List of Appendices

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. The CCHS operates on a two-year collection cycle. The first year of the survey cycle ".1" is a large sample, general population health survey, designed to provide reliable estimates at the health region level. The second year of the survey cycle ".2" is a smaller survey designed to provide provincial level results on specific focused health topics.

This Microdata File contains data collected in the first year of collection for the CCHS (Cycle 1.1). Information was collected between September 2000 and November 2001, for 136 health regions, covering all provinces and territories. The CCHS (Cycle 1.1) collects responses from persons aged 12 or older, living in private occupied dwellings. Excluded from the sampling frame are individuals living on Indian Reserves and on Crown Lands, institutional residents, full-time members of the Canadian Armed Forces, and residents of certain remote regions.

This document has been produced to facilitate the manipulation of the CCHS (Cycle 1.1) cross-sectional microdata files, which are described in detail in the following text and appendices.

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

- For technical/general data support call:
  Electronic Products Help Line: 1-800-949-9491

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

- For remote access support call 1-613-951-1653
  E-mail: cchs-escc@statcan.ca
  Fax: 1-613-951-4198

2. Background

In 1991, the National Task Force on Health Information cited a number of issues and problems with the health information system. These problems were that: data was fragmented; data was incomplete; data could not be easily shared; data was not being analysed to the fullest extent; and the results of research are not consistently reaching Canadians[1]. In responding to the needs, the Canadian Institute for Health Information (CIHI), Statistics Canada and Health Canada have joined forces to create a National Health Information Roadmap.

The Roadmap is a direct response to the concerns and desires of more than 500 individuals representing a wide range of organizations and interest groups. Early in 1999, the three national organizations listed above conducted a broadly based national consultation on health information needs. Participants stressed that national agencies must work together to strengthen Canada's health information system, and must build on and contribute to the considerable investments and expertise at local, regional, and provincial/territorial levels.[2]

The Roadmap represents an important contribution to building a comprehensive national health information system and infrastructure to provide Canadians with the information they need to maintain and improve Canada's health system and the population's health.[3] What is needed is a co-ordinated plan of action. No single government or organization can combat the above–noted problems alone. Co-operation at all levels – national, provincial, territorial, regional and local health organizations – is a prerequisite for success.[4]

The plan of action starts by seeking answers to two crucial questions[5]:

1. How healthy is the health care system?
2. How healthy are Canadians?

The first question encompasses the effectiveness, efficiency and responsiveness of the health care system. Generally, an effective, efficient and responsive health care system is one that offers the quality of care Canadians expect.[6]

---

[1] 1999. Health Information Roadmap Responding to Needs, Health Canada, Statistics Canada. p.3.
[2] 1999. Ibid. p.1.
[3] 1999. Ibid. p.1.
[4] 1999. Ibid. p.3.
[5] 1999. Ibid. p.3.
[6] 1999. Ibid. p.3.

The second question is broader, and addresses the basic objective of the system: is the health of Canadians improving? To answer this, a strong health information system is needed.[7] This information system must embrace six principle characteristics[8].
The information system must be:

- secure and respectful of Canadians' privacy
- consistent
- relevant
- integrable
- flexible
- user-friendly and accessible

This health information system needs to be timely, provide person-oriented information, and have common data standards with other Canadian health surveys, such as the National Population Health Survey (NPHS). The new system must also provide: new or expanded data sets; data on health services; data on outcomes, health status and non-medical determinants of health; data on outcomes of selected health interventions; implement special studies involving priority issues; data on costs per service; information exchange protocols; expanded analytical and dissemination capacity, and public reports on the health care system.[9]

Given 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, federal, provincial and community health region stakeholders to determine their data requirements[10].

The purpose of this publication, the Public Use Microdata File, is to follow through on the mandate of collecting reliable, relevant information on health services, health status, and health issues of importance to Canadians - at the regional, provincial and national level - and disseminating this information to the public.

---

[7] 1999. Ibid. p.5.
[8] Expansion on these characteristics is described in Health Information Roadmap: Responding to Needs, 1999, Canadian Institute for Health Information. ISBN 1-895581-30-3. (http://www.cihi.ca)
[9] 1999. Ibid. p.11-14.
[10] 1999. Roadmap Initiative … Launching the Process. Canadian Institute for Health Information / Statistics Canada. ISBN 1-895581-70-2. p.19.

3. Objectives

The primary objectives of the CCHS are to:

- provide timely, reliable, cross-sectional estimates of health determinants, health status and health system utilization across Canada

- gather data at the sub-provincial levels of geography

- create a flexible survey instrument that:
  - meets specific health region data gaps
  - develops focused survey content for key data
  - deals with emerging health and health care issues as they arise

As a key component of the Population Health Surveys Program of Statistics Canada, the CCHS helps fulfil broader requirements of health issues in Canada. These are:

- aid in the development of public policy

- provide data for analytic studies that will assist in understanding the determinants of health

- collect data on the economic, social, demographic, occupational and environmental correlates of health

- increase the understanding of the relationship between health status and health care utilization

4. Survey Content

The first sub-section of this section provides a general discussion of the consultation process used in survey content development and gives a summary of the final content selected for inclusion in this study. The second sub-section describes the common content in detail. A sub-section illustrating the optional module content of the CCHS (Cycle 1.1) follows this.

4.1 Consultation Processes

One of the main CCHS objectives is to address priority health determinants, health status and health system utilization data gaps at the health region level. Topic selection for the content of the CCHS (Cycle 1.1) was conducted through a process of extensive consultations with regional, provincial, federal representatives and the research community. A Canada-wide consultation process with key users of health information was undertaken during June 1999. This process comprised a series of day-long workshops and topic ranking worksheets. Workshop participants included data users at the health-region, provincial health ministry and federal levels, as well as university and hospital researchers and special interest or agency groups. In total more than 17 workshops were conducted with over 225 participants.

Consultations revealed considerable agreement, but also some variation in the content priorities and data needs of stakeholders at different levels of representation. All possible survey topics and sub-topics were ultimately categorized as high, moderate or low priority based on the workshops, worksheets and other discussions with experts in health survey research.

The end result was a questionnaire composed of common and optional content. The common content was made up of survey topics identified as high priority in all health regions. The optional content portion of the questionnaire consisted of survey topics identified as local data needs and were chosen by each individual health region. Both the common and optional survey topics are discussed in more detail in the following sub-sections.

4.2 Common Content

Topics that make up the common content are varied, ranging from Alcohol, Alcohol Dependence & Abuse through to Physical Activities and Two-week Disability. The following table outlines the common content for the CCHS (Cycle 1.1) for the first year of the CCHS, as identified in the cross-Canada content consultation that took place in June 1999.

These common content topics, transformed into survey questions, were asked of all respondents in all health regions. This provides a Canada wide database of health information, which, when used with the appropriate sampling weights, provides the opportunity for a cross-sectional look at health concerns in Canada.

Table 4.1: Common Content Modules

• Alcohol • Alcohol dependence / abuse • Blood pressure check • Breastfeeding • Chronic conditions • Contacts with mental health professionals • Exposure to second hand smoke • Food insecurity • Fruit & vegetable consumption • General health • Health car utilization • Health Utility Index (HUI) • Height / weight • Injuries	• Mammography • PAP smear test • Physical activities • PSA test • Restriction of activities • Smoking • Tobacco alternatives • Two-week disability • Household composition and housing • Income • Labour force • Socio-demographic characteristics • Administration

4.3 Optional Content

The topic content of the optional modules also emerged from the consultation process (see Table 4.2). However, topics were designated as optional so that regions with a need for data or interest in the topics would be able to select the specific topic module for inclusion in the CCHS (Cycle 1.1) in their own region. The advantage of this approach is that health regions can expand the health topic coverage tailored to the characteristics of the regions. The disadvantage is that, unlike the topic modules contained in the common content, the resulting data from the optional content modules is not easily generalized across Canada. Therefore, the size and characteristics of the regions in which the modules are used limit comparison of the results between regions.

Table 4.2: Optional Topic Modules

• Breast examinations • Breast self examinations • Changes made to improve health • Child and adult stressors (traumas) • Dental visits • Depression • Distress • Driving under influence • Drug use • Eye examinations • Flu shots • Home care • Mastery	• Mood • Ongoing problems • Physical check-up • Recent life events • Sedentary activities • Self-esteem • Sexual behaviours • Smoking cessation aids • Social support • Spirituality • Suicidal thoughts and attempts • Use of protective equipment • Work stress

5. Sample Design

5.1 Target Population

The CCHS (Cycle 1.1) targets persons aged 12 years or older who are living in private dwellings in the ten provinces and the three territories. Persons living on Indian Reserves or Crown lands, clientele of institutions, full-time members of the Canadian Armed Forces and residents of certain remote regions are excluded from this survey. The CCHS (Cycle 1.1) covered approximately 98% of the Canadian population aged 12 or older.

5.2 Health Regions

For administrative purposes, each province is divided into health regions (HR) and each territory was designated as a single HR (Table 5.1). Statistics Canada, in consultation with the provinces, has made minor changes to the boundaries of some of the HRs to correspond to the geography of the 1996 Census. Cycle 1.1 of the CCHS collected data in 133 HRs in the ten provinces, in addition to one HR per territory, totalling 136 HRs.

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

Province	Number of HRs	Total sample size (targeted)
Newfoundland	6	4,010
Prince Edward Island	2	2,000
Nova Scotia	6	5,040
New Brunswick	7	5,150
Quebec	16	24,280
Ontario	37	42,260
Manitoba	11	8,000
Saskatchewan	11	7,720
Alberta	17	14,200
British Columbia	20	18,090
Yukon	1	850
Northwest Territories	1	900
Nunavut	1	800
Canada	**136**	**133,300**

5.3 Sample Size and Allocation

To provide reliable estimates to the 136 HRs, and given the budget allocated to the CCHS (Cycle 1.1) component, a sample of 133,300 respondents was desired. Although producing reliable estimates at the HR level 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 two steps, the sample was allocated among the provinces according to their respective populations and the number of HRs they contain (Table 5.1). In the third step, each province's sample was allocated among its HRs proportionally to the square root of the estimated population in each HR.

This three-step approach guaranteed each HR sufficient sample with minimal disturbance to the provincial allocation of sample sizes. The sample sizes were enlarged before data collection to take into account out-of-scope and vacant dwellings and anticipated non-response. (For the complete list of HRs and achieved sample sizes see Section 9 on data quality.)

Note that the three territories were not part of the above allocation strategy as they were dealt with separately. The Yukon was attributed 850 sample units, 900 for Northwest Territories and 800 for Nunavut.

5.4 Frames, Household Sampling Strategies

The CCHS (Cycle 1.1) used three sampling frames to select the sample of households. The majority of the sample of households (83%) came from an area frame. In some HRs, a Random Digit Dialling (RDD) sampling frame or a list frame of telephone numbers was also used. Approximately 7% of the sample of households came from the RDD frame while the list frame generated almost 10% of the sample.

5.4.1 Sampling of Households from the Area Frame

The CCHS (Cycle 1.1) used the area frame designed for the Canadian Labour Force Survey (LFS) as its primary frame. The sampling plan of the LFS is a multistage stratified cluster design in which the dwelling is the final sampling unit[11]. In the first stage, homogeneous strata were formed and independent samples of clusters were

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

drawn from each stratum. In the second stage, dwelling lists were prepared for each cluster and dwellings, or households, were selected from the lists.

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

The other cities and rural regions of each province are stratified first on a geographical basis, then according to socio-economic characteristics. In the majority of strata, six clusters (usually census EAs) are selected using the PPS method. Where there is low population density, a three-step plan is used whereby two or three primary sampling units (PSU), which normally correspond to groups of EAs, are selected and divided into clusters, six of which are sampled. The selection is made at each step using the PPS method.

Once the new clusters are listed, the sample is obtained using a systematic sampling of dwellings. Table 5.2 gives an overview of the types of PSUs used for the entire LFS sample. The *yield* is the number of households selected within the framework of the LFS for a given month. As the sampling rates are determined in advance, there is frequently a difference between the expected sample size and the numbers that are obtained. The yield of the sample, for example, is sometimes excessive. This especially happens in sectors where there is an increase in the number of dwellings due to new construction, for example. To reduce the cost of collection, an excessive output is corrected by eliminating, from the beginning, a part of the units selected and by modifying the weight of the sample design. Such an operation, usually conducted at an aggregate level, is called *sample stabilization*. Moreover, one increases the required size of the sample by households to account for dwellings, experience having shown that 12% of all dwellings are not occupied by households that are part of the field of observation (certain dwellings are vacant or occupied seasonally, others are occupied by households that are not targeted by the survey).

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

Area	Primary Sampling Unit (PSU)	Size (households per PSU)	Yields (sampled households)
Toronto, Montréal, Vancouver	Cluster	200-250	6
Other cities	Cluster	150-200	8
Apartment frame	Apartment	Varies	5
Most rural areas / small urban centres	Enumeration area	300	10

Requirements specific to CCHS (Cycle 1.1) led to some modifications to this sampling strategy[12]. To get a base sample of 97,000 households for CCHS (Cycle 1.1), 123,000 dwellings must be selected from the area frame (to account for vacant dwellings and non-responding households). On an on-going monthly basis the LFS design provides approximately 68,000 dwellings distributed across the various economic regions in Canada whereas the CCHS (Cycle 1.1) required a total of 123,000 dwellings distributed in the HRs, which have different geographic boundaries from those of the LFS economic regions. Overall, the CCHS (Cycle 1.1) required close to double the number of dwellings than those generated by the LFS selection mechanism, or an *adjustment factor* of 1.8 (123,000/68,000). At the HR level, however, the adjustment factors varied from 0.6 to 6.0, which required certain adjustments.

