Canadian Community Health Survey (CCHS)

Annual component

User guide

2019-2020 Microdata file

April 2022

Statistics Canada Statistique Canada

Canada

TABLE OF CONTENT

WHAT'S NEW IN THE 2020 CANADIAN COMMUNITY HEALTH SURVEY (CCHS)?

The COVID-19 pandemic had major impacts on the data collection operations for CCHS 2020. The collection was stopped mid-March, towards the end of the first collection period, and did not resume until September. The second, third and fourth quarterly samples were collected during very short collection periods, each of about five weeks, from September to December. The impossibility of conducting in-person interviews, the shorter collection periods and collection capacity issues resulted in a significant decrease in the response rates.

As for previous CCHS cycles, survey weights were adjusted to minimise any potential bias that could arise from survey non-response; non-response adjustments and calibration using available auxiliary information were applied and are reflected in the survey weights provided with the data file. Extensive validations of survey estimates were also performed and examined from a bias analysis perspective. Despite these rigorous adjustments and validations, the high non-response increases the risk of a remaining bias and the magnitude with which such a bias could impact estimates produced using the survey data. Therefore, users are advised to use the CCHS 2020 data with caution, especially when creating estimates for small sub-populations or when comparing to other CCHS years.

Content

The following are some of the modifications that were made to the Canadian Community Health Survey (CCHS) – Annual component questionnaire in 2020.

- **Childhood Experiences (CEX), Consultations about mental health (CMH), Health Insurance Coverage (INS), Patient Experiences (PEX), Social Provisions (SPS), Sources of Stress (STS)** and **Contact with Dental Professionals (CP3)** were offered as optional modules in the provinces (they were theme in 2019) and kept as a two-year theme modules (2019-2020) in the territories.
- **Food Guide Use (FGU)**, **Physical activities - adults 18 years and older (PAA)**, **Physical activities for youth (PAY)** and **Sexual Behaviors (SXB)** were asked as theme modules in the provinces only (they were optional in 2019).
- **Fruit and vegetable consumption (FVC)** was asked as a theme module in the provinces (it was optional in 2019) and kept as optional in the territories.
- **Loss of Productivity (LOP)** was added as a theme module in the provinces only in 2020.
- **Barriers to physical activity (BPA)** and **Pain relief medication (PRM)** were added as optional modules.
- A new module on **COVID-19 (COV)** was added as a theme module for the 2020 - September to December collection period in the provinces only. See section 4.6 for important note about data utilisation.
- The module **Food Security (FSC)** was part of the optional content component for 2019-2020 with all provinces and territories selecting it except for British Columbia. In order to support analysis of impacts from the pandemic on food insecurity, this module was added to the

collection plan for September to December 2020. See section 4.3 for an important note about data utilisation.

For more information on CCHS content, please see Appendix A for a selection of optional content by provinces and territories.

Collection

Due to the COVID-19 pandemic, collection for the 2020 collection year came to a halt in March. Collection resumed again in September until December under strict safety protocols and best practices. All interviewers, including those previously working in centralised collection centres, worked from home. Also in-person interviews were suspended and only conducted over the telephone.

Sampling

The coverage in the territories was exceptionally limited to the capitals (Whitehorse, Yellowknife and Iqaluit) during the last three collection periods of 2020 given that in-person interviews were not possible for most of the year due to the COVID-19 pandemic.

Documentation

- In order to reflect the July 2016 federal government changes to the Canadian Child Tax Benefit (CCTB) we now refer to the current system Canadian Child Benefit (CCB) throughout the documentation
- Errata: Errors that are discovered in the CCHS data files and products after dissemination are communicated to users in a cumulative error log called “CCHS Errata". The errata provides details on:
    - the products affected (e.g. master or share files or PUMF);
    - the years affected;
    - suggested corrections (if possible); and
    - steps to carry out to apply the correction

  Since the release of the 2019 CCHS, there have been 0 additional entries to the errata documents.
  For your convenience, the errata document has an index on the first page. It lists modules affected and the years affected. To see if a particular variable is affected, follow the hyperlinks on the associated errata item numbers.

  An up to date copy of the CCHS Errata can be obtained by contacting the Centre for Population Health Data at 613-951-1746 or statcan.hd-ds.statcan@canada.ca.

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. The CCHS underwent a major redesign that started in 2012 and was implemented in collection in 2015. Details of this redesign are provided in Chapter 3.

The survey's objectives are as follows:

- support health surveillance programs by providing health data at the national, provincial and health region 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.

The CCHS data is collected from persons aged 12 and over living in private dwellings in over 100 health regions covering all provinces and territories. Excluded from the sampling frame are individuals living on Indian Reserves and on Crown Lands, institutional residents (health institutions, prisons, religious institutions, convents, etc.), full-time members of the Canadian Forces, youth aged 12 to 17 living in foster homes, and residents of certain remote regions. The CCHS covers approximately 98% of the Canadian population aged 12 and over.

The purpose of this document is to facilitate the manipulation of the CCHS microdata files and to describe the methodology used. The CCHS produces three types of microdata files: master files, share files and public use microdata files (PUMF). The characteristics of each of these files are presented in this guide. The PUMF is released after two years of data collection and contains two years of data. The next PUMF file (release date to be determined) will include the data collected for the years 2019 and 2020.
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, Centre for Population Health Data:	613-951-1746
E-mail:	statcan.HD-DS.statcan@canada.ca
For remote access support:	613-951-1746
E-mail:	statcan.cchssm-esccds.statcan@canada.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 in 1999. 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] The survey started data collection in 2000.

To meet many data requirements, the CCHS had a two-year data collection cycle. Up to 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 focused on specific health topics.

Starting in 2007, the regional component of the CCHS program began collection on an ongoing basis. The x.1 cycles of the CCHS were renamed "the annual component" of the CCHS and are now collected every year. The full title of a given annual cycle is "The Canadian Community Health Survey – Annual component, 20XX" and the short title is simply "CCHS –20XX".

The focused content component of the survey have been 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”).

In 2012, a major redesign began which changed the survey sampling and collection strategy. Survey content was qualitatively tested to ensure validity and reliability in preparation for a new 8 year content plan beginning in 2015.

3. CCHS REDESIGN IN 2015

In 2012, the redesign began with extensive consultations across Canada with key experts and federal, provincial and health region stakeholders. The goal was to gather and input proposed changes and collect detailed information on the data requirements and products of the various partners.

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

Below are the main changes resulting from the CCHS redesign:

- Content changes were made to most modules, but mainly to the core and theme modules. Roughly 70% of the modules that existed before the redesign have seen some level of change. Those changes vary, and have involved either minor tweaks or major changes to concepts, vocabulary, or response categories. Some new modules were created to allow for emerging concepts or data gaps identified by data sharing partners such as Health Canada, the Public Health Agency of Canada and the provincial and territorial ministries of health.

- Variable conventions have changed in 2015. The naming convention used for variables in the 2015 CCHS use a maximum of eight characters. As before, positions 1 to 3 make up the module or questionnaire section name, while position 4 indicates the variable type (underscore, D, F, C or G). Survey elements in the questionnaire (e.g. E, C, D, Q, N) now follow a uniform convention that increases in increments of five (e.g. CCC_005, CCC_010, CCC_015).

- All the derived variables were reviewed and specifications were updated with the new variable names. Content and conceptual changes were also reviewed and new derived variables were created for new content modules. The annual data collection strategy changed for 2015. Prior to 2015, there were six two-month collection periods. Starting in 2015, survey collection was divided into four non-overlapping three-month periods.

- Starting in 2015, changes to data collection in the North resulted in approximately half of the communities in each territory visited annually. As a result, the data is only representative in the territories after two years.

4. CONTENT STRUCTURE OF THE CCHS

In addition to sociodemographic and administrative data, the CCHS includes three content components, each of which addresses a different need: Common content (Core content and Theme content), Optional content, and Rapid Response content. Appendix A lists the modules included in the 2019-2020 questionnaire and their type of component.

The average length of a CCHS interview is estimated between 40 and 45 minutes.

Table 4.1 Length of survey by component

CCHS COMPONENT		**Length**
Common content	Core content	**19 minutes**
	Theme content	**8 minutes**
Optional content		**8 minutes**
Rapid Response content (cost recovery basis)		**2 minutes**

4.1 Core content

The CCHS core content component includes questions asked of respondents in all provinces and territories (unless otherwise specified). These questions will remain relatively stable in the questionnaire for a period of about six years up until 2021.

4.2 Theme content

The theme content is asked of all survey respondents and comprises modules, or in some cases groups of questions within modules, that are related to a specific topic. There are two types of theme collected annually: two-year theme modules which are asked over a period of two years and one-year theme modules that are asked over a one year period. The one-year theme changes annually. Combined, the one-year and two-year theme content takes about 8 minutes of interview time. Themes are 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.

4.3 Optional content

The optional content component is designed to give provinces and territories the opportunity to select content that addresses their own public health priorities. Each province and territory selects modules for up to 8 minutes of content. Survey participants will be asked the optional content modules that are selected for the province or territory in which they live. Optional content is selected two years at a time. It should be noted that, unlike the modules included in the core or theme content, the resulting data from the optional content modules should not be generalized across Canada[3].

Appendix A presents the selection results of the optional content for 2019-2020 by province or territory of residence.

In 2020, the module Food security (FSC) was offered as optional and selected by all provinces and territories, except British Columbia. Given that this content represents an area of growing concern for public health researchers and the expected risks of food insecurity related to the ongoing COVID-19 pandemic, this module was added to the collection questionnaire for British Columbia for the period of September to December in order to allow for a national rate of food security to be estimated.

Important note: to facilitate analysis of the impacts of the pandemic on food security, the responses for respondents from British Columbia for the collection periods of September to December have been included on the 2020 annual file, but not on the 2019-2020 files.

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

4.4 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 one or two collection periods (three to six months) and are asked of all CCHS respondents during that period, excluding the Territories.

4.5 Content included in data files

The survey produces different data files:

- Sub-annual data file
- One-year data file
- Two-year combined data file
- Rapid response data file (when applicable)

Table 4.2 provides clarification about the data files available for the 2019 and 2020 CCHS.

Sub-annual data files

In 2020, the collection of the CCHS was halted at the end of March due to the COVID-19 pandemic. Collection did not resume until September. The CCHS has produced two sub-annual files for 2020: the first one covers the collection period from January to March, and the second, from September to December.

One-year data files

The survey produces data files every year. The 2020 annual file includes respondents from the 2020 data collection and variables from the core and theme content, as well as optional content. The territories are excluded from the one-year data file as the sample is only representative of the territories after two years.

Two-year data files

With each release of an even year data file, for example 2018, a file combining two years of data is released. The next two year file is scheduled to be released in 2021, and will include both the 2019 and 2020 reference years of collection.

The two-year data file includes all respondents and questions that were in the survey over the two year reference period. Unless otherwise specified, it comprises the questions from the annual core content and two-year theme content, as well as optional content that was selected over the two year period. The one-year theme and optional content selected for only one year are not available in the two-year data file.

Table 4.2 Content components for the 2019 and 2020 data files

Files		Annual core content	2019 one year theme content	2020 COVID-19 Module	2020 one year theme content	2019-2020 two-year theme content	Optional content[4]
2019	**Main**	Yes	Yes[5]	N/A	N/A	Yes[5]	Yes
2020 January to March	**Sub-annual**	Yes	N/A	No	Yes[5]	Yes[5]	Yes
2020 September to December	**Sub-annual**	Yes	N/A	Yes	Yes[5]	Yes[5]	Yes
2020	**Main**	Yes	N/A	Yes[6]	Yes	Yes	Yes
2019 - 2020	**Main**	Yes	No	No	No	Yes	Yes

4.6 New module on COVID-19 (COV)

When collection resumed in September, a new module was added with content related to the pandemic: changes in mental health, precautions taken to protect against COVID-19, experiences of symptoms, getting tested for COVID, and willingness to get a vaccine.

Important note: the responses for the COV module for the collection periods of September to December have been included on the 2020 annual file, but not on the 2019-2020 files. Respondents from the January to March collection period have the COV variables all set to 'Not applicable'. Since the person-level weight (WTS_M) on the 2020 file is designed for the full year, users should not produce population totals for the COV content. Users should opt to use the September to December sub-annual file for this analysis.

5. SAMPLE DESIGN

5.1 Target population

The CCHS covers the population 12 years of age and over living in the ten provinces. The three territories are covered over two years of collection. Excluded from the survey's coverage are: persons living on reserves and other Aboriginal settlements in the provinces; full-time members of the Canadian Forces; youth aged 12 to 17 living in foster homes; the institutionalized population; and persons living in the Quebec health regions of Nunavik and Terres-Cries-de-la-Baie-James. Altogether, these exclusions represent less than 3% of the target population.

The coverage in the territories was exceptionally limited to the capitals (Whitehorse, Yellowknife and Iqaluit) during the last three collection periods of 2020 given that in-person interviews were

[4] Optional content will be included in the 2019-2020 data file if it is asked of respondents in a province or territory 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 one-year theme content module is selected for the optional content of a jurisdiction during the second year, the module will be included in the two-year data file and will be processed as optional content.

[5] One-year theme content is not available for the territories unless they have selected that module in their optional content selection for the other year.

[6] The 2020 new module on COVID-19 was only asked from September to December. Respondents from the January to March collection period will have these variables set to 'Not Applicable'.

not possible for most of the year due to the COVID-19 pandemic. Considering the regions that could not be visited, the CCHS 2019-2020 covered about 82% of the targeted population in the Yukon, 70% in the Northwest Territories and 56% in Nunavut[7].

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 2019-2020, data was collected in over 100 HRs in the ten provinces, as well as in one HR per territory (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. Starting in 2015, the samples for the youth population aged 12 to 17 and the adult population aged 18 and over are treated separately. Based on the historical sample sizes, 120,000 of the 130,000 respondents were allocated to the adults' population and about 10,000 to the youth population.

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 the first step, provinces and territories are treated separately. A sample of 117,000 respondents is allocated to the provinces and 3,000 respondents are allocated to the territories. Within each group, the sample is allocated using a 0.75 power allocation based on the population size. Table 5.1 gives the targeted sample sizes for a two year period, as determined at the time of the CCHS redesign in 2015. Note that the sample size and allocation for the territories only allow for representativeness over a two-year period.

[7] The coverage of the target population was estimated using total population counts from the 2016 Census.

