language as a barrier to conducting interviews, each of the Statistics Canada Regional Offices recruited interviewers with a wide range of language competencies. When necessary, cases were transferred to an interviewer with the language competency needed to complete an interview.

e) Youth interviews

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

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

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

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

f) Proxy interviews

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

6.5 Field operations

The majority of the 2007/2008 sample was divided into twelve non-overlapping two-month collection periods. Regional collection offices were instructed to use the first 4 weeks of each collection period to resolve the majority of the sample, with next 4 weeks being used finalise the remaining sample and to follow up on outstanding non-response cases. All cases were to have been attempted by the second week of each collection period.

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

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

At the end of data collection, a national response rate of 76% was achieved. Complete details regarding the response rates can be found in Appendix E.

6.6 Quality control and collection management

During the 2007 and 2008 collection years, several methods were used to ensure data quality and to optimize collection. These included using internal measures to verify interviewer performance and the use of a series of ongoing reports to monitor various collection targets and data quality.

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

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

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

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

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

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

7. DATA PROCESSING

7.1 Editing

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

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

7.2 Coding

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

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

7.3 Creation of derived variables

To facilitate data analysis and to minimize the risk of error, a number of variables on the file have been derived using items found on the CCHS questionnaire. Derived variables generally have a "D", "G" or "F" in the fourth character of the variable name. In some cases, the derived variables are straightforward, involving collapsing of response categories. In other cases, several variables have been combined to create a new variable. The *Derived Variables Documentation (DV)* provides details on how these more complex variables were derived. For more information on the naming convention, please go to Section 12.5.

7.4 Weighting

The principle behind estimation in a probability sample such as 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.

Details of the method used to calculate sampling weights are presented in Section 8.

7.5 Conversion of CCHS 2007-2008 Master File to Public Use Microdata File (PUMF)

The approach for creating a PUMF is intended to balance the requirements for maintaining respondent confidentiality by minimising disclosure risks, while providing the most useful data at the level of geography of the health region. The following paragraphs outline some of the procedures applied to convert the CCHS master file into a PUMF.

Health regions: Health regions fall under provincial jurisdiction. As mentioned in Section 5.2, at the time of the design of the 2007-2008 sample, there were 121 HRs across Canada. Thus, the master data file provides representative information for 121 HRs.

Some health regions have population sizes that were considered to be too small to appear individually in the PUMF. These health regions were thus collapsed with other(s). The approach for keeping or collapsing health region geography in the PUMF entails applying a minimum value of 70,000 on the population size. This resulted in:

i) collapsing of 29 HRs, in all provinces except Quebec, Nova Scotia, Alberta, and British Columbia into 13 health region groupings,
ii) grouping of the three territories into a single entity (results in suppression of all optional content modules),
iii) one Quebec HR (Nord-du-Québec) is excluded from the PUMF because its small population size and demographic characteristics did not lend itself well to being collapsed with other HRs,
iv) all sample design variables were excluded.

After collapsing, the CCHS PUMF comprised 102 geographic HRs/HR groupings across the country. HRs before and after collapsing are shown in Appendix F.

Optional content suppression: As a result of the grouping of HRs, the optional content that was not common to each health region within a grouping would have to be suppressed. Although optional content selection could vary from region to region within a province, in each of 2007 and 2008 all health regions within each province selected the same optional content; therefore this suppression was not necessary for any provinces. However, because the three territories were grouped, suppression of optional content occurs in the territories if only one or two of the territories selected the content.

Beginning in 2007, provinces and regions were given the possibility of changing their optional content selection every year. The 2007-2008 master and PUMF files only contain optional content data for modules that were asked in both years, which applies to the vast majority of optional content data.

Disclosure control: As mentioned earlier, the 2007-2008 PUMF is designed to preserve analytical value of data as much as possible while ensuring the potential for identifying individual respondents is minimal.

Several controls were implemented in creating the PUMF. Restriction methods such as removing direct identifiers (e.g., sample ID, name, telephone number), reducing, recoding, and/or suppressing detail based on small frequencies given specific socio-demographic characteristics were used. Examples of some master file variables not being included in the PUMF due to a high risk of disclosure (either because the variable is a risk on its own or is so in combination with other variables) include: attempted suicide in the past 12 months, has skin cancer, and whether the respondent is presently pregnant.

Some response categories deemed to be at possible risk of disclosure were regrouped and included on the PUMF. Examples include: body part affected by most serious injury, number of times consulting health professionals, number of years since stopping smoking and main source of household income. Thus, the PUMF contains both fewer variables and less detail compared with the 2007-2008 master file, but in a way that preserves analytical value to the data.

Age of respondent: Respondents' age is provided in the CCHS 2007-2008 PUMF in age groupings. Most are found in 5-year age groups, from age 20-24 to age 75-79. All respondents aged 80 and over were also grouped. Prior to the CCHS 3.1 PUMF, youth age groups included those aged 12-14 and 15-19. Beginning with the CCHS 3.1 PUMF, the older youth age group has been split between those aged 15-17 and those aged 18-19. The older youth age groupings were changed for two principal reasons:

1) Body Mass Index (BMI) has emerged as a top priority in the public health realm. New BMI calculations are newly available for youths aged 12-17 but would not have been fully available to users via the older youth age grouping;
2) in prior releases we had to suppress data for those aged 18-19 for variables which were only asked to respondents aged 18 and over.

Therefore, the proposed new youth age groupings have at least three important benefits:

i) BMI data can be presented for those aged 12-17;
ii) data for those aged 18-19 is longer automatically suppressed for variables that were only

asked to those aged 18 and over;
iii) users can still compare youth data from previous PUMFs by grouping the 15-17 year olds with the 18-19 year olds to create the age group 15-19.

Although information on some variables were collected for people of a specific age bracket, some data for certain ages still had to be suppressed to ensure confidentiality. For example, labour force data were collected for respondents aged 15 to 75. Because there is an age group 75-79 on the PUMF, publishing labour force data for 75-year olds would reveal their exact age by default. Therefore, labour force data (LBS), work stress (WST), and physical activities – facilities at work (PAF) are only available on the PUMF for respondents aged 15-74. Similar suppression was done for education (EDU) data of 14-year olds and maternal experience (MEX, MXS, and MXA) data for 55-year old women.

Special suppression was done for the three modules on maternal experiences (MEX, MXS, and MXA), data for women aged 15-19 because there was concern for a high risk of disclosure for women in this age group. As such, though MEX, MXS, and MXA data on the master file are shown for all women aged 15 to 55, the PUMF only presents data for women aged 20 to 54.

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.

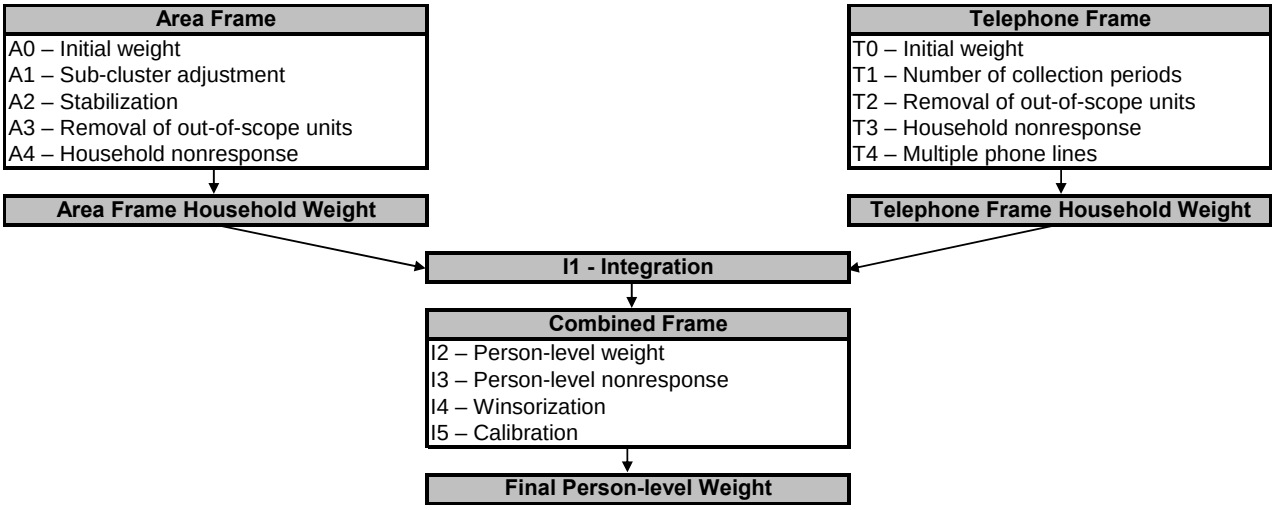
8.1 Overview

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, 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 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)[9].