The changes made to the selection mechanism in a HR varied depending on the size of the adjustment factors. For HRs that had a factor smaller than or equal to 1, a simple stabilisation, as described above, was applied to the sample of dwellings. For those with a factor greater than 1 but smaller than or equal to 2, the sampling process of dwellings within a PSU was repeated for all selected PSUs that were part of the same HR. For HRs with a factor greater than 2 but smaller than or equal to 4, the PSU sampling process, as well as that of dwellings in a PSU, was repeated. For HRs with a factor between 4 and 6, the PSU sampling process was repeated not once but twice while that of dwellings was repeated only once. Where the chosen approach created an unnecessary surplus of dwellings, stabilisation was performed.

It should be noted that the changes made to the LFS mechanism resulted in, at most, tripling the number of PSUs selected and, at most, doubling the number of dwellings selected in the PSUs, which explained the maximum adjustment factor of 6.0. At the HR

[12] Morano M., Lessard, S. and Béland, Y. (2000). Creation of a dual frame for the Canadian Community Health Survey, *2000 Proceedings of the Survey Methods* Section, Ottawa: Statistical Society of Canada, 249-254.

level, adjustment factors were purposely capped at 6.0 for two reasons: to limit the listing of clusters (each new selected PSU requires a listing), and to avoid possible cluster effects created by too great a number of dwellings selected in a single PSU. This limit to the adjustment factor of certain HRs has consequently dictated the number of households required from the telephone frames.

Sampling of Households from the Area Frame in the Three Territories

For operational reasons the area frame sample design implemented in the three northern territories had one additional stage of selection. For each territory, in-scope communities were first stratified based on various characteristics (population, geography, percent Inuit and/or Aboriginal and median household income). There were five design strata in Yukon, ten in the Northwest Territories and ten in Nunavut. Then the first stage of selection consisted of randomly selecting one community with a probability proportional to population size within each design stratum. From that point on, the household sampling strategy from the area frame within the selected community was identical as the one described above.

It is worth mentioning that the frame for CCHS (Cycle 1.1) covered 90% of the private households in Yukon, 97% in the Northwest Territories and 90% in Nunavut.

5.4.2 Sampling of Households from the RDD Frame of Telephone Numbers

In some HRs and for some collection months, a Random Digit Dialling (RDD) sampling frame of telephone numbers was used in addition to the area frame. The sampling of households from the RDD frame used the Elimination of Non-Working Banks (ENWB) method, a procedure adopted by the General Social Survey[13]. A hundreds bank (the first eight digits of a ten-digit telephone number) is considered to be non-working if it does not contain any residential telephone numbers. The frame begins as a list of all possible hundreds banks and, as non-working banks are identified, they are eliminated from the frame. It should be noted that these banks are eliminated only when there is evidence from various sources that they are non-working. When there is no information about a bank it is left on the frame. The Canada Phone directory was used in conjunction with various internal administrative files to eliminate non-working banks.

Using available geographic information (postal codes) the banks on the frame were regrouped to create RDD strata to encompass, as closely as possible, the HR areas. Within each RDD stratum, a bank was randomly chosen and a number between 00 and

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

99 was generated at random to create a complete, ten-digit telephone number. This procedure was repeated until the required number of telephone numbers within the RDD stratum was reached. Frequently, the number generated is not in service or is out-of-scope, and therefore many additional numbers must be generated to reach the targeted sample size. This success rate is referred to as the *hit rate* and varies from region to region. Within CCHS (Cycle 1.1) the hit rates ranged from 15% to 61% at the regional level.

To different extents, the RDD frame was used in more than 60 HRs to complement the area frame, and it was solely used in five HRs.

5.4.3 Sampling of Households from the List Frame of Telephone Numbers

As for the RDD frame a list frame of telephone numbers was also used in some HRs for some collection months to complement the area frame. The Canada Phone directory, a commercially available CD-ROM consisting of names, addresses and telephone numbers from telephone directories in Canada, was linked to internal administrative conversion files to obtain postal codes. These were mapped to HRs to create list frame strata. There was one list frame stratum per HR. Within each stratum the required number of telephone numbers were selected using a simple random sampling process from the list. As for the RDD frame, additional telephone numbers were selected to account for the numbers not in service or out-of-scope. The hit rates observed under the list frame approach were much higher than those for the RDD frame as they varied from 50% to 65%.

It is important to mention that the coverage of the list frame is less than the one for the RDD as unlisted numbers do not have a chance of being selected. Nevertheless as the list frame was only used in HRs where the area frame was the main source for the sample the impact of the undercoverage of the list frame was minimal and was dealt with in weighting.

To different extents, the list frame was used in more than 50 HRs to complement the area frame.

5.5 Sampling of Interviewees

Selection of individual respondents was designed to ensure over-representation of youths (12 to 19) and seniors (65 or older). The selection strategy was designed to consider user needs, cost, design efficiency, response burden and operational constraints[14].

Among the households from the area frame, one person aged 12 or older was randomly selected from among 82% of the sampled households and two persons (12 or older) were randomly chosen in the remaining 18%. The rule for selecting persons from households in the area frame was defined as a function of the household composition. The Table 5.3 describes the rule for selecting persons within the area frame sampled households.

Table 5.3. Selection Strategy based on Household Composition - Area Frame Sample

Number of persons aged 12-19	Number of persons aged 20 and over					
	0	1	2	3	4	5+
0	-	A	A	A	A	B
1	A	A	C	C	C	B
2	A	C	C	C	C	C
3+	A	C	C	C	C	C

A: random selection of one person aged 12 and over
B: random selection of two persons aged 12 and over
C: random selection of one person in the age group 12-19 **and** random selection of one person aged 20 and over

For all households from the telephone frames, a single person aged 12 and older was randomly chosen from among all members of the household.

[14] Béland, Y., Bailie, L., Catlin, G. and Singh, M.P. CCHS and NPHS – An Improved Health Survey Program at Statistics Canada, *2000 Proceedings of the American Statistical Association Meeting, Survey Research Methods* Section, Indianapolis: American Statistical Association, 677-682.

5.6 Sample Allocation over the Collection Period

In order to balance interviewer workload and to minimize possible seasonal effects on certain key characteristics such as physical activity, the initial sample of dwellings/ telephone numbers was equally allocated at random, within each HR, over the 12 months of data collection. To start with, each PSU selected in the first stage from the area frame was randomly assigned to a collection quarter (Q1: September to November 2000, Q2: December 2000 to February 2001, Q3: March to May 2001 and Q4: June to August 2001). Within each collection quarter the selected dwellings were then randomly allocated to a collection month. For the telephone frames, independent samples were selected each month. This strategy ensured that each quarterly sample was representative of the Canadian population in scope.

5.7 Supplementary Buy-in Sample in Prince Edward Island

During the course of the data collection, the provincial government of Prince Edward Island provided extra funds so that a larger sample of dwellings could be selected. The purpose of this buy-in was to get sufficient sample size to provide reliable estimates for five sub-provincial areas. The original CCHS (Cycle 1.1) sample design considered two sub-provincial areas. The buy-in sample was combined with the main sample to produce one large file of data. Due to confidentiality reasons, however, only rural and urban regions are reported.

The entire buy-in sample was selected from the list frame of telephone numbers. The Canada Phone directory was linked to internal administrative files in order to stratify the listed telephone numbers in five sub-provincial areas (West Prince, East Prince, Queens, Southern Kings and Eastern Kings). The sample size per sub-provincial area was based upon the funding available and the requirements of the province to obtain reliable estimates by sub-provincial area. An extra sample of 1,300 sample units was added to the planned sample of 2,000 units in Prince Edward Island. The allocation of the 3,300 sample units among the sub-provincial areas was performed using the root-N approach. This allocation scheme balances the reliability requirements at provincial and sub-provincial levels. Table 5.4 gives the sample allocation by sub-provincial area. The data for those extra sample units were collected between May and October 2001.

Table 5.4. Final Sample Allocation including Extra Units in Prince Edward Island

Sub-provincial Area	Sample Size
West Prince	525
East Prince	780
Queens	1,055
Southern Kings	520
Eastern Kings	420
Total	**3,300**

6. Data Collection

6.1 Questionnaire Design and Data Collection Method

The CCHS (Cycle 1.1) questionnaire was administered using computer-assisted interviewing (CAI). Sample units selected from the area frame were interviewed using the Computer-Assisted Personal Interviewing (CAPI) method while units selected from the Random Digit Dialling (RDD) and telephone list frames were interviewed using the Computer-Assisted Telephone Interviewing (CATI) method.

CAI offers a number of data quality advantages over other collection methods. First, question text, including reference periods and pronouns, is customised automatically based on factors such as the age and sex of the respondent, the date of the interview and answers to previous questions.

Second, edits to check for inconsistent answers or out-of-range responses are applied automatically and on-screen prompts are shown when an invalid entry is recorded. Immediate feedback is given to the respondent and the interviewer is able to correct any inconsistencies.

Third, questions that are not applicable to the respondent are skipped automatically.

6.2 Supervision and Control

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

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

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

6.3 Field Tests

Separate CAPI and CATI field tests were conducted during the late spring and early summer of 2000. The test included each of Statistics Canada's 5 Regional Offices.

The main objectives of the CAPI test were to evaluate respondent reaction to the questions and to obtain estimates of completion times for the various sections of the questionnaire. Field operations procedures, interviewer training and the computer-assisted interviewing application were also tested.

The objectives of the CATI test were similar to those of the CAPI test. In addition, the technical infrastructure of the CATI offices and procedures unique to CATI interviewing were tested.

6.4 Interviewing

In all selected dwellings, a knowledgeable household member was asked to supply basic demographic information on all residents of the dwelling. Depending on the composition of the household, and on whether the interviewing method was CAPI or CATI, either one or two members were then selected for a more in-depth interview.

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

In cases where the selected respondent or respondents were either absent for an extended period of time or 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. Therefore, every effort was taken to keep proxy interviews to a minimum.

6.5 Minimising Non-response

Prior to the first contact by an interviewer, an introductory letter and brochure were delivered to each selected dwelling. These explained the importance of the survey and provided examples of how CCHS (Cycle 1.1) data would be used.

Interviewers were instructed to make all reasonable attempts to obtain CCHS (Cycle 1.1) interviews. When the timing of the interviewer's call (or visit) was inconvenient, an

appointment was made to call back at a more convenient time. If no one was home, numerous call-backs were made. For individuals who at first refused to participate in the CCHS (Cycle 1.1), a letter was sent from the Regional Office to the respondent, stressing the importance of the survey and the household's collaboration. This was followed by a second call (or visit) from a senior interviewer, a project supervisor or another interviewer to try to convince respondents of the importance of participating in the survey. During the final months of data collection, non-response cases and selected persons who had previously refused were again approached and encouraged to participate in the survey. This diligence in contact may have resulted in stronger survey results by maximising the response rate.

To remove language as a barrier to conducting interviews, each of the Statistics Canada Regional Offices has recruited interviewers with a wide range of language competencies. When necessary, cases were transferred to an interviewer with the language competency needed to complete an interview. In addition, the survey questions were translated into the following languages: Chinese, Punjabi and Inuktitut.

6.6 Special Circumstances during CCHS (Cycle 1.1) Collection Operations

The initial plan called for data collection between September 2000 and early October 2001, a period of 13 months. This plan was carefully designed to ensure that the survey's quality objectives would be met. To even out the interviewers' workload and eliminate any seasonal effects, the final sample was randomly divided in 12 so that each month of the year would be properly represented for each HR. A 13th month of collection was planned to provide interviewers with an opportunity for a final attempt to convert non-responding cases.

For most of Statistics Canada's household surveys, collection operations proceed smoothly and within the established parameters. For CCHS (Cycle 1.1), the total workload imposed by the large sample size proved to be a challenge for the data collection infrastructure in place. To ensure the success of collection operations, a number of established procedures were altered, some more than others[15]. Among those procedures, adding a 14th month of collection, transferring caseloads from a CATI office to another and introducing a new sampling approach to select telephone numbers were the most important. At the end of data collection, a national response rate of 84.7% was achieved. The reader will find complete details regarding the response rates in Section 9.

[15] Béland, Y. , Dufour, J. and Hamel, M. (2001). Preventing non-response in the Canadian Community Health Survey, *Proccedings of Statistics Canada's Symposium 2001*, Statistics Canada.

7. Data Processing

7.1 Editing

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

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

7.2 Coding

Pre-coded answer categories were supplied for all suitable variables.

Several questions in the CCHS (Cycle 1.1) questionnaire allow write-in responses. For some of these questions, write-in responses were coded into one of the existing listed categories if the write-in information duplicated a listed category.

7.3 Creation of Derived and Grouped Variables

To facilitate data analysis, a number of variables on the file have been derived using items found on the CCHS (Cycle 1.1) questionnaire. Derived variables generally have a "D" or "G" in the fifth character of the variable name. In some cases, the derived variables are straightforward, involving collapsing of response categories. In other cases, several variables have been combined to create a new variable. Appendix D provides details on how these more complex variables were derived.

