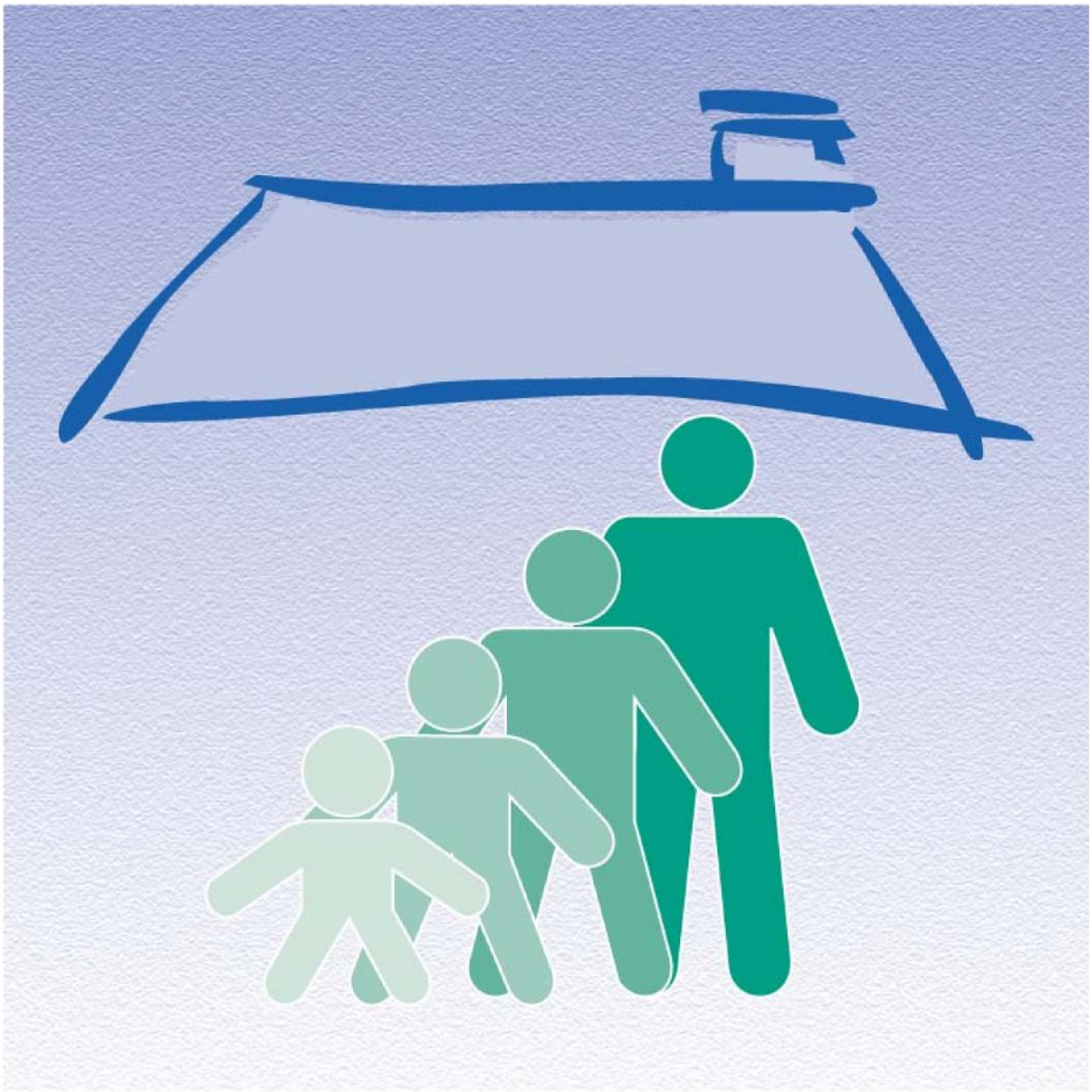
Canadian Community Health Survey (CCHS) - Healthy Aging

User Guide

December 2010



Statistics Canada Statistique Canada

Canada

TABLE OF CONTENTS

1.0 Introduction

The Canadian Community Health Survey (CCHS) - Healthy Aging is a cross-sectional survey that collected information about the factors, influences and processes that contribute to healthy aging through a multidisciplinary approach focusing on health, social and economic determinants.

The CCHS - Healthy Aging collected responses from persons aged 45 and over living in private dwellings in the ten provinces. Excluded from the sampling frame were residents of the three territories, persons living on Indian reserves or Crown lands, persons living in institutions, full-time members of the Canadian Forces and residents of some remote regions. Data were collected between December 2008 and November 2009.

The purpose of this document is to facilitate the manipulation of the CCHS - Healthy Aging microdata files and to describe the methodology used. The CCHS - Healthy Aging survey produces three types of microdata files: master files, share files and public use microdata files (PUMF). The characteristics of each of these files are presented in this guide. The PUMF will be released in 2011.