A1 – Sub-cluster adjustment

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

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

A2 – Stabilization

In some HRs, the increase of the sample size as described in section 5, results in a larger sample than necessary. Stabilization is used to bring the sample size back down to the desired level. The stabilization process consists of randomly sub-sampling dwellings at the HR level from the dwellings originally 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 representative of, in-scope dwellings or households. These in-scope dwellings that remain maintain the same weight as in the previous step, which is now called A3.

A4 – Household nonresponse

During collection, a certain proportion of sampled households inevitably result in nonresponse. This usually occurs when a household refuses to participate in the survey, provides unusable data, or cannot be reached for an interview. Weights of the nonresponding households are redistributed to responding households within response homogeneity groups (RHGs). In order to create the response groups, a scoring method based on logistic regression models is used to determine the propensity to respond and these response probabilities are used to divide the sample into groups with similar response properties. The information available for nonrespondents is limited so the regression model uses characteristics such as the collection period and geographic information, as well as paradata 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. 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[10] for more details). Therefore, the probability of selection is the ratio between the number of sampled units and one hundred times the number of banks within the RDD stratum.

For the list frame, telephone numbers are 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 the inverse of the number of samples being combined (i.e. the number of collection periods). Following this adjustment, the entire list frame sample corresponds to the average over the entire combined collection period. The initial weights are multiplied by this adjustment factor to produce weight T1.

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

T2 - Removal of out-of-scope numbers

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

T3 – Household nonresponse

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

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

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[11]. 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. 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.

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

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 was used in the treatment of household nonresponse, the adjustment was applied within response homogeneity groups. In this process, the scoring method was 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 were available for the estimation of the response probabilities as well as geographic information and some paradata. The probabilities were grouped into response groups and the following adjustment factor was calculated:

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

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

I4 – Winsorization

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

The population estimates are based on the most recent Census counts and counts of birth, death, immigration and emigration since that time. The average of these monthly estimates for each of the HR-age-sex post-strata by collection period is used to calibrate. The weight I4 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.

8.6 Particular aspects of the weighting in the three territories

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

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

For the weighting of the telephone frame units, it should be noted that only the RDD frame is used and exclusively in the Yukon and Northwest Territories capitals. All of the telephone frame adjustments are applied to derive a final weight for the telephone units.
The two sets of weights (area and telephone) are subsequently integrated and post-stratified in a similar way to what is done for the provinces, with 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.

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

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. These controls ensure that the proportion of the estimates represented by these different groups is consistent with proportions indicated by the 2006 Census.

9. DATA QUALITY

9.1 Response rates

In total, 84,973 of the selected units in the CCHS 2007 were in-scope for the survey[13]. Out of these, 71,922 households accepted to participate in the survey resulting in an overall household-level response rate of 84.6%. Among these responding households, 71,922 individuals (one per household) were selected to participate to the survey, out of which a response was obtained for 65,946 individuals, resulting in an overall person-level response rate of 91.7%. At the Canada level, this yields a combined response rate of 77.6% for the CCHS 2007. 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 9.1: 2007-2008 response rate by health region and frames

(see Appendix E)

Table 9.2: Reponse rate by Local Health Integrated Network (LHIN) and frames in the CCHS 2007-2008 in Ontario

(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 =		# of responding households in both frames
		all in-scope households in both frames

Person-level response rate

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

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

Combined response rate = HHRR x PPRR

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

$$\text{HHRR} = \frac{35{,}748 + 36{,}174}{40{,}668 + 44{,}310} = \frac{71{,}922}{84{,}978} = 0.846$$

$$\text{PPRR} = \frac{33{,}473 + 32{,}473}{35{,}748 + 36{,}174} = \frac{65{,}946}{71{,}922} = 0.917$$

Combined response rate = 0.846 x 0.917

= 0.776

= **77.6%**

9.2 Survey Errors

The estimates derived from this survey are based on a sample of individuals. Somewhat different figures might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. than those actually used. The difference between the estimates obtained from the sample and the results from a complete count under similar conditions is called the *sampling error* of the estimate.

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

9.2.1 Non-sampling Errors

Over a large number of observations, randomly occurring errors will have little effect on estimates derived from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Considerable time and effort was made to reduce non-sampling errors in the CCHS 2007. Quality assurance measures were implemented at each step of data collection and processing to monitor the quality of the data. These measures included the use of highly skilled interviewers, extensive training with respect to the survey procedures and questionnaire, and the observation of interviewers to detect problems. Testing of the CAI application and field tests were also essential procedures to ensure that data collection errors were minimized.

A major source of non-sampling errors in surveys is the effect of *non-response* on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. Partial non-response to the CCHS 2007 was minimal; once the questionnaire was started, it tended to be completed with very little non-response. Total non-response occurred either because a person refused to participate in the survey or because the interviewer was unable to contact the selected person. Total non-response was handled by adjusting the weight of persons who responded to the survey to compensate for those who did not respond. See Section 8 for details on the weight adjustment for non-response.

9.2.2 Sampling Errors

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error. The basis for measuring the potential size of sampling errors is the standard deviation of the estimates derived from survey results. However, because of the large variety of estimates that can be produced from a survey, the standard deviation of an estimate is usually expressed relative to the estimate to which it pertains. This resulting measure, known as the coefficient of variation (CV) of an estimate, is obtained by dividing the standard deviation of the estimate by the estimate itself and is expressed as a percentage of the estimate.

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

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

Statistics Canada commonly uses CV results when analyzing data and urges users producing estimates from the CCHS 2007 data files to also do so. For details on how to determine CVs, see Section 11. For guidelines on how to interpret CV results, see the table at the end of Sub-section 10.4.

10. GUIDELINES FOR TABULATION, ANALYSIS AND RELEASE

This section of the documentation outlines the guidelines to be used by users in tabulating, analyzing, publishing or otherwise releasing any data derived from the survey files. With the aid of these guidelines, users of microdata should be able to produce figures that are in close agreement with those produced by Statistics Canada and, at the same time, will be able to develop currently unpublished figures in a manner consistent with these established guidelines.

10.1 Rounding guidelines

In order that estimates for publication or other release derived from the data files (Master, Share or PUMF) correspond to those produced by Statistics Canada, users are urged to adhere to the following guidelines regarding the rounding of such estimates:

a) Estimates in the main body of a statistical table are to be rounded to the nearest hundred units using the normal rounding technique. In normal rounding, if the first or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is raised by one. For example, in normal rounding to the nearest 100, if the last two digits are between 00 and 49, they are changed to 00 and the preceding digit (the hundreds digit) is left unchanged. If the last digits are between 50 and 99 they are changed to 00 and the proceeding digit is incremented by 1;

b) Marginal sub-totals and totals in statistical tables are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units using normal rounding;

c) Averages, proportions, rates and percentages are to be computed from unrounded components (i.e., numerators and/or denominators) and then are to be rounded themselves to one decimal using normal rounding. In normal rounding to a single digit, if the final or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is increased by 1;

d) Sums and differences of aggregates (or ratios) are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units (or the nearest one decimal) using normal rounding;

e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released that differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s);

f) Under no circumstances are unrounded estimates to be published or otherwise released by users. Unrounded estimates imply greater precision than actually exists.

10.2 Sample weighting guidelines for tabulation

The sample design used for this survey was not self-weighting. That is to say, the sampling weights are not identical for all individuals in the sample. When producing simple estimates, including the production of ordinary statistical tables, users must apply the proper sampling weight. If proper weights are not used, the estimates derived from the data file cannot be considered to be representative of the survey population, and will not correspond to those produced by Statistics Canada.

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

10.2.1 Definitions: categorical estimates, quantitative estimates

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

Categorical estimates:

Categorical estimates are estimates of the number or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of individuals who smoke daily is an example of such an estimate. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Example of categorical question:

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

Quantitative estimates:

Quantitative estimates are estimates of totals or of means, medians and other measures of central tendency of quantities based upon some or all of the members of the surveyed population.