7.4 Imputation

Because of their private or sensitive nature, many CCHS (Cycle 1.1) questions/questionnaire modules were appropriate for self-response only, and were skipped when the questionnaire was answered by proxy respondents. During data collection, an unexpectedly high proportion of interviews were completed by proxy. Proxy interviews were allowed only if it was confirmed that the selected respondent would not be present for the entire collection period, in cases of mental or physical incapacity preventing an interview to take place, or for language barrier. At the end of

data collection, 6.3% of all interviews were completed by proxy; the rates varied from 2 to 23% at the health region level. Consequently, important information was missing for the individuals represented in those interviews. This represented approximately one third of the questionnaire. Among the common questionnaire modules ten were entirely skipped and two were partially skipped. Among the list of optional questionnaire modules, 21 were skipped.

To fill in these missing responses, values were imputed using the "nearest neighbour" imputation method[16]. This method was only used to fill in the proxy interviews; it was not used for cases of total or partial non-responses obtained in non-proxy interviews. Data from a "non-proxy respondent" with similar characteristics was used as a donor and information from that record was copied to the record with missing data. This method was applied within defined imputation classes. The nearest neighbour was found based on a specific distance function which used relevant information available for both proxy and non-proxy respondents. In cases where data quality could not be improved through imputation, responses were left coded as missing.

The following modules were entirely imputed:

- Blood pressure
- Dental visits
- Eye examinations
- Contact with mental health professionals
- Alcohol dependence
- Driving under influence
- Social support
- Depression
- Suicide thoughts and attempts
- Sexual behaviours
- Fruit and vegetable consumption

The following modules were partially imputed:

- PAP smear test (PAPA_020 only)
- PSA test (PSAA_170 only)
- Mammography (MAMA_30, MAMA_37 and MAMA_38)

[16] St-Pierre, M. and Béland, Y. (2002). Imputation of proxy respondents in the Canadian Community Health Survey, *2002 Proceedings of the Survey Methods Section,* Statistical Society of Canada. In press.

- Flu shots (FLUA_160 only)
- Breast examinations (BRXA_110 only)
- Breast self-examinations (BSXA_120 only)
- Height and weight (HWTA_4 only)

The following modules which were skipped during a proxy interview were not imputed:

- Physical check-up
- Smoking cessation aids
- General health
- Self-esteem
- Mastery
- Spirituality
- Mood
- Distress
- Work stress
- Physical activities
- Sedentary activities
- Use of protective equipment
- Changes made to improve health
- Breastfeeding
- Patient satisfaction

7.5 Weighting

The principle behind estimation in a probability sample such as the CCHS (Cycle 1.1) is that each person in the sample "represents", besides himself or herself, several other persons not in the sample. For example, in a simple random 2% sample of the population, each person in the sample represents 50 persons in the population. In the terminology used here, it can be said that each person has a weight of 50.

The weighting phase is a step that calculates, for each person, his or her associated sampling weight. This weight appears on the microdata file, and must be used to derive meaningful estimates from the survey. For example, if the number of individuals who smoke daily is to be estimated, it is done by selecting the records referring to those individuals in the sample having that characteristic and summing the weights entered on those records.

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

7.6 Suppression of Confidential Information

It should be noted that the 'Public Use' microdata files described above differ in a number of important respects from the survey 'master' files held by Statistics Canada. These differences are the result of actions taken to protect the anonymity of individual survey respondents. Protection of respondents is assured through suppression of individual values, variable grouping, and variable capping. Users requiring access to information excluded from the microdata files have three options: to purchase custom tabulations, use one of the Research Data Centres[17], or use the remote access option. (See sub-section 12.3)

[17] The most current information about the Research Data Centres can be found at www.statcan.ca

8. Weighting

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

As described in Section 5, CCHS (Cycle 1.1) had recourse to three sampling frames for its sample selection: an area frame acting as the primary frame, and two frames formed of telephone numbers complementing the area frame. Since only minor differences differentiate the two frames formed of telephone numbers in terms of weighing, they are treated together. They are referred to as being part of the telephone frame.

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

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

Diagram A: Weighting Strategy Overview

Area Frame	Telephone Frame
A0 – Initial weight	T0 – Initial weight
A1 – Sample increase	T1 – Coverage of the list frame
A2 – Stabilization	T2 – Number of months
A3 – Removal of out-of-scope units	T3 – Removal of out-of-scope numbers
A4 – Household nonresponse	T4 – Combination of RDD and list frames
A5 – Creation of person level weight	T5 – Household nonresponse
A6 – Person nonresponse	T6 – Households without telephone
Final area frame weight	T7 – Creation of person level weight
↘	T8 – Person nonresponse
	T9 – Multiple lines
	Final telephone frame weight
	↙
I1 – Integration	
I2 – Seasonal effect	
I3 - Post-stratification	
Final CCHS (Cycle 1.1) weight	

8.1 File weighting

As mentioned previously, units from both area and telephone frames are treated separately up to the integration step (I1). Sub-section 8.1.1 provides details on the weighting strategy for the area frame, while sub-section 8.1.2 deals with the strategy for the telephone frame. The integration of the two frames is discussed in 8.1.3. This is followed by the two last weighting steps, that is, the adjustment controlling for the seasonal effect and the post-stratification, which are explained in sub-sections 8.1.4 and 8.1.5 respectively.

Although these two frames were used to cover the three territories, some modifications had to be done relative to their use. These modifications affected the weighting of these three regions substantially, and they are reported in sub-section 8.1.6.

8.1.1 Weighting of the area frame sample

A0 – Initial weight

Since the mechanism established for the Labour Force Survey (LFS) was used to select the area frame sample, the initial weights had to be computed with respect to that

mechanism. First, within each stratum defined by the LFS, clusters (primary units) are selected with probabilities proportional to population sizes (based on 1991 Census counts). Next, dwellings are sampled within each selected cluster using a systematic sampling. The product of the probabilities for each of these selections represents the overall probability of selection, and the inverse of that probability is used as the CCHS (Cycle 1.1) initial weight. For more details about the selection mechanism, as well as a more complete definition of strata and clusters, refer to Statistics Canada (1998)[18].

A1 – Sample increase

Some modifications were made to the default LFS mechanism at the time of sample selection for CCHS (Cycle 1.1). The LFS design provides approximately 68,000 dwellings nationally, while CCHS (Cycle 1.1) requirements in terms of sample size were almost twice that number. Modifications made in order to obtain the needed sample within a HR consisted, in summary, of repeating the sampling process of dwellings within all selected clusters of the HR. This modification had the effect of boosting the sample and had to be accounted for in the weighting to correctly represent the probability of selection. An adjustment factor A1 representing the sample increase rate was calculated. The initial weight calculated in A0 is multiplied by this adjustment factor, which results in the weight A1.

A2 – Stabilization

In some HRs, increasing the sample as described in the previous paragraph resulted in a significantly larger sample than necessary. Stabilization was therefore instituted to bring the sample size back down to the desired level. The stabilization process consisted of randomly subsampling dwellings at the HR level. An adjustment factor representing the effect of this stabilization was therefore calculated to adjust the probability of selection appropriately. This factor, multiplied by the weight A1, produces the weight A2.

A3 – Removal of out-of-scope units

Among all dwellings sampled, a certain proportion of them is identified during collection as being out-of-scope. Dwellings demolished or in construction, vacant, seasonal or secondary dwellings, and institutions are examples of out-of-scope cases for CCHS (Cycle 1.1). Records for these dwellings were simply removed from the sample, leaving only in-scope dwellings. They kept the same weight as in the previous step, which is now called weight A3.

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

A4 – Household nonresponse

During collection, a certain proportion of interviewed households inevitably resulted 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 nonresponding households were distributed to respondents using response propensity classes. The CHAID (Chi-Square Automatic Interaction Detector) algorithm available in Knowledge Seeker[19] was used to identify which characteristics best split the sample into groups that were dissimilar with respect to response/nonresponse. Note that groups were formed independently within each HR. Since the information available for nonrespondents is limited, only characteristics such as province, collection period and a rural/urban indicator could be used in the creation of the classes. Results actually showed that only the collection period characteristic (with 5 periods; Sept. to Nov. 2000 / Dec. 2000 to Feb. 2001 / March to May 2001 / June to August 2001 / Sept. to Oct. 2001) was significant in the creation of classes for each HR. An adjustment factor was therefore calculated within each class as follows:

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

The weight A3 for responding households was multiplied by this factor to produce the weight A4. Nonresponding households were dropped out of the process at this point.

A5 – Creation of person level weight

Since the ultimate sampling unit for the CCHS is a person, the household level weights computed up to this point need to be converted down to the person-level. The factor calculated at this step also incorporates the adjustment necessary to account for the fact that one or two persons could have been selected within each household. The adjustment factor A5 was therefore based on the number of persons in the household (this information is collected from the roster of all household members), the distribution of this number among the 12-19 and 20+ age groups, and the number of persons selected. Consult table 5.3 to obtain more details on the algorithm used to determine the number of persons to be selected within each household.

For selected persons from households where only one person was selected, the adjustment simply consists of the number of household members over 12 years old. For

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

cases where one person in the 12-19 age group and one in the 20+ age group were selected, the adjustments comprise the number of household members in the 12-19 age group and the number in the 20+ age group respectively. Finally, for cases where two people were selected without constraints on the age (but still 12+), the adjustment for each person was half the number of household members over 12 years old. The household level weight obtained in A4 was multiplied by the adjustment factor derived here to result into the person level weight A5.

A6 – Person nonresponse

A CCHS (Cycle 1.1) interview can be seen as a two-part process. First the interviewer gets the complete roster of the people living within the responding household. Second, (s)he interviews the selected person(s) within the household. 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 cases are defined as person nonresponse and an adjustment factor must be applied to the weights of respondents to overcome this nonresponse. As for the treatment of household nonresponse, the adjustment was applied within classes based on characteristics available for both respondents and nonrespondents. All characteristics collected when rostering all household members were in fact available for the creation of the classes. The CHAID algorithm was once again used to define the classes and the final result presented definitions that varied from one HR to the other. Depending on the HR, the following characteristics were used to form the adjustment classes: sex, age group, urban/rural indicator, education, marital status and the number of persons selected in the household. As a result, an adjustment factor is calculated as follows:

$$\frac{\textit{Sum of weight A5 for all selected persons}}{\textit{Sum of weight A5 for all responding selected persons}}$$

The weight A5 of responding persons was therefore multiplied by this adjustment factor to produce the weight A6. Nonresponding persons are dropped out of the weighting process from this point onward.

Since this adjustment was the last one necessary for the sample drawn from the area frame, the weight A6 represents the ***final area frame weight***. This weight was later integrated with the final weight of the telephone frame to create the final CCHS (Cycle 1.1) weight.

8.1.2 Weighting of the telephone frame sample

As mentioned previously, the telephone frame is composed of two frames: a random digit dialling frame (RDD), and a list frame. However, units coming from these two frames are treated together and therefore are subject to the same adjustments. There are two exceptions; first, since the probability of selection is relative to the frame used for the selection, this probability will be slightly different depending on whether the unit is from the RDD frame or the list frame. The other exception concerns the adjustment T1. Details about these two exceptions are given in the sub-sections presenting the adjustments implicated.

Another aspect particular to units coming from the telephone frame affects the way the sample was weighted. This particularity concerns the geographical location of sampled units. The geography used to select the sample for the telephone frame did not perfectly replicate the HR geography, which caused some units to be selected from one location while the information collected at the time of the interview was locating them in a neighbouring region. This particularity was dealt with in the weighting by applying all adjustments relative to the HR assigned at the time of sample selection. However, since all units had to belong to their right HR, that is, the HR identified during collection, all unit weights were adjusted according to the correct HR from sample selection. This adjustment was incorporated in the post-stratification (I3), which is described later in this section.

T0 –Initial weight

The initial weight is computed slightly differently between the RDD and List frame samples. Both are defined as the inverse probability of selection, but the methods of selection, and therefore the probabilities, differ. For the RDD, the selection of numbers is done within each RDD stratum. A RDD stratum is an aggregation of area code prefixes (ACP; the first six digits of a 10-digit number), each containing valid banks of one hundred numbers (see Norris and Paton[20] for more details). Therefore, the probability of selection is the ratio of the number of sampled units to one hundred times the number of banks within the RDD stratum.

For the list frame, telephone numbers are selected among all numbers available on the list, within the HR for which the unit is selected. Hence, the probability of selection corresponds to the ratio of the number of sampled units to the number of telephone

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

numbers in the list within the HR. The inverse of these probabilities represents the initial weight T0.

T1 – Coverage of the list frame

Since the list frame does not cover some phone numbers, which are actually covered by the RDD frame, an adjustment had to be applied to the initial weights of the list frame units to make both frames comparable in terms of coverage. The adjustment consisted of inflating the weights of the list frame units by the amount of undercoverage, individually for each HR. Estimating the undercoverage was one of the most challenging tasks and was done using the data collected from the CCHS (Cycle 1.1) area frame sample. For all people interviewed via the area frame, the questionnaire included a set of questions verifying if the household had a telephone, how many residential lines it had, and the phone number for each line. The desired coverage rate was derived by simply computing the percentage of all collected numbers that were present in the list frame. The inverse of this rate represents the factor used for this adjustment. The factor, once multplied by the initial weight T0, resulted in the weight T1.