Any questions about the data set or its use should be directed to:

Electronic Products Help Line:	1-800-949-9491
For custom tabulations or general data support:	
Client Custom Services, Health Statistics Division:	613-951-1746
E-mail:	hd-ds@statcan.gc.ca
For remote access support:	613-951-1653
E-mail:	cchs-escc@statcan.gc.ca
Fax:	613-951-0792

2.0 Background and objectives

2.1 Canadian Community Health Survey (CCHS) background

In 1991, the National Task Force on Health Information cited a number of issues and problems with the health information system. The members felt that data was fragmented; incomplete, could not be easily shared, was not being analysed to the fullest extent, and the results of research were not consistently reaching Canadians.[1]

In responding to these issues, the Canadian Institute for Health Information (CIHI), Statistics Canada and Health Canada joined forces to create a Health Information Roadmap. From this mandate, the Canadian Community Health Survey (CCHS) was conceived. The format, content and objectives of the CCHS evolved through extensive consultation with key experts and federal, provincial and community health region stakeholders to determine their data requirements.[2]

To meet many data requirements, the CCHS had a two-year data collection cycle. Until a redesign in 2007, the first year of the survey cycle, designated by ".1", was a general population health survey, designed to provide reliable estimates at the health region level. The second year of the survey cycle, designated by ".2", had a smaller sample and was designed to provide provincial level results on specific health topics.

New designations for Cycles .1 and .2

As of 2007, the regional CCHS survey began collection on an ongoing basis. To avoid confusion with the health focused surveys, the two components stopped using the ".1" and ".2" designations to distinguish them. Henceforth, the x.1 cycles of the CCHS are designated as "the annual component" of the CCHS. The full title is "The Canadian Community Health Survey – Annual component, 2008" and the short title is simply "CCHS – 2008".

The focused content component of the survey remains unchanged. It will continue to examine in greater detail more specific topics or populations. It will be designated by the name of the survey followed by the topic of the themes covered by each survey (e.g., "Canadian Community Health Survey - Healthy Aging", "CCHS – Healthy Aging" or "CCHS-HA").

2.2 CCHS – Healthy Aging background

The Canadian Community Health Survey – Healthy Aging (CCHS – Healthy Aging) is the third CCHS focused content cycle. Previous cycles include Mental Health in 2002 and Nutrition in 2004. The Canadian Health Measures Survey (CHMS) replaced the 2006 cycle of the survey.

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

In January 2006, Health Statistics Division surveyed primary stakeholders to determine their data needs for 2008-2009 and future focus content cycles of CCHS. The consultation process involved stakeholder prioritization of the various themes in terms of relevance, issues, timing, and topics within themes. Participating stakeholders included Health Canada, the Public Health Agency of Canada, representatives from the provinces and health regions, and members of the research community. As a result of this consultation, healthy aging was selected for the 2009 CCHS focus content survey.

2.3 Objectives

The objectives of the survey were as follows:

- to better understand the aging process of Canadians aged 45 and over by collecting data on various aspects of their health and well being, use of health care services, social support and participation and work and retirement transitions;

- to examine how lifestyle determinants affect health as people age;

- to examine the links among healthy aging and social, demographic, geographic and economic variables or characteristics using a multidisciplinary approach; and

- to provide information on successful aging by age group and sex.

3.0 Survey content

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 the CCHS – Healthy Aging.

The CCHS – Healthy Aging content was developed based on a multi-stage consultation process between the Health Statistics Division at Statistics Canada, Health Canada and the Public Health Agency of Canada, with significant input from the experts conducting the Canadian Longitudinal Study on Aging (CLSA). Consultations also included stakeholders from Human Resources and Social Development Canada and provincial and territorial health ministries.

Questionnaire design started in the fall of 2006 and involved extensive consultations with stakeholders. Content revisions were on-going until June 2008.

Content selection was based on the following criteria:

- Issue(s) identified as data gaps from the stakeholder consultations;
- Strong evidence of aging as an important influence on health and quality of life, morbidity and premature mortality;
- Significant number of people affected by the targeted issue;
- Significant impact on family, community, and health care costs;
- Data have potential for health improvement with policy intervention;
- Issue(s) identified as priority for the support/development of programs and policy, surveillance requirements and/or research; and
- Issues(s) identified as relevant to the Healthy Aging Framework.

3.1 Qualitative testing

Questionnaire review and qualitative testing were conducted in English and French through Statistics Canada‘s Questionnaire Design Resource Centre (QDRC). The questionnaire was subjected to three phases of qualitative testing which took place in March, May and June, 2007 and consisted of one-on-one interviews and focus groups.

A total of 52 qualitative one-on-one interviews were completed in Ottawa, Gatineau and Montreal. The questionnaires were administered face-to-face with respondents. These one-on-one interviews explored the four steps in the cognitive process of responding to the questionnaire: understanding the question and response categories,

recalling/searching for the requested information, thinking about the answer and making a judgment about what to report, and reporting the answer. Results of each qualitative testing phase were instrumental in revising questionnaire content for the subsequent round of qualitative testing.