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

Example of quantitative question:

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

10.2.2 Tabulation of categorical estimates

Estimates of the number of people with a certain characteristic can be obtained from the data file by summing the final weights of all records possessing the characteristic of interest.

Proportions and ratios of the form $\hat{X} / \hat{Y}$ are obtained by:

a) summing the final weights of records having the characteristic of interest for the numerator ($\hat{X}$);

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

c) dividing the numerator estimate by the denominator estimate.

10.2.3 Tabulation of quantitative estimates

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

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

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

c) dividing the numerator estimate by the denominator estimate.

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

10.3 Guidelines for statistical analysis

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

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

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

10.4 Release guidelines

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

Table 10.1 Sampling variability guidelines

Type of Estimate	CV (in %)	Guidelines
Acceptable	$0.0 \leq CV \leq 16.5$	Estimates can be considered for general unrestricted release. Requires no special notation.
Marginal	$16.6 < CV \leq 33.3$	Estimates can be considered for general unrestricted release but should be accompanied by a warning cautioning subsequent users of the high sampling variability associated with the estimates. Such estimates should be identified by the letter E (or in some other similar fashion).
Unacceptable	$CV > 33.3$	Statistics Canada recommends not to release estimates of unacceptable quality. However, if the user chooses to do so then estimates should be flagged with the letter F (or in some other fashion) and the following warning should accompany the estimates: "The user is advised that . . .(specify the data) . . . do not meet Statistics Canada's quality standards for this statistical program. Conclusions based on these data will be unreliable and most likely invalid. These data and any consequent findings should not be published. If the user chooses to publish these data or findings, then this disclaimer must be published with the data."

11. APPROXIMATE SAMPLING VARIABILITY TABLES

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

The coefficients of variation (CV) are derived using the variance formula for simple random sampling and incorporating a factor which reflects the multi-stage, clustered nature of the sample design. This factor, known as the ***design effect,*** was determined by first calculating design effects for a wide range of characteristics and then choosing, for each table produced, a conservative value among all design effects relative to that table. The value chosen was then used to generate a table that applies to the entire set of characteristics.

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

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

11.1 How to use the CV tables for categorical estimates

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

Rule 1: Estimates of numbers possessing a characteristic (aggregates)

The coefficient of variation depends only on the size of the estimate itself. On the appropriate Approximate Coefficients of Variations Table, locate the estimated number in the left-most column of the table (headed "Numerator of Percentage") and follow the asterisks (if any) across to the first figure encountered. Since not all the possible values for the estimate are available, the smallest value which is the closest must be taken (as an example, if the estimate is equal to 1,700 and the two closest available values are 1,000 and 2,000, the first has to be chosen). This figure is the approximate coefficient of variation.

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

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

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

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

Rule 3: Estimates of differences between aggregates or percentages

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

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

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

Rule 4: Estimates of ratios

In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example, to the case where the denominator is the number of individuals who smoke at all and the numerator is the number of individuals who smoke daily out of those who smoke at all.

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

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

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

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

Rule 5: Estimates of differences of ratios

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

11.2 Examples of using the CV tables for categorical estimates

The following "real life" examples are included to assist users in applying the foregoing rules.

Example 1: Estimates of numbers possessing a characteristic (aggregates)

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

1) Refer to the CANADA level CV table.

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

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

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

Example 2 : Estimates of proportions or percentages possessing a characteristic

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

1) Refer to the CANADA level CV table.

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

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

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

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

Example 3 : Estimates of differences between aggregates or percentages

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

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

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

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

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

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

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

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

Example 4 : Estimates of ratios

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

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

2) Refer to the CANADA level CV table.

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

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

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

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

That is,

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

$$= 0.025$$

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

11.3 How to use the CV tables to obtain confidence limits

Although coefficients of variation are widely used, a more intuitively meaningful measure of sampling error is the confidence interval of an estimate. A confidence interval constitutes a statement on the level of confidence that the true value for the population lies within a specified range of values. For example a 95% confidence interval can be described as follows: if sampling of the population is repeated indefinitely, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.

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

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

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

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

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

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

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

11.4 Example of using the CV tables to obtain confidence limits

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

$\hat{X}$ = 0.777

z = 2

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

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

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

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

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

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

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

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

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

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

11.7 Exact variances/coefficients of variation

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

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

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

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

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

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

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

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

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

11.8 Release cut-offs for the CCHS

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

12. MICRODATA FILES: DESCRIPTION, ACCESS AND USE

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

12.1 Master files

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

12.1.1 Research Data Centre

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

12.1.2 Custom tabulations

Another way to access the master files is to offer all users the option of having staff in Client Services of the Health Statistics Division prepare custom tabulations. This service is offered on a cost-recovery basis. It allows users who do not possess knowledge of tabulation software products to get custom results. The results are screened for confidentiality and reliability concerns before release. For more information, please contact Client Services at 613-951-1653 or by e-mail at hd-ds@statcan.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 core content component of the theme content collected over two years and the selected optional content for two years.

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

12.4 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 available because of the quality associated with limited sample sizes.

The CCHS recommends having a Coefficient of Variation of less than 33% and having at least 10 respondents in the domain with the characteristic before publishing an estimate. 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 **Food Security** module measures concepts that apply not only to the respondent's situation, but also to that of the respondent's entire household. Depending on the level of analysis, the analysis of the variables may require use of a weight calculated to represent the number of Canadian households, rather than the number of persons. This weight variable **WTS_HH** is found in a separate file (HS_HHWT.txt). It can be used in place of the variable **WTS_M** for household analyses at the national and provincial levels. The weight variable **WTS_M** must be used for regional analyses.

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

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

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

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

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

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

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

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

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

In general, the last four positions (5 to 8) follow the variable numbering used on the questionnaire. The letter "Q" used to represent the word "question" is removed, and all question numbers are presented in a two or three digit format. For example, question Q01A in the questionnaire becomes simply 01A, and question Q15 becomes simply 15.

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

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

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

12.7 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 cycles 1.1 to 3.1.

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

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

Appendix A

Appendix A - Canadian community health survey (2007-2008)

<table>
<tr><th colspan="5">Core content (all regions)</th></tr>
<tr><td colspan="2">• Age of respondent (ANC)
• Alcohol use (ALC)
• Chronic conditions (CCC)
• Exposure to second-hand smoke (ETS)
• Flu shots (FLU)
• General health (GEN)
• Health care utilization (HCU)
• Pain and discomfort (HUP)</td><td colspan="2">• Height and weight – Self-reported (HWT)
• Maternal experiences - Breastfeeding (MEX)
• Fruit and vegetable consumption (FVC)
• Physical activities (PAC)
• Restriction of activities (RAC)
• Smoking (SMK)</td><td>Administration and socio-demographic information
• Administrative information (ADM)
• Dwelling characteristics (DWL)
• Education (EDU)
• Income (INC)
• Labour force (LBS)
Socio-demographic characteristics</td></tr>
<tr><th colspan="5">Theme content (all regions)</th></tr>
<tr><th colspan="2">Theme for 2007-2008: Healthy living</th><th colspan="2">Theme for 2007: Health Service and Access Survey</th><th>Theme for 2008: Prevention of Chronic Illnesses</th></tr>
<tr><td>• Changes made to improve health (CIH)
• Food security (FSC)
• Oral health 1 (OH1)</td><td>• Physical activities - Facilities at work (PAF)
• Sedentary activities (SAC)</td><td colspan="2">• Access to health care services (ACCZ)
• Patient satisfaction - Health care services (PASZ)
• Wait times (WTMZ)</td><td>• Blood test (BLT)
• PAP smear test (PAP)
• Mammography (MAM)</td><td>• Spirometry (SPI)
• Physical check-up (PCU)
• Colorectal cancer screening (CCS)</td></tr>
</table>