Table 5.1 Number of health regions and targeted sample sizes by province/territory, 2019-2020

		Target sample size 2019-2020	
Province or territory	**Number of HRs**	**Adults**	**Youths**
Newfoundland and Labrador	4	3,172	313
Prince Edward Island	1	2,116	221
Nova Scotia	4	4,952	385
New Brunswick	7	4,182	328
Quebec	16	24,804	1,896
Ontario	35	36,354	3,032
Manitoba	5	5,804	522
Saskatchewan	11	5,128	453
Alberta	5	13,984	1,175
British Columbia	16	16,388	1,259
Yukon	1	899	166
Northwest Territories	1	979	166
Nunavut	1	722	166
Canada	**107**	**119,484**	**10,082**

Then, within each province, the sample was allocated to the HRs using a 0.35 power allocation based on the size of the population of the HR. The number of respondents allocated to each HR was then inflated to compensate for expected non-response and out-of-scope units. This inflated sample size is called the raw sample size and corresponds to the number of dwellings that need to be sampled and sent to collection to obtain the targeted number of respondents. The provincial-level raw sample sizes obtained at the time of the 2015 redesign were kept stable for the following cycles, while the HR-level allocation within provinces was reviewed at each cycle to use more up-to-date historical information about in-scope and response rates for the CCHS. The number of units sent to collection by province/territory for the CCHS 2019-2020 are available in Appendix D.

5.4 Frames, household sampling strategies

The CCHS sample is selected using different frames according to the age group. For the adult population (18 years and older), the sample of households is selected from an area frame. For the youth population (12 to 17 years old) a list frame is used to select persons.

5.4.1 Sampling of households from the area frame for the adult population

The area frame used by the Labour Force Survey (LFS) is used as a sampling frame for the adult population. The LFS uses a two-stage sample design[8]. In the first stage, a sample of primary sampling units (PSUs), corresponding to geographical regions called clusters, is selected. In each selected PSU, a sample of dwellings is drawn at the second stage. The details about the area frame and the LFS sample design can be found here:
https://www150.statcan.gc.ca/n1/en/catalogue/71-526-x

Sampling of households from the area frame in the provinces

The provinces are divided into clusters, geographic regions generally formed of 100 to 600 dwellings. Each cluster is assigned to one of the six LFS rotation groups, and clusters are grouped into homogeneous, mutually exclusive, geographic or socio-economic groups called strata. The area frame contains all of the PSUs and their corresponding stratum.

In the first stage of sampling, a sample of clusters is drawn from each stratum. In the second stage, dwelling lists are prepared for each selected cluster, and dwellings are selected from these lists through systematic sampling.

Each cluster is divided into a set of starts, to allow for systematic sampling of dwellings within each cluster. For example, a cluster of 300 dwellings could be divided into 20 starts. On a list where the 300 dwellings are sorted according to the location of the dwellings, the first start would correspond to dwellings #1, #21, #41, …, #281, the second start would correspond to dwellings #2, #22, #42,…, #282, and so on. The sample size provided by a given start (15 dwellings in this case) is called the yield of the start.

The division of clusters into starts was done at the time of the latest LFS redesign in 2015, and the expected yield of each start was computed at that time. When a cluster is first selected by a survey, an updated list of the dwellings found in that cluster is defined. By applying the sampling rate that had been determined at the time of the LFS redesign to the latest count of dwellings, a more precise expected yield can then be obtained for the starts of the cluster.

Knowing the expected yield of each start, the objective of the CCHS sample reservation process is to determine how many starts have to be selected in each LFS stratum, in order for the expected yield to be large enough to obtain the allocated sample size for each of the CCHS strata. When selecting starts from a given LFS stratum, the same number of starts has to be selected within each of the six LFS rotation groups, meaning that the number of selected starts is always a multiple of six[9].
The LFS strata and the CCHS strata have different geographic boundaries, meaning that starts selected from a given LFS stratum can provide sample to different CCHS strata. Since we ensure

[8] Except for Prince Edward Island where a stratified Simple Random Sample design is used.
[9] A few LFS strata have two or three clusters for each rotation group, meaning that starts will come in groups of 12 or 18.

that the total expected yield within each CCHS stratum will be at least the required sample size, and that starts are selected in groups of six, the total expected yield obtained from the selected starts will, aside from exceptions, typically exceed what is needed for each of the CCHS strata. When this is the case, a sub-sample of dwellings is selected within each CCHS stratum through a process called stabilization, in order to have a final sample size that corresponds to the target sample allocation for each CCHS stratum.

The stabilization process was not run for the last two collection periods of 2020, a larger raw sample being necessary given the expected drop in response rate.

Sampling of households from the area frame in the three territories

For operational reasons, the LFS area frame sample design for the three territories was different. For each territory, the larger communities each have their own stratum while smaller communities are grouped into strata based on various characteristics (population, geographical information, proportion of Inuit and/or Aboriginal persons, and median household income). The LFS defined six design strata in the Yukon, ten in the Northwest Territories and ten in Nunavut. For strata consisting of a group of communities, the first stage of selection consisted of randomly selecting one community with a probability proportional to population size within each design stratum. Then, within the selected community, the second stage consisted of selecting households using the same sampling strategy as the one described above. The CCHS selected its sample from the same communities sampled by the LFS, while ensuring that different dwellings were selected. If too many or too few dwellings were available for a community within a stratum, another community was selected for the CCHS. For larger communities with their own stratum, only one stage design was necessary where households were selected directly using the same sampling strategy described above.

It is worth mentioning that the frame for the CCHS covered 94% of the targeted population in the Yukon, 96% in the Northwest Territories and 93% in Nunavut. Due to the absence of in-person interviews after mid-March during the CCHS 2020, CCHS interviews were exceptionally conducted only in the three capitals (Whitehorse, Yellowknife and Iqaluit) during the last three collection periods, resulting in a lower coverage than in a normal two-year cycle of the CCHS.

5.4.2 Sampling of persons from the list frame for the youth population

To sample persons for the youth population between the ages of 12 and 17 years old, the CCHS uses a list frame created from the Canadian Child Benefit (CCB) files. The CCB files contain a list of all program beneficiaries with their names, addresses and phone numbers. This list is used to select directly the youths who will be interviewed over the phone.

5.5 Sample allocation over the collection periods

In order to balance interviewer workload and to minimize possible seasonal effects on estimates, the initial sample size for each frame is allocated equally over the four collection periods.

In the area frame (adults), each PSU selected within each HR was randomly assigned to a collection period accounting for a number of constraints related to field operations or weighting, while maintaining a uniform size for each period.

For the sample of youth selected from a list, independent samples were selected in each collection period. This strategy ensures that each sample is representative of the in scope Canadian population aged 12 to 17 years old for each period.

In 2020, all collection activities stopped from mid-March to the end of August. As a result, the first collection period was conducted from January to mid-March, while the last three collection periods were conducted in September, October and November-December respectively.

5.6 Sampling of interviewees

Since the sample for the youth population is selected from a list of individuals, the selection of the interviewee is automatically handled at sampling and no extra step is required at collection. The selection of an interviewee at collection is now only required for the adult population.

The selection strategy that was adopted accounted for user needs, cost, design efficiency, response burden and operational constraints. For a household selected as part of the 18+ sample, 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 5.2 gives the person-level selection factors used to determine the probabilities of selection of individuals in sampled households by age group. For example, for a three-person household formed of two people between the ages of 35 and 64 and one 19-year-old, the 19-year-old would have a 1/2 chance of being selected (i.e., 20/(20+10+10)) while each of the 35 to 64 year-olds would have a 1/4 chance of being selected. To avoid extreme sampling weights, there is one exception to this rule: if the size of the household is greater than or equal to five then the person-level selection factor equals 1 for each individual in the household. Consequently, all people in that household have the same probability of being selected.

Table 5.2 Person-level selection factors, by age

Person-level selection Factors				
Age	**18-34**	**35-49**	**50-64**	**65+**
Factor	20	10	10	20

5.7 Supplementary buy-in sample in Ontario

In 2019-2020, the province of Ontario is divided into 14 Local Health Integration Networks (LHINs), which are each subdivided into sub-LHIN (76 sub-LHINs in total). The Province of Ontario requested a sample increase in order to produce estimates at the sub-LHIN level over a time frame of two years. In order to do so, the CCHS stratification had to be adjusted.

The Ontario HR and sub-LHIN boundaries intersect each other. To preserve the initial HR-level CCHS allocation, the adjusted stratification was set up so that each stratum mapped to a single HR (that is, no CCHS stratum corresponds to multiple HRs). Some of these HRs had to be subdivided into multiple CCHS strata, according to the HR*sub-LHIN intersections, for a total of 96 strata.

In some cases, the regular CCHS sample yielded enough units to obtain the targeted number of respondents required to produce estimates at the sub-LHIN level. For the sub-LHINs where the target was not achieved, the sample in the CCHS strata corresponding to the deficient sub-LHINs was increased in an attempt to reach the sub-LHIN targets. In a number of sub-LHINs, the target was not achievable due to low population count and/or the limitations of the area frame in providing enough sample in specific small geographical areas.

In terms of person-level sampling, the additional units are treated the same way as the other CCHS units selected from the area frame. That is, the selection of interviewees is done via the use of a roster and selection factors as described in section 5.6. The main difference between how both sets of units are treated is in terms of survey content: the buy-in units are not eligible to receive either the CHSS supplemental content or any rapid response content.

5.8 CCHS oversample for the purposes of the Canadian Health Survey on Seniors

The Canadian Health Survey on Seniors (CHSS) supplement to the 2019 and 2020 CCHS was designed to collect health information for the Canadian population aged 65 and over. CHSS estimates were to be produced at a provincial level for 65+ year-olds as well as national-level estimates for 85+ year-olds. To produce CHSS estimates with a targeted level of precision in all provinces, an oversample was required in eight provinces.

Note that units selected for this oversample were not eligible for any rapid response content.

5.8.1 Frame used for the CHSS

Many sampling strategies were considered for the CHSS. The advantages and disadvantages of selecting people directly, as opposed to first selecting dwellings and then people, as well as using a variety of different frames were weighed. Ultimately, for reasons of sample availability, coverage, design effect, and quality of contact information (in a CATI-only environment), three tables from the Household Survey Frame Service (HSFS) were chosen to be the basis for a frame. The three key linked HSFS tables are: the Dwelling Universe File (DUF) to provide addresses, the Socio-Economic indicators File (SEF) to obtain demographic information for the residents of the dwellings on the DUF, and the Residential Telephone File (RTF) to obtain a list of phone numbers associated with each dwelling on the DUF.

The final frame used for the CCHS Oversample for the purposes of the CHSS (referred to as the CHSS frame throughout this document) was created by linking these three files from the HSFS and deleting all dwellings that were not linked to a 65+ year-old person, as per the data on the SEF.

The resulting CHSS frame contained a list of Canadian dwellings in which it would be likely to find a person at least 65 years of age. A simple random sample of dwellings within each eligible province was selected, according to the sample allocation described in Section 5.8.2.

5.8.2 Sample size determination for the CHSS

The sample size for CHSS had to be sufficiently large to achieve survey objectives and was based on target prevalence rates, expected levels of precision, expected in-scope and response rates, and provincial population sizes. The number of people in Ontario and Quebec aged 65+ that typically respond to the CCHS in a given year was sufficient to hit provincial targets. As a result, oversamples were only required in the eight other provinces.

The final sample sizes by province can be found in Appendix D.

5.8.3 Selection of CHSS interviewees

As is done for any CCHS unit sampled from the area frame, a roster of household members is completed for any dwelling selected from the CHSS frame. If a dwelling turned out to have no seniors in it, then the unit was considered to be out-of-scope. If there was at least one senior in the dwelling, then one had to be selected at random to respond to CCHS and CHSS. This was done by setting the person-level selection factors in the roster to be equal to 1 for each eligible 65+ year-old in the dwelling, meaning that each senior in a CHSS-sampled dwelling had the same chance of being selected.

6. DATA COLLECTION

6.1 Computer-assisted interviewing

The CCHS uses two separate computer assisted interview (CAI) applications to collect data, one for telephone interviews (CATI) and one for personal interviews (CAPI). This is done in order to customise each application's functionality to the type of interview being conducted. Each application consists of an entry component, a health content component, and an exit component. CAPI interviews which are attempted in person first, were only conducted for the first 3 months of the collection year. All collection activities were suspended at the end of March due to the COVID-19 pandemic. Collection did not resume again until September, at which point all interviews were exclusively conducted by phone to respect best practices and safety protocols. Each of these applications are also part of a CATI or CAPI collection platform, or system, used to transmit data to and from interviewer's computers.

During the 2020 collection year, approximately 42,132 valid interviews were conducted using CAI. Approximately 6% of these completed cases were conducted in person using CAPI. Note these in-person interviews were only conducted during the first 3 months of the 2020 collection year prior to the cessation of all collection related activities due to the COVID-19 pandemic. The remaining 94% of interviews were conducted over the phone using both the CAPI and CATI platforms.

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 for the CATI collection platform.

The case management system routes the questionnaire applications and sample files from Statistics Canada's main office to the interviewers laptops. Data returning to the main office takes the reverse route. To ensure confidentiality, the data is encrypted before transmission. The data are then decrypted 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

Entry and exit components were developed for CATI and CAPI interviewing. These components contain standard sets of questions designed to guide the interviewer through contact initiation, collection of important sample information, respondent selection and determination of cases status. The health content consists of the health modules themselves and made up the bulk of the applications. This includes common modules asked of all respondents and optional modules which differed by provinces and territories. 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 standalone 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 demographic information such as age, sex, and economic family status are passed from the entry to the health content and exit components of the applications. It also ensures that variables affecting skip patterns and logic flows are correctly passed between modules within the health content.

Since, at this stage the applications essentially function as they will in the field, all possible scenarios faced by interviewers are simulated to ensure proper functionality. These scenarios test various aspects of the entry and exit components including, establishing contact, collecting contact information, determining whether a case is in scope, rostering households, creating appointments and selecting respondents. The applications are also tested to ensure that during an interview, correct modules are triggered reflecting optional content selections.