In addition to one-on-one interviews, the QDRC also held focus group discussions. A total of eight (8) focus groups were conducted in English and French in Ottawa, Calgary and Montreal. Focus group participants were asked for their reactions and feedback to the proposed survey. Overall, focus group participants had favourable reactions and attitudes towards a healthy aging survey. In addition, participants did not consider any of the possible survey topics to be particularly sensitive.

All qualitative interviews were conducted by trained interviewers from QDRC and observed by members of the CCHS – Healthy Aging project team. Some of the key findings from the qualitative testing are discussed below.

Key findings from the qualitative testing

Overall, participants reported the questionnaire to be straightforward and easy to answer but lengthy. In general, respondents did not find the questions to be overly sensitive except questions on social networks, assets and inheritance and loneliness. Qualitative testing also indicated that it was feasible to assess cognition in a population survey. Participants showed willingness to complete the tasks. Results showed that even if the memory task and the alternation task were somewhat challenging for some respondents, they did not create too much stress and anxiety.

In addition, qualitative testing provided valuable guidance on how to adapt some instruments, such as the Physical Activity Scale of the Elderly (PASE) and the Older American Resources and Services (OARS) Multidimensional Functional Assessment Questionnaire, to a computer assisted personal interviewer. During qualitative testing, the sections on Care giving and Care receiving were considered as very long and demanding. Modifications were made to simplify these modules and add clarification on what was to be reported.

3.2 Pilot survey

A pilot survey was conducted in November and early December 2007 in Halifax, Montreal, Toronto and Winnipeg. The 2005 Canadian Community Health Survey (CCHS 3.1) frame was used to select approximately 1,000 respondents. Respondents were interviewed in their dwelling using a computer-assisted personal interviewing (CAPI) collection method. The objectives of the pilot test were to observe respondent reaction to the survey, to obtain estimates of time for the various content modules, to determine the effectiveness of the training of interviewers and the communication strategy, to test the training procedures and material, to provide a preliminary indication of response rates, to further test the computerized questionnaire and to assess the feasibility of using the CCHS – Healthy Aging as the inception cohort of a longitudinal survey.

The number and length of questionnaire modules developed for the pilot survey greatly exceeded Statistics Canada's capacity to administer a questionnaire within the planned interview length. For the pilot survey, survey content was split into two (2) questionnaires (Pilot A and Pilot B). In consideration of a possible longitudinal survey, all proposed content was tested with the intention that, following the pilot survey, some modules would be selected for the main survey while the remaining modules could be deferred to future waves of a longitudinal survey.

The selection of the final content modules for the CCHS – Healthy Aging main survey was based on several factors, including:

- Analysis of results from the pilot survey, including feedback received from the interviewers and field operation personnel;
- Constraints associated with the administration of the questionnaire;
- Input from experts and feedback received from stakeholders, including Health Canada, the Public Health Agency of Canada and the Canadian Longitudinal Study on Aging who provided specific input and priorities; and
- The final decision that the main survey would be designed as a cross-sectional survey.

Following the Pilot test, decisions were made to exclude the longer, complex or more sensitive modules. The questions on social networks, assets and inheritances, work history, spousal retirement and self-reported cognition were removed to avoid respondent fatigue and respondent burden.

3.3 Final questionnaire content

This section outlines the modules comprising the content of the CCHS – Healthy Aging questionnaire. The questionnaire was made up of 37 modules, excluding the entry and exit modules. Table 3.1 provides a summary description of each module.

Table 3.1 Summary description of modules

	Module	Module Summary Description & Source	Additional comments
1	Proxy interview (GR)	This module collected information about interviews being completed by proxy and contained questions about the identity of the proxy respondent and reasons for the proxy interview. Source: Canadian Community Health Survey 2008.	A proxy interview was permitted with another household member if the selected respondent was unable to participate due to a mental and/or physical health issue.
2	General health (GEN)	The general health module was used to collect data on self-perceived health, self-perceived mental health, self-perceived stress and sense of belonging to local community. Source: Canadian Community Health Survey 2008.	The module was modified to utilize a different life satisfaction question (GEN_Q02AA).
3	Sleep 2 (SL2)	In this module, respondents were asked one question on trouble going to sleep or staying asleep. Source: Canadian Community Health Survey 2008.	
4	Height and weight – self-reported (HWT)	In this module, respondents were asked about their height and weight, and whether they considered themselves overweight, underweight or just about right. The data serve to calculate the body mass index (BMI), which can be used to identify whether a respondent is underweight or obese. Source: Canadian Community Health Survey 2008.	
5	Chronic	This group of questions dealt	This module was modified