Optional content (certain regions)			
• Access to health care services (ACC)[ii] • Alcohol use - Dependence (ALD) • Alcohol use – Former drinkers (ALN) • Alcohol use during the past week (ALW) • Blood pressure check (BPC) • Breast examination (BRX) • Breast self-examination (BSX) • Stress – Childhood and adult stressors (CST) • Colorectal cancer screening (CCS) • Consultations about mental health (CMH) • Dental visits (DEN) • Depression (DEP) • Diabetes care (DIA) • Dietary supplement use – Vitamins and minerals (DSU) • Distress (DIS) • Driving and safety (DRV)	• Eye examinations (EYX) • Food choices (FDC) • Health care system satisfaction (HCS) • Health utility index (HUI) • Health status (SF-36) (SFR) • Home care services (HMC) • Home safety (HMS) • Illicit drugs use (DRG) • Injuries (INJ) • Insurance coverage (INS) • Mammography (MAM) • Mastery (MAS) • Maternal experiences – Alcohol use during pregnancy (MXA) • Maternal experiences – Smoking during pregnancy (MXS) • Smoking – Nicotine dependence (NDE)	• Oral health 2 (OH2) • PAP smear test (PAP) • Patient satisfaction - Health care services (PAS)[ii] • Patient satisfaction - Community-based care (PSC) • Problem gambling (CPG) • Prostate cancer screening (PSA) • Psychological well-being (PWB) • Stress – Recent life events (RLE) • Satisfaction with life (SWL) • Self-esteem (SFE) • Sexual behaviours (SXB) • Sleep (SLP) • Smoking cessation methods (SCA) • Smoking – Physician counselling (SPC)	• Smoking – Stages of change (SCH) • Social support - Availability (SSA) • Social support - Utilization (SSU) • Spiritual values (SPR) • Stress - Coping with stress (STC) • Stress – Sources (STS) • Suicidal thoughts and attempts (SUI) • Sun safety behaviours (SSB) • Smoking – Other tobacco products (TAL) • Use of protective equipment (UPE) • Voluntary organizations - Participation (ORG) • Stress - Work stress (WST) • Smoking - Youth smoking (YSM)

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

Appendix B

Appendix B - Selection of optional content by province or territory (2007-2008)

Optional Modules	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.	Nvt.
Access to health care services				●									
Alcohol use – Dependence								●		●			
Alcohol use – Former drinkers	●									●			
Alcohol use during the past week	●		●			●				●			●
Blood pressure check									●			●	
Breast examinations				●								●	
Breast self examinations	●			●									
Stress - Childhood and adult stressors												●	
Colorectal cancer screening	●	●				●		●				●	●
Consultations about mental health		●	●	●		●			●			●	●
Dental visits	●								●			●	
Depression			●	●	●				●		●		●
Diabetes care	●	●	●	●		●				●	●	●	
Dietary supplement use – Vitamins and minerals											●	●	
Distress			●		●								
Driving and safety							●		●			●	
Eye examinations						●							
Food choices		●					●		●	●		●	
Health care system satisfaction						●					●	●	
Health utility index					●								
Home care services						●							
Home safety												●	
Illicit drugs use			●							●			
Injuries			●							●			
Insurance coverage				●									

Optional Modules	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.	Nvt.
Mammography	●		●	●		●		●				●	
Mastery							●						
Maternal experiences – Alcohol use during pregnancy						●				●		●	
Maternal experiences - Smoking during pregnancy						●				●		●	●
Smoking - Nicotine dependence									●			●	●
Oral health 2	●												
PAP smear test				●		●		●				●	
Patient satisfaction – Health care services											●	●	
Patient satisfaction – Community-based care				●				●			●	●	
Problem gambling					●	●		●					●
Prostate cancer screening	●	●										●	
Psychological well-being				●									
Stress - Recent life events											●		
Satisfaction with life					●				●				●
Self-esteem													●
Sexual behaviours			●	●		●		●				●	●
Health status (SF-36)	●	●					●						
Sleep			●								●		
Smoking - Physician counselling									●				
Smoking - Stages of change						●							
Smoking cessation methods						●							
Social support - Availability			●		●					●	●		●
Social support - Utilization										●	●		●
Spiritual values								●					●
Stress – Coping with stress											●		●
Stress - Sources											●		●
Suicidal thoughts and attempts						●			●	●		●	

Optional Modules	**N.L.**	**P.E.I.**	**N.S.**	**N.B.**	**Que.**	**Ont.**	**Man.**	**Sask.**	**Alta.**	**B.C.**	**Y.T.**	**N.W.T.**	**Nvt.**
Sun safety behaviours		●											
Smoking - Other tobacco products						●			●				
Use of protective equipment			●					●	●				
Voluntary organizations - Participation			●									●	
Stress - Work stress									●				
Smoking - Youth smoking										●			

Appendix C

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

3504	Hamilton Niagara Haldimand Brant Health Integration Network
3505	Central West Health Integration Network
3506	Mississauga Halton Health Integration Network
3507	Toronto Central Health Integration Network
3508	Central Health Integration Network
3509	Central East Health Integration Network
3510	South East Health Integration Network
3511	Champlain Health Integration Network
3512	North Simcoe Muskoka Health Integration Network
3513	North East Health Integration Network
3514	North West Health Integration Network
35	Ontario by Health Unit
3526-C	District of Algoma Health Unit
3527-A	Brant County Health Unit
3530-B	Durham Regional Health Unit
3531-E	Elgin-St. Thomas Health Unit
3533-E	Grey Bruce Health Unit
3534-E	Haldimand-Norfolk Health Unit
3535-E	Haliburton, Kawartha, Pine Ridge District Health Unit
3536-B	Halton Regional Health Unit
3537-A	City of Hamilton Health Unit
3538-A	Hastings and Prince Edward Counties Health Unit
3539-E	Huron County Health Unit
3540-A	Chatham-Kent Health Unit
3541-A	Kingston, Frontenac and Lennox and Addington Health Unit
3542-A	Lambton Health Unit
3543-E	Leeds, Grenville and Lanark District Health Unit
3544-A	Middlesex-London Health Unit
3546-A	Niagara Regional Area Health Unit
3547-C	North Bay Parry Sound District Health Unit
3549-H	Northwestern Health Unit
3551-B	City of Ottawa Health Unit
3552-E	Oxford County Health Unit
3553-B	Peel Regional Health Unit
3554-E	Perth District Health Unit
3555-A	Peterborough County-City Health Unit
3556-H	Porcupine Health Unit
3557-E	Renfrew County and District Health Unit
3558-E	Eastern Ontario Health Unit
3560-E	Simcoe Muskoka District Health Unit
3561-C	Sudbury and District Health Unit
3562-C	Thunder Bay District Health Unit
3563-C	Timiskaming Health Unit
3565-B	Waterloo Health Unit
3566-B	Wellington-Dufferin-Guelph Health Unit
3568-B	Windsor-Essex County Health Unit
3570-B	York Regional Health Unit
3595-G	City of Toronto Health Unit

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

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

Appendix D

Appendix D – 2007-2008 Sample allocation by health region and frame

		Area Frame		Phone frames		Combined	
Prov./Terr.	**Health Region**	**# expected respondents**	**raw sample size**	**# expected respondents**	**raw sample size**	**# expected respondents**	**raw sample size**
Canada	Total	66507	96086	65567	110846	132074	206932
N.L.	Total	2006	2838	2006	2975	4012	5813
	1011	810	1136	810	1190	1620	2326
	1012	470	682	470	702	940	1384
	1013	426	546	426	632	852	1178
	1014	300	474	300	451	600	925
P.E.I.	Total	1241	1894	1239	2108	2480	4001
	1101	177	312	177	321	354	633
	1102	640	980	648	1088	1288	2068
	1103	424	602	414	699	838	1300
N.S.	Total	2522	3826	2522	3707	5044	7533
	1201	396	620	396	595	792	1215
	1202	320	457	320	459	640	916
	1203	360	488	360	544	720	1032
	1204	350	634	350	546	700	1180
	1205	420	606	420	611	840	1217
	1206	676	1021	676	952	1352	1973
N.B.	Total	2578	3784	2578	4433	5156	8217
	1301	500	744	500	873	1000	1617
	1302	486	765	486	862	972	1628
	1303	470	711	470	820	940	1531
	1304	270	394	270	486	540	879
	1305	250	347	250	435	500	782
	1306	346	489	346	513	692	1002
	1307	256	334	256	444	512	778
Que.	Total	11742	16659	12542	22259	24284	38918
	2401	600	792	600	934	1200	1726
	2402	628	826	628	931	1256	1757
	2403	927	1261	928	1494	1855	2755
	2404	802	1047	802	1227	1604	2274
	2405	618	951	618	980	1236	1931
	2406	1552	2241	1552	2787	3104	5028
	2407	646	996	646	1047	1292	2043
	2408	600	876	600	883	1200	1759
	2409	600	848	600	1030	1200	1878
	2410	0	0	800	3214	800	3214
	2411	600	839	600	954	1200	1793
	2412	722	973	722	1150	1444	2123
	2413	670	950	670	1099	1340	2049
	2414	718	958	718	1155	1436	2113
	2415	760	1360	758	1389	1518	2749
	2416	1300	1740	1300	1985	2600	3725
Ont.	Total	22230	31901	22230	38160	44460	70061
	3526	426	688	426	696	852	1384