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

6.3 Interviewer training

Project managers, senior interviewers and interviewers from regional collection offices were sent self-study training packages before the start of collection. These packages were prepared by the CCHS project team and were used by existing experienced CCHS interviewers to reinforce their previous training. Project managers and senior interviewers also conducted customised training sessions for new CCHS interviewing staff as needed. There were also specific training sessions to deal with various topics related to CCHS collection on a monthly basis. The focus of the training sessions were to get interviewers comfortable using the CCHS 2020 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
- 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 were selected from an area frame for the population aged 18 and older, and the Canadian Child Benefit (CCB) frame for the population aged 12 to 17. Approximately 60% of the respondents selected from the area frame and all of the respondents selected from the CCB frames were interviewed using the CATI collection platform. About 40% of the area frame respondents were interviewed using the CAPI collection platform. All interviewers worked independently from their homes using laptop computers and were supervised from a distance by senior interviewers. The variable CASEMODE on the microdata file indicates which frame a case was selected from, as well as the collection application and platform that was used to complete it. The following shows the breakdown of CASEMODE:

- CASEMODE = 0: completed using CAPI collection platform
- CASEMODE = 1: completed using CATI collection platform

Prior to the suspension of all interviews due to the COVID-19 pandemic, 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_N040 on the microdata files indicates whether the CAPI interview (CASEMODE = 0) was completed face-to-face, by telephone or using a combination of the two techniques. When collection resumed in September 2020, all interviews were conducted by phone only.

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

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_N045) and whether the interviewer felt that the respondent's answers were influenced by the presence of the other person (ADM_N050).

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 and brochures 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. Numerous call-backs were made at different times on different days. When the timing of the interviewer's call was inconvenient, an appointment was made to call back at a more convenient time.

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 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 needed to obtain verbal permission from parents/guardians to interview youths between the ages of 12 to 14 who were selected for interviews. This information was collected in the Parental/Guardian Consent (PGC) block in the survey entry component. Several procedures were followed by interviewers to alleviate potential parental concerns and to ensure a completed interview.

If a parent/guardian asked to see the actual questions; interviewers were instructed to immediately have the regional office send a copy of the questionnaire.

If privacy could not be obtained to interview the selected youth over the phone (another person listening in) the interview was coded a refusal.

The Person Most Knowledgeable (PMK) block collected household level information found at the end of the survey (Insurance coverage, Food Security, Income and Administration) from the most knowledgeable person in the household. This block is initiated when the selected respondent is between the ages of 12 to 17. The block formalizes the process of identifying a person in the household who is likely better able to answer these household level questions than the young selected respondent. If a PMK is found, then the interview moves from the younger selected respondent between the ages of 12 and 17, to a household member who finishes the rest of the interview after the PMK block.

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.

Since 2010, the Proxy interview (GR) block has been modified to prompt the interviewer to specifically identify whether the proxy interview is being conducted due to a physical or mental condition. Interviewers are then asked to record the specific condition for either case. The variable ADM_PRX indicates whether a case was completed by proxy.

6.5 Field operations

The 2019 sample was divided into 4 non-overlapping three-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 8 weeks being used finalise the remaining sample and to follow up on outstanding non-response cases. The first collection period of 2020 (January to March) was treated in the same way as the 2019 sample. The CCHS 2020 - September to December sample was divided into three non-overlapping six-week collection periods. Regional collection offices were instructed to use the first few weeks of each collection period to resolve the majority of the sample, and use the remaining weeks to finalise the rest of the sample and to follow up on outstanding non-response cases. All CATI 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. A series of dummy cases were included with each 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 interviewer assignments for those using the CAPI collection platform. Wherever possible, assignments were generally no larger than 15 cases per interviewer for these interviewers.

For final response rates, refer to Appendix E.

6.6 Quality control and collection management

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

A system of validation was used whereby interviewers had their work validated on a regular basis by the senior collection staff. Each collection period, randomly selected cases were flagged in the sample. Validation procedures generally occurred during the first few weeks of a collection period to ensure that any issues were detected promptly. Interviewers were provided feedback by their supervisors on a regular basis.

In 2014, a new collection management approach was implemented to improve the efficiency of the CATI data collection: Responsive Collection Design. Responsive Collection Design (RCD) is a new collection strategy that allows collection approaches to be adjusted throughout the collection period. In basic terms, it allows the grouping cases according to a number of factors, and then target the groups that are of particular interest. In an ordinary survey using the CATI collection platform, the interviewer groups remain constant throughout collection. However, during Responsive Collection Design, these groups change during each phase of collection in order to maximize response rates, productivity and the representativeness of the collected sample.

With the 2015 CCHS redesign, a dynamic telephone number strategy was developed. Starting in 2015, the 18+ years old sample can have up to 2 phone numbers, while the CCB samples can have up to 4 telephone numbers. This increases the number of available telephone numbers interviewers can call to maximize their potential ability to reach the selected households or respondents.

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 Flows

When data collected in the field is returned to Head Office, any questions that were skipped will appear as blank on the preliminary data files. During processing, every blank variable is given one of two processing code values:

‘Valid Skip’ – VS
For variables of length 1, the value for VS will simply be 6. Otherwise, these values will appear on the file as a string of 9’s, followed by a 6, according to the length of the numerical variable or text field.

The code VS is applied whenever a question is skipped because it does not apply to the respondent (a population exclusion). The two main reasons for this code being used are optional module variables for respondents in provinces where that content was not selected; and questions that are skipped by a flow in the questionnaire since it did not apply, as with the Maternal Experiences questions for male respondents.

'Not Stated' – NS

For a variable of length 1, the value for NS will simply be 9. Otherwise, these will appear on the file as a string of 9's, according to the length of the numerical variable or text field.

The code NS is applied whenever a question is skipped, but could have been asked. These are respondents that are in the universe, but without an actual response. The two main reasons for this code being used are questions that are skipped because the interview is being completed by proxy; and questions that are skipped because a preceding question was answered with a 'Don't Know', 'Refuse', or else was itself 'Not Stated'. For example, if a respondent refuses to answer CCC_Q095 (Do you have diabetes?), the questions that ask about the age of the diagnosis, insulin use, etc., are all set to 'Not Stated' since it is not known whether or not the respondent has diabetes.

7.3 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. In 2019-2020, variables that were OSLA coded include GDR_010, SDC_037, CAN_025, CAN_035, STS_015, PG_01 and LOP_025. Other variables with other-specify responses for this cycle were not selected for coding. For all questions, the 'Other-specify' responses are taken into account when refining the answer categories for future cycles.

When write-in responses ("Other-specify") are coded into an existing category during head office processing, it is possible that other questions in the questionnaire would become in scope to the respondent. Since those questions were never asked during the interview, the missing answers were set to 'Not Stated' in processing. For example, in the Injuries (INJ) module, a question asks what the respondent was doing at the time of their injury (INJ_Q065). If the interviewer collected an answer in the "Other Specify" that indicated the respondent was working, then the variable INJ_065 would be coded to the category 'Working at a job or business (excluding travel to and from work)' and assigned a value of '04'. Had the interviewer used this category in the interview, the respondent would then get the Workplace injury (INW) block of questions if it were survey content selected as optional content. All of the questions in INW in this case will have been set to 'Not Stated'.

7.4 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 the collapsing of response categories. In other cases, several variables have been combined to create a new variable. The *Derived Variables Documentation*

(DV) provides details on how these more complex variables were derived. For more information on the naming convention, please go to Section 12.6.

7.5 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 microdata files, and must be used to derive meaningful estimates from the survey. For example, if the number of individuals who smoke daily is to be estimated, it is done by selecting the records referring to those individuals in the sample having that characteristic and summing the weights entered on those records.

7.6 Income variables for 2020 - Linkage and Imputation

Since 2016, income variables in the CCHS master files are populated from one of three sources: tax records, respondent provided data, imputed data. Where respondents did not object to a link to their tax data and where a link to the appropriate tax records could be found, the income source variables (INC_005A-O, INC_Q025A-O), the main source of income (INC_015, INC_035), if they received a supplement for people with a disability (INC_010, INC_030) and the total income amounts (INC_021, INC_041) were determined from tax records. Where linkage to tax data was not feasible, respondent data was used. If neither linked tax data nor respondent reported data was available, an imputation of the total personal and total household income variables was conducted. The sources of incomes (INC_005A-O, INC_010, INC_015, INC_025A-O, INC_030, INC_035) were not imputed.

The variable INCFIMPP indicates the source used for the personal income variables and INCFIMPH indicates the source for household income variables. Starting in 2019, any tax source will be indicated as imputed. For more information concerning income on the master file, please refer to the document “CCHS_2019-2020_Master_File_Income” available in the ‘Documentation’ folder of the 2019-2020 CCHS. For more information concerning income on the share file, please refer to the document “CCHS_2019-2020_Share_File_Income”.

7.7 Postal code imputation

A sampled unit’s postal code is used together with the Postal Code Conversion File, to derive the rest of the geographical variables that are available on the CCHS data file. It is therefore important that all sampled units have a valid postal code. If a sampled unit’s postal code is missing or invalid, it is usually imputed through a donor imputation process, although other imputation methods are sometimes used. The donor is chosen from the same geographical area, with as much precision as possible, as the unit with the missing or invalid postal code.

For the 2019-2020 CCHS, approximately 3.6% of sampled units from the area frame (those aged 18 and over) had their postal codes imputed. For sampled units from the CCB frame (aged 12-17), approximately 0.9% had their postal codes imputed in 2019-2020. For respondents from the CHSS frame (those aged 65 and over), 3.8% had their postal codes imputed in 2019-2020.

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 respondent included in the final sample. 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 uses three sampling frames for its sample selection: an area frame for the Canadian population aged 18 and over, an additional list frame of dwellings for people aged 65 and over, and a frame of telephone numbers from Canada Child Benefit (CCB) records for the 12-17 population.

The frame used to get extra sample for people aged 65 and over (referred to as seniors throughout this document) is for the Canadian Health Survey on Seniors (CHSS) supplement to the CCHS. People selected via this oversample were asked all of the CCHS questions, and then were offered the possibility to complete the CHSS content. CCHS will therefore include seniors coming from both the CCHS sample and CHSS oversample.

The weighting strategy treats both the area and the CHSS frames independently up to and including the adjustment for household nonresponse; then, it integrates both samples (adjusting the weights to account for the overlap in population) to end up with a person-level weight for the adult population. The CCB frame is weighted independently of the two adult frames up to and including person-level nonresponse. The person-level weights from the integrated adult and CCB frames are then combined into a single set of weights; they jointly undergo some more adjustments (including being matched to known population totals) and become the final person-level weights.

8.1 Overview

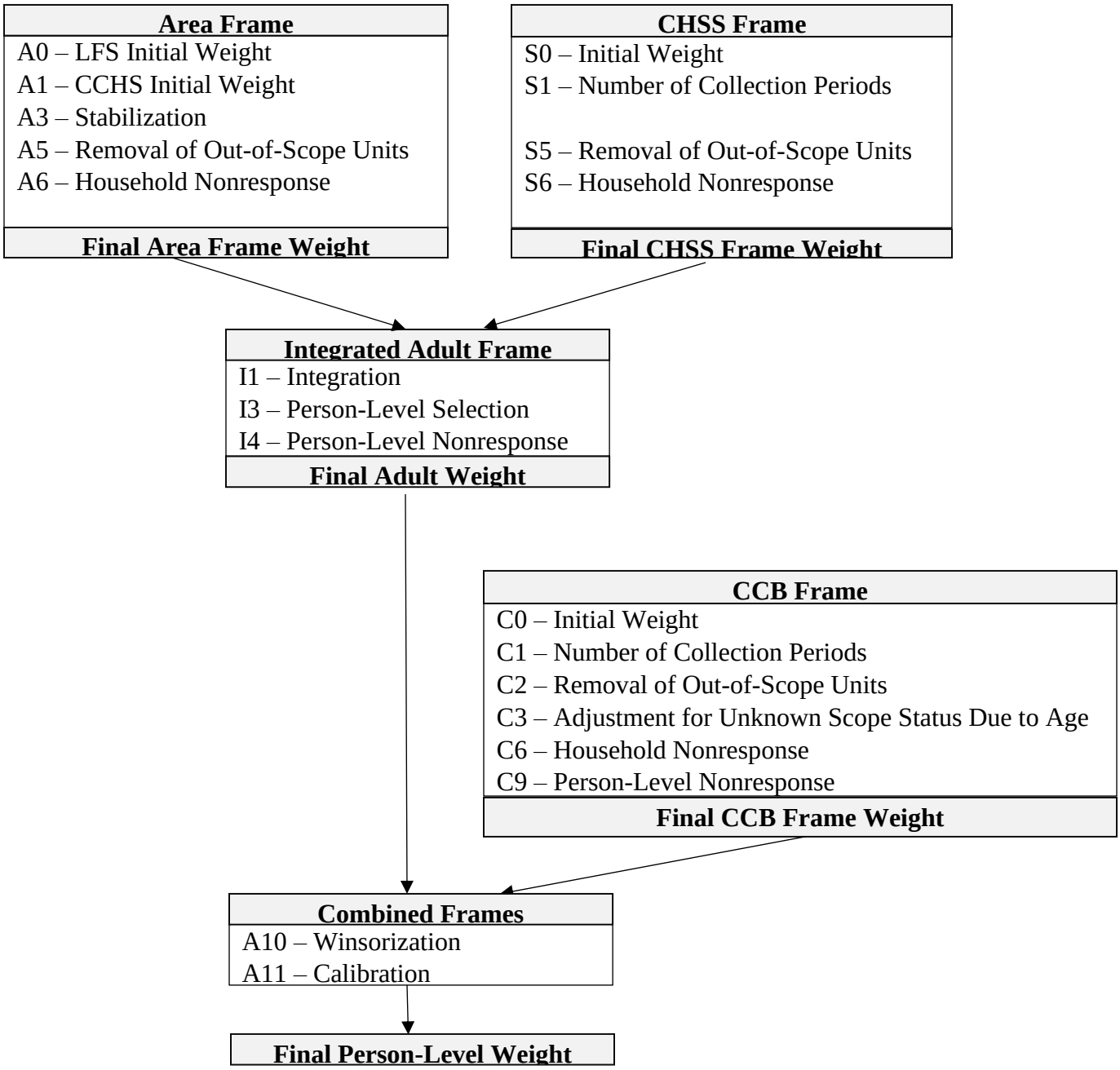
The following sections describe the weighting process for the provinces. Sub-section 8.2 provides details on the weighting strategy for the area frame, while sub-section 8.3 deals with the strategy for the CHSS frame. Subsection 8.4 provide details on the integration of the area and CHSS frames. Subsection 8.5 deals with the strategy for the CCB frame. The final weighting steps of Winsorization and calibration, which involve both frames, are discussed in sub-section 8.6.