	Module	Module Summary Description & Source	Additional comments
	conditions (CCC)	with long-term health conditions, diagnosed by a health professional, that have lasted or are expected to last at least 6 months. Source: Canadian Community Health Survey 2008.	to also include conditions more frequently present in older populations: angina, osteoporosis, Parkinson's Disease, cataracts and thyroid disease.
6	Healthy utility index (HUI)	This module contained seven sets of questions. Each one dealt with different aspects of health: vision, hearing, speech, mobility, dexterity, emotion and cognition (memory and thinking). The respondent's answers to the seven sets of questions plus the addition of the HUP questions on pain and discomfort are combined to obtain a score that indicates overall functional health. Source: Canadian Community Health Survey 2008.	
7	Pain and discomfort (HUP)	This module contained three questions about pain and discomfort. These questions represent an eighth dimension for the Health Utility Index. Source: Canadian Community Health Survey 2008.	
8	Satisfaction with life scale (SLS)	This module asked respondents how satisfied they were with their life overall. Answers to this module are used to calculate a respondent's life satisfaction score. Source: The Satisfaction With Life Scale was developed by Dr. Ed Diener (University of Illinois).	
9	Cognition	The purpose of the cognition	Separate coding and

	Module	Module Summary Description & Source	Additional comments
	(COG)	module was to assess the cognitive functioning of respondents as well as to identify gradual changes with aging over time. The cognition module contained four timed tasks: • The first task was a memory test. The respondent was asked to recall a list of words. • The second task tested executive functioning. It required the respondent to name as many animals as they could in a minute. • The third task, the Mental Alternation Test, included three sub-tasks: counting from 1 to 20, reciting the alphabet, and alternating numbers and letters (1A, 2B, 3C, etc.). • The fourth task asked respondents to recall the list of words heard in task one. Sources: • The memory test used the "REY word list", a modified version of the Rey Auditory Verbal Learning Test (RAVLT). • The Animals test is a common test used in many standardized batteries for the measure of cognition. • The Mental Alternation Test (MAT) © is modeled on the Trail making test, which is a common measure used by neuropsychologists. The MAT was developed by Dr. Evelyn	weighting procedures were required for this module. For further information, see Section 6.2.1 "Coding of cognition module data".

	Module	Module Summary Description & Source	Additional comments
		Teng (University of Southern California). Statistics Canada received permission from Dr Teng for the use of this instrument in this survey.	
10	Physical activities 2 (PA2)	The questions in this module were slightly modified from the Physical Activity Scale of the Elderly (PASE) and asked about light, moderate and strenuous activities and exercise in the past seven days, as well as household and volunteer or work physical activities during the past seven days. Answers to these questions were used to calculate a respondent's physical activity score. Additional questions were added following the PASE instrument (starting at PA2_Q11A) and asked whether or not the activity reported in the last seven days represented normal routine activity over the last 12 months, whether the respondent wanted to participate more and, if yes, why were they not able to. Source: This module is a modification of the Physical Activity Scale for the Elderly (PASE)© New England Research Institutes. Statistics Canada secured the rights for the use of this instrument in this survey. The French version was based on the wording used in the Quebec Longitudinal Study on Nutrition and Aging	During data processing, it was discovered that several responses in the "Other-Specify" field were miscoded. An extensive review was done of the responses. For further information, see Section 6.1.1 "Editing of physical activities".

	Module	Module Summary Description & Source	Additional comments
		(NuAge).	
11	Nutritional risk (NUR)	This module used the Seniors in the Community: Risk Evaluation for Eating and Nutrition (SCREEN) screening tool to ask respondents about weight change; eating habits; difficulty eating; fruit, vegetable and fluid consumption; and meal satisfaction. Answers were used to calculate a nutritional risk score. Source: This module is a modification of the SCREEN© developed by Dr. Heather Keller (University of Guess) and is used with the permission of the author.	
12	Oral health 3 (OH3)	In this module, respondents were asked to describe the health of their mouth, denture wearing, tooth brushing and flossing habits, and to report how often they avoided eating particular foods and experienced discomfort when eating any food because of problems with their mouth. Questions about dental visits were also included. Source: The majority of questions were taken from the Canadian Health Measures Survey (CHMS) (OHM module) and the Canadian Community Health Survey 2008 (OH2 and DEN modules).	
13	Medication use (MED)	This module asked respondents about prescription and non-prescription medications taken	

	Module	Module Summary Description & Source	Additional comments
		in the past month. One follow-up question was added to determine if respondents used specific medication on a daily basis. Source: Canadian Community Health Survey 2005.	
14	Dietary supplement use – Vitamins and minerals (DS2)	The questions in this module asked about the frequency of use of multivitamins, calcium, vitamin D and B12 supplements. Source: New module based on Canadian Community Health Survey 2008 and Canadian Health Measures Survey.	
15	Smoking (SMK)	This module included a series of questions about current and past smoking habits, including frequency of smoking, number of cigarettes smoked in a day and the age at which the respondent began smoking daily. Source: Canadian Community Health Survey 2008 and the Canadian Health Measures Survey.	
16	Alcohol use (ALC)	This module asked respondents about alcohol use and frequency of heavy drinking. Source: Canadian Community Health Survey 2008.	
17	Changes made to improve health (CIH)	This module asked respondents a series of questions related to changes in behaviour that they have made in the past year to improve their health, including losing weight and stopping smoking. The module also	