		Area Frame		Phone frames		Combined	
Prov./Terr.	**Health Region**	**# expected respondents**	**raw sample size**	**# expected respondents**	**raw sample size**	**# expected respondents**	**raw sample size**
	3527	406	576	406	640	812	1216
	3530	816	1182	816	1289	1632	2471
	3531	340	458	340	563	680	1021
	3533	474	705	474	852	948	1557
	3534	386	579	386	680	772	1259
	3535	476	782	476	901	952	1683
	3536	706	921	706	1118	1412	2040
	3537	826	1203	826	1354	1652	2557
	3538	470	669	470	811	940	1480
	3539	296	400	296	505	592	905
	3540	400	487	400	659	800	1146
	3541	506	781	506	852	1012	1632
	3542	436	628	436	746	872	1374
	3543	476	699	476	778	952	1478
	3544	750	1163	750	1256	1500	2420
	3546	766	1009	766	1268	1532	2277
	3547	400	601	400	732	800	1333
	3549	328	535	328	722	656	1257
	3551	1026	1488	1026	1597	2052	3084
	3552	376	516	376	542	752	1058
	3553	1314	1817	1314	2345	2628	4162
	3554	326	445	326	490	652	935
	3555	426	652	426	772	852	1424
	3556	376	558	376	602	752	1159
	3557	376	547	376	659	752	1206
	3558	520	711	520	799	1040	1510
	3560	1142	1748	1142	2065	2284	3813
	3561	540	755	540	924	1080	1679
	3562	717	1006	717	1246	1434	2252
	3563	250	373	250	410	500	782
	3565	766	1049	766	1150	1532	2199
	3566	564	746	564	886	1128	1632
	3568	716	976	716	1193	1432	2169
	3570	942	1213	942	1690	1884	2903
	3595	2169	3233	2169	4371	4338	7604
Man.	Total	3754	5188	3754	5658	7508	10846
	4610	1056	1413	1056	1494	2112	2907
	4615	280	391	280	434	560	825
	4620	250	364	250	480	500	844
	4625	300	409	300	442	600	851
	4630	346	566	346	526	692	1092
	4640	400	503	400	544	800	1047
	4645	356	490	356	504	712	994
	4660	266	377	266	386	532	763
	4670	250	355	250	448	500	803
	4685	250	319	250	400	500	719

		Area Frame		Phone frames		Combined	
Prov./Terr.	**Health Region**	**# expected respondents**	**raw sample size**	**# expected respondents**	**raw sample size**	**# expected respondents**	**raw sample size**
Sask.	Total	3612	5161	4112	7780	7724	12941
	4701	300	387	300	440	600	827
	4702	300	423	300	456	600	879
	4703	266	388	266	396	532	784
	4704	620	871	620	872	1240	1743
	4705	310	446	310	458	620	904
	4706	660	889	660	948	1320	1837
	4707	270	456	270	376	540	832
	4708	260	389	260	360	520	749
	4709	326	522	326	518	652	1040
	4710	300	391	300	454	600	845
	4714	0	0	500	2502	500	2502
Alta.	Total	6104	8750	6104	9284	12208	18034
	4820	510	699	510	762	1020	1461
	4821	416	509	416	600	832	1109
	4822	1396	2048	1396	2072	2792	4120
	4823	700	1024	700	1058	1400	2082
	4824	446	595	446	666	892	1261
	4825	1310	1901	1310	1926	2620	3827
	4826	540	830	540	890	1080	1720
	4827	466	669	466	752	932	1421
	4828	320	475	320	558	640	1033
B.C.	Total	8048	11864	8050	13250	16098	25114
	5911	304	422	304	490	608	912
	5912	310	468	310	488	620	956
	5913	590	742	590	958	1180	1700
	5914	500	761	500	816	1000	1577
	5921	520	701	520	824	1040	1525
	5922	760	1018	760	1280	1520	2298
	5923	798	1161	800	1314	1598	2475
	5931	426	595	426	734	852	1329
	5932	800	1292	800	1500	1600	2792
	5933	546	885	546	894	1092	1779
	5941	676	955	676	1068	1352	2023
	5942	526	766	526	798	1052	1564
	5943	264	363	264	400	528	763
	5951	326	587	326	586	652	1173
	5952	426	663	426	646	852	1309
	5903	276	484	276	454	552	938
Terr.	60	950	1459	250	679	1200	2138
	61	1020	1631	180	553	1200	2184
	62	700	1132	0	0	700	1132

Sample allocation by Local Health Integrated Network and frames in the CCHS 2007-2008 in Ontario.

		Area Frame		Phone frames		Combined	
Prov./Terr.	**LHIN**	**# expected respondents**	**raw sample size**	**# expected respondents**	**raw sample size**	**# expected respondents**	**raw sample size**
Ont.	Total	22230	31900	22230	38161	44460	70061
	3501	1552	2092	1552	2598	3104	4689
	3502	2568	3696	2568	4217	5136	7913
	3503	1230	1659	1230	1886	2460	3545
	3504	2624	3671	2624	4308	5248	7979
	3505	1062	1502	1062	1934	2124	3436
	3506	1136	1534	1136	1945	2272	3480
	3507	1085	1641	1085	2117	2170	3758
	3508	1404	1893	1404	2676	2808	4569
	3509	2076	3136	2076	3694	4152	6829
	3510	1344	2003	1344	2270	2688	4273
	3511	2058	2939	2058	3280	4116	6219
	3512	1058	1626	1058	1914	2116	3540
	3513	1992	2975	1992	3363	3984	6338
	3514	1041	1534	1041	1959	2082	3493

Appendix E

Appendix E – 2007-2008 response rate by health region and frames

		Area frame							Phone frames							
Prov. Terr.	**Health Region**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**Combined resp. rates**
Canada	**Total**	81837	70879	86.6	70879	66114	93.3	80.8	90872	73623	81.0	73623	65845	89.4	72.5	76.4
N.L.	**Total**	2387	2181	91.4	2181	2022	92.7	84.7	2624	2307	87.9	2307	2076	90.0	79.1	81.8
	1011	972	888	91.4	888	827	93.1	85.1	1032	903	87.5	903	807	89.4	78.2	81.5
	1012	540	500	92.6	500	463	92.6	85.7	645	568	88.1	568	512	90.1	79.4	82.3
	1013	455	431	94.7	431	413	95.8	90.8	563	497	88.3	497	450	90.5	79.9	84.8
	1014	420	362	86.2	362	319	88.1	76.0	384	339	88.3	339	307	90.6	79.9	77.9
P.E.I.	**Total**	1454	1264	86.9	1264	1184	93.7	81.4	1571	1314	83.6	1314	1183	90.0	75.3	78.2
	1101	159	134	84.3	134	126	94.0	79.2	125	103	82.4	103	91	88.3	72.8	76.4
	1102	880	749	85.1	749	700	93.5	79.5	957	799	83.5	799	727	91.0	76.0	77.7
	1103	415	381	91.8	381	358	94.0	86.3	489	412	84.3	412	365	88.6	74.6	80.0
N.S.	**Total**	3113	2752	88.4	2752	2540	92.3	81.6	3289	2843	86.4	2843	2612	91.9	79.4	80.5
	1201	460	438	95.2	438	415	94.7	90.2	522	452	86.6	452	419	92.7	80.3	84.9
	1202	390	341	87.4	341	316	92.7	81.0	384	336	87.5	336	314	93.5	81.8	81.4
	1203	429	381	88.8	381	359	94.2	83.7	445	374	84.0	374	347	92.8	78.0	80.8
	1204	438	382	87.2	382	341	89.3	77.9	471	409	86.8	409	375	91.7	79.6	78.8
	1205	530	490	92.5	490	459	93.7	86.6	545	467	85.7	467	425	91.0	78.0	82.2
	1206	866	720	83.1	720	650	90.3	75.1	922	805	87.3	805	732	90.9	79.4	77.3
N.B.	**Total**	3060	2707	88.5	2707	2549	94.2	83.3	3737	3221	86.2	3221	2960	91.9	79.2	81.1
	1301	623	539	86.5	539	495	91.8	79.5	736	622	84.5	622	582	93.6	79.1	79.2
	1302	575	492	85.6	492	470	95.5	81.7	722	621	86.0	621	568	91.5	78.7	80.0
	1303	536	480	89.6	480	449	93.5	83.8	664	577	86.9	577	523	90.6	78.8	81.0
	1304	343	297	86.6	297	283	95.3	82.5	395	348	88.1	348	320	92.0	81.0	81.7