Following the redesign of the CCHS in 2015, due to operational constraints, collection of a representative sample will be spread over a two year period. This means that estimates in the three territories will only be representative over two years. As such, respondents in the territories are omitted from the one-year files and will only be included on the two-year files (discussed in sub-section 8.6).

In 2020, the CCHS collection strategy and methodology had to be adapted due to the challenges brought by the COVID-19 pandemic. The collection was interrupted from mid-March to August. It resumed in September, but via telephone only. The response rate was also lower compared to previous years. For these reasons, some of the weighting steps had to be modified. Details on these modifications will be described at each of the steps below. For the territories, the weighting process is discussed in section 8.8, and changes to the process due to the pandemic are discussed in section 8.9. Analysis using the two year file should be done with caution.

Many adjustments to unit weights are made throughout the weighting process. Adjustments applied to units from the area frame are enumerated from A0 to A5, adjustments applied to units from the CHSS frame are enumerated from S0 to S5, while adjustments applied to units from the CCB frame are enumerated from C0 to C9. The adjustments to the weights of units from the integrated frame are enumerated from I1 to I4. The final adjustments after combining the adult and CCB frames are called A10 and A11. Diagram A presents an overview of the most important adjustments that are part of the weighting strategy.

Diagram A Weighting Strategy Overview



The subsections 8.2 to 8.6 present the details behind all the steps shown in Diagram A.

8.2 Weighting of the area frame sample

A0 – LFS Initial Weight

The weighting process 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, and the LFS initial weights reflect the probabilities of selection within LFS strata.

A1 – CCHS Initial Weight

In the initial adjustment A1, the LFS initial weight is adjusted to take into consideration the fact that the CCHS selects a sample of sufficient size for estimation at the Health Region level. To do so, the CCHS selects a different number of clusters than the LFS and can repeat the sampling of dwellings within the selected clusters. The resulting weight is called weight A1. For more details about the selection mechanism, as well as a more complete definition of LFS strata and clusters, refer to Statistics Canada (2017)[10].

A3 – Stabilization

In some HRs, the increase in 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. In order to minimize the dispersion of weights and thus reduce variability of estimates, stabilization is done in one of two ways within each HR. If the initial weights in each cluster within an HR are the same, the process consists of randomly sub-sampling dwellings at the HR level from the dwellings originally selected within each cluster. However, if the initial weights are not the same in each cluster within an HR, the process is done iteratively. First, the cluster with the lowest initial weight is selected and one unit from that cluster is removed. Then the weights are recalculated to adjust for the removed unit. Again the cluster with the lowest weight is selected and one unit is removed. This process is repeated until enough the desired sample size is achieved. 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 A3.

Exceptionally for the collection periods of October 2020 and November 2020, no stabilization was done to the sample. This was done in an attempt to get as many respondents as possible given that the collection was done only over the phone for these periods. For these two periods, the adjustment factor was equal to 1 for all records.

A5 – Removal of Out-of-Scope Units

Among all dwellings sampled, a certain proportion are identified during collection as being out-of-scope. Dwellings that are demolished or under construction, vacant, seasonal or secondary, as well as institutions are examples of out-of-scope cases for the CCHS. Dwellings in which all residents are out-of-scope, for example, all are full-time members of the Armed Forces or visitors

[10] Statistics Canada, 2017. *Guide to the Labour Force Survey*, Statistics Canada. Cat. No. 71-543-G.

to Canada are also out-of-scope. All these out-of-scope dwellings and their associated weights are simply removed from the sample, as it is assumed that the weighted out-of-scope dwellings in the sample are representative of out-of-scope dwellings in the population. For dwellings whose scope status is unknown (not enough information was gathered to determine scope status), their weights are reduced by the out-of-scope rate of their HR to account for some proportion of them being out-of-scope. This leaves a sample that is representative of in-scope dwellings or households. In-scope dwellings maintain the same weight as in the previous step, A3, which is now called weight A5.

For CCHS 2020, for the collection periods of September to December the scope rates were different compared to previous years. The effect of the COVID-19 pandemic, lower response rates and the fact that the collection was done only over the phone could explain such differences. It was evaluated that the observed scope rates for these collection periods were not representing reality. Instead, the scope rates of CCHS 2019 were used to adjust the weights of units with unknown scope for the respondents from September to December, 2020.

A6 – 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), which are formed within province. In order to create the RHGs, 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 models use characteristics such as the collection period, geographic information and variables from Statistics Canada's Household Survey Frame, 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. Separate models are created for dwellings that were contacted by telephone and for dwellings that were contacted in-person. An adjustment factor is calculated within each RHG as follows:

$$\frac{Sum\ of\ weights\ A5\ for\ all\ households}{Sum\ of\ weights\ A5\ for\ all\ responding\ households}$$

Weight A5 is multiplied by this adjustment factor to produce weight A6 for the responding households. Non-responding households are dropped from the process at this point.

For the collection period of September to December 2020, the collection periods were shorter than the regular three-month collection periods. The response rate was also lower than usual. Fewer RHGs were created than usual and the correction of the non-response bias is probably not as efficient as for other years.

8.3 Weighting of the CHSS frame sample

S0 – Initial Weight

The selection of senior dwellings for the CHSS oversample is done from the CHSS frame within each province.

The probability of selection of dwellings within each province corresponds to the ratio of the number of sampled senior dwellings to the total number of senior dwellings on the frame within the province. The ratio is based on the total set of units available on the frame and the number of units selected for the collection period in question. The probability of selection can therefore change depending on the sample allocation and frame updates. The inverse of this probability represents the initial weight S0.

S1 – Number of Collection Periods

Samples are drawn for each collection period on the CHSS frame. Each of the sampled units comes with an initial weight that allows each sample to be representative of the sampled population at the provincial level. To ensure that the total sample represents the population only once, an adjustment factor is applied to reduce the weights of the units in each sample. This adjustment factor is equal to the inverse of the number of samples being combined (i.e. the number of collection periods, 4). The initial weights (S0) are multiplied by this adjustment factor (1/4) to produce weight S1. Following this adjustment, the entire sample corresponds to the average population on the CHSS frame over the entire combined collection period.

S5 – Removal of Out-of-Scope Units

Among all sampled dwellings, a certain proportion are identified during collection as being out-of-scope. Institutions and dwellings that are seasonal or secondary are examples of out-of-scope cases for the CHSS. Dwellings in which all residents are out-of-scope for CHSS, for example, not containing any seniors, or where all seniors are full-time members of the Armed Forces or visitors to Canada, are also out-of-scope. Similar to the methods used on the area frame, out-of-scope dwellings and their weights are removed from the sample; dwellings with unknown scope have their weights reduced by the provincial out-of-scope rate; and in-scope dwellings maintain the same weight as in the previous step. Following this step, the resulting weight, S5, is representative of in-scope dwellings or households.

S6 – Household Nonresponse

The adjustment applied here to compensate for the effect of household nonresponse is identical to the one applied for the area frame units (adjustment A6). The adjustment factor calculated within each response homogeneity group is obtained as follows:

$$\frac{Sum\ of\ weights\ S5\ for\ all\ households}{Sum\ of\ weights\ S5\ for\ all\ responding\ households}$$

The weight S5 of responding households is multiplied by this adjustment factor to produce weight S6 for the responding households. Nonresponding households are dropped from the process at this point.

For the collection period of September to December 2020, the collection periods were shorter than the regular three-month collection periods. The response rate was also lower than usual. Fewer RHGs were created than usual and the correction of the non-response bias is probably not as efficient as for other years.

8.4 Weight adjustments with area and CHSS frames combined

I1 – Integration

This step consists of integrating the weights for senior households common to the area and CHSS frames by applying a method of integration[11]. Two types of households appearing on the integrated frame do not have their weight adjusted: households that are not common to both frames, and households on the area frame not containing any seniors. For all other units, a provincial adjustment factor, α_p, between 0 and 1 is applied to the weights. For each province (p), α_p is determined in such a way as to represent the relative importance of each sample (area frame sample or CHSS oversample) in the total provincial sample being integrated. The term "importance" here refers to the estimated effective sample size from each frame – the ratio of sample size and estimated design effect. The weight of the area frame units is multiplied by this factor α_p, while the weight of the CHSS frame units is multiplied by 1- α_p. Note that in the provinces only covered by the area frame (Ontario and Quebec) the adjustment is equal to 1 for all households. The product between the adjustment derived here and the final household weight calculated earlier (A6 or S6, depending on which frame the unit belongs to) gives the integrated household weight I1.

I3 - Person-level Selection

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 integrated household weight by the inverse of the probability of selection of the person selected in the household. This gives the weight I3. The probability of selection for an individual living in a dwelling selected from the area frame changes depending on the number of people in the household and their ages (see subsection 5.6 for more details). On the CHSS frame, all in-scope seniors within a household have an equal probability of selection; this probability is equal to the inverse of the number of in-scope seniors in the given household (see subsection 5.8.3).

For the collection period of September to December 2020, it was noticed that the proportion of respondents in households with a single person was higher compared to previous years. This is probably due to shorter collection periods (1 month instead of 3 months). The reduced amount of time to collect data, impacted the households where the person selected for the survey was not the person on the phone at the time of selection. Note that due to the overrepresentation of single-person households in the final master file, a potential bias could be present in the estimates for health characteristics that are correlated with the household size.

I4 - Person-level Nonresponse

[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.

A CCHS interview for the integrated adult frame dwellings can be seen as a two-part process. First, the interviewer gets the complete roster of the people within the household. Second, the selected person is interviewed. In some cases, interviewers can only get through the first part, either because they cannot get in touch with the selected person, or because that selected person refuses to be interviewed. Such individuals are defined as person nonrespondents and an adjustment factor must be applied to the weights of person respondents to account for this nonresponse. Using the same methodology that is used in the treatment of household nonresponse, the adjustment is applied within response homogeneity groups. In this process, the scoring method is used to define a response probability based on characteristics available for both respondents and nonrespondents. In addition to characteristics previously used to model household-level nonresponse (geographic information and paradata), all characteristics collected when creating the roster of household members are available for the estimation of response probabilities. The probabilities are grouped into response homogeneity groups and the following adjustment factor is calculated within each group:

$$\frac{Sum\ of\ weights\ I3\ for\ all\ selected\ persons}{Sum\ of\ weights\ I3\ for\ all\ responding\ selected\ persons}$$

Weight I3 for responding persons is multiplied by the above adjustment factor to produce weight I4. Nonresponding persons are dropped from the weighting process from this point onward.

For the collection period of September to December 2020, the collection periods were shorter then the regular three-month collection periods. The response rate was also lower than usual. Fewer RHGs were created than usual and the correction of the non-response bias is probably not as efficient as for other years.

8.5 Weighting of the CCB frame sample

C0 – Initial Weight

The initial design weight is defined as the inverse of the probability of selection. For the CCB frame, the selection of children is done from the CCB list frame within each CCHS stratum.

For the CCB frame, children 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 children on the list within the HR. The ratio is based on the frame available and the number of units selected for the particular 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 C0.

C1 – Number of Collection Periods

Samples are drawn for every collection period on the CCB frame. 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 collection period sample. The adjustment factor applied to each collection period sample is equal to the inverse of the number of samples being combined

(i.e. the number of collection periods). Following this adjustment, the entire list frame sample corresponds to the average population over the entire combined collection period. The initial weights (C0) are multiplied by this adjustment factor to produce weight C1.

C2 - Removal of Out-of-Scope Units

Selected children who are full-time members of the Canadian Armed Forces, are part of a First Nations reserve, or are deceased, are all examples of out-of-scope cases for the CCB frame. Similar to the methods used on the area frame, these cases are simply removed from the process, leaving only in-scope children in the sample. These in-scope children keep the same weight as in the previous step, now called weight C2.

C3 – Adjustment for Unknown Scope Status Due to Age

Some of the selected children have birthdays that fall within a collection period. For a proportion of these children, their birthday can change their scope status due to age. Some selected children may be eleven at the beginning of the collection period and therefore out-of-scope, but become in-scope when they turn twelve. Other children may be in-scope at seventeen but become out-of-scope when they turn eighteen. The age of these children, and therefore scope status, can be resolved (and treated accordingly) when contact is made and the age at time of contact is established. However, the scope status is unknown when children cannot be contacted or asked their age. They have their weights adjusted by their modeled probability of being in-scope had they been contacted and asked their age at contact. The rate of in-scope children turning 12 and 18 amongst known scope cases is used as the adjustment factor for the weights of these unknown scope cases. Those found out of scope due to their age have their weights set to zero similar to other out-of-scope units, in-scope children keep their weights and those with unknown scope and birthdays that could affect their scope have their weights adjusted, the resulting weights are called C3.

C6 – Household Nonresponse

The adjustment applied here to compensate for the effect of household nonresponse is identical to the one applied for the area frame units that were contacted by telephone (adjustment A6). For the CCB frame the selection was done at the person-level. However a household response is defined as a household where the roster was completed, similar to the area frame. The adjustment factor calculated within each response homogeneity group is obtained as follows:

$$\frac{\textit{Sum of weights C3 for all households}}{\textit{Sum of weights C3 for all responding households}}$$

The weight C3 of responding households is multiplied by this adjustment factor to produce the weight C6. Nonresponding households are removed from the process at this point.

For the collection period of September to December 2020, the collection periods were shorter than the regular three-month collection periods. The response rate was also lower than usual. Fewer RHGs were created than usual and the correction of the non-response bias is probably not as efficient as for other years.

C9 – Person-level Nonresponse

The adjustment applied here to compensate for the effect of person nonresponse is identical to the one applied for the area frame (adjustment A9). The adjustment factor calculated within each response homogeneity group is obtained as follows:

$$\frac{Sum\ of\ \ weights\ C6\ \ for\ all\ \ persons}{Sum\ of\ \ weights\ C6\ \ for\ all\ responding\ \ persons}$$

The weight C6 of responding persons is multiplied by this adjustment factor to produce the weight C9. Nonresponding persons are removed from the process at this point.

For the collection period of September to December 2020, the collection periods were shorter than the regular three-month collection periods. The response rate was also lower than usual. Fewer RHGs were created than usual and the correction of the non-response bias is probably not as efficient as for other years.