	Module	Module Summary Description & Source	Additional comments
		asked what stopped them from making a change, including lack of time and lack of will power. Source: Canadian Community Health Survey 2008.	
18	Falls (FAL)	This module asked about falls in the past 12 months in which the respondent has been hurt enough to limit some or all of their normal activities. More detailed questions were then asked on follow-up care and fear of falling. Source: New module. Similar content in Canadian Community Health Survey 2008 (Injuries module), the US Health and Retirement Study and the Survey on Healthy Aging and Retirement in Europe. The questions concerning fear or concern of falling are based on the original work of Tennstedt S, et al., 1998.	This module is only asked **to respondents aged 65 years and over.**
19	Instrumental activities of daily living (IAL)	This module was a modified version of the Older American Resources and Services (OARS) Multidimensional Functional Assessment Questionnaire. Respondents were asked a series of questions about their ability to independently perform a series of daily activities such as using the telephone, traveling, shopping, cooking, doing housework, taking medicine and handling money. Source: Modified version of	

	Module	Module Summary Description & Source	Additional comments
		the Older American Resources and Services (OARS) Multidimensional Functional Assessment Questionnaire©. Permission for the use of the instrument was granted by Dr. Gerda G. Fillenbaum (Duke University).	
20	Basic activities of daily living (ADL)	This modified version of the Older American Resources and Services (OARS) Multidimensional Functional Assessment Questionnaire assessed respondents' ability to perform basic daily activities such as feeding and dressing, taking care of their appearance, walking around, getting in and out the bed, bathing and incontinence problems. Source: See IAL above.	
21	Health case utilization 2 (HC2)	This module asked respondents a series of questions related to the use of health care services, including having a regular medical doctor, use of medical specialists, and use of alternative medicine providers. Source: Modified from the Canadian Community Health Survey 2008 (HCU module).	
22	Care receiving 1 (CR1)	This module asked respondents whether they received home care services provided by professionals due to a health problem that affects their daily activities. Respondents were asked to report assistance provided by paid workers or organizations for various activities, such as	"Professional home care services" mean health care, homemaker or other support services received at home from paid workers or volunteer organizations due to a health condition or limitation, regardless of the funding for these services. CCHS 2008 asked about

	Module	Module Summary Description & Source	Additional comments
		medical care, personal care, managing care, housework, transportation and meal preparation or delivery. Respondents are also asked about home care they needed but did not receive. Source: New module, similar content in the General Social Survey (GSS), Cycle 21 and Canadian Community Health Survey 2008 (HMC module).	health care, homemaker or other support services received at home with the cost being entirely or partially funded by government.
23	Care receiving 2 (CR2)	This module asked respondents about assistance provided by family members, friends or neighbours (informal caregivers) due to a health problem that affects daily activities. Respondents were asked to report assistance provided for various activities, such as medical care, personal care, managing care, housework, transportation and meal preparation or delivery. The module included a range of questions related to informal home care including the identity of the person providing assistance, the duration and level of intensity of the care received. Source: New module, similar content in General Social Survey (GSS) (Cycles 16 and 21).	
24	Social support availability (SSA)	This module included a series of detailed questions about the availability of social support.	

	Module	Module Summary Description & Source	Additional comments
		The answers were combined to determine how likely it is that the respondent has access to social support. Source: Canadian Community Health Survey 2008.	
25	Social participation (SPA)	The module asked about the type and amount of participation in different social activities such as sporting activities, religious services, and limitations to such participation. Source: New module.	
26	Care giving (CAG)	This module asked respondents whether they provided assistance to others because of a health condition or limitation during the past 12 months. Questions in this module covered topics related to assisting others such as types of assistance provided, total number of people to whom assistance was provided as well information about the person the caregiver helped the most, the impact of providing assistance on work and health and the positive and negative aspects of providing assistance. Source: New module, similar content in General Social Survey (GSS) (Cycles 16 and 21).	
27	Care giving expenses (CGE)	This sub-module of Care Giving (CAG) asked respondents about out-of-pocket expenses due to care giving, including expenses for	

	Module	Module Summary Description & Source	Additional comments
		purchasing items, paying for services, transportation, and increased expenses such as additional housing costs or food. Source: New module, similar content in the Participation and Activity Limitation Survey (PALS).	
28	Depression (DEP)	This module consisted of a series of detailed questions about depression. Source: Canadian Community Health Survey 2008.	It is recommended that analysis be restricted to examination of depression as a correlate of other health behaviours and characteristics. Use of the data to calculate prevalence is discouraged.
29	Loneliness (LON)	This module was taken from the 3-item Loneliness Score. Respondents are asked about the lack of companionship, feeling left out, and feeling isolated. A loneliness score was then calculated. Source: The 3-item Loneliness Score developed by Dr. Mary Elizabeth Hugues (Johns Hopkins Bloomberg School of Public Health). Permission to use the instrument was granted by the author.	
30	Transportation (TRA)	This module asked respondents about having a driver's license, most common method of transportation, and access to public transportation. Source: new module, similar module in CCHS 2001 and the General Social Survey (GSS) Cycle 19.	
31	Labour force	This module collected	This module is only asked