		Area frame							Phone frames							
	1305	282	251	89.0	251	239	95.2	84.8	384	329	85.7	329	299	90.9	77.9	80.8
	1306	423	399	94.3	399	376	94.2	88.9	461	400	86.8	400	368	92.0	79.8	84.2
Prov. Terr.	**Health Region**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**Combined resp. rates**
	1307	278	249	89.6	249	237	95.2	85.3	375	324	86.4	324	300	92.6	80.0	82.2
Que.	**Total**	14508	12663	87.3	12663	11958	94.4	82.4	17405	13965	80.2	13965	12485	89.4	71.7	76.6
	2401	681	646	94.9	646	627	97.1	92.1	775	664	85.7	664	606	91.3	78.2	84.7
	2402	721	638	88.5	638	595	93.3	82.5	785	666	84.8	666	597	89.6	76.1	79.2
	2403	1207	1031	85.4	1031	963	93.4	79.8	1299	1039	80.0	1039	930	89.5	71.6	75.5
	2404	934	831	89.0	831	792	95.3	84.8	1130	949	84.0	949	862	90.8	76.3	80.1
	2405	794	657	82.7	657	619	94.2	78.0	803	666	82.9	666	609	91.4	75.8	76.9
	2406	1952	1636	83.8	1636	1526	93.3	78.2	2475	1797	72.6	1797	1549	86.2	62.6	69.5
	2407	855	703	82.2	703	659	93.7	77.1	872	693	79.5	693	626	90.3	71.8	74.4
	2408	769	721	93.8	721	687	95.3	89.3	739	644	87.1	644	595	92.4	80.5	85.0
	2409	724	647	89.4	647	618	95.5	85.4	792	649	81.9	649	580	89.4	73.2	79.0
	2410	.	.	.	.	.	.	.	1224	1010	82.5	1010	898	88.9	73.4	73.4
	2411	675	625	92.6	625	604	96.6	89.5	779	649	83.3	649	569	87.7	73.0	80.7
	2412	824	764	92.7	764	734	96.1	89.1	922	749	81.2	749	680	90.8	73.8	81.0
	2413	885	739	83.5	739	680	92.0	76.8	1044	790	75.7	790	678	85.8	64.9	70.4
	2414	890	764	85.8	764	719	94.1	80.8	977	774	79.2	774	702	90.7	71.9	76.1
	2415	971	863	88.9	863	803	93.0	82.7	1018	810	79.6	810	725	89.5	71.2	76.8
	2416	1626	1398	86.0	1398	1332	95.3	81.9	1771	1416	80.0	1416	1279	90.3	72.2	76.9
Ont.	**Total**	27904	23784	85.2	23784	21914	92.1	78.5	31813	25065	78.8	25065	22044	87.9	69.3	73.6
	3526	609	512	84.1	512	478	93.4	78.5	564	477	84.6	477	417	87.4	73.9	76.3
	3527	534	467	87.5	467	406	86.9	76.0	536	435	81.2	435	391	89.9	72.9	74.5

		Area frame							Phone frames							
	3530	1049	879	83.8	879	795	90.4	75.8	1158	928	80.1	928	810	87.3	69.9	72.7
	3531	398	339	85.2	339	313	92.3	78.6	447	356	79.6	356	325	91.3	72.7	75.5
	3533	534	476	89.1	476	459	96.4	86.0	636	515	81.0	515	473	91.8	74.4	79.7
	3534	487	425	87.3	425	376	88.5	77.2	554	432	78.0	432	387	89.6	69.9	73.3
	3535	649	555	85.5	555	525	94.6	80.9	643	540	84.0	540	489	90.6	76.0	78.5
	3536	859	736	85.7	736	668	90.8	77.8	984	792	80.5	792	684	86.4	69.5	73.4
Prov. Terr.	**Health Region**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**Combined resp. rates**
	3537	1055	862	81.7	862	776	90.0	73.6	1153	876	76.0	876	764	87.2	66.3	69.7
	3538	585	505	86.3	505	459	90.9	78.5	639	531	83.1	531	465	87.6	72.8	75.5
	3539	392	366	93.4	366	356	97.3	90.8	437	351	80.3	351	310	88.3	70.9	80.3
	3540	434	414	95.4	414	392	94.7	90.3	569	446	78.4	446	394	88.3	69.2	78.4
	3541	655	537	82.0	537	488	90.9	74.5	649	528	81.4	528	477	90.3	73.5	74.0
	3542	484	433	89.5	433	402	92.8	83.1	573	445	77.7	445	389	87.4	67.9	74.8
	3543	562	502	89.3	502	470	93.6	83.6	639	541	84.7	541	494	91.3	77.3	80.3
	3544	1032	796	77.1	796	742	93.2	71.9	1100	886	80.5	886	775	87.5	70.5	71.2
	3546	866	769	88.8	769	718	93.4	82.9	1080	857	79.4	857	758	88.4	70.2	75.8
	3547	499	431	86.4	431	388	90.0	77.8	563	465	82.6	465	419	90.1	74.4	76.0
	3549	397	347	87.4	347	314	90.5	79.1	484	388	80.2	388	339	87.4	70.0	74.1
	3551	1366	1086	79.5	1086	978	90.1	71.6	1387	1123	81.0	1123	989	88.1	71.3	71.4
	3552	473	427	90.3	427	395	92.5	83.5	471	400	84.9	400	353	88.3	74.9	79.2
	3553	1714	1458	85.1	1458	1339	91.8	78.1	2024	1530	75.6	1530	1281	83.7	63.3	70.1
	3554	416	375	90.1	375	348	92.8	83.7	439	366	83.4	366	332	90.7	75.6	79.5
	3555	533	446	83.7	446	411	92.2	77.1	552	451	81.7	451	417	92.5	75.5	76.3
	3556	502	443	88.2	443	380	85.8	75.7	494	406	82.2	406	379	93.3	76.7	76.2

		Area frame							Phone frames							
	3557	427	398	93.2	398	362	91.0	84.8	508	409	80.5	409	376	91.9	74.0	78.9
	3558	656	568	86.6	568	537	94.5	81.9	688	576	83.7	576	498	86.5	72.4	77.0
	3560	1507	1276	84.7	1276	1190	93.3	79.0	1641	1353	82.4	1353	1206	89.1	73.5	76.1
	3561	621	533	85.8	533	494	92.7	79.5	743	606	81.6	606	557	91.9	75.0	77.1
	3562	830	704	84.8	704	656	93.2	79.0	989	809	81.8	809	726	89.7	73.4	76.0
	3563	248	224	90.3	224	213	95.1	85.9	336	268	79.8	268	231	86.2	68.8	76.0
	3565	930	778	83.7	778	705	90.6	75.8	1008	811	80.5	811	731	90.1	72.5	74.1
	3566	684	627	91.7	627	584	93.1	85.4	773	637	82.4	637	581	91.2	75.2	80.0
	3568	842	718	85.3	718	679	94.6	80.6	1047	822	78.5	822	703	85.5	67.1	73.2
Prov. Terr.	**Health Region**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**Combined resp. rates**
	3570	1128	970	86.0	970	897	92.5	79.5	1508	1108	73.5	1108	930	83.9	61.7	69.3
	3595	2947	2402	81.5	2402	2221	92.5	75.4	3797	2601	68.5	2601	2194	84.4	57.8	65.5
Man.	**Total**	4452	4000	89.8	4000	3789	94.7	85.1	4674	4042	86.5	4042	3731	92.3	79.8	82.4
	4610	1322	1132	85.6	1132	1073	94.8	81.2	1342	1129	84.1	1129	1039	92.0	77.4	79.3
	4615	342	305	89.2	305	281	92.1	82.2	365	312	85.5	312	284	91.0	77.8	79.9
	4620	310	282	91.0	282	271	96.1	87.4	352	312	88.6	312	299	95.8	84.9	86.1
	4625	353	329	93.2	329	314	95.4	89.0	372	336	90.3	336	299	89.0	80.4	84.6
	4630	414	375	90.6	375	351	93.6	84.8	400	347	86.8	347	318	91.6	79.5	82.2
	4640	452	417	92.3	417	394	94.5	87.2	491	434	88.4	434	405	93.3	82.5	84.7
	4645	410	379	92.4	379	362	95.5	88.3	416	371	89.2	371	350	94.3	84.1	86.2
	4660	288	268	93.1	268	264	98.5	91.7	325	277	85.2	277	255	92.1	78.5	84.7
	4670	275	259	94.2	259	240	92.7	87.3	325	278	85.5	278	264	95.0	81.2	84.0
	4685	286	254	88.8	254	239	94.1	83.6	286	246	86.0	246	218	88.6	76.2	79.9
Sask.	**Total**	4220	3850	91.2	3850	3653	94.9	86.6	5383	4575	85.0	4575	4166	91.1	77.4	81.4