8.6 Final weight adjustments with integrated and CCB frames combined

At this point in the process the sample from the two frames can be combined in anticipation of the final weight adjustments of Winsorization and calibration. It is not necessary to perform an additional weight adjustment in combining the frames because they cover mutually exclusive populations. The selection probability and subsequent weighting of a unit on a given frame is in no way impacted by the structure or selection of units from the other frame.

A10 – 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 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.

A11 – Calibration

The last step necessary to obtain the final CCHS weight is calibration (A11). Calibration is done using Statistics Canada's generalized estimation system, G-EST, to ensure that the sum of the final weights corresponds to the population estimates defined at the HR level, for all 12 age-sex groups of interest. The six age groups are 12-17, 18-34, 35-49, 50-64, 65-74 and 75+, for both males and females. Starting in 2009, additional controls at more detailed geographic levels were introduced for HRs where additional information is available. A minimum size of 20 respondents is required to calibrate at the HR by age and sex. When getting less than 20 respondents, some collapsing is done within province and / or within gender. At the same time, weights are adjusted to ensure that each collection period 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 at the time of sampling.

Exceptionally for 2020, because the collection was interrupted from April to August, instead of the weights being adjusted for each collection period, they were adjusted for each of two periods: the months covering the pre-pandemic period (January to March) and for the remainder of the year (April to December).

The population estimates are projections based on Census counts and counts of birth, death, immigration and emigration since that time. In 2020, the base Census year used to create these population projections changed; 2019 projections are based on the 2011 Census, and 2020 projections are based on the 2016 Census. The average of these monthly estimates for each of the HR-age-sex post-strata by collection period is used to calibrate. The weight A10 is adjusted using G-EST to obtain the final weight A11. Weight A11 corresponds to the ***final CCHS person-level weight*** and can be found on the master file with the variable name WTS_M.

8.7 Creation of a share weight

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

$$\frac{Sum\ of\ weights\ I4\ (or\ C9)\ for\ all\ respondents}{Sum\ of\ weights\ I4\ (or\ C9) for\ all\ respondents\ agreeing\ to\ share\ their\ data}$$

Similar to the nonresponse adjustments, this factor is calculated within response homogeneity groups, where in this case, individuals with similar estimated propensity to share will be grouped together. This share adjustment is calculated separately for respondents from the adult integrated frame and from the CCB frame. Weight I4 (or C9 for the CCB respondents) for respondents who agreed to share is multiplied by the above adjustment factor to produce a share weight. As with the Master weighting process, the adult integrated and CCB frames are combined at this point, and Winsorization and calibration (similar to adjustments A10 and A11) are applied to the share units. The final weight after these adjustments is called WTS_S.

8.8 Particular aspects of the weighting in the three territories

Starting in 2015-2016, the territories have been sampled on a two-year basis. Prior to 2015, each stratum was sampled in a year but to reduce collection costs, strata such as the capitals were sampled annually while other strata sampled just once every two years. Since characteristics of the population are quite variable from one community to the next, the sample is now only representative on a two year basis.

In addition to the two-year sampling scheme, 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 subsection summarises the changes applied to the steps described in subsections 8.1 to 8.7. For the area frame, 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). The initial weight is calculated annually. Once the initial weight is calculated, the same series of adjustments (A1 to A5) is applied to the area frame units. The stabilisation adjustment (A3) in general is performed at the community level. The out-of-scope adjustment (A5) is performed at the community level and by mode of collection. Furthermore, since some strata are sampled annually and others biannually, an adjustment to combine the strata into one biannual sample is performed after the adjustment for out-of-scope. Household-level and person level nonresponse adjustment classes are built in the same way as for the provinces, using the same set of variables plus the variable year (year of collection).

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 the Yukon. A similar control based on Inuit status was introduced for Nunavut starting in 2009. These controls ensure that the proportion of the estimates represented by these different groups is consistent with proportions indicated by the 2011 Census. Since the 2019-20 cycle, these proportions were calculated using the 2016 census.

Prior to 2013, CCHS only covered the 10 largest communities in Nunavut. The population counts used in calibration were adjusted to take this undercoverage into consideration. Starting in 2013, CCHS increased its coverage to match the Labour Force Survey where 93% of the population is covered. Therefore, the population counts used for calibration are based on the total population and no longer adjusted for this undercoverage.

8.9 Changes to weighting in the three territories due to COVID-19

Due to the COVID-19 pandemic, collection by in-person interviews (CAPI) was stopped in March 2020. For remote areas in the north, CAPI is the only mode of collection available, and in fact, the only collection in the territories in the final three quarters of 2020 was done in the territorial capitals (Whitehorse, Yellowknife and Iqaluit). Since each community excepting capital cities is only visited in one period of the two year cycle, each territory was not completely covered. Communities that were scheduled to be collected between April and December 2020 represented 8% of the population in Yukon, 25% of the population in NWT, and 34% of the population in Nunavut, for 22% in the territories overall.

As a result of these collection stoppages, several changes were made to the weighting process for the territories as detailed in section 8.8.

a) No paradata was available for units collected by in person interview in the territories in 2020, including those collected between January and March 2020. As a result, no household or person-level non-response modelling was done for these units, and the non-response adjustment was calculated using a simple weight ratio in each territory.

b) Bootstrap weights could not be coordinated with previous cycles. Because some communities were missing from the file entirely, it was necessary to recreate the bootstrap weights for the remaining towns. This impacts variance calculations for pooled estimates

obtained from combining this file with previous cycles. The variance as calculated will be higher than it would be if the bootstrap weights had remained coordinated.

c) To account for the difference between the sampled area and the collected area, calibration was performed using population projections obtained for the collected area only. Original population projections were adjusted downwards by the proportion of the population living in collected communities. As a result, the weights of North respondents total to only 78,616 instead of the projected 100,519 target population of the North. Extrapolating from analysis done on the collected units in order to make conclusions about each territory as a whole is therefore not recommended.

d) It was also necessary to change some of the calibration requirements. As mentioned in section 8.8, the calibration step controls for the proportion of Indigenous individuals living in the Yukon and Northwest Territories. This process requires reliable projections of the Indigenous population twelve and older living in those territories. However, because collection in 2019-20 did not cover each territory, these projections could not be used. Instead, projections for the collected areas had to be derived from other data sources, which were not available broken down by age. As a result, the calibration of the indigenous population of these territories was done using projections for the whole population, instead of the population twelve and older. This would also have been the case in Nunavut. However, Nunavut could not be calibrated on the Inuit population at all due to the low number of respondents in 2020. Extreme caution should be taken when performing analysis on these groups due to potential bias. Lastly, it should be noted that north calibration was performed using five age groups instead of six, like the provinces. Due to the low number of respondents, it was necessary to collapse the groups for 65 to 74 and 75+.

8.10 Weighting for a two-year file

The territories are weighted as specified in section 8.8 with the modifications specified in section 8.9. They are weighted to be representative of the two years. The provinces have been weighted at an annual level. When two years of data are combined to create a two-year file, new weights are calculated by halving the annual weights. This ensures that the sum of the final weights is equal to the average population size over the two years. In some cases the adjustment is a little more complex where groups had been collapsed during calibration for the one year files. For more information on combining multiple years, please refer to the article “Combining cycles of the Canadian Community Health Survey” published in the Statistics Canada Health Reports publication (82-003) at the following link:

https://www150.statcan.gc.ca/n1/pub/82-003-x/2009001/article/10795-eng.pdf

It should also be noted that calibration on collection period was performed differently depending on the year and region. Calibration for 2019 was done for four 3-month collection periods, while for 2020, it was done in two groups: pre-COVID, and during COVID. For the north, calibration was done using four 3-month collection periods, which group together units collected in the same period, but different years; for example, units collected in the first collection period of 2019 were calibrated together with units in the first collection period of 2020.

Note also that CHSS had no sample in the territories. As well, no extra adjustment was required to account for CHSS for the provinces.

Note that combining cycles of CCHS data from before and after the redesign (e.g. combining 2014 and 2015 Annual files) is NOT recommended because of the numerous changes in content and methodology.

9. DATA QUALITY

9.1 Response rates

In total, 264,677 of the selected units in the CCHS 2019-2020 and in the CHSS 2019-2020 were in-scope for the survey[12], out of which a response was obtained for 108,604 individuals, resulting in a response rate of 41.0%. Appendix E provides the counts of in-scope units, the counts of respondents and the resulting response rates, by province/territory. These figures are provided for the adults and for the youths separately, as well as for the complete CCHS sample in Table 9.1 (Appendix E). The figures for the CHSS are provided in the table 9.2 (Appendix E).

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. 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.

[12] Among the units selected, some are not in-scope for the survey. They are, for example, dwellings selected from the area frame but that turned out to be vacant, demolished or non-residential dwellings, or youths that have moved outside Canada. 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. This response rate includes the three territories.

A major source of non-sampling errors in surveys is the effect of ***non-response*** on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. Partial non-response to the CCHS was minimal; once the questionnaire was started, it tended to be completed with very little non-response. Total non-response occurred either because a person refused to participate in the survey or because the interviewer was unable to contact the selected person. The response rates were lower for the CCHS 2020 than for a normal cycle of the CCHS, due to the impact of the COVID-19 pandemic on the collection activities. For units that did not have a telephone number available on the sampling frame, the contact rates were very low since the usual in-person visits by interviewers were not allowed. For units that could be contacted by phone, having shorter collection periods and reduced interviewer capacity for CCHS data collection were the main reasons for the drop in response rates. 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.

Users are advised to pay particular attention to the note on data quality of the CCHS 2019-2020 in the box on page 4 of this guide

9.2.2 Sampling Errors

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error. The basis for measuring the potential size of sampling errors is the standard deviation of the estimates derived from survey results. However, because of the large variety of estimates that can be produced from a survey, the standard deviation of an estimate is usually expressed relative to the estimate to which it pertains. This resulting measure, known as the coefficient of variation (CV) of an estimate, is obtained by dividing the standard deviation of the estimate by the estimate itself and is expressed as a percentage of the estimate.
For example, suppose hypothetically that it is estimated that 25% of Canadians aged 12 and over are regular smokers and that this estimate is found to have a standard deviation of 0.003. Then the CV of the estimate is calculated as:

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

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

10. GUIDELINES FOR TABULATION, ANALYSIS AND RELEASE

This section of the documentation offers guidelines to users for tabulating, analyzing, publishing or otherwise releasing any estimates 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. They will also be able to develop currently unpublished figures in a manner consistent with these established guidelines. Methods to measure precision and quality are also described along with release guidelines to help decide when an estimate should be used in publication.

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[13];

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

[13] CCHS tables for 2015 onwards use an algorithm that round totals, then solves internally within the table to match the totals. Due to this method of rounding, tables generated using CCHS microdata may not replicate totals found in the data tables

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. If options are available, users should ensure that they specify that the weight is a sample weight rather than a frequency weight.

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_005**)*
__ 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_045)***
|_|_| 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_045** and the weight **WTS_M.** Next, sum this value over those records with a value of "daily" to the variable **SMK_005**. The denominator ($\hat{Y}$) is obtained by summing the final weight of those records with a value of "daily" to the variable **SMK_005**. 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. The use of the survey weights is the first step to ensuring that the proper results are obtained.

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. The end result is that, while in many cases the estimates produced by the packages are correct, the variances and the statistics based on those variances are almost meaningless. To get the proper results, the weights must be correctly interpreted by the software package and the proper variance estimation techniques must be used.

To ensure that results from this complex design are meaningful, there are several options. The first and most appropriate option is to use the variance calculation options outlined in Chapter 11 based on the bootstrap methodology. Second, survey specific procedures exist in many packages that properly interpret the weight. For example, in SAS, PROC SURVEYMEANS would provide results that are more appropriate as the bootstrap weights are used for the variance estimates than the results produced from PROC MEANS. PROC MEANS can use the survey weights to correctly adjust estimates, taking into account the unequal weighting design, but it tends to underestimate the variances in not using the bootstrap weights and assuming the design is simple random sampling. Please consult the bootstrap variance estimation document found in the CCHS release package for more information. It is recommended that data users use software packages and functions that can incorporate both sample weights and bootstrap weights while performing their analyses.

Only the options outlined in Chapter 11 based on the bootstrap methodology will take the stratification, clustering and multiple frame design into account when calculating the variance.

10.4 Release guidelines

Since the estimates obtained from the survey are based on a sample, there is variability in the values obtained in the sense that a different sample could result in different results. To take this into consideration, users should first ensure that there are enough observations to properly estimate the statistic and also to estimate the variance. Once the variance is obtained, users should ensure that the variance is reasonable enough that the estimate can properly be interpreted as being near the true population value.

Master or share file

For users of the master or share files, it is recommended to have at least 10 observations with the characteristic of interest and 20 in the domain if a proportion is being calculated. With enough observations, the user can proceed to calculating the variance and the coefficients of variation using the bootstrap weights provided with the data along with the appropriate software to do the analysis. The variances can be used to calculate the CVs which aid in assessing the reliability of an estimate in regard to the sampling variability. The CV provides a relative measure of the sampling error as a proportion of the estimate. Estimates should be vetted using the guidelines in Table 10.1. CVs will increase as the variability of an estimate increases, and decrease as an estimate is more precise. However, an estimate can be precise while still having a large CV, specifically if the estimate has a value close to 0. Examining the confidence interval of the estimate will provide further indication of the quality of the estimate in terms of the variability. Long

confidence intervals indicate less precision in the estimate while smaller confidence intervals indicate greater precision. When assessing the trustworthiness of sample proportions, the confidence intervals of estimates should be taken into account.

Table 10.1 Sampling variability guidelines

Type of Estimate	CV (in %)	Guidelines
Acceptable	CV $\leq$ 15.0	Estimates can be considered for general unrestricted release. Requires no special notation.
Marginal	15.0 < CV $\leq$ 35.0	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 > 35.0	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 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

For a quick approximation of coefficients of variation that will be applicable to a wide variety of categorical estimates produced from the Share file, a set of Approximate Sampling Variability Tables will be produced. 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. Table 11.1 shows the CCHS 2017 look-up table at the Canada level as an example.