	Module	Module Summary Description & Source	Additional comments
	(LBF)	information on respondents' work life, including whether they were employed, unemployed or retired. Questions about occupation, reasons for not working and usual work schedule were also asked. Source: Modified from Canadian Community Health Survey 2005.	of **respondents aged between 45 and 74 years of age**. A question was added to take into account those who have never worked.
32	Reasons for retirement (RET)	The questions in this module asked about age at retirement, main reasons for retirement, labour force participation and ability to work. This module also asked about partial retirement for those respondents who may have retired officially but have continued working or who are taking a gradual retirement. Source: New module, similar content in the General Social Survey (GSS) (Cycles 16 and 21).	The module is only asked of **respondents between 45 and 85 years of age.**
33	Retirement planning (RPL)	This module asked about planning for retirement: desired age of retirement, pensions and other preparations and planning for retirement. Source: New module, similar content in the General Social Survey (GSS) (Cycles 16 and 21).	The module is only asked of **respondents between 45 and 85 years of age who were not retired or only partly retired.**
34	Home owner (OWN)	This module included questions modified from the Survey of Household Spending, Survey on Ageing and Independence and the	

	Module	Module Summary Description & Source	Additional comments
		General Social Survey, Cycle 21. Respondents were asked to provide information about their home ownership status, the value of their home, and the value of their mortgage. Source: New module. Modified from the Survey of Household Spending, Survey on Ageing and Independence and General Social Survey, (GSS) Cycle 21.	
35	Income (INC)	In this module, respondents were asked to provide their individual and household incomes in order to identify all sources of household and personal income, as well as the main source for each. Follow up questions were asked about Canada or Quebec pension plan benefits. Respondents were also asked to estimate the value of their savings and investments and whether their income covered their basic expenses. Source: Modified from the Canadian Community Health Survey 2008 with the addition of new questions.	
36	Socio-demographic characteristics (SDC)	This module collected social and demographic information, including immigrant status, country of birth, ethnic origin, and language. Source: Canadian Community Health Survey 2008.	
37	Administration information (ADM)	This module collected administrative information including permission for data	The CLS sub-module was only asked of **respondents between 45 and 85 years**

	Module	Module Summary Description & Source	Additional comments
		linkage and data sharing. There was a sub-module called 'Consent to Share' (CLS). Respondents were given an information package and consent form to review and complete regarding sharing contact and/or survey information with three universities who lead an initiative called 'The Canadian Longitudinal Study on Aging'. Source: Canadian Community Health Survey 2008 and new sub-module (CLS).	**of age.**

4.0 Sample design

4.1 Target population

The CCHS – Healthy Aging targets persons aged 45 years and older who are living in private dwellings in the ten provinces. Residents of the three territories, persons living on Indian Reserves or Crown lands, those residing in institutions, full-time members of the Canadian Forces and residents of certain remote regions are excluded from this survey.

4.2 Sample size and allocation

The original target population was 55 and older. The additional 5,000 respondents aged 45-54 were funded by the Canadian Longitudinal Study on Aging. To meet the survey objectives of estimating healthy aging among Canadians aged 45 and older for specific domains of interest for each province, and given the budget allocated to the survey, a sample of 32,000 responding units was desired over a period of one year. Of this total, 5,000 respondents would be between ages 45–54 and 27,000 would be of age 55 or older. The goal was to produce reliable estimates by province for five age groups (45–54, 55–64, 65–74, 75–84 and 85+) and by sex.

A two-step strategy was used to allocate the sample to the provinces. First, 125 sample units were allocated to each domain of interest (10 age/sex groups) in each province. Thus, 1,250 units were assigned to each province in the first step for a total of 12,500. The remaining 19,500 units were allocated to the provinces using a power allocation method with power $q=0.7$[3]. The total sample size of any given province was found by adding the sizes obtained in the two steps.

In order to have a good urban and rural representation in each province, the sample was subsequently allocated to two strata: urban and rural, as derived from the 2006 Census Blocks. The provincial sample was proportionally allocated to the urban and rural strata according to the number of dwellings having people aged 45 and older in each stratum. Sample sizes were then inflated before data collection to take into account anticipated non-response and out-of-scope dwellings. The latter was particularly important in this survey since selected households would be rejected if there was no one in the target population (age 45+) living there. The raw sample size required to obtain 32,000 respondents was estimated at 52,010. Table 4.1 gives the raw sample sizes for the CCHS – Healthy Aging by province and urban/rural status.

[3] Bankier, M.D. (1988). Power Allocations: Determining Sample Sizes for Subnational Areas. *The American Statistician*, 42, 174-177.