		Area frame							Phone frames							
	4701	341	325	95.3	325	307	94.5	90.0	349	307	88.0	307	278	90.6	79.7	84.8
	4702	339	313	92.3	313	297	94.9	87.6	402	333	82.8	333	303	91.0	75.4	81.0
	4703	308	280	90.9	280	264	94.3	85.7	362	284	78.5	284	264	93.0	72.9	78.8
	4704	786	703	89.4	703	663	94.3	84.4	876	736	84.0	736	667	90.6	76.1	80.0
	4705	305	284	93.1	284	273	96.1	89.5	351	291	82.9	291	276	94.8	78.6	83.7
	4706	822	742	90.3	742	701	94.5	85.3	946	797	84.2	797	723	90.7	76.4	80.5
	4707	259	239	92.3	239	230	96.2	88.8	335	292	87.2	292	262	89.7	78.2	82.8
	4708	306	283	92.5	283	272	96.1	88.9	259	216	83.4	216	199	92.1	76.8	83.4
	4709	405	357	88.1	357	339	95.0	83.7	511	439	85.9	439	403	91.8	78.9	81.0
	4710	349	324	92.8	324	307	94.8	88.0	344	306	89.0	306	279	91.2	81.1	84.6
	4714	.	.	.	.	.	.	.	648	574	88.6	574	512	89.2	79.0	79.0
Alta.	**Total**	7448	6323	84.9	6323	5836	92.3	78.4	8372	6825	81.5	6825	6089	89.2	72.7	75.4
	4821	608	556	91.4	556	520	93.5	85.5	664	542	81.6	542	495	91.3	74.5	79.8
Prov. Terr.	**Health Region**	207	183	88.4	183	169	92.3	81.6	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**Combined resp. rates**
	4822	420	381	90.7	381	359	94.2	85.5	529	450	85.1	450	408	90.7	77.1	80.8
	4823	1818	1503	82.7	1503	1369	91.1	75.3	1856	1528	82.3	1528	1366	89.4	73.6	74.4
	4824	895	755	84.4	755	688	91.1	76.9	951	807	84.9	807	709	87.9	74.6	75.7
	4825	521	474	91.0	474	445	93.9	85.4	604	504	83.4	504	453	89.9	75.0	79.8
	4826	1645	1333	81.0	1333	1240	93.0	75.4	1965	1549	78.8	1549	1374	88.7	69.9	72.4
	4827	613	548	89.4	548	512	93.4	83.5	711	573	80.6	573	499	87.1	70.2	76.4
	4828	569	494	86.8	494	449	90.9	78.9	665	531	79.8	531	479	90.2	72.0	75.2
	4829	359	279	77.7	279	254	91.0	70.8	427	341	79.9	341	306	89.7	71.7	71.2
B.C.	**Total**	9972	8377	84.0	8377	7848	93.7	78.7	11450	8977	78.4	8977	8055	89.7	70.3	74.2
	5911	353	323	91.5	323	307	95.0	87.0	455	354	77.8	354	321	90.7	70.5	77.7

		Area frame							Phone frames							
	5912	364	330	90.7	330	316	95.8	86.8	413	322	78.0	322	295	91.6	71.4	78.6
	5913	702	629	89.6	629	607	96.5	86.5	859	681	79.3	681	615	90.3	71.6	78.3
	5914	607	533	87.8	533	503	94.4	82.9	657	542	82.5	542	488	90.0	74.3	78.4
	5921	590	516	87.5	516	478	92.6	81.0	707	585	82.7	585	535	91.5	75.7	78.1
	5922	897	726	80.9	726	688	94.8	76.7	1105	867	78.5	867	767	88.5	69.4	72.7
	5923	997	873	87.6	873	800	91.6	80.2	1165	907	77.9	907	799	88.1	68.6	74.0
	5931	507	441	87.0	441	423	95.9	83.4	639	483	75.6	483	432	89.4	67.6	74.6
	5932	1092	829	75.9	829	793	95.7	72.6	1243	894	71.9	894	790	88.4	63.6	67.8
	5933	639	543	85.0	543	500	92.1	78.2	769	574	74.6	574	515	89.7	67.0	72.1
	5941	865	723	83.6	723	668	92.4	77.2	883	704	79.7	704	644	91.5	72.9	75.1
	5942	638	549	86.1	549	518	94.4	81.2	702	576	82.1	576	529	91.8	75.4	78.1
	5943	314	259	82.5	259	250	96.5	79.6	373	321	86.1	321	294	91.6	78.8	79.2
	5951	460	350	76.1	350	324	92.6	70.4	485	365	75.3	365	322	88.2	66.4	68.4
	5952	607	486	80.1	486	433	89.1	71.3	616	495	80.4	495	438	88.5	71.1	71.2
	5953	340	267	78.5	267	240	89.9	70.6	379	307	81.0	307	271	88.3	71.5	71.1
Y.T.	**6001**	1211	1073	88.6	1073	1013	94.4	83.6	325	289	88.9	289	262	90.7	80.6	83.0
Prov. Terr.	**Health Region**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**Combined resp. rates**
N.W.T.	**6101**	1219	1100	90.2	1100	1049	95.4	86.1	229	200	87.3	200	182	91.0	79.5	85.0
Nvt.	**6201**	889	805	90.6	805	759	94.3	85.4	.	.	.	.	.	.	.	85.4

Reponse rate by Local Health Integrated Network (LHIN) and frames in the CCHS 2007-2008 in Ontario

		Area frame / Base aréolaire							Phone frames / Bases téléphoniques							
Prov. Terr.	**LHIN**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**# in scope HH**	**# resp. HH**	**HH resp. rates**	**# pers. select.**	**# resp.**	**Pers. resp. rates**	**Resp. rates**	**Combined resp. rates**
Ont.	**Total**	27904	23784	85.2	23784	21914	92.1	78.5	31813	25065	78.8	25065	22044	87.9	69.3	73.6
	3501	1760	1565	88.9	1565	1473	94.1	83.7	2189	1713	78.3	1713	1486	86.7	67.9	74.9
	3502	3244	2781	85.7	2781	2613	94.0	80.5	3504	2847	81.3	2847	2543	89.3	72.6	76.4
	3503	1500	1305	87.0	1305	1202	92.1	80.1	1657	1338	80.7	1338	1220	91.2	73.6	76.7
	3504	3233	2764	85.5	2764	2501	90.5	77.4	3644	2860	78.5	2860	2533	88.6	69.5	73.2
	3505	1393	1178	84.6	1178	1088	92.4	78.1	1655	1252	75.6	1252	1038	82.9	62.7	69.8
	3506	1446	1250	86.4	1250	1127	90.2	77.9	1728	1334	77.2	1334	1122	84.1	64.9	70.9
	3507	1453	1137	78.3	1137	1043	91.7	71.8	1794	1216	67.8	1216	1053	86.6	58.7	64.6
	3508	1753	1497	85.4	1497	1392	93.0	79.4	2410	1721	71.4	1721	1443	83.8	59.9	68.1
	3509	2751	2324	84.5	2324	2149	92.5	78.1	3019	2385	79.0	2385	2099	88.0	69.5	73.6
	3510	1650	1409	85.4	1409	1293	91.8	78.4	1752	1454	83.0	1454	1297	89.2	74.0	76.1
	3511	2624	2209	84.2	2209	2021	91.5	77.0	2785	2277	81.8	2277	2024	88.9	72.7	74.8
	3512	1391	1171	84.2	1171	1089	93.0	78.3	1503	1249	83.1	1249	1118	89.5	74.4	76.3
	3513	2479	2143	86.4	2143	1953	91.1	78.8	2700	2222	82.3	2222	2003	90.1	74.2	76.4
	3514	1227	1051	85.7	1051	970	92.3	79.1	1473	1197	81.3	1197	1065	89.0	72.3	75.4