Table 11.1, CCHS Approximate Sampling Variability Tables, 2017 Share

NUMERATOR OF PERCENTAGE ('000)	ESTIMATED PERCENTAGE													
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	126.8	126.2	125.5	123.6	120.3	116.9	113.4	109.8	106.1	102.2	98.2	89.7	69.5	40.1
2	89.6	89.2	88.8	87.4	85.1	82.7	80.2	77.7	75.0	72.3	69.5	63.4	49.1	28.4
3	73.2	72.9	72.5	71.4	69.5	67.5	65.5	63.4	61.3	59.0	56.7	51.8	40.1	23.2
4	63.4	63.1	62.8	61.8	60.2	58.5	56.7	54.9	53.1	51.1	49.1	44.8	34.7	20.1
5	56.7	56.4	56.1	55.3	53.8	52.3	50.7	49.1	47.5	45.7	43.9	40.1	31.1	17.9
6	51.7	51.5	51.3	50.5	49.1	47.7	46.3	44.8	43.3	41.7	40.1	36.6	28.4	16.4
7	47.9	47.7	47.5	46.7	45.5	44.2	42.9	41.5	40.1	38.6	37.1	33.9	26.3	15.2
8	44.8	44.6	44.4	43.7	42.5	41.3	40.1	38.8	37.5	36.1	34.7	31.7	24.6	14.2
9	42.3	42.1	41.8	41.2	40.1	39.0	37.8	36.6	35.4	34.1	32.7	29.9	23.2	13.4
10	40.1	39.9	39.7	39.1	38.0	37.0	35.9	34.7	33.6	32.3	31.1	28.4	22.0	12.7
11	38.2	38.0	37.9	37.3	36.3	35.3	34.2	33.1	32.0	30.8	29.6	27.0	20.9	12.1
12	36.6	36.4	36.2	35.7	34.7	33.8	32.7	31.7	30.6	29.5	28.4	25.9	20.1	11.6
13	35.2	35.0	34.8	34.3	33.4	32.4	31.5	30.5	29.4	28.4	27.2	24.9	19.3	11.1
14	33.9	33.7	33.6	33.0	32.2	31.2	30.3	29.4	28.4	27.3	26.3	24.0	18.6	10.7
15	32.7	32.6	32.4	31.9	31.1	30.2	29.3	28.4	27.4	26.4	25.4	23.2	17.9	10.4
16	31.7	31.5	31.4	30.9	30.1	29.2	28.4	27.5	26.5	25.6	24.6	22.4	17.4	10.0
17	30.7	30.6	30.4	30.0	29.2	28.4	27.5	26.6	25.7	24.8	23.8	21.7	16.8	9.7
18	29.9	29.7	29.6	29.1	28.4	27.6	26.7	25.9	25.0	24.1	23.2	21.1	16.4	9.5
19	29.1	28.9	28.8	28.4	27.6	26.8	26.0	25.2	24.3	23.5	22.5	20.6	15.9	9.2
20	28.3	28.2	28.1	27.6	26.9	26.1	25.4	24.6	23.7	22.9	22.0	20.1	15.5	9.0
21	27.7	27.5	27.4	27.0	26.3	25.5	24.8	24.0	23.2	22.3	21.4	19.6	15.2	8.8
22	27.0	26.9	26.8	26.4	25.7	24.9	24.2	23.4	22.6	21.8	20.9	19.1	14.8	8.6
23	26.4	26.3	26.2	25.8	25.1	24.4	23.7	22.9	22.1	21.3	20.5	18.7	14.5	8.4
24	25.9	25.8	25.6	25.2	24.6	23.9	23.2	22.4	21.7	20.9	20.1	18.3	14.2	8.2
25	25.4	25.2	25.1	24.7	24.1	23.4	22.7	22.0	21.2	20.4	19.6	17.9	13.9	8.0
30	23.1	23.0	22.9	22.6	22.0	21.3	20.7	20.1	19.4	18.7	17.9	16.4	12.7	7.3
35	21.4	21.3	21.2	20.9	20.3	19.8	19.2	18.6	17.9	17.3	16.6	15.2	11.7	6.8
40	20.0	20.0	19.9	19.5	19.0	18.5	17.9	17.4	16.8	16.2	15.5	14.2	11.0	6.3
45	18.9	18.8	18.7	18.4	17.9	17.4	16.9	16.4	15.8	15.2	14.6	13.4	10.4	6.0
50	17.9	17.8	17.8	17.5	17.0	16.5	16.0	15.5	15.0	14.5	13.9	12.7	9.8	5.7
55	17.1	17.0	16.9	16.7	16.2	15.8	15.3	14.8	14.3	13.8	13.2	12.1	9.4	5.4
60	16.4	16.3	16.2	16.0	15.5	15.1	14.6	14.2	13.7	13.2	12.7	11.6	9.0	5.2
65	15.7	15.7	15.6	15.3	14.9	14.5	14.1	13.6	13.2	12.7	12.2	11.1	8.6	5.0
70	15.2	15.1	15.0	14.8	14.4	14.0	13.6	13.1	12.7	12.2	11.7	10.7	8.3	4.8
75	14.6	14.6	14.5	14.3	13.9	13.5	13.1	12.7	12.3	11.8	11.3	10.4	8.0	4.6
80	14.2	14.1	14.0	13.8	13.5	13.1	12.7	12.3	11.9	11.4	11.0	10.0	7.8	4.5
85	13.7	13.7	13.6	13.4	13.0	12.7	12.3	11.9	11.5	11.1	10.7	9.7	7.5	4.3
90	13.4	13.3	13.2	13.0	12.7	12.3	12.0	11.6	11.2	10.8	10.4	9.5	7.3	4.2
95	13.0	12.9	12.9	12.7	12.3	12.0	11.6	11.3	10.9	10.5	10.1	9.2	7.1	4.1
100	12.7	12.6	12.6	12.4	12.0	11.7	11.3	11.0	10.6	10.2	9.8	9.0	6.9	4.0
125	11.3	11.3	11.2	11.1	10.8	10.5	10.1	9.8	9.5	9.1	8.8	8.0	6.2	3.6
150	10.3	10.3	10.3	10.1	9.8	9.5	9.3	9.0	8.7	8.3	8.0	7.3	5.7	3.3
200	9.0	8.9	8.9	8.7	8.5	8.3	8.0	7.8	7.5	7.2	6.9	6.3	4.9	2.8
250	8.0	8.0	7.9	7.8	7.6	7.4	7.2	6.9	6.7	6.5	6.2	5.7	4.4	2.5
300	7.3	7.3	7.2	7.1	6.9	6.8	6.5	6.3	6.1	5.9	5.7	5.2	4.0	2.3
350	********	6.7	6.7	6.6	6.4	6.2	6.1	5.9	5.7	5.5	5.3	4.8	3.7	2.1
400	********	6.3	6.3	6.2	6.0	5.8	5.7	5.5	5.3	5.1	4.9	4.5	3.5	2.0
450	********	5.9	5.9	5.8	5.7	5.5	5.3	5.2	5.0	4.8	4.6	4.2	3.3	1.9
500	********	5.6	5.6	5.5	5.4	5.2	5.1	4.9	4.7	4.6	4.4	4.0	3.1	1.8
750	**********		4.6	4.5	4.4	4.3	4.1	4.0	3.9	3.7	3.6	3.3	2.5	1.5
1000	**********		4.0	3.9	3.8	3.7	3.6	3.5	3.4	3.2	3.1	2.8	2.2	1.3
1500	**********		3.2	3.2	3.1	3.0	2.9	2.8	2.7	2.6	2.5	2.3	1.8	1.0
2000	**********		2.8	2.7	2.6	2.5	2.5	2.4	2.3	2.2	2.0	1.6	0.9	
3000	**********		2.3	2.2	2.1	2.1	2.0	1.9	1.9	1.8	1.6	1.3	0.7	
4000	**********		1.9	1.8	1.8	1.7	1.7	1.6	1.6	1.4	1.1	0.6		
5000	**********		1.7	1.6	1.6	1.5	1.4	1.4	1.3	1.0	0.6			
6000	**********		1.5	1.5	1.4	1.4	1.3	1.3	1.2	0.9	0.5			
7000	**********		1.4	1.3	1.3	1.2	1.2	1.1	0.8	0.5				
8000	**********		1.2	1.2	1.1	1.1	1.0	0.8	0.4					
9000	**********		1.2	1.1	1.1	1.0	0.9	0.7	0.4					
10000	**********		1.1	1.0	0.9	0.7	0.4							
12500	**********		0.9	0.8	0.6	0.4								
15000	**********		0.8	0.7	0.6	0.3								

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

The Approximate Sampling Variability Tables, along with the design effects, the sample sizes and the population counts that were used to produce them, are provided in the document *Approximate*

Sampling Variability Tables, which is available to the share file users. All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Options concerning the computation of exact coefficients of variation are discussed in sub-section 11.7. Analysis using bootstrap weights is the method that should be employed for more precise results.

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 10, 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. Using the tables to estimate coefficients of variation are even less precise than using the bootstrap weights, so in this case, a minimum of 30 observations is required.

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. Unrounded numbers are used in the examples to facilitate the step by step instructions. These examples use the 2017 CCHS CV tables (Table 11.1). Users should still follow the rounding guidelines outlined in section 10.1 before publishing numbers.

Example 1: Estimates of numbers possessing a characteristic (aggregates)
Suppose that a user estimates that 3,665,449 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 (3,665,449) 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 3,000,000.
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, 2.3%.
4) So the approximate coefficient of variation of the estimate is 2.3%. According to the Sampling Variability Guidelines presented in Section 10.4, the finding that there were 3,665,449 individuals who smoke daily is publishable after applying the rounding rules.

Example 2: Estimates of proportions or percentages possessing a characteristic
Suppose that the user estimates that 3,665,449/5,151,237=71.2% 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 (71.2%) and the numerator portion of the percentage (3,665,449) in determining the coefficient of variation.
3) The numerator (3,665,449) 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 3,000,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the smallest figure closest to it, 70.0%.
4) The figure at the intersection of the row and column used, namely 1.3% is the coefficient of variation (expressed as a percentage) to be used.
5) So the approximate coefficient of variation of the estimate is 1.3%. According to the Sampling Variability Guidelines presented in Section 10.4, the finding that 71.2% 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, 2,067955 /15,115,880 = 13.7% smoke daily (estimate 1), while for women, this percentage is estimated at 1,597,494 / 15,528,483 = 10.3% (estimate 2). How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the CANADA level CV table in the same manner as described in example 2 gives the CV for estimate 1 as 2.7% (expressed as a percentage), and the CV for estimate 2 as 3. 1% (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 $\hat{d}$ = (0.137 - 0.103) = 0.034 is:

$$\sigma_{\hat{d}} = \sqrt{[(0.137)(0.027)]^2 + [(0.103)(0.031)]^2}$$

= 0.00489

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.00489/0.034 = 0.144.
4) So the approximate coefficient of variation of the difference between the estimates is 14.6% (expressed as a percentage). According to the Sampling Variability Guidelines presented in Section 10.4, this estimate can be published.

Example 4: Estimates of ratios

Suppose that the user estimates that 3,665,449 individuals smoke daily, while 1,485,788 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 1,485,788. 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, 4.0%.
4) The denominator of this ratio estimate is 3,665,449. The figure closest to it is 3,000,000. 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%.
5) So the approximate coefficient of variation of the ratio estimate is given by rule 4:

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

That is,

$$\alpha_{\hat{R}} = \sqrt{(0.040)^2 + (0.023)^2}$$

= 0.046

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 1,485,788/3,665,449 which is 0.41:1. The coefficient of variation of this estimate is 4.6% (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:

$\hat{X}$ = 0.712

z = 2

$\alpha_{\hat{X}}$ = 0.013 is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.712 - (2) (0.712) (0.013), 0.712 + (2) (0.712) (0.013)}

$CI_{\hat{x}}$ = {0.693, 0.731}

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 two 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 5% 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 = 0.00489. Hence,

$$z = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.137 - 0.103}{0.0048995} = \frac{0.034}{0.0048995} = 6.87$$

Since z = 6.95 is greater than 2, it must be concluded that there is a significant difference between the two estimates at the 5% level of significance. Note that the two sub-groups compared are considered as being independent, so the test is valid.

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.

Many statistical packages allow for analyses using the bootstrap weights. More details are provided in section 10.3 and in the document on bootstrap variance estimation.

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.

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.

12. MICRODATA FILES: DESCRIPTION, ACCESS AND USE

The CCHS produces three types of microdata files: master files, share files and public use microdata files (PUMF). Table 12.1 includes the list of all available 2020 collection period data files.

12.1 Master files

The master files contain all variables and all records from the survey for an associated time frame. 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.

For 2020, because collection only took place in half the communities for each territory, a single year of data is not representative of the territories. Because of this, the 2020 master file will only contain the data of respondents from the ten provinces. The data collected in 2020 for respondents from the territories will first be released on the 2019-2020 master data file.

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: https://www.statcan.gc.ca/eng/rdc/index

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 Centre for Population Health Data prepare custom tabulations. This service is offered on a cost-recovery basis. It allows users who do not possess knowledge of tabulation software products to get custom results. The results are screened for confidentiality and reliability concerns before release. For more information, please contact Client Services at 613-951-1746 or by e-mail at statcan.hd-ds.statcan@canada.ca.

12.1.3 Remote access

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 but have an approved research proposal. A researcher can be supplied with a synthetic or 'dummy' master file and a corresponding record layout. With these tools, the researcher can develop his or her own set of analytical computer programs. The code for the custom tabulations is then sent via e-mail to statcan.hd-ds.statcan@canada.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.

12.1.4 Real Time Remote Access (RTRA)

Users can access CCHS data using the Real Time Remote Access (RTRA) system, enabling fast access to Statistics Canada microdata. The RTRA system is an on-line remote access facility allowing users to run SAS programs, in real-time, against micro-data sets located in a central and secure location. Researchers using the RTRA system do not gain direct access to the micro-data and cannot view the content of the micro-data file. Instead, users submit SAS programs to extract results in the form of frequency tables. As RTRA researchers cannot view the micro-data, becoming a deemed employee of Statistics Canada is no longer necessary. This relationship is the basis that allows the RTRA to service its clients rapidly. Please see https://www.statcan.gc.ca/eng/rtra/rtra for more information.

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 Health Canada and the Public Health Agency of Canada. 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. Due to the risk of disclosure household weights are not included on the PUMF.

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

There is no charge to access the PUMF in a post-secondary educational institution that is part of the Data Liberation Initiative, a partnership between post-secondary institutions and Statistics Canada for improving access to Canadian data resources. For more information on the Data Liberation Initiative please see http://www.statcan.gc.ca/eng/dli/dli.