T2 – Number of months

Contrary to the area frame where the entire sample was selected at the beginning of the sampling process, samples were drawn monthly for the telephone frame. Each of these monthly samples came with an initial weight that made each sample representative at the HR level. However, to ensure that the total sample would represent the population only once, an adjustment factor had to be applied to reduce the weights of each monthly sample. The adjustment factor applied to each monthly sample was equal to the proportion the monthly sample represented among the total sample. Note that this adjustment was done separately for both RDD and list frames, which means that the sample from each of the two frames represented the total population. To correct this situation, RDD and list samples were later combined (in step T4) in such a way that the telephone frame total sample would represent the total population only once. Therefore, the weight T2 was obtained by multiplying the weight T1 by the factor defined above.

T3 - Removal of out-of-scope numbers

Telephone numbers leading to businesses, institutions or other out-of-scope dwellings, as well as numbers not in service or any other non-working numbers, are all examples of out-of-scope cases for the telephone frame. As for the area frame, these cases were simply removed from the process, leaving only in-scope dwellings in the sample. These in-scope dwellings kept the same weight as in the previous step, now called weight T3.

T4 - Combination of RDD and list frames

Up to this step, the RDD and list frames samples both represent the entire population of the HR where they were used. They both had to be combined so that together they would represent the total population only once. An adjustment factor was applied to do so, and it was based exclusively on the size of the samples used in each frame. For RDD units, the factor represented the proportion of the telephone frame sample coming from the RDD frame. The complement of this proportion represented the factor used for the list frame units. These factors were calculated and applied independently within each HR where the two frames were used. Consequently, the weight T4 was obtained by multiplying the weight T3 by the combining factor.

T5 – Household nonresponse

The adjustment applied here to compensate for the effect of household nonresponse is identical to the one applied for the area frame (adjustment A4). As for A4, the only significant characteristic explaining the nonresponse was the collection period, which was then used to define the adjustment classes. The adjustment factor calculated within each class was obtained as follows:

$$\frac{\textit{Sum of weights T4 for all households}}{\textit{Sum of weights T4 for all responding households}}$$

The weight T4 of responding households was multiplied by this factor to produce the weight T5. Nonresponding households are removed from the process at this point.

T6 - Households without telephone

A certain proportion of the Canadian population does not have access to a private residential telephone line. As explained in step T1, information about the presence of a telephone was collected for the area frame sample, which was used here to estimate the proportion of households without a phone line at the HR level. Similarly to T1, the telephone frame sample weights were inflated based on proportions observed using the area frame data, adjusting the weights for the undercoverage of the frame for this uncovered sub-population. The factor used for this adjustment corresponded to the inverse of the estimated proportion, and once multiplied by the weight T5, resulted in weight T6.

T7 – Creation of person level weight

As for adjustment A5, this adjustment converts the household level weight to a person level weight. Unlike the area frame, only one person was selected per household for the telephone frame, hence the adjustment factor was relatively simple; it represented the total number of in-scope persons within the household of the person selected. This factor, multiplied by the weight T6, gave the weight T7.

T8 - Person nonresponse

This adjustment was similar to the adjustment A6 used for the area frame. It consisted of compensating for the effect of nonresponse at the person level. As for A6, an approach based on adjustment classes was used, where classes were defined from variables available for all selected persons, respondent or not (see A6 for the list of variables available). Within each class, an adjustment factor was calculated as follows:

$$\frac{\textit{Sum of weights T7 for all selected persons}}{\textit{Sum of weights T7 for all responding selected persons}}$$

The weight T7 of responding persons was therefore multiplied by this adjustment factor to produce the weight T8. Nonresponding persons were dropped out of the weighting process at this point.

T9 – Multiple lines

Some households can possess more than one residential telephone line. This has an impact on the weighting; having more lines translates into having a higher probability of being selected. Therefore, the weights needed to be adjusted for the number of residential telephone lines the household had. Note that this information was obtained during the early stage of the interview from the selected person. The adjustment factor represented the inverse of the number of lines. The weight T9 was therefore obtained by multiplying this factor by the weight T8.

Since this adjustment was the last one for the sample drawn from the telephone frame, the weight T9 represents the ***final telephone frame weight***. This weight was later integrated, in step I1, with the final area frame weight to finally create the final CCHS (Cycle 1.1) weight.

8.1.3 Integration of the area and telephone frames(I1)

This step consisted in integrating the final area and telephone frame sampling weights created until now, into a single weight, by applying a method of integration[21]. An adjustment factor between 0 and 1 was determined in such a way that it represented the relative importance of each sample in the total sample. This relative importance was measured in terms of sample size and design effect. The larger the proportion a sample represented in the total sample was, the higher was its relative importance in the total sample. For the design effect, the relative importance was bigger for units coming from the frame that had the smallest design effect. To obtain the integration adjustment factor, a factor α was first calculated as follows:

$$\alpha = \frac{n_A}{R} \Big/ \left(\frac{n_A}{R} + n_T \right)$$

where n_A and n_T represent the area and telephone frames sample sizes respectively, while R represents the median ratio of the design effects observed for each frame. The weight of the area frame units was multiplied by this factor α, while the weight of the telephone frame units was multiplied by 1- α. Note that in the case where a HR was covered by only one frame, the adjustment factor was equal to 1. The product between the factor derived here and the final weight calculated earlier (A6 or T9 depending on which frame the unit belongs to), gave the integrated weight I1.

8.1.4 Seasonal effect(I2)

The CCHS (Cycle 1.1) had initially planned to allocate the data collection equally throughout ten months of the survey's reference year, partly to control for the seasonal effect in the data collected. However, some events affected these plans, with the result that an additional adjustment had to be added to ensure that there was no seasonal effect in the estimates produced using CCHS (Cycle 1.1) data.[22] The adjustment applied in I2 was done so that the sum of the weights of all units interviewed during one of the four seasons would represent exactly 25 % of the total sum of weights. In other words, after applying the adjustment, the portion of the sample interviewed each season represented 25 % of the total population for each HR.

---

[21] 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.

[22] Béland, Y. , Dufour, J. and Hamel, M. (2001). Preventing non-response in the Canadian Community Health Survey, *Proccedings of Statistics Canada's Symposium 2001*, Statistics Canada.

The four seasons defined for the CCHS (Cycle 1.1) are the periods covering September to November, December to February, March to May, and June to August. The adjustment factor I2 used to control the seasonal effect for a person interviewed during season *S*, is defined as:

$$\frac{\textit{Sum of weights I1 for the total sample}}{4 \times \textit{sum of weights I1 for the sample interviewed during season S}}$$

This seasonal adjustment applied to the weight I1 results in the weight I2.

8.1.5 Post-stratification(I3)

The final step necessary to obtain the final CCHS (Cycle 1.1) weight was the post-stratification. Post-stratification is done to ensure that the sum of the final weights corresponds to the population estimates defined at the HR level, for all 10 age-sex groups of interest, that is, the five age groups 12-19, 20-29, 30-44, 45-64, 65+, for both males and females. Note that for British Columbia, the post-stratification was done using a revised geography that contained 16 regions instead of the 20 initially used at the design stage and throughout data collection.

The population estimates were based on the 1996 Census counts and estimates of birth, death, immigration and emigration counts. The average of these monthly estimates for each of the HR-age-sex post-strata was used to post-stratify. The weight I2 was therefore adjusted to obtain the final weight I3 with the help of the adjustment factor I3 defined as follows:

$$\frac{\textit{Population estimate for the HR - age - sex group of the respondent}}{\textit{Sum of weights I2 for the HR - age - sex group of the respondent}}$$

Consequently, the weight I3 corresponds to the ***final CCHS (Cycle 1.1) weight*** that can be found on the data file with the variable name WTSAM.

8.1.6 Particular aspects of the weighting in the three territories

As described in Section 5, the sampling frame used in the three territories was somewhat different from the one used in the ten provinces. Therefore, the weighting strategy had to be adapted to comply with these differences. This section summarises the changes applied to the steps described in sub-sections 8.1.1 to 8.1.5.

For the area frame, as mentioned in sub-section 5.4.1, an additional stage of selection was added in the territories. Each territory was initially stratified into groupings of communities, where one community was selected within each group. Note that the capital of each territory formed a stratum on its own, and was consequently automatically selected at this first stage. This particularity only had an effect in the computation of the probability of selection, and therefore in the value of the initial weight (A0). Once the initial weight was calculated, the same series of adjustments (A1 to A6) was applied to the area frame units. Household-level and person-level nonresponse adjustment classes were built in the same way as for the provinces, using the same set of variables available. Only the definition of the collection periods was modified to better reflect the collection process that started in November 2000 in the territories. The four periods used were defined as November 2000 to February 2001, March to May 2001, June to August 2001, and September to October 2001.

For the weighting of the telephone frame units, let us first mention that only the RDD frame was used for the territories, and exclusively in the capitals. Consequently, this eliminated the need of adjustments T1 (coverage of the list frame) and T4 (combination of RDD and list frames). All other adjustments were applied. Similarly to the area frame part, the definition of the collection periods was modified for the nonresponse adjustments. Finally, adjustment A6 (household without telephone lines) was also subject to a slight modification since the RDD frame was used only in the capitals. The proportions of households without telephone lines were derived, as for the provinces, using the area frame data, but by excluding the data from households located outside the capitals from the calculations.

The two sets of weights (area and telephone) were subsequently integrated, then adjusted for the seasonal effect, and finally poststratified in a similar way to what was done for the provinces, with the exception of two details. First, the integration was applied only to units located in the capitals, the other communities having been covered only by the telephone frame. The second detail relates to the seasonal adjustment. Because a strong concentration of the interviews was conducted during a short period of time in the Nunavut territory, the seasonal effect adjustment could not be applied efficiently. Estimations produced for the Nunavut using these weights will therefore not account for a possible seasonal effect in the data.

9. Data Quality

9.1 Response Rates

In total and after removing the out-of-scope units, 136,937 households were selected to participate in the CCHS (Cycle 1.1). Out of these selected households a response was obtained for 125,159 which results in an overall household-level response rate of 91.4%. Among these responding households 142,421 individuals were selected to participate in the CCHS (Cycle 1.1) out of which a response was obtained for 130,827 which results in an overall person-level response rate of 91.9%. At the Canada level, this would yield a combined response rate of **84.7%** for the CCHS (Cycle 1.1). It should be noted that because of the selection of two persons in some households the combined response rate is not obtained by multiplying the household and the person-level response rates. Table 9.1 provides combined response rates as well as relevant information for calculation of them by health region or combined health region.

In British Columbia, CCHS (Cycle 1.1) collection was conducted using health region boundaries as they existed in 2000-2001, there were 20 of them. A subsequent reorganization of boundaries in this province resulted in 16 new health regions being reported in this PUMF which are different than those under which collection took place in 2000-2001. As a result, it is not appropriate to report response rates for the new regions.

It is also important to note that, for the other provinces/territories, there might be discrepancies between the figures reported in Table 9.1 and the actual record counts found on the PUMF. Response rates are reported based on the design geography and some units might have been re-located in a different health region during data processing.

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

Household-level response rate

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

Person-level response rate among the 2-person area frame households

$$PPRR/A2 = \frac{\text{\# of responding persons in 2-person area frame households}}{\text{all selected persons in 2-person area frame households}}$$

Person-level response rates among the 1-person area and phone frames households

$$\text{PPRR/A1, PPRR/P1} = \frac{\text{\# of responding persons in 1-person households}}{\text{all selected persons in 1-person households}}$$

In order to accurately compute a combined response rate one should calculate the following three ratios (these ratios represent the "importance", at the household level, of each component in the combined response rate).

Ratio for the 2-person area frame households

$$\text{R/A2} = \frac{\text{\# of responding households (2-person) of the area frame}}{\text{all responding households of both frames}}$$

Note: it is of importance to mention that the "*# of responding households (2-person) of the area frame*" is obtained by dividing by 2 the number of selected persons of the 2-person area frame households found in Table 9.1.

Ratio for the 1-person area frame households

$$\text{R/A1} = \frac{\text{\# of responding households (1-person) of the area frame}}{\text{all responding households of both frames}}$$

Note: the "*# of responding households (1-person) of the area frame*" is the same as the number of selected persons of the 1-person area frame households found in Table 9.1.

Ratio for the phone frames households

$$\text{R/P} = \frac{\text{\# of responding households of the phone frames}}{\text{all responding households of both frames}}$$

Once all the above components have been calculated one would obtain the combined response rate using the following formulae.

