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