The CCHS PUMF files are also free of charge to any researcher that contacts Client Services at 613-951-1746 or by e-mail at statcan.hd-ds.statcan@canada.ca.

Table 12.1 2020 CCHS data files

Reference Period	Files	File Type	File name	Sampling weight	Bootstrap weights file	Variables included	Records included
2020	Main file	Master	HS.txt	WTS_M	bsw.txt	All core and all optional / theme modules, including the new COV module	All respondent records in the ten provinces
		Share		WTS_S			Records of all respondents who agreed to share their data
	Rapid response file (SCH2/TAV)	Master	HS_SCH2.txt	WTS_M	bsw.txt	All core modules and questions in the SCH2 & TAV rapid response	All respondent records in the ten provinces (Excluding oversample)
		Share		WTS_S			Records of all respondents who agreed to share their data (Excluding oversample)
January To March 2020	Main file (Pre-COVID) Excluding CHSS cases	Master	HS.txt	WTS_M	bsw.txt	All core and all optional / theme modules	All respondent records in the ten provinces (Excluding CHSS cases)
		Share		WTS_S			Records of all respondents who agreed to share their data (Excluding CHSS cases)
	Main file (Pre-COVID) Including CHSS cases	Master		WTS_M			All respondent records in the ten provinces (Including CHSS cases)
		Share		WTS_S			Records of all respondents who agreed to share their data (Including CHSS cases)
	Rapid response file (SCH2/TAV)	Master	HS_SCH2.txt	WTS_M	bsw.txt	All core modules and questions in the SCH2 & TAV rapid response	All respondent records in the ten provinces (Excluding oversample)
		Share		WTS_S			Records of all respondents who agreed to share their data (Excluding oversample)
	Rapid response file (HLV)	Master	HS_HLV.txt	WTS_M	bsw.txt	All core modules and questions in the HLV rapid response	All respondent records in the ten provinces (Excluding oversample)
		Share		WTS_S			Records of all respondents who agreed to share their data (Excluding oversample)
September to December 2020	Main file (During-COVID) Excluding CHSS cases	Master	HS.txt	WTS_M	bsw.txt	All core and all optional / theme modules, including the new COV module	All respondent records in the ten provinces (Excluding CHSS cases)
		Share		WTS_S			Records of all respondents who agreed to share their data (Excluding CHSS cases)

	Main file (During-COVID) Including CHSS cases	Master		WTS_M			All respondent records in the ten provinces (Including CHSS cases)
		Share		WTS_S			Records of all respondents who agreed to share their data (Including CHSS cases)
	Rapid response file (SCH2/TAV)	Master	HS_SCH2.txt	WTS_M	bsw.txt	All core modules and questions in the SCH2 & TAV rapid response	All respondent records in the ten provinces (Excluding oversample)
		Share		WTS_S			Records of all respondents who agreed to share their data (Excluding oversample)
	Rapid response file (TBI)	Master	HS_TBI.txt	WTS_M	bsw.txt	All core modules and questions in the TBI rapid response.	All respondent records in the ten provinces (Excluding oversample)
		Share		WTS_S			Records of all respondents who agreed to share their data (Excluding oversample)

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

Users who have access to share files or master files have the choice of using one-year or two-year data files. Decisions about which period to use in a given data analysis should be guided by the level of detail and the quality required. With a one-year file, estimates will not always be available because of the quality associated with limited sample sizes. Since the territories are not available on the one-year file, users wanting true national estimates should use the two-year file.
Before interpreting and using a CCHS estimate, it is recommended to make sure that the estimates meets the following rules:

- Coefficient of Variation 35.0% or less
- a minimum of 10 respondents in the domain with the characteristic and
- total domain of interest includes at least 20 respondents (for proportions or ratios)

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. For more information on combining CCHS cycles into a multi-year file, please see https://www150.statcan.gc.ca/n1/pub/82-003-x/2009001/article/10795-eng.pdf

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

12.5 Use of weight variable

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

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

12.6 Variable naming convention

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.

A 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., 2015 and 2017) will be the same. However, some minor changes have been made to variable names since the 2015 redesign.

The naming convention used for variables beginning with the 2015 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 in increments of five (e.g. 005, 010, 015)

Example 1 shows that the structure of the variable name for question 005, Smoking Module, is SMK_005 :

Positions 1 to 3 :	SMK	Smoking module
Position 4 :	_	(underscore = collected data)
Position 5 to 8 :	005	Question number

Example 2 shows the structure of the variable name for question 2 of the Contacts with health professionals – part 1 Module (CHP_015), which is a multi-response question:

Positions 1 to 3 :	CHP	Health care utilization module
Position 4 :	_	(underscore = collected data)
Position 5 to 8 :	015	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. It should be noted that some module acronyms consist of four characters. Variables names from these modules are still bound by a maximum of 8 characters.
Position 4 designates the variable type based on whether it is a variable collected directly from a questionnaire question ("_"), derived ("D"), flag ("F"), or grouped ("G") 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 Q005A in the questionnaire becomes simply 005A, and question Q010 becomes simply 010.

Table 12.2 Designation of codes used in the 4th position of the CCHS variable names

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

For questions that have more than one response option, the final position in the variable naming sequence is represented by a letter. For this type of question, new variables were created to differentiate between a "yes" or "no" answer for each response option. For example, if Q010 had 4 response options, the new questions would be named Q010A for option 1, Q010B for option 2, Q010C for option 3, and Q010D for option 4. If only options 2 and 3 were selected, then Q010A = No, Q010B = Yes, Q010C = Yes and Q010D = No.

12.7 Data dictionaries

Separate data dictionary reports, including variable names, concepts, universe statements and frequencies, are provided for the main master and share files and each of the rapid response files (if applicable). All variables on these files will appear in the data dictionary report, though some may not include a frequency (SampleID for example).

When a variable includes the frequencies for responses, the report will show four columns:

1. Content – The text label for each of the response categories.
2. Code – The numerical value associated with each category.
3. Sample – The unweighted frequency on the file of respondents with each response.
4. Population – The weighted frequency for the response.

For each module in the data dictionary reports, a flag is used to indicate the availability of the module for each respondent. When the flag is equal to 2 (No), all variables in the module have “not applicable” values. For example, the DOWST variable indicates if the Work stress module applies to a given respondent.

12.8 Differences in calculation of core and theme content variables using different files

Variables from core content modules can be estimated using either of the two data files provided, when a one year and a two-year data file is available. Depending on which file is used, very small differences will be observed.

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

APPENDIX A - CANADIAN COMMUNITY HEALTH SURVEY (CCHS) - CONTENT OVERVIEW – 2019-2020

Content availability by province and territory – 2019-2020

		2019-2020												
Acronym	**Module Name**	**NL (10)**	**PEI (11)**	**NS (12)**	**NB (13)**	**QC (24)**	**ON (35)**	**MB (46)**	**SK (47)**	**AB (48)**	**BC (59)**	**YK[1] (60)**	**NWT[1] (61)**	**NU[1] (62)**
ADM	Administration information	C	C	C	C	C	C	C	C	C	C	C	C	C
ALC	Alcohol use	C	C	C	C	C	C	C	C	C	C	C	C	C
ALW	Alcohol use during the past week					O	O							O
AMU	Antibiotic medication use	O												
ANC1	Age of respondent	C	C	C	C	C	C	C	C	C	C	C	C	C
BPC	Blood pressure check			O										
CAN	Cannabis use	C	C	C	C	C	C	C	C	C	C	C	C	C
CCC[2]	Chronic conditions	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T	C/T
CCT	Colorectal cancer testing									O				O
CEX	Childhood experiences		O			O	O			O		T	T	T
CIH	Changes made to improve health		O			O						O		
CMH	Consultations about mental health	O		O		O	O	O		O	O	T	T	T
CP2	Contacts with health professionals - extended block					O						O		O
CP3	Contact with dental professionals			O								C-S	C-S	C-S
DEP	Depression							O					O	
DRG	Drug use							O	O	O	O			
DRV	Driving and safety										O	O		
DWI	Driving while under the influence						O	O			O			
ETS	Exposure to second hand smoke				O	O	O		O		O			
FGU	Canada's Food Guide use	O		O		O								
FLU	Flu shots	C	C	C	C	C	C	C	C	C	C	C	C	C
FSC	Food security	O	O	O	O	O	O	O	O	O		O	O	O
FVC	Fruit and vegetable consumption		O	O					O	O			O	
GAM	Gambling	O												
GEN	General health	C	C	C	C	C	C	C	C	C	C	C	C	C
GDR	Sex and Gender	C	C	C	C	C	C	C	C	C	C	C	C	C
GR	Proxy interview	C	C	C	C	C	C	C	C	C	C	C	C	C
HMC	Home care services - with palliative care	T	T	T	T	T	T	T	T	T	T	T	T	T
HUI	Health utility index	T	T	T	T	T	T	T	T	T	T	T	T	T
HWT	Height and weight - self-reported	C	C	C	C	C	C	C	C	C	C	C	C	C

		2019-2020												
Acronym	**Module Name**	**NL (10)**	**PEI (11)**	**NS (12)**	**NB (13)**	**QC (24)**	**ON (35)**	**MB (46)**	**SK (47)**	**AB (48)**	**BC (59)**	**YK[1] (60)**	**NWT[1] (61)**	**NU[1] (62)**
INC	Income (sub-block of ADM)	C	C	C	C	C	C	C	C	C	C	C	C	C
INJ	Injuries							O						
INS	Health insurance coverage			O				O	O			T	T	T
LBF	Labour force	C	C	C	C	C	C	C	C	C	C	C	C	C
MAC	Main activity	C	C	C	C	C	C	C	C	C	C	C	C	C
MAM	Mammography					O	O							O
MDA	Medical doctor attachment	O		O	O			O						
MEX	Maternal experiences				O						O			O
MXA	Alcohol use during maternal experience										O			
NDE	Nicotine dependence													O
OHT	Oral health											O	O	
PAA	Physical activities - adults 18 years and older		O						O	O				
PAP	Pap smear test													O
PAY	Physical activities for youth		O						O	O				
PCN2	Prescriptions - cost-related non-adherence - short version											O		
PEX	Patient experiences			O						O		T	T	T
PHC	Primary health care	C	C	C	C	C	C	C	C	C	C	C	C	C
PMK	Person most knowledgeable about household situation	C	C	C	C	C	C	C	C	C	C	C	C	C
PNC	Perceived need for care	T	T	T	T	T	T	T	T	T	T	T	T	T
PSC	Patient satisfaction - Community-based care			O										
REL	Relationship matrix	C	C	C	C	C	C	C	C	C	C	C	C	C
SBE	Sedentary behaviours										O			
SCA	Smoking cessation methods								O					O
SDC	Socio-demographic characteristics	C	C	C	C	C	C	C	C	C	C	C	C	C
SDS	Severity of cannabis dependence scale	T	T	T	T	T	T	T	T	T	T	T	T	T
SMK	Smoking	C	C	C	C	C	C	C	C	C	C	C	C	C
SPC	Consultations with health professionals on smoking											O		
SPI	Spirometry											O		
SPS	Social provisions (reduced)	O			O		O					T	T	T
STS	Sources of stress						O				O	T	T	T

		2019-2020												
Acronym	**Module Name**	**NL (10)**	**PEI (11)**	**NS (12)**	**NB (13)**	**QC (24)**	**ON (35)**	**MB (46)**	**SK (47)**	**AB (48)**	**BC (59)**	**YK[1] (60)**	**NWT[1] (61)**	**NU[1] (62)**
SUI	Suicidal thoughts and attempts	T	T	T	T	T	T	T	T	T	T	T	T	T
SWL	Satisfaction with life	O												
SXB	Sexual behaviours	O	O				O	O			O			
TAL	Tobacco products alternatives					O					O			O
UCN	Unmet health care needs	O		O	O	O	O	O			O			O
YAC	Youth access to cigarettes												O	

Legend - Content Type	
C	Core
C-S	Core-Selected
T	Theme
O	Optional
(blank)	Optional - Not selected

1 Although they are collected every year, data of respondents living in the territories are only included on two year microdata files.

2 In 2020, CCC core content was asked in the provinces and territories. A shorter version of theme 1 content was asked in the provinces and territories, theme 2 content was not asked. For details, please see the CCC table in appendix B.

APPENDIX B - CANADIAN COMMUNITY HEALTH SURVEY (CCHS) - CHRONIC CONDITIONS (CCC) – 2019-2020

In 2019 and 2020, CCC core content was asked in the provinces and territories. A shorter version of theme 1 content was asked in the provinces and territories, theme 2 content was not asked.
In 2020, one new question (CCC_190) was asked in the provinces and territories. It is not included on two year microdata files.