Combined response rate

COMB/RR = HHRR * [(R/A2 * PPRR/A2) + (R/A1 * PPRR/A1) + (R/P * PPRR/P1)]

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

$$\text{HHRR} = \frac{100{,}396 + 24{,}763}{109{,}315 + 27{,}622} = \frac{125{,}159}{136{,}937} = 0.914$$

$$\text{PPRR/A2} = \frac{29{,}777}{34{,}524} = 0.863$$

$$\text{PPRR/A1} = \frac{78{,}129}{83{,}134} = 0.940$$

$$\text{PPRR/P1} = \frac{22{,}921}{24{,}763} = 0.926$$

$$\text{R/A2} = \frac{(34{,}524 \div 2)}{100{,}396 + 24{,}763} = \frac{17{,}262}{125{,}159} = 0.138$$

$$\text{R/A1} = \frac{83{,}134}{100{,}396 + 24{,}763} = \frac{83{,}134}{125{,}159} = 0.664$$

$$\text{R/P} = \frac{24{,}763}{100{,}396 + 24{,}763} = \frac{24{,}763}{125{,}159} = 0.198$$

then COMB/RR = 0.914 * [(0.138 * 0.863) + (0.664 * 0.940) + (0.198 * 0.926)]

= 0.914 * [0.1191 + 0.6242 + 0.1833]

= 0.8469

= **84.7%**.

Table/Tableau 9.1		Area Frame / Base aréolaire										Phone Frames / Bases téléphoniques						All cases Tous les cas
					2 persons select. 2 personnes sélect.			1 person select. 1 personne sélect.										
Prov. Terr.	Health Region Région socio sanitaire	# In Scope HH # Mén. cibles	# Resp HH # Mén. rép.	HH Resp. Rates Taux de rép. mén.	# Pers. Select. # Pers. sélect.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	# Pers. Select. # Pers. sélect.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	Resp. Rates Taux de rép.	# In Scope HH # Mén. cibles	# Resp HH # Mén. rép.	HH Resp. Rates Taux de rép. mén.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	Resp. Rates Taux de rép.	Combined Resp. Rates Taux de rép. Combiné
CA	Total	109315	100396	91.8	34524	29777	86.3	83134	78129	94	85.1	27622	24763	89.6	22921	92.6	83.0	84.7
NL	Total	3453	3267	94.6	1384	1211	87.5	2575	2385	92.6	86.6	307	295	96.1	274	92.9	89.3	86.8
	10901	847	803	94.8	284	250	88	661	642	97.1	90.6	.	.	.	.	.	.	90.6
	10902	754	721	95.6	310	281	90.6	566	529	93.5	88.8	.	.	.	.	.	.	88.8
	10903	689	645	93.6	280	251	89.6	505	458	90.7	84.7	.	.	.	.	.	.	84.7
	10904*	1163	1098	94.4	510	429	84.1	843	756	89.7	83.4	307	295	96.1	274	92.9	89.3	84.7
PE	Total	1689	1581	93.6	552	483	87.5	1305	1239	94.9	87.7	2335	2074	88.8	1929	93	82.6	84.7
	11901	955	874	91.5	242	210	86.8	753	731	97.1	87.5	514	455	88.5	429	94.3	83.5	86.1
	11902	734	707	96.3	310	273	88.1	552	508	92	87.8	1821	1619	88.9	1500	92.6	82.4	83.9
NS	Total	4540	4265	93.9	1482	1313	88.6	3524	3373	95.7	88.8	709	670	94.5	633	94.5	89.3	88.8
	12901	856	809	94.5	270	244	90.4	674	651	96.6	90.3	113	106	93.8	105	99.1	92.9	90.6
	12902	406	393	96.8	122	109	89.3	332	312	94	90.3	306	286	93.5	267	93.4	87.3	89.0
	12903	742	711	95.8	248	229	92.3	587	570	97.1	92.3	.	.	.	.	.	.	92.3
	12904	654	614	93.9	242	214	88.4	493	469	95.1	88.1	3	3	100	3	100	100.0	88.1
	12905	554	526	94.9	194	160	82.5	429	405	94.4	87.5	287	275	95.8	258	93.8	89.9	88.3
	12906	1328	1212	91.3	406	357	87.9	1009	966	95.7	86.2	.	.	.	.	.	.	86.2
NB	Total	4743	4461	94.1	1534	1342	87.5	3694	3521	95.3	88.4	144	141	97.9	133	94.3	92.4	88.5
	13901	949	897	94.5	288	253	87.8	753	728	96.7	90.0	.	.	.	.	.	.	90.0
	13902	890	831	93.4	304	265	87.2	679	649	95.6	87.8	.	.	.	.	.	.	87.8
	13903	907	844	93.1	266	214	80.5	711	659	92.7	84.5	.	.	.	.	.	.	84.5
	13904*	925	871	94.2	330	309	93.6	706	682	96.6	90.4	81	78	96.3	73	93.6	90.1	90.4

Table/Tableau 9.1		Area Frame / Base aréolaire										Phone Frames / Bases téléphoniques						All cases Tous les cas
					2 persons select. 2 personnes sélect.			1 person select. 1 personne sélect.										
Prov. Terr.	Health Region Région socio sanitaire	# In Scope HH # Mén. cibles	# Resp HH # Mén. rép.	HH Resp. Rates Taux de rép. mén.	# Pers. Select. # Pers. sélect.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	# Pers. Select. # Pers. sélect.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	Resp. Rates Taux de rép.	# In Scope HH # Mén. cibles	# Resp HH # Mén. rép.	HH Resp. Rates Taux de rép. mén.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	Resp. Rates Taux de rép.	Combined Resp. Rates Taux de rép. Combiné
	13905*	1072	1018	95	346	301	87	845	803	95	88.9	63	63	100	60	95.2	95.2	89.3
QC	Total	20999	19316	92	5656	4970	87.9	16488	15512	94.1	85.7	1773	1587	89.5	1477	93.1	83.3	85.5
	24901	1019	997	97.8	304	261	85.9	845	779	92.2	89.3	95	87	91.6	85	97.7	89.5	89.3
	24902	1117	1046	93.6	426	338	79.3	833	776	93.2	84.6	.	.	.	.	.	.	84.6
	24903	1765	1643	93.1	336	279	83	1475	1372	93	85.6	.	.	.	.	.	.	85.6
	24904	1553	1477	95.1	378	352	93.1	1288	1241	96.4	91.2	.	.	.	.	.	.	91.2
	24905	1208	1114	92.2	302	271	89.7	963	903	93.8	86.0	.	.	.	.	.	.	86.0
	24906	3090	2622	84.9	606	540	89.1	2319	2180	94	79.3	.	.	.	.	.	.	79.3
	24907	1204	1085	90.1	320	298	93.1	925	872	94.3	84.8	.	.	.	.	.	.	84.8
	24908	1061	991	93.4	326	305	93.6	828	804	97.1	90.2	172	160	93	151	94.4	87.8	89.8
	24909	875	796	91	278	256	92.1	657	626	95.3	86.2	278	256	92.1	243	94.9	87.4	86.5
	24911	840	809	96.3	256	229	89.5	681	622	91.3	87.7	399	373	93.5	351	94.1	88.0	87.8
	24912	1347	1268	94.1	374	347	92.8	1081	1054	97.5	91.1	10	8	80	8	100	80.0	91.0
	24913	870	773	88.9	244	170	69.7	651	530	81.4	70.7	432	377	87.3	341	90.5	78.9	73.4
	24914	1274	1213	95.2	388	345	88.9	1019	972	95.4	89.8	242	209	86.4	189	90.4	78.1	88.0
	24915	1308	1203	92	414	365	88.2	996	949	95.3	86.5	145	117	80.7	109	93.2	75.2	85.4
	24916	2468	2279	92.3	704	614	87.2	1927	1832	95.1	86.7	.	.	.	.	.	.	86.7
ON	Total	32024	29117	90.9	10398	8678	83.5	23918	22175	92.7	82.8	10593	9151	86.4	8425	92.1	79.5	82.0
	35926	646	607	94	184	163	88.6	515	488	94.8	88.2	224	181	80.8	162	89.5	72.3	84.1
	35927	583	509	87.3	182	147	80.8	418	385	92.1	78.6	282	240	85.1	222	92.5	78.7	78.7
	35930	1457	1313	90.1	496	388	78.2	1065	987	92.7	81.1	.	.	.	.	.	.	81.1

Table/Tableau 9.1		Area Frame / Base aréolaire										Phone Frames / Bases téléphoniques						All cases Tous les cas
					2 persons select. 2 personnes sélect.			1 person select. 1 personne sélect.										
Prov. Terr.	Health Region Région socio sanitaire	# In Scope HH # Mén. cibles	# Resp HH # Mén. rép.	HH Resp. Rates Taux de rép. mén.	# Pers. Select. # Pers. sélect.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	# Pers. Select. # Pers. sélect.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	Resp. Rates Taux de rép.	# In Scope HH # Mén. cibles	# Resp HH # Mén. rép.	HH Resp. Rates Taux de rép. mén.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	Resp. Rates Taux de rép.	Combined Resp. Rates Taux de rép. Combiné
	35931	430	407	94.7	146	127	87	334	325	97.3	90.3	381	331	86.9	304	91.8	79.8	85.4
	35933	543	518	95.4	158	145	91.8	439	421	95.9	90.9	366	312	85.2	292	93.6	79.8	86.4
	35934	579	538	92.9	196	181	92.3	440	417	94.8	87.7	171	140	81.9	128	91.4	74.9	84.7
	35935	536	484	90.3	160	141	88.1	404	382	94.6	84.4	548	478	87.2	459	96	83.8	84.1
	35936	913	861	94.3	326	234	71.8	698	618	88.5	80.5	513	452	88.1	409	90.5	79.7	80.2
	35937	1478	1251	84.6	448	364	81.3	1027	935	91	75.6	42	28	66.7	27	96.4	64.3	75.3
	35938	307	287	93.5	82	65	79.3	246	229	93.1	85.2	702	633	90.2	583	92.1	83.0	83.7
	35939*	613	587	95.8	212	171	80.7	481	447	92.9	86.9	786	690	87.8	643	93.2	81.8	84.0
	35940	682	664	97.4	236	212	89.8	546	532	97.4	93.5	388	344	88.7	318	92.4	82.0	89.3
	35941	586	519	88.6	190	173	91.1	424	402	94.8	83.4	465	401	86.2	367	91.5	78.9	81.4
	35942	660	612	92.7	234	201	85.9	495	471	95.2	86.6	229	204	89.1	185	90.7	80.8	85.1
	35943	696	608	87.4	180	156	86.7	518	495	95.6	82.3	334	282	84.4	256	90.8	76.6	80.5
	35944	1413	1185	83.9	412	346	84	979	919	93.9	77.3	.	.	.	.	.	.	77.3
	35945	396	367	92.7	116	96	82.8	309	293	94.8	86.1	474	415	87.6	387	93.3	81.6	83.7
	35946	1364	1206	88.4	376	326	86.7	1018	946	92.9	81.3	.	.	.	.	.	.	81.3
	35947*	990	923	93.2	252	209	82.9	797	746	93.6	85.9	632	560	88.6	531	94.8	84.0	85.2
	35949	288	257	89.2	74	62	83.8	220	212	96.4	84.4	523	480	91.8	445	92.7	85.1	84.8
	35951	1878	1741	92.7	596	543	91.1	1443	1382	95.8	88.0	.	.	.	.	.	.	88.0
	35952	698	624	89.4	236	194	82.2	506	473	93.5	81.7	81	54	66.7	50	92.6	61.7	79.6
	35953	1579	1453	92	750	535	71.3	1078	980	90.9	79.0	425	366	86.1	316	86.3	74.4	78.0
	35955	229	199	86.9	44	31	70.5	177	159	89.8	76.2	801	708	88.4	660	93.2	82.4	81.0

Table/Tableau 9.1		Area Frame / Base aréolaire										Phone Frames / Bases téléphoniques						All cases Tous les cas
					2 persons select. 2 personnes sélect.			1 person select. 1 personne sélect.										
Prov. Terr.	Health Region Région socio sanitaire	# In Scope HH # Mén. cibles	# Resp HH # Mén. rép.	HH Resp. Rates Taux de rép. mén.	# Pers. Select. # Pers. sélect.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	# Pers. Select. # Pers. sélect.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	Resp. Rates Taux de rép.	# In Scope HH # Mén. cibles	# Resp HH # Mén. rép.	HH Resp. Rates Taux de rép. mén.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	Resp. Rates Taux de rép.	Combined Resp. Rates Taux de rép. Combiné
	35956	363	333	91.7	128	99	77.3	269	249	92.6	82.2	468	421	90	397	94.3	84.8	83.7
	35957	656	615	93.8	188	172	91.5	521	504	96.7	89.9	72	43	59.7	40	93	55.6	86.5
	35958	631	580	91.9	214	197	92.1	473	453	95.8	87.4	460	379	82.4	345	91	75.0	82.2
	35960	1215	1145	94.2	438	378	86.3	926	851	91.9	85.6	146	111	76	100	90.1	68.5	83.8
	35961	957	883	92.3	268	229	85.4	749	701	93.6	85.2	71	50	70.4	47	94	66.2	83.9
	35962	858	823	95.9	270	253	93.7	688	659	95.8	91.6	65	49	75.4	44	89.8	67.7	89.9
	35965	1363	1247	91.5	462	366	79.2	1016	936	92.1	82.1	.	.	.	.	.	.	82.1
	35966	994	959	96.5	360	335	93.1	779	736	94.5	90.9	127	102	80.3	95	93.1	74.8	89.1
	35968	1257	1146	91.2	414	354	85.5	939	878	93.5	83.9	26	21	80.8	16	76.2	61.5	83.5
	35970	1044	975	93.4	542	481	88.7	704	660	93.8	86.3	791	676	85.5	597	88.3	75.5	81.6
	35995	3142	2691	85.6	828	604	72.9	2277	1904	83.6	70.2	.	.	.	.	.	.	70.2
MB	Total	7379	6960	94.3	2322	2111	90.9	5799	5586	96.3	90.0	909	824	90.6	773	93.8	85.0	89.5
	46910	2095	1923	91.8	552	498	90.2	1647	1567	95.1	86.7	.	.	.	.	.	.	86.7
	46915*	1475	1401	95	446	409	91.7	1178	1152	97.8	92.0	350	317	90.6	308	97.2	88.0	91.2
	46920*	897	874	97.4	354	321	90.7	697	674	96.7	93.0	339	310	91.4	283	91.3	83.5	90.4
	46930	685	632	92.3	252	222	88.1	506	490	96.8	87.7	57	53	93	47	88.7	82.5	87.3
	46940	772	743	96.2	268	237	88.4	609	585	96.1	91.1	.	.	.	.	.	.	91.1
	46960*	1455	1387	95.3	450	424	94.2	1162	1118	96.2	91.4	163	144	88.3	135	93.8	82.8	90.5
SK	Total	6885	6379	92.7	2200	1912	86.9	5279	5037	95.4	87.0	1242	1154	92.9	1060	91.9	85.3	86.8
	47901*	1638	1514	92.4	506	452	89.3	1261	1182	93.7	86.0	249	227	91.2	213	93.8	85.5	85.9
	47904	1187	1080	91	382	328	85.9	889	841	94.6	84.7	.	.	.	.	.	.	84.7