APPENDIX F: Health regions

Health regions for dissemination, before PUMF		Health regions for PUMF	
Newfoundland		**Newfoundland**	
1011	Eastern Health Authority	1011	Eastern Health Authority
1012	Health and Community Services Central Region	1012	Health and Community Services Central Region
1013	Health and Community Services Western Region	1013	Western Region/Labrador-Grenfell
1014	Labrador-Grenfell Health Authority		
Prince Edward Island		**Prince Edward Island**	
1101	West Prince Health Region	1101	Prince Edward Island
1102	East Prince Health Region		
1104	Kings Health Region		
1103	Queens Health Region		
Nova Scotia		**Nova Scotia**	
1201	Zone 1	1201	Zone 1
1202	Zone 2	1202	Zone 2
1203	Zone 3	1203	Zone 3
1204	Zone 4	1204	Zone 4
1205	Zone 5	1205	Zone 5
1206	Zone 6	1206	Zone 6
New Brunswick		**New Brunswick**	
1301	Region 1	1301	Region 1
1302	Region 2	1302	Region 2
1303	Region 3	1303	Region 3
1304	Region 4	1304	Region 4/5
1305	Region 5		
1306	Region 6	1306	Region 6/7
1307	Region 7		
Quebec		**Quebec**	
2401	Région du Bas-Saint-Laurent	2401	Région du Bas-Saint-Laurent
2402	Région du Saguenay - Lac-Saint-Jean	2402	Région du Saguenay - Lac-Saint-Jean
2403	Région de Québec	2403	Région de Québec
2404	Région de la Mauricie et du Centre-du-Québec	2404	Région de la Mauricie et du Centre-du-Québec
2405	Région de l'Estrie	2405	Région de l'Estrie
2406	Région de Montréal-Centre	2406	Région de Montréal-Centre
2407	Région de l'Outaouais	2407	Région de l'Outaouais
2408	Région de l'Abitibi-Témiscamingue	2408	Région de l'Abitibi-Témiscamingue
2409	Région de la Côte-Nord	2409	Région de la Côte-Nord
2410	Région du Nord-du-Québec		
2411	Région de la Gaspésie - Îles-de-la-Madeleine	2411	Région de la Gaspésie - Îles-de-la-Madeleine
2412	Région de la Chaudière-Appalaches	2412	Région de la Chaudière-Appalaches
2413	Région de Laval	2413	Région de Laval
2414	Région de Lanaudière	2414	Région de Lanaudière
2415	Région des Laurentides	2415	Région des Laurentides
2416	Région de la Montérégie	2416	Région de la Montérégie

Health regions for dissemination, before PUMF		Health regions for PUMF	
Ontario		**Ontario**	
3526	Algoma	3526	Algoma
3527	Brant	3527	Brant
3530	Durham	3530	Durham
3531	Elgin-St. Thomas	3531	Elgin-St. Thomas
3533	Grey Bruce	3533	Grey Bruce
3534	Haldimand-Norfolk	3534	Haldimand-Norfolk
3535	Haliburton, Kawartha, Pine Ridge	3535	Haliburton, Kawartha, Pine Ridge
3536	Halton	3536	Halton
3537	Hamilton	3537	Hamilton
3538	Hastings and Prince Edward	3538	Hastings and Prince Edward
3539	Huron	3539	Huron/Perth
3554	Perth		
3540	Chatham-Kent	3540	Chatham-Kent
3541	Kingston, Frontenac and Lennox and Addington	3541	Kingston, Frontenac and Lennox and Addington
3542	Lambton	3542	Lambton
3543	Leeds, Grenville and Lanark	3543	Leeds, Grenville and Lanark
3544	Middlesex-London	3544	Middlesex-London
3546	Niagara	3546	Niagara
3547	North Bay Parry Sound	3547	North Bay Parry Sound/Timiskaming
3563	Timiskaming		
3549	Northwestern	3549	Northwestern
3551	Ottawa	3551	Ottawa
3552	Oxford	3552	Oxford
3553	Peel	3553	Peel
3555	Peterborough	3555	Peterborough
3556	Porcupine	3556	Porcupine
3557	Renfrew	3557	Renfrew
3558	Eastern Ontario	3558	Eastern Ontario
3560	Simcoe Muskoka	3560	Simcoe Muskoka
3561	Sudbury	3561	Sudbury
3562	Thunder Bay	3562	Thunder Bay
3565	Waterloo	3565	Waterloo
3566	Wellington-Dufferin-Guelph	3566	Wellington-Dufferin-Guelph
3568	Windsor-Essex	3568	Windsor-Essex
3570	York	3570	York
3595	City of Toronto	3595	City of Toronto
Manitoba		**Manitoba**	
4610	Winnipeg RHA	4610	Winnipeg RHA
4615	Brandon RHA	4615	Brandon/Assiniboine
4645	Assiniboine RHA		
4620	North Eastman RHA	4620	North Eastman/South Eastman
4625	South Eastman RHA		
4630	Interlake RHA	4630	Interlake RHA
4640	Central RHA	4640	Central RHA
4660	Parkland RHA	4660	Parkland/Norman/Burntwood-Churchill
4670	Norman RHA		
4685	Burntwood RHA/Churchill RHA		

Health regions for dissemination, before PUMF		Health regions for PUMF	
Saskatchewan		**Saskatchewan**	
4701	Sun Country RHA	4701	Sun Country/Five Hills/Cypress
4702	Five Hills RHA		
4703	Cypress RHA		
4704	Regina Qu'Appelle RHA	4704	Regina Qu'Appelle RHA
4705	Sunrise RHA	4705	Sunrise/Kelsey Trail
4708	Kelsey trail RHA		
4706	Saskatoon RHA	4706	Saskatoon RHA
4707	Heartland RHA	4707	Heartland/Prairie North
4710	Prairie North RHA		
4709	Prince Albert Parkland RHA	4709	Prince Albert Parkland/Mamawetan-Keewatin Yatthé-Athabaska
4714	Mamawetan Churchill River RHA/Keewatin Yatthé RHA/Athabaska HA		
Alberta		**Alberta**	
4821	Chinook Regional Health Authority	4821	Chinook Regional Health Authority
4822	Palliser Health Region	4822	Palliser Health Region
4823	Calgary Health Region	4823	Calgary Health Region
4824	David Thompson Regional Health Authority	4824	David Thompson Regional Health Authority
4825	East Central Health	4825	East Central Health
4826	Capital Health	4826	Capital Health
4827	Aspen Regional Health Authority	4827	Aspen Regional Health Authority
4828	Peace Country Health	4828	Peace Country/Northern Lights
4829	Northern Lights Health Region	4829	Northern Lights Health Region
British Columbia		**British Columbia**	
5911	East Kootenay	5911	East Kootenay
5912	Kootenay-Boundary	5912	Kootenay-Boundary
5913	Okanagan	5913	Okanagan
5914	Thompson/Cariboo	5914	Thompson/Cariboo
5921	Fraser East	5921	Fraser East
5922	Fraser North	5922	Fraser North
5923	Fraser South	5923	Fraser South
5931	Richmond	5931	Richmond
5932	Vancouver	5932	Vancouver
5933	North Shore/Coast Garibaldi	5933	North Shore/Coast Garibaldi
5941	South Vancouver Island	5941	South Vancouver Island
5942	Central Vancouver Island	5942	Central Vancouver Island
5943	North Vancouver Island	5943	North Vancouver Island
5951	Northwest	5951	Northwest
5953	Northeast	5953	Northeast
5952	Northern Interior	5952	Northern Interior
Territories		**Territories**	
6001	Yukon Territory	6001	Yukon Territory/Northwest Territories/Nunavut
6101	Northwest Territories		
6201	Nunavut		