Chronic conditions (CCC)		Provinces and Territories			
Question	**Concept**	**Content type**	**2019**	**2020**	**2019-2020**
CCC_005	Joint pain – 30 d	T1	N	N	N
CCC_010	Joint pain symptoms – 3 mo	T1	N	N	N
CCC_015	Has asthma	C	Y	Y	Y
CCC_020	Asthma – had symptoms / attacks – 12 mo	C	Y	Y	Y
CCC_025	Asthma – took medication – 12 mo	C	Y	Y	Y
CCC_030	Has a COPD	C	Y	Y	Y
CCC_035	Has sleep apnea	T1	Y	Y	Y
CCC_040	Has scoliosis	T1	N	N	N
CCC_045	Has fibromyalgia	T1	Y	Y	Y
CCC_050	Has arthritis (e.g. osteoarthritis, rheumatoid arthritis, gout)	C	Y	Y	Y
CCC_055	Has back problems (excluding scoliosis, fibromyalgia and arthritis)	T1	N	N	N
CCC_060	Has osteoporosis	T1	Y	Y	Y
CCC_065	Has high blood pressure	C	Y	Y	Y
CCC_070	High blood pressure – took medication – 1 mo	C	Y	Y	Y
CCC_075	Has high blood cholesterol / lipids	C	Y	Y	Y
CCC_080	High blood cholesterol / lipids - took medication – 1 mo	C	Y	Y	Y
CCC_085	Has heart disease	C	Y	Y	Y
CCC_087	Has been diagnosed with heart disease - lifetime	C	Y	Y	Y
CCC_090	Suffers from the effects of a stroke	C	Y	Y	Y
CCC_095	Has diabetes	C	Y	Y	Y
CCC_100	Diabetes – age first diagnosed	C	Y	Y	Y
CCC_105	Diabetes – diagnosed – during pregnancy	C	Y	Y	Y
CCC_110	Diabetes – diagnosed – other than during pregnancy	C	Y	Y	Y
CCC_115	Diabetes – time between diagnosis and start of insulin	C	Y	Y	Y
CCC_120	Diabetes – currently takes insulin	C	Y	Y	Y
CCC_125	Diabetes – took pills to control blood sugar – 1 mo	C	Y	Y	Y
CCC_130	Has cancer	C	Y	Y	Y
CCC_135	Has been diagnosed with cancer – lifetime	C	Y	Y	Y
CCC_140	Has migraine headaches	T1	N	N	N
CCC_145	Has Alzheimer's Disease or any other dementia	C	Y	Y	Y
CCC_150	Has intestinal or stomach ulcers	T2	N	N	N
CCC_155	Has a bowel disorder (Crohn's, ulcerative colitis, IBS, incontinence)	T2	N	N	N
CCC_160	Type of bowel disease	T2	N	N	N
CCC_165	Has urinary incontinence	T2	N	N	N
CCC_170	Has allergies – was told by a health professional	T2	N	N	N
CCC_175	Source(s) of allergy	T2	N	N	N
CCC_180	Had hay fever or nasal allergy – lifetime	T2	N	N	N
CCC_185	Has chronic fatigue syndrome	T1	Y	Y	Y
CCC_190	Suffers from multiple chemical sensitivities	T1	N	Y	N
CCC_195	Has a mood disorder (depression, bipolar, mania, dysthymia)	C	Y	Y	Y
CCC_200	Has an anxiety disorder (phobia, OCD, panic)	C	Y	Y	Y

APPENDIX C - AVAILABLE GEOGRAPHY IN THE MASTER AND SHARE FILES AND THEIR CORRESPONDING CODES: CANADA, PROVINCES, HEALTH REGIONS AND PEER GROUPS

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

0	Canada	
10	Newfoundland and Labrador	
1011-C		Eastern Regional Integrated Health Authority
1012-E		Central Regional Integrated Health Authority
1013-E		Western Regional Integrated Health Authority
1014-E		Labrador-Grenfell Regional Integrated Health Authority
11	Prince Edward Island	
1100-C		Prince Edward Island
12	Nova Scotia	
1201-E		Zone 1 - Western
1202-C		Zone 2 - Northern
1203-E		Zone 3 - Eastern
1204-A		Zone 4 - Central
13	New Brunswick	
1301-C		Zone 1 (Moncton area)
1302-A		Zone 2 (Saint John area)
1303-C		Zone 3 (Fredericton area)
1304-E		Zone 4 (Edmundston area)
1305-E		Zone 5 (Campbellton area)
1306-E		Zone 6 (Bathurst area)
1307-E		Zone 7 (Miramichi area)
24	Quebec	
2401-C		Région du Bas–Saint–Laurent
2402-C		Région du Saguenay — Lac–Saint–Jean
2403-A		Région de la Capitale–Nationale
2404-C		Région de la Mauricie et du Centre–du–Québec
2405-C		Région de l'Estrie
2406-G		Région de Montréal
2407-B		Région de l'Outaouais
2408-C		Région de l'Abitibi–Témiscamingue
2409-C		Région de la Côte–Nord
2410-C		Région du Nord–du–Québec
2411-E		Région de la Gaspésie — Îles–de–la–Madeleine
2412-D		Région de la Chaudière–Appalaches
2413-A		Région de Laval

2414-B		Région de Lanaudière
2415-B		Région des Laurentides
2416-B		Région de la Montérégie
35	Ontario by Local Health Integration Network	
3501		Erie St. Clair
3502		South West
3503		Waterloo Wellington
3504		Hamilton Niagara Haldimand Brant
3505		Central West
3506		Mississauga Halton
3507		Toronto Central
3508		Central
3509		Central East
3510		South East
3511		Champlain
3512		North Simcoe Muskoka
3513		North East
3514		North West
35	Ontario by Health Unit	
3526-E		The District of Algoma Health Unit
3527-A		Brant County Health Unit
3530-B		Durham Regional Health Unit
3533-D		Grey Bruce Health Unit
3534-D		Haldimand-Norfolk Health Unit
3535-D		Haliburton, Kawartha, Pine Ridge District Health Unit
3536-B		Halton Regional Health Unit
3537-A		City of Hamilton Health Unit
3538-C		Hastings and Prince Edward Counties Health Unit
3539-D		Huron County Health Unit
3540-E		Chatham-Kent Health Unit
3541-C		Kingston, Frontenac and Lennox and Addington Health Unit
3542-C		Lambton Health Unit
3543-D		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-C		Northwestern Health Unit

3551-B		City of Ottawa Health Unit
3553-H		Peel Regional Health Unit
3554-D		Perth District Health Unit
3555-C		Peterborough County–City Health Unit
3556-C		Porcupine Health Unit
3557-D		Renfrew County and District Health Unit
3558-D		The Eastern Ontario Health Unit
3560-B		Simcoe Muskoka District Health Unit
3561-C		Sudbury and District Health Unit
3562-C		Thunder Bay District Health Unit
3563-C		Timiskaming Health Unit
3565-B		Waterloo Health Unit
3566-B		Wellington–Dufferin-Guelph Health Unit
3568-A		Windsor–Essex County Health Unit
3570-H		York Regional Health Unit
3575-D		Oxford Elgin St. Thomas Health Unit
3595-G		City of Toronto Health Unit
46	Manitoba	
4601-A		Winnipeg Regional Health Authority
4602-D		Prairie Mountain Health
4603-D		Interlake-Eastern Regional Health
4604-F		Northern Regional Health Authority
4605-B		Southern Health
47	Saskatchewan	
4701-D		Sun Country Regional Health Authority
4702-D		Five Hills Regional Health Authority
4703-D		Cypress Regional Health Authority
4704-B		Regina Qu'Appelle Regional Health Authority
4705-D		Sunrise Regional Health Authority
4706-B		Saskatoon Regional Health Authority
4707-D		Heartland Regional Health Authority
4708-C		Kelsey Trail Regional Health Authority
4709-C		Prince Albert Parkland Regional Health Authority
4710-B		Prairie North Regional Health Authority
4714-F		Mamawetan/Keewatin/Athabasca Regional Health Authorities[14]

[14] For most data sources (with the exception of Census and Demographic population estimates), health region level data are not available for some northern health regions in Saskatchewan which have small populations. To avoid suppression in these areas where small numbers or sample size impact on data quality, data have been grouped with neighbouring regions, as follows: Athabasca Health Authority, Saskatchewan (4713) is combined with Mamawetan

48	Alberta	
4831-B		South Zone
4832-B		Calgary Zone
4833-B		Central Zone
4834-B		Edmonton Zone
4835-I		North Zone
59	British Columbia	
5911-D		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-B		Fraser East Health Service Delivery Area
5922-H		Fraser North Health Service Delivery Area
5923-H		Fraser South Health Service Delivery Area
5931-H		Richmond Health Service Delivery Area
5932-G		Vancouver Health Service Delivery Area
5933-A		North Shore/Coast Garibaldi Health Service Delivery Area
5941-A		South Vancouver Island Health Service Delivery Area
5942-C		Central Vancouver Island Health Service Delivery Area
5943-C		North Vancouver Island Health Service Delivery Area
5951-C		Northwest Health Service Delivery Area
5952-C		Northern Interior Health Service Delivery Area
5953-I		Northeast Health Service Delivery Area
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	

Churchill River Regional Health Authority (4711) and Keewatin Yatthé Regional Health Authority (4712) and referred to as "Mamawetan/Keewatin/Athabasca Regional Health Authorities (4714)"

APPENDIX D - SAMPLE ALLOCATION BY PROVINCE/TERRITORY (CCHS 2019-2020 AND CHSS 2019-2020)

Appendix D – Sample allocation by province/territory (CCHS 2019-2020 and CHSS 2019-2020)[15]

Province/Territory	CCHS		CHSS	Total
	Adults	Youths	Seniors	
Canada	228,739	16,900	49,068	294,707
N.L.	5,949	553	9,741	16,243
P.E.I.	3,348	471	7,688	11,507
N.S.	8,877	649	4,698	14,224
N.B.	7,409	587	5,818	13,814
Que.	40,942	2,712	0	43,654
Ont.	83,341	5,012	0	88,353
Man.	10,215	897	5,648	16,760
Sask.	9,222	830	7,104	17,156
Alta.	25,802	1,999	5,760	33,561
B.C.	30,443	2,418	2,611	35,472
Y.T.	1,182	288	0	1,470
N.W.T.	1,234	274	0	1,508
Nvt.	775	210	0	985

[15] These raw sample sizes correspond to the number of units that were sent to collection.

APPENDIX E - RESPONSE RATES BY PROVINCE/TERRITORY (CCHS 2019-2020 AND CHSS 2019-2020)

Appendix E – Table 9.1 Response rates by province/territory (CCHS 2019-2020)

Province/ Territory	Adults			Youth			Combined
	# in-scope	# respondents	Response rates	# in scope	# respondents	Response rates	Response rates
Canada	201,834	83,112	41.2	16,343	6,526	39.9	41.1
N.L.	5,046	2,118	42.0	533	156	29.3	40.8
P.E.I.	2,553	1,094	42.9	463	147	31.7	41.1
N.S.	7,603	3,265	42.9	644	242	37.6	42.5
N.B.	6,232	2,701	43.3	577	228	39.5	43.0
Que.	35,134	17,972	51.2	2,658	1,495	56.2	51.5
Ont.	76,632	28,783	37.6	4,899	1,811	37.0	37.5
Man.	8,750	3,644	41.6	812	343	42.2	41.7
Sask.	7,128	3,062	43.0	772	297	38.5	42.5
Alta.	22,886	8,704	38.0	1,940	766	39.5	38.1
B.C.	27,175	10,488	38.6	2,360	794	33.6	38.2
Y.T.	1,066	471	44.2	266	114	42.9	43.9
N.W.T.	1,011	491	48.6	240	89	37.1	46.4
Nvt.	618	319	51.6	179	44	24.6	45.5

Appendix E – Table 9.2 Response rates by province (CHSS 2019-2020)

		Main CCHS Content			Supplementary CHSS Content				
Frame	**Prov.**	**# in scope pers.**	**# resp[16]**	**Resp. rates**	**# in scope seniors[17]**	**# resp[18]**	**Resp. rates.**	**# Sharers[19]**	**Share rates**
Both	**Canada**	**245,639**	**100,797**	**41.0%**	**45,863**	**41,635**	**90.8%**	**38,522**	**92.5%**
Both	**N.L.**	14,387	5,182	36.0%	3,879	3,425	88.3%	3,157	92.2%
Both	**P.E.I.**	9,873	4,008	40.6%	3,319	2,996	90.3%	2,766	92.3%
Both	**N.S.**	12,054	5,405	44.8%	3,474	3,092	89.0%	2,904	93.9%
Both	**N.B.**	11,713	5,154	44.0%	3,545	3,120	88.0%	2,906	93.1%
Both	**Que.**	35,134	17,972	51.2%	6,368	5,808	91.2%	5,370	92.5%
Both	**Ont.**	76,632	28,783	37.6%	8,538	7,461	87.4%	6,983	93.6%
Both	**Man.**	14,074	6,123	43.5%	3,676	3,502	95.3%	3,213	91.7%
Both	**Sask.**	13,835	5,932	42.9%	3,818	3,635	95.2%	3,319	91.3%
Both	**Alta.**	28,301	10,856	38.4%	4,528	4,303	95.0%	3,968	92.2%
Both	**B.C.**	29,636	11,382	38.4%	4,718	4,293	91.0%	3,936	91.7%
Area	**Canada**	**199,139**	**81,831**	**41.1%**	**26,897**	**24,357**	**90.6%**	**22,642**	**93.0%**
Area	**N.L.**	5,046	2,118	42.0%	815	753	92.4%	703	93.4%
Area	**P.E.I.**	2,553	1094	42.9%	405	375	92.6%	347	92.5%
Area	**N.S.**	7,603	3,265	42.9%	1334	1194	89.5%	1119	93.7%
Area	**N.B.**	6,232	2,701	43.3%	1092	1006	92.1%	949	94.3%
Area	**Que.**	35,134	17,972	51.2%	6,368	5,808	91.2%	5,370	92.5%
Area	**Ont.**	76,632	28,783	37.6%	8,538	7,461	87.4%	6,983	93.6%
Area	**Man.**	8,750	3,644	41.6%	1197	1125	94.0%	1042	92.6%
Area	**Sask.**	7,128	3,062	43.0%	948	908	95.8%	843	92.8%
Area	**Alta.**	22,886	8,704	38.0%	2,376	2,251	94.7%	2,093	93.0%
Area	**B.C.**	27,175	10,488	38.6%	3,824	3,476	90.9%	3,193	91.9%
CHSS	**Canada**	**46,500**	**18,966**	**40.8%**	**18,966**	**17,278**	**91.1%**	**15,880**	**91.9%**
CHSS	**N.L.**	9,341	3,064	32.8%	3,064	2,672	87.2%	2,454	91.8%
CHSS	**P.E.I.**	7,320	2,914	39.8%	2,914	2,621	89.9%	2,419	92.3%
CHSS	**N.S.**	4,451	2,140	48.1%	2,140	1,898	88.7%	1,785	94.0%
CHSS	**N.B.**	5,481	2,453	44.8%	2,453	2,114	86.2%	1,957	92.6%
CHSS	**Man.**	5,324	2,479	46.6%	2,479	2,377	95.9%	2,171	91.3%
CHSS	**Sask.**	6,707	2,870	42.8%	2,870	2,727	95.0%	2,476	90.8%
CHSS	**Alta.**	5,415	2,152	39.7%	2,152	2,052	95.4%	1,875	91.4%
CHSS	**B.C.**	2461	894	36.3%	894	817	91.4%	743	90.9%

[16] Includes all CCHS adult respondents

[17] Includes all CCHS senior respondents with the exception of 1,980 Ontario Buy-in senior respondents (units in the Ontario Buy-in sample were not eligible to receive the CHSS supplement)

[18] CCHS senior respondents who completed CHSS questionnaire and agreed to link their CCHS responses to CHSS

[19] CHSS respondents who agreed to share their data with certain partners