Table/Tableau 9.1		Area Frame / Base aréolaire										Phone Frames / Bases téléphoniques						All cases Tous les cas
					2 persons select. 2 personnes sélect.			1 person select. 1 personne sélect.										
Prov. Terr.	Health Region Région socio sanitaire	# In Scope HH # Mén. cibles	# Resp HH # Mén. rép.	HH Resp. Rates Taux de rép. mén.	# Pers. Select. # Pers. sélect.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	# Pers. Select. # Pers. sélect.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	Resp. Rates Taux de rép.	# In Scope HH # Mén. cibles	# Resp HH # Mén. rép.	HH Resp. Rates Taux de rép. mén.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	Resp. Rates Taux de rép.	Combined Resp. Rates Taux de rép. Combiné
	47905*	1158	1094	94.5	362	335	92.5	913	899	98.5	92.1	182	165	90.7	157	95.2	86.3	91.3
	47906	1261	1163	92.2	408	352	86.3	959	919	95.8	86.8	.	.	.	.	.	.	86.8
	47907*	995	934	93.9	354	286	80.8	757	706	93.3	85.3	279	260	93.2	246	94.6	88.2	85.9
	47909*	646	594	92	188	159	84.6	500	490	98	88.2	532	502	94.4	444	88.4	83.5	86.0
AB	Total	10961	10201	93.1	3776	3200	84.7	8313	7735	93	85.2	4148	3844	92.7	3521	91.6	84.9	85.1
	48901	858	807	94.1	242	215	88.8	686	663	96.6	89.8	12	11	91.7	10	90.9	83.3	89.7
	48902	539	510	94.6	196	163	83.2	412	397	96.4	88.8	202	179	88.6	166	92.7	82.2	87.0
	48903*	1017	950	93.4	348	297	85.3	776	728	93.8	86.2	376	338	89.9	307	90.8	81.6	85.0
	48904	2147	1956	91.1	680	579	85.1	1616	1505	93.1	83.6	.	.	.	.	.	.	83.6
	48906	926	882	95.2	288	265	92	738	706	95.7	90.6	.	.	.	.	.	.	90.6
	48907	691	650	94.1	264	227	86	518	471	90.9	84.6	114	110	96.5	105	95.5	92.1	85.7
	48908*	815	759	93.1	266	231	86.8	626	539	86.1	80.3	557	515	92.5	469	91.1	84.2	81.9
	48910	2097	1932	92.1	666	574	86.2	1599	1503	94	85.4	.	.	.	.	.	.	85.4
	48911	386	363	94	178	147	82.6	274	258	94.2	85.9	525	485	92.4	440	90.7	83.8	84.7
	48912	483	457	94.6	216	156	72.2	349	294	84.2	77.0	438	411	93.8	377	91.7	86.1	81.3
	48913	472	446	94.5	184	146	79.3	354	339	95.8	87.3	368	344	93.5	317	92.2	86.1	86.8
	48914*	530	489	92.3	248	200	80.6	365	332	91	81.5	1556	1451	93.3	1330	91.7	85.5	84.5
BC	Total	14329	12872	89.8	4254	3749	88.1	10745	10150	94.5	83.9	5077	4716	92.9	4403	93.4	86.7	84.7
	59901	295	288	97.6	94	86	91.5	241	239	99.2	95.6	382	354	92.7	331	93.5	86.6	90.5
	59902	304	286	94.1	78	73	93.6	247	236	95.5	89.6	458	427	93.2	400	93.7	87.3	88.3
	59903	469	436	93	88	80	90.9	392	365	93.1	86.4	508	474	93.3	446	94.1	87.8	87.1

Table/Tableau 9.1		Area Frame / Base aréolaire										Phone Frames / Bases téléphoniques						All cases Tous les cas
					2 persons select. 2 personnes sélect.			1 person select. 1 personne sélect.										
Prov. Terr.	Health Region Région socio sanitaire	# In Scope HH # Mén. cibles	# Resp HH # Mén. rép.	HH Resp. Rates Taux de rép. mén.	# Pers. Select. # Pers. sélect.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	# Pers. Select. # Pers. sélect.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	Resp. Rates Taux de rép.	# In Scope HH # Mén. cibles	# Resp HH # Mén. rép.	HH Resp. Rates Taux de rép. mén.	# Resp. # Rép.	Pers. Resp. Rates Taux de rép. pers.	Resp. Rates Taux de rép.	Combined Resp. Rates Taux de rép. Combiné
	59904	986	923	93.6	300	273	91	773	740	95.7	88.9	49	47	95.9	45	95.7	91.8	89.0
	59905	487	452	92.8	164	155	94.5	370	353	95.4	88.4	545	510	93.6	485	95.1	89.0	88.7
	59906	1059	986	93.1	380	349	91.8	796	766	96.2	88.8	.	.	.	.	.	.	88.8
	59907	1424	1273	89.4	552	482	87.3	997	948	95.1	83.5	.	.	.	.	.	.	83.5
	59908	1062	957	90.1	316	257	81.3	799	732	91.6	81.0	199	182	91.5	170	93.4	85.4	81.7
	59909	201	178	88.6	38	35	92.1	159	148	93.1	82.3	513	482	94	443	91.9	86.4	85.2
	59910	987	894	90.6	270	248	91.9	759	730	96.2	86.5	106	104	98.1	96	92.3	90.6	86.9
	59911	470	431	91.7	128	115	89.8	367	354	96.5	87.6	320	299	93.4	283	94.6	88.4	87.9
	59912	308	285	92.5	108	96	88.9	231	223	96.5	88.0	394	364	92.4	350	96.2	88.8	88.5
	59913*	1092	982	89.9	388	344	88.7	788	742	94.2	83.7	197	187	94.9	173	92.5	87.8	84.3
	59915	662	581	87.8	176	139	79	493	463	93.9	80.4	284	265	93.3	255	96.2	89.8	83.2
	59916	1453	1244	85.6	282	222	78.7	1103	998	90.5	76.3	44	41	93.2	40	97.6	90.9	76.8
	59917	674	587	87.1	204	166	81.4	485	444	91.5	78.2	348	320	92	289	90.3	83.0	79.8
	59918	612	500	81.7	170	161	94.7	415	404	97.3	79.2	340	298	87.6	279	93.6	82.1	80.2
	59919	497	443	89.1	224	202	90.2	331	314	94.9	83.5	390	362	92.8	318	87.8	81.5	82.6
	59920	1287	1146	89	294	266	90.5	999	951	95.2	84.2	.	.	.	.	.	.	84.2
Terr.	60901*	2313	1977	85.5	966	808	83.6	1494	1416	94.8	78.7	385	307	79.7	293	95.4	76.1	78.3

* = collapsed health region

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. as those actually used. The difference between the estimates obtained from the sample and the results from a complete count under similar conditions is called the sampling error of the estimate.

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

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 (Cycle 1.1). Quality assurance measures were implemented at each step of data collection and processing to monitor the quality of the data. These measures included the use of highly skilled interviewers, extensive training with respect to the survey procedures and questionnaire, and the observation of interviewers to detect problems. Testing of the CAI application and field tests were also essential procedures to ensure that data collection errors were minimized.

A major source of non-sampling errors in surveys is the effect of *non-response* on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. Partial non-response to CCHS (Cycle 1.1) was minimal; once the questionnaire was started, it tended to be completed with very little non-response. Total non-response occurred either because a respondent refused to participate in the survey, or because the interviewer was unable to contact the selected respondent. Total non-response was handled by adjusting the weight of persons who responded to the survey to compensate for those who did not respond. See Section 8 for details of the weight adjustment for non-response.

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 one estimates that 25% of Canadians aged 12 and over are regular smokers and that this estimate is found to have a standard deviation of .003. Then the CV of the estimate is calculated as:

$$(.003/.25) \times 100\% = 1.20\%.$$

Statistics Canada commonly uses CV results when analyzing data, and urges users producing estimates from CCHS (Cycle 1.1) data files to do so also. For details on how to determine CVs, see Section 11. For guidelines on how to interpret CV results, see the table at the end of sub-section 10.4.

10. Guidelines for Tabulation, Analysis and Release

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

10.1 Rounding Guidelines

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

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

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

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

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

e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released that differ from corresponding estimates published by

Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s);

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

10.2 Sample Weighting Guidelines for Tabulation

The sample design used for the CCHS was not self-weighting. That is to say, the sampling weights are not identical for all individuals in the sample. When producing simple estimates, including the production of ordinary statistical tables, users must apply the proper sampling weight.

If proper weights are not used, the estimates derived from the microdata files cannot be considered to be representative of the survey population, and will not correspond to those produced by Statistics Canada.

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

10.2.1 Definitions: Categorical Estimates, Quantitative Estimates

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

Categorical Estimates:

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

SMKA_202: At the present do/does ... smoke cigarettes daily, occasionally or not at all?
__ 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:

SMKA_204: How many cigarettes do/does you/he/she smoke each day now?

|_|_| 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 microdata files by summing the final weights of all records possessing the characteristic of interest.

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

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

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

c) dividing the numerator estimate by the denominator estimate.

10.2.3 Tabulation of Quantitative Estimates

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

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

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

c) dividing the numerator estimate by the denominator estimate.

For example, to obtain an estimate of the average number of cigarettes smoked each day by individuals who smoke daily, multiply the value of variable SMKA_204[23] by the weight, **WTSAM**, then sum this value over those records with a value of "daily" to the **variable SMKA_202** to obtain the numerator ($\hat{X}$). Sum the final weight of those records with a value of "daily" to the variable SMKA_202 to obtain the denominator ($\hat{Y}$). 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 (Cycle 1.1) is based upon a complex design, with stratification and multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used.

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

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

[23] See Section 12.2 for variable naming convention

In order to provide a means of assessing the quality of tabulated estimates, Statistics Canada has produced a set of Approximate Coefficients of Variations Tables (commonly referred to as "CV Tables") for the CCHS. These tables can be used to obtain approximate coefficients of variation for categorical-type estimates and proportions. See Section 11 for more details.

10.4 Release Guidelines

Before releasing and/or publishing any estimate from these microdata files, users should first determine the number of sampled respondents who contribute to the calculation of the estimate. If this number is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the rounded estimate and follow the guidelines below.

Table 10.1: Sampling Variability Guidelines

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

11. Approximate Sampling Variability Tables

In order to supply coefficients of variation that would be applicable to a wide variety of categorical estimates produced from this microdata file and that could be readily accessed by the user, a set of Approximate Sampling Variability Tables has been 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.

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 tables below show the design effects, sample sizes and population counts used to produce the Approximate Sampling Variability Tables.