Table 4.1 Raw sample sizes by province and urban/rural status

Province	Urban	Rural	Total
Newfoundland and Labrador	1,885	1,320	3,205
Prince Edward Island	1,397	1,160	2,557
Nova Scotia	2,201	1,560	3,761
New Brunswick	2,016	1,440	3,456
Quebec	7,490	1,316	8.806
Ontario	9,762	1,280	11,042
Manitoba	3,045	740	3,785
Saskatchewan	2,555	1,079	3,634
Alberta	4,585	675	5,260
British Columbia	5,904	600	6,504
Canada	**40,840**	**11,170**	**52,010**

4.3 Sample design

A three-stage design was used to select the sample of CCHS – Healthy Aging respondents. First, geographical areas called clusters were selected. Households were then selected within each sampled cluster and, finally, one respondent per household was randomly selected. Each stage is described below.

4.3.1 Selection of geographical areas

In order to control cost and ensure a manageable distribution of interviews in the field since personal interviews were used, groups of households were created based on geographical proximity. These groups are called clusters. The clusters were created so that each and every dwelling in Canada is part of one and only one cluster. 2006 Census Blocks (CB) were used as the building blocks for the clusters. Adjoining CBs were grouped to create larger areas, based on the following criteria (using the 2006 Census data):

- At least 100 dwellings per cluster;
- At least a certain number of dwellings with members aged 75–84 and 85+ (this number varied by province and urban/rural status);
- No more than a certain land area for each cluster (again varying by province and urban/rural status);
- No grouping of urban and rural CBs in the same cluster, as well as no cluster crossing provincial boundaries;
- Emphasis on accessibility (being able to travel from one point in the cluster to another point without going out of it) and compactness.

A different level of importance could be assigned to each of these criteria. Not all clusters created satisfied all criteria but the final set of clusters most closely matched those requirements. In total, close to 17,000 clusters were created.

The number of dwellings to select in each cluster had to be determined in order to decide on the number of clusters to sample. This was done in an effort to balance cost effectiveness and precision. Fewer clusters selected with more interviews per cluster meant a lower collection cost, while more clusters with fewer interviews in each meant better precision because it reduces the "cluster effect" caused by the fact that dwellings in the same cluster tend to be more similar than dwellings in general. It was determined that a raw sample size of 35 cases per urban cluster and 20 per rural cluster was reasonable.

To achieve that, the clusters were divided into two groups (a process called stratification) in each province: urban and rural clusters. In each province, the total sample was divided among urban and rural based on the proportion of persons aged 85 and older in each of the two groups. Then, using the target of 35 and 20 cases per urban and rural cluster, the required number of clusters to sample in the urban and rural areas in each province was derived.

The clusters were then selected using a sampling with probability proportional to size approach. Under this method, the higher the number of persons aged 45 and older in a particular cluster, the higher the chance of this cluster being selected in the sample. This method allows for a better control of the variability that could be introduced by the sampling of clusters of varying sizes.

Finally, each cluster was randomly allocated to a collection period. The survey was in the field from December 2008 to November 2009, but this was divided into six collection periods of two months. The sample for the first two collection periods (December 2008 to January 2009 and February to March 2009) was reduced because of decreased interviewer resources. It was increased in the other periods to make up for that adjustment.

4.3.2 Selection of households within each cluster

Because the targeted sample size for older Canadians (the 75–84 and 85+ age groups) was much more than their proportional share of the population, getting enough in the sample was one of the main challenges of this design. Several sources were considered, but it was decided that the 2006 Census of Population would be used as the sampling frame from which dwellings would be selected. The Census information could then be used to target dwellings that were more likely to include people aged 45 and older.

Only dwellings where there was at least one person aged 45 and older according to the Census household composition (with the age adjusted to reflect age at the time of collection) were considered.

In each selected cluster, the dwellings were divided into three strata:

- Those with at least one person aged 85 and older;
- Those with only people under age 55 (but at least one 45–54);
- All other dwellings.

The first two strata were created to ensure that the number of people selected in each age group was appropriate. Without this stratification, the sample would have included too many people in the 45–54 age group and not enough in the 85+ group.

The sample of 35 dwellings in urban areas and 20 in rural areas was then allocated to each of those three strata, with an allocation that varied by province. One challenge of this method was that the sample was selected based on the household composition at the time of the Census, which took place 2.5 to 3.5 years prior to collection for this survey. The composition of many households will have changed during that time. To determine how the sample should be allocated to the three strata, simulations were run using the household composition at the time of the Census and then, changes in household composition were imputed based on what was observed in the pilot survey. The sample for the pilot was drawn from respondents to the CCHS Cycle 3.1 (2005) but took place 2 to 2.5 years later. This allocation was later revised based on what was observed in the first two collection periods.

Within each stratum, the required number of dwellings in the sample was obtained using simple random sampling.

4.3.3 Sampling of interviewees within each household

One person in each dwelling was chosen to be the respondent for the CCHS – Healthy Aging survey. Upon visiting a selected dwelling, the household composition at the time of the survey was obtained. Households with no eligible respondents (only people aged 44 and younger or those not in the target population) were classified as out of scope. For the other households, one respondent was selected at random among all eligible respondents according to varying selection probabilities.