Input Data for Health Region Level Sampling Variability Tables

HEALTH REGION	DESIGN EFFECT	SAMPLE SIZE	POPULATION (aged 12 & over)
Health and Community Services St John's Region (10901)	1.37	892	158,173
Health and Community Services Eastern Region (10902)	1.65	810	103,604
Health and Community Services Central Region (10903)	1.51	711	90,416
Hlth & Com. Serv. Western Region / Grenfell Reg. Hlth Serv. Board / Health Labrador Corp. (10904)	1.93	1,457	108,911
Region 1 - Prince Edward Island (11901)	1.92	1,389	54,439
Region 2 - Prince Edward Island (11902)	2.05	2,262	61,887
Zone 1 - Nova Scotia (12901)	1.58	956	108,292
Zone 2 - Nova Scotia (12902)	1.55	711	70,499
Zone 3 - Nova Scotia (12903)	1.52	801	88,937
Zone 4 - Nova Scotia (12904)	1.63	691	83,250
Zone 5 - Nova Scotia (12905)	1.39	820	112,272
Zone 6 - Nova Scotia (12906)	1.40	1,340	324,722
Region 1 - New Brunswick (13901)	1.37	985	157,931
Region 2 - New Brunswick (13902)	1.35	915	149,340
Region 3 - New Brunswick (13903)	1.49	873	137,485

HEALTH REGION	DESIGN EFFECT	SAMPLE SIZE	POPULATION (aged 12 & over)
Region 4 / Region 5 - New Brunswick (13904)	1.52	1,061	73,071
Region 6 / Region 7 - New Brunswick (13905)	1.52	1,162	116,438
Région du Bas-Saint-Laurent (24901)	1.45	1,127	174,533
Région du Saguenay - Lac-Saint-Jean (24902)	1.51	1,122	241,654
Région de Québec (24903)	1.65	1,653	556,346
Région de la Mauricie-Bois-Francs (24904)	1.90	1,622	408,383
Région de l'Estrie (24905)	2.13	1,180	244,455
Région de Montréal-Centre (24906)	1.44	2,721	1,569,344
Région de l'Outaouais (24907)	1.45	1,185	267,803
Région de l'Abitibi-Témiscaminque (24908)	1.55	1,253	123,875
Région de la Côte-Nord (24909)	2.21	1,098	77,721
Région de la Gaspésie-Îles-de-la-Madeleine (24911)	1.55	1,184	85,308
Région de la Chaudière-Appalaches (24912)	1.69	1,427	330,976
Région de Laval (24913)	1.43	1,045	297,108
Région de Lanaudière (24914)	1.57	1,494	330,528
Région de Laurentides (24915)	1.63	1,440	394,808
Région de la Montérégie (24916)	2.09	2,461	1,113,880
Algoma (35926)	1.48	812	105,357
Brant (35927)	1.54	756	105,865
Durham (35930)	2.11	1,383	427,780
Elgin-St Thomas (35931)	1.70	742	69,966
Bruce-Grey-Owen Sound (35933)	1.64	860	133,720
Haldimand-Norfolk (35934)	1.58	723	93,186
Haliburton (35935)	1.47	967	144,537
Halton (35936)	1.48	1,257	321,355
Hamilton-Wentworth (35937)	1.58	1,326	423,505
Hastings and Prince Edward (35938)	1.51	889	132,650
Huron / Perth (35939)	1.41	1,242	114,057
Kent-Chatham (35940)	1.55	1,059	93,445
Kingston (35941)	1.56	938	149,440
Lambton (35942)	1.60	866	108,811
Leeds (35943)	1.60	901	137,717
Middlesex-London (35944)	1.50	1,282	348,789
Muskoka-Parry Sound (35945)	1.39	763	72,021
Niagara (35946)	1.48	1,275	362,313
North Bay / Timiskaming (35947)	1.57	1,484	109,218

HEALTH REGION	DESIGN EFFECT	SAMPLE SIZE	POPULATION (aged 12 & over)
Northwestern (35949)	1.51	710	55,555
Ottawa Carleton (35951)	1.49	1,936	664,036
Oxford (35952)	1.77	713	85,976
Peel (35953)	1.43	1,837	857,586
Peterborough (35955)	1.42	842	109,112
Porcupine (35956)	1.45	755	74,595
Renfrew (35957)	1.70	722	82,501
Eastern Ontario (35958)	1.50	982	163,078
Simcoe (35960)	1.81	1,338	316,951
Sudbury (35961)	1.51	979	165,778
Thunder Bay (35962)	1.50	959	130,084
Waterloo (35965)	1.42	1,304	378,067
Wellington-Dufferin-Guelph (35966)	1.63	1,170	203,929
Windsor-Essex (35968)	1.49	1,250	324,756
York (35970)	1.56	1,732	634,669
City of Toronto (35995)	1.48	2,524	2,176,887
Winnipeg (46910)	1.39	2,070	535,694
Brandon / Marquette / South Westman (46915)	1.64	1,863	97,676
North Eastman / South Eastman (46920)	1.72	1,271	72,950
Interlake (46930)	1.99	762	60,785
Central (46940)	1.54	827	75,577
Parkland / Norman / Burntwood / Churchill (46960)	2.01	1,677	64,811
Weyburn (A) Service Area / Moose Jaw (B) Service Area / Swift Current (C) Service Area (47901)	1.60	1,855	131,237
Regina (D) Service Area (47904)	1.35	1,171	199,000
Yorkton (E) Service Area / Melfort (H) Service Area (47905)	1.68	1,391	84,013
Saskatoon (F) Service Area (47906)	1.45	1,274	231,775
Rosetown (G) Service Area / North Battleford (J) Service Area (47907)	1.60	1,236	87,992
Prince Albert (I) Service Area / Northern Health Services Branch (K) Service Area (47909)	1.75	1,082	71,976
Chinook Regional Health Authority (48901)	1.78	890	120,653
Palliser Regional Health Authority (48902)	1.66	726	77,508
Headwaters Regional Health Authority / Regional Health Authority #5 (48903)	1.76	1,324	106,187

HEALTH REGION	DESIGN EFFECT	SAMPLE SIZE	POPULATION (aged 12 & over)
Calgary Regional Health Authority (48904)	1.54	2,092	809,818
David Thompson Regional Health Authority (48906)	1.56	973	162,697
East Central Regional Health Authority (48907)	1.63	802	85,574
Westview Regional Health Authority / Crossroads Regional Health Authority (48908)	2.09	1,229	111,313
Capital Health Authority (48910)	1.54	2,111	700,227
Aspen Regional Health Authority (48911)	1.69	761	73,940
Lakeland Regional Health Authority (48912)	1.58	814	84,599
Mistahia Regional Health Authority (48913)	1.44	799	73,189
Peace / Keeweetinok / Northern Lights & Northwestern Reg. Hlth Auth. (48914)	1.73	1,935	75,863
East Kootenay (59911)	1.49	645	67,119
Kootenay-Boundary (59912)	1.39	705	69,608
Okanagan (59913)	1.63	1,671	257,796
Thompson / Cariboo (59914)	1.57	1,668	175,967
Fraser Valley (59921)	1.63	1,125	198,477
Simon Fraser (59922)	1.47	2,036	457,268
South Fraser (59923)	1.53	1,437	496,419
Richmond (59931)	1.53	828	149,581
Vancouver (59932)	1.57	1,285	508,699
North Shore / Coast Garibaldi(59933)	1.55	1,475	222,778
South Vancouver Island (59941)	1.43	1,522	337,941
Central Vancouver Island / North Vancouver Island (59942)	1.69	1,526	247,858
Norhtwest / Northeast (59951)	1.58	1,261	109,769
Northern Interior (59952)	1.55	1,118	122,391
Yukon / Northwest Territories / Nunavut (60901)	1.12	2,517	76,928

Input Data for Provincial and Canada Level Sampling Variability Tables

PROVINCE	DESIGN EFFECT	SAMPLE SIZE	POPULATION (age 12 & over)
Newfoundland	1.63	3,870	461,104
Prince Edward Island	2.02	3,651	116,326
Nova Scotia	1.63	5,319	787,972
New Brunswick	1.51	4,996	634,264
Quebec	2.08	22,012	6,216,722
Ontario	2.31	39,278	9,877,292
Manitoba	2.44	8,470	907,493
Saskatchewan	1.82	8,009	805,993
Alberta	2.39	14,456	2,481,568
British Columbia	1.77	18,302	3,421,671
Territories	1.12	2,517	76,928
CANADA	2.34	130,880	25,787,333

Input Data for Age Group (Canada) Sampling Variability Tables

AGE GROUP	DESIGN EFFECT	SAMPLE SIZE	POPULATION
12-19	2.13	17,557	3,240,646
20-29	2.20	16,326	4,137,811
30-44	2.17	35,614	7,477,907
45-64	2.44	37,150	7,283,951
65+	2.70	24,233	3,647,018

All coefficients of variation in the Approximate Sampling Variability Tables are *approximate* and, therefore, unofficial. Options concerning the computation of exact coefficients of variation are discussed in sub-section 11.7.

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

11.1 How to Use the CV Tables for Categorical Estimates

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

Rule 1: Estimates of Numbers Possessing a Characteristic (Aggregates)

The coefficient of variation depends only on the size of the estimate itself. On the appropriate Sampling Variability 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. This figure is the approximate coefficient of variation.

Rule 2: Estimates of Proportions or Percentages 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 CV's 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). 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 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.

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 CV's for the two ratios are first determined using Rule 4, and then the CV of their difference is found using Rule 3.

11.2 Examples of Using the CV Tables for Categorical Estimates

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

Example 1 : Estimates of Numbers Possessing a Characteristic (Aggregates)

Suppose that a user estimates that 5,529,104 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 (5,529,104) does not appear in the left-hand column (the "Numerator of Percentage" column), so it is necessary to use the figure closest to it, namely 6,000,000.

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, 0.8%.

4) So the approximate coefficient of variation of the estimate is 0.8%. The finding that there were 5,529,104 individuals who smoke daily is publishable with no qualifications.

Example 2 : Estimates of Proportions or Percentages Possessing a Characteristic

Suppose that the user estimates that 5,529,104/6,677,374=82.8% 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 (82.8%) and the numerator portion of the percentage (5,529,104) in determining the coefficient of variation.

3) The numerator (5,529,104) does not appear in the left-hand column (the "Numerator of Percentage" column) so it is necessary to use the figure closest to it, namely 6,000,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the figure closest to it, 90.0%.

4) The figure at the intersection of the row and column used, namely 0.3% is the coefficient of variation (expressed as a percentage) to be used.

5) So the approximate coefficient of variation of the estimate is 0.3%. The finding that 82.8% 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, 2,985,871/12,697,959 = 23.5% smoke daily (estimate 1), while for women, this percentage is estimated at 2,543, 234 / 13,089,375 = 19.4% (estimate 2). How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the CANADA level CV table in the same manner as described in example 2 gives the CV for estimate 1 as 1.1% (expressed as a percentage), and the CV for estimate 2 as 1.1% (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 $\hat{d}$ = (.235 - .194) = .041 is :

$$\sigma_{\hat{d}} = \sqrt{[(.235)(.011)]^2 + [(.194)(.011)]^2}$$
$$= .003$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = .003/.041 = 0.07.

4) So the approximate coefficient of variation of the difference between the estimates is 7.0% (expressed as a percentage). This estimate can be published with no qualifications.

Example 4 : Estimates of Ratios

Suppose that the user estimates that 5,529,104 individuals smoke daily, while 1,148,270 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 1,148,270. The figure closest to it is 1,000,000. The coefficient of variation for this estimate (expressed as a percentage) is found by referring to the first non-asterisk entry on that row, namely, 2.1%.

4) The denominator of this ratio estimate is 5,529,104. The figure closest to it is 6,000,000. 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, 0.8%.

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

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

That is,

$$\alpha_{\hat{R}} = \sqrt{(.021)^2 + (.008)^2}$$

$$= 0.022$$

where α_1 et 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 1,148,270/5,529,104 which is 0.21:1. The coefficient of variation of this estimate is 2.2% (expressed as a percentage), which is releasable with no qualifications.

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 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 ($\hat{X}$ -k, $\hat{X}$ +k), 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 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:

$\hat{X}$ = 0.828

z = 2

$\alpha_{\hat{X}}$ = 0.003 is the coefficient of variation of this estimate as determined from the tables.

CI_X = {0.828 - (2) (0.828) (0.003) , 0.828 + (2) (0.828) (0.003)}

CI_X = {0.823 , 0.833}

11.5 How to Use the CV Tables to do a Z-test

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

Let $\hat{X}_1$ and $\hat{X}_2$ be sample estimates for 2 characteristics of interest. Let the standard error on the difference $\hat{X}_1 - \hat{X}_2$ be $\sigma_{\hat{d}}$. If the ratio of $\hat{X}_1 - \hat{X}_2$ over $\sigma_{\hat{d}}$ is between -2

and 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level.

11.6 Example of Using the CV Tables to do a Z-test

Let us suppose we wish to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of men who smoke daily AND the proportion of women who smoke daily. From example 3, sub-section 11.2, the standard error of the difference between these two estimates was found to be = .003. Hence,

$$z = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{.235 - .194}{.003} = \frac{.041}{.003} = 13.7$$

Since z = 13.7 is greater than 2, it must be concluded that there is a significant difference between the two estimates at the 0.05 level of significance. Note that the two sub-groups compared are considered as being independent, so the test is correct.

11.7 Exact Variances/Coefficients of Variation

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

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

The computation of coefficients of variation (or any other measure of precision) with the use of the bootstrap method requires access to information that is considered confidential and not available on the public use microdata file. This computation must be done via other avenues, such as remote access. Remote access, as well as other alternatives to obtain exact coefficients of variation, is discussed in sub-section 12.3.

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

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

Firstly, if a user desires estimates at a geographic level other than the health region 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 coefficients 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 (Cycle 1.1) are primarily categorical in nature, this has not been done. Thus, users wishing to obtain coefficients of variation for quantitative variables can do so through the exact variance program. As a general rule, however, the coefficient of variation of a quantitative total will be larger than the coefficient of variation of the corresponding category estimate (i.e., the estimate of the number of persons contributing to the quantitative estimate). If the corresponding category estimate is not releasable, the quantitative estimate will not be either. For example, the coefficient of variation of the estimate of the total number of cigarettes smoked each day by individuals who smoke daily would be greater than the coefficient of variation of the corresponding estimate of the number of individuals who smoke daily. Hence if the coefficient of variation of the latter is not releasable, then the coefficient of variation of the corresponding quantitative estimate will also not be releasable.

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

11.8 Release Cut-offs for the CCHS

The following tables present the release cut-offs for estimates of totals at the health region, provincial and Canada levels, as well as for some age groups (at the Canada level only). Estimates smaller than the value given in the "Marginal" column may not be released under any circumstances.

Table of Release Cut-offs for Health Region Level Totals

HEALTH REGION	CV	
	CV BETWEEN 0% AND 16.5%	CV BETWEEN 16.5% AND 33.3%
	ACCEPTABLE	MARGINAL
Health and Community Services St John's Region (10901)	8,500	2,000
Health and Community Services Eastern Region (10902)	7,000	2,000
Health and Community Services Central Region (10903)	6,500	1,500
Hlth & Com. Serv. Western Region / Grenfell Reg. Hlth Serv. Board / Health Labrador Corp. (10904)	5,000	1,500
Region 1 - Prince Edward Island (11901)	2,500	500
Region 2 - Prince Edward Island (11902)	2,000	500
Zone 1 - Nova Scotia (12901)	6,000	1,500
Zone 2 - Nova Scotia (12902)	5,000	1,500
Zone 3 - Nova Scotia (12903)	6,000	1,500
Zone 4 - Nova Scotia (12904)	6,500	1,500
Zone 5 - Nova Scotia (12905)	6,500	1,500
Zone 6 - Nova Scotia (12906)	12,000	3,000
Region 1 - New Brunswick (13901)	7,500	2,000
Region 2 - New Brunswick (13902)	7,500	2,000
Region 3 - New Brunswick (13903)	8,000	2,000
Region 4 / Region 5 - New Brunswick (13904)	3,500	1,000
Region 6 / Region 7 - New Brunswick (13905)	5,500	1,500
Région du Bas-Saint-Laurent (24901)	8,000	2,000
Région du Saguenay - Lac-Saint-Jean (24902)	11,500	3,000
Région de Québec (24903)	19,500	5,000
Région de la Mauricie-Bois-Francs (24904)	17,000	4,500
Région de l'Estrie (24905)	15,000	4,000

HEALTH REGION	CV	
	CV BETWEEN 0% AND 16.5%	CV BETWEEN 16.5% AND 33.3%
	ACCEPTABLE	MARGINAL
Région de l'Outaouais (24907)	11,500	3,000
Région de l'Abitibi-Témiscaminque (24908)	5,500	1,500
Région de la Côte-Nord (24909)	5,500	1,500
Région de la Gaspésie-Îles-de-la-Madeleine (24911)	4,000	1,000
Région de la Chaudière-Appalaches (24912)	14,000	3,500
Région de Laval (24913)	14,000	3,500
Région de Lanaudière (24914)	12,500	3,000
Région de Laurentides (24915)	16,000	4,000
Région de la Montérégie (24916)	33,500	8,500
Algoma (35926)	6,500	1,500
Brant (35927)	7,500	2,000
Durham (35930)	22,500	6,000
Elgin-St Thomas (35931)	5,500	1,500
Bruce-Grey-Owen Sound (35933)	9,000	2,500
Haldimand-Norfolk (35934)	7,000	2,000
Hamilton-Wentworth (35937)	18,000	4,500
Hastings and Prince Edward (35938)	8,000	2,000
Huron / Perth (35939)	4,500	1,000
Kent-Chatham (35940)	5,000	1,000
Kingston (35941)	8,500	2,000
Lambton (35942)	7,000	2,000
Leeds (35943)	8,500	2,000
Middlesex-London (35944)	14,500	3,500
Muskoka-Parry Sound (35945)	4,500	1,000
Niagara (35946)	15,000	4,000
North Bay / Timiskaming (35947)	4,000	1,000
Northwestern (35949)	4,000	1,000
Ottawa Carleton (35951)	18,500	4,500
Oxford (35952)	7,000	2,000
Peel (35953)	24,000	6,000
Peterborough (35955)	6,500	1,500
Porcupine (35956)	5,000	1,500
Renfrew (35957)	6,500	1,500

HEALTH REGION	CV	
	CV BETWEEN 0% AND 16.5%	CV BETWEEN 16.5% AND 33.3%
	ACCEPTABLE	MARGINAL
Eastern Ontario (35958)	8,500	2,000
Simcoe (35960)	15,000	4,000
Sudbury (35961)	9,000	2,500
Thunder Bay (35962)	7,000	2,000
Waterloo (35965)	14,500	3,500
Wellington-Dufferin-Guelph (35966)	10,000	2,500
Windsor-Essex (35968)	13,500	3,500
York (35970)	20,500	5,000
City of Toronto (35995)	46,000	11,500
Winnipeg (46910)	13,000	3,000
Brandon / Marquette / South Westman (46915)	3,000	1,000
North Eastman / South Eastman (46920)	3,500	1,000
Interlake (46930)	5,500	1,500
Central (46940)	5,000	1,000
Parkland / Norman / Burntwood / Churchill (46960)	2,500	500
Weyburn (A) Service Area / Moose Jaw (B) Service Area / Swift Current (C) Service Area (47901)	4,000	1,000
Regina (D) Service Area (47904)	8,000	2,000
Yorkton (E) Service Area / Melfort (H) Service Area (47905)	3,500	1,000
Saskatoon (F) Service Area (47906)	9,500	2,500
Rosetown (G) Service Area / North Battleford (J) Service Area (47907)	4,000	1,000
Prince Albert (I) Service Area / Northern Health Services Branch (K) Service Area (47909)	4,000	1,000
Chinook Regional Health Authority (48901)	8,500	2,000
Palliser Regional Health Authority (48902)	6,000	1,500
Headwaters Regional Health Authority / Regional Health Authority #5 (48903)	5,000	1,500
Calgary Regional Health Authority (48904)	21,500	5,500
David Thompson Regional Health Authority (48906)	9,000	2,500
East Central Regional Health Authority (48907)	6,000	1,500
Westview Regional Health Authority / Crossroads Regional Health Authority (48908)	6,500	1,500

HEALTH REGION	CV	
	CV BETWEEN 0% AND 16.5%	**CV BETWEEN 16.5% AND 33.3%**
	ACCEPTABLE	**MARGINAL**
Capital Health Authority (48910)	18,500	4,500
Aspen Regional Health Authority (48911)	5,500	1,500
Lakeland Regional Health Authority (48912)	5,500	1,500
Mistahia Regional Health Authority (48913)	4,500	1,000
Peace / Keeweetinok / Northern Lights & Northwestern Reg. Hlth Auth. (48914)	2,500	500
East Kootenay (59911)	5,000	1,500
Kootenay-Boundary (59912)	4,500	1,000
Okanagan (59913)	9,000	2,000
Thompson / Cariboo (59914)	6,000	1,500
Fraser Valley (59921)	10,000	2,500
Simon Fraser (59922)	12,000	3,000
South Fraser (59923)	18,500	4,500
Richmond (59931)	9,500	2,500
Vancouver (59932)	22,000	5,500
North Shore / Coast Garibaldi(59933)	8,500	2,000
South Vancouver Island (59941)	11,500	3,000
Central Vancouver Island / North Vancouver Island (59942)	9,500	2,500
Norhtwest / Northeast (59951)	5,000	1,000
Northern Interior (59952)	6,000	1,500
Yukon / Northwest Territories / Nunavut (60901)	1,000	500

Table of Release Cut-offs for provincial and Canada level totals

PROVINCE OR TERRITORIES	CV	
	CV BETWEEN 0% AND 16.5%	CV BETWEEN 16.5% AND 33.3%
	ACCEPTABLE	MARGINAL
Newfoundland	7,000	1,500
Prince Edward Island	2,500	500
Nova Scotia	9,000	2,000
New Brunswick	7,000	1,500
Quebec	21,500	5,500
Ontario	21,500	5,000
Manitoba	9,500	2,500
Saskatchewan	6,500	1,500
Alberta	15,000	3,500
British Columbia	12,000	3,000
Territories	1,000	500
CANADA	17,000	4,000

Table of Release Cut-offs for estimates of totals by age group

AGE GROUP	CV	
	CV BETWEEN 0% AND 16.5%	CV BETWEEN 16.5% AND 33.3%
	ACCEPTABLE	MARGINAL
12-19	14,500	3,500
20-29	20,500	5,000
30-44	16,500	4,000
45-64	17,500	4,500
65+	15,000	3,500

12. File Usage

This section starts with a discussion of the ***weight variables*** and explains how they should be used when doing tabulations on the public use microdata files. This is followed by an explanation of the variable naming convention that is employed for the CCHS. The last part of the section discusses alternate approaches to data access available to analysts.

12.1 Use of Weights

Only one weight, WTSAM, appears on the file. This weight is applicable to all age groups, regions, provinces and territories. VARIABLES ON THE FILE SHOULD BE ANALYZED USING THIS WEIGHT.

(For a more detailed explanation on the creation of this weight, see Section 8 of the documentation on weighting.)

12.2 Variable Naming Convention

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

12.2.1 Variable Name Component Structure in CCHS

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

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

For example: The variable from question 8B, Depression Module, CCHS Cycle 1.1 **(DPSA_08B)**:

Position 1-3:	**DPS**	depression module
Position 4:	**A**	Cycle 1.1
Position 5:	**_**	underscore (_ = collected data)
Position 6-8:	**08B**	question number & answer option

12.2.2 Positions 1-3: Variable / Questionnaire Section Name

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

ADM	Administration	INC	Income
ALC	Alcohol	INJ	Injury
ALD	Alcohol dependence / abuse	LBF	Labour force
BPC	Blood pressure	MAM	Mammography
BRF	Breastfeeding	MAS	Mastery
BRX	Breast examinations	MDB	Mood
BSX	Breast self examinations	PAC	Physical activities
CCC	Chronic conditions	PAP	PAP smear test
CIH	Changes made to improve health	PCU	Physical check-up
CMH	Contacts with mental health professionals	PSA	PSA test
DEN	Dental visits	RAC	Restriction of activities
DHH	Demographics and household	REP	Repetitive strain
DIS	Distress	SAC	Sedentary activities
DPS	Depression	SAM	Sample Identifiers
DRG	Drug use	SAT	Patient Satisfaction
DUI	Driving under influence	SCA	Smoking cessation aids
EDU	Education	SDC	Socio-demographics
ETS	Exposure to second hand smoke	SFE	Self-esteem
EYX	Eye examination	SMK	Smoking
FIN	Food insecurity	SPR	Spirituality
FLU	Flu shots	SSM	Social support
FVC	Fruit and vegetable consumption	SUI	Suicidal thoughts and attempts
GEN	General health	SXB	Sexual Behaviour
GEO	Geographic identifiers (methodology)	TAL	Tobacco alternatives
HCU	Health care utilization	TWD	Two-week disability
HMC	Home care	UPE	Use of protective equipment
HUI	Health Utility Index (HUI)	WTS	Sample weights
HWT	Height and Weight	WST	Work stress

12.2.3 Position 4: Cycle

Cycle **Description**

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

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

12.2.4 Position 5: Variable Type

_	Collected variable	A variable that appeared directly on the questionnaire
C	Coded variable	A variable coded from one or more collected variables (e.g., SIC, Standard Industrial Classification code)
D	Cross-sectional 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)

12.2.5 Positions 6-8: Variable Name

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

For questions which have more than one response option, the final position in the variable naming sequence is represented by a letter. For this type of question, new variables were created to differentiate between a “yes” or “no” answer for each response

option. For example, if Q2 had 4 response options, the new questions would be named Q2A for option 1, Q2B for option 2, Q2C for option 3, etc. If only options 2 and 3 were selected, then Q2A = No, Q2B = Yes, Q2C = Yes and Q2D = No.

12.3 Access to Master Files data

In order to protect the confidentiality of respondents participating in the survey, microdata files 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 microdata file goes through a formal review process to ensure that an individual cannot be identified. Rare values in variables that may lead to identification of an individual are suppressed on the file or are collapsed to broader categories so that individual disclosure is minimized. Frequently, these are the variables that are most critical for doing a complete and comprehensive analysis of the survey data. Since a significant amount of resources is spent on collecting these data, ensuring that the microdata files reach their full analytical potential is important for a complete return on the statistical investment.

Remote access to the survey master file is one way to have access to these data. Each purchaser of the microdata product can be supplied with a 'dummy' test master file and a corresponding record layout. With this, the user can spend time developing a set of analytical computer programs using the test file to confirm that the program commands are functioning correctly. At that point, the code for the custom tabulations is then sent via e-mail to *cchs-escc@statcan.ca*. The code will then be transferred into Statistics Canada's internal secured network and processed using the appropriate master file of CCHS data. Remote access allows computer programs to be submitted by users for processing at Statistics Canada. 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, once these have been addressed, the output is returned to the client. There is no charge for this service.

A second approach for any client is the production of custom tabulations done by the Client Custom Services staff in Health Statistics Division. This service allows users who do not possess knowledge of tabulation software products to get custom results. As with remote access, the results are screened for confidentiality and reliability concerns before release. Unlike remote access, there is a charge for this service.

Finally, the Research Program allows researchers to submit to Statistics Canada, a research project that uses data from the Master Files. These projects are accepted based on a set of specific rules. When the project is accepted, the researcher is designated as a "deemed employee" of Statistics Canada for the duration of the research, and given access to the Master Files data from designated Statistics Canada sites.