Every household member aged 45 and older was assigned a selection probability factor according to the five age groups for which estimates are required (45–54, 55–64, 65–74, 75–84, 85+). The selection probabilities varied by province in order to achieve the targeted number of respondents in each age group. A simulation study using household roster information from the pilot survey was used to create the selection probabilities.

5.0 Data collection

Collection for the Canadian Community Health Survey - Healthy Aging took place between December 2008 and November 2009. Over the collection period, a total of 30,865 valid interviews were conducted using a type of computer-assisted interviewing (CAI) called computer assisted personal interviewing (CAPI).

5.1 Computer-assisted interviewing

Computer-assisted interviewing (CAI) offers two main advantages over other collection methods. First, CAI offers a case management system and data transmission functionality. This case management system automatically records important management information for each attempt on a case and provides reports for the management of the collection process.

The case management system routes the questionnaire applications and sample files from Statistics Canada's main office to the regional offices. Data returning to the main office take the reverse route. To ensure confidentiality, the data are encrypted before transmission. The data are then unencrypted when they are on a separate secure computer with no remote access.

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

- Questions that are not applicable to the respondent are skipped automatically;
- Edits to check for inconsistent answers or out-of-range responses are applied automatically and on-screen prompts are shown when an invalid entry is recorded;
- Immediate feedback is given to the respondent and the interviewer is able to correct any inconsistencies; and
- 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.

5.2 CCHS – Healthy Aging application development

For the CCHS – Healthy Aging, a computer assisted personal interviewing (CAPI) application was utilized. The application consisted of entry, health content (known as the C2), and exit components.

Entry and exit components contain standard sets of questions designed to guide the interviewer through contact initiation, collection of important sample information, respondent selection and determination of cases status. The C2 consists of the health modules and made up the bulk of the application. The application underwent three stages of testing: block testing, integrated testing and end-to-end testing.

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

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

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

5.3 Interviewer training

In October and November 2008, representatives from Statistics Canada's Collection Planning and Management Division visited the four regional offices (Halifax, Montreal, Toronto and Edmonton) coordinating data collection for the CCHS – Healthy Aging. The purpose of the visits was to train the regional office project managers and senior interviewers for the CCHS – Healthy Aging. Members of the CCHS – Healthy Aging project team from Health Statistics Division also attended the training sessions to present information about the background and development of the survey, and to offer additional support and clarify any questions or concerns that arose.

The focus of these sessions was to get interviewers comfortable using the CCHS – Healthy Aging application and familiarise interviewers with survey content. The training sessions focused on:

- the goals and objectives of the survey;
- survey methodology;
- application functionality;
- review of the questionnaire content and exercises;
- the administration of the cognition module-exercises on recording audio responses;

- use of mock interviews to simulate difficult situations and practise potential non-response situations;
- survey management; and
- transmission procedures.

5.4 The interview

Units selected from the area frame were interviewed by decentralized field interviewers using CAPI. CAPI interviewers worked independently from their homes using laptop computers and were supervised from a distance by senior interviewers. CAPI interviewers were trained to make an initial personal contact with each sampled dwelling. Every effort was made to conduct the interviews face-to-face, and 94% of interviews were conducted exclusively in person. Collection by telephone was authorized only when a respondent requested an interview in the other official language but no bilingual interviewer was available in the area, or when the respondent spoke neither official language but another interviewer was available to translate for the respondent.

In all selected dwellings, a knowledgeable household member was asked to supply basic demographic information on all residents of the dwelling. One member of the household aged 45 years or older was then selected for a more in-depth interview, which is referred to as the C2 interview.

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

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

a) Introductory letters

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

b) Initiating contact

Interviewers were instructed to make all reasonable attempts to obtain interviews. When the timing of the interviewer's visit was inconvenient, an appointment was made to conduct the interview at a more convenient time. If no one was home on first visit, a notice of visit and intention to make contact was left at the door. Numerous repeat visits to the dwelling were made at different times on different days.

c) Refusal conversion

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

d) Language barriers

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

e) Proxy interviews

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

5.5 Field operations

The CCHS – Healthy Aging sample was divided into six non-overlapping two-month collection periods. Regional collection offices were instructed to use the first 4 weeks of each collection period to resolve the majority of the sample, with the next 4 weeks being used to finalise the remaining sample and to follow up on outstanding non-response cases. All cases were to have been attempted by the second week of each collection period. Cases initially coded to “No contact” or “Absent for duration of survey” were resent in two additional collection periods: 1) August 15 to November 30, 2009 and 2) October 15 to November 30, 2009.

Sample files were sent approximately six weeks before the start of each collection period to the four regional collection offices for workload planning. A series of dummy cases were included with each CAPI sample. These cases were completed by senior interviewers for the purposes of ensuring that all data transmission procedures were working through the collection cycle. Once the samples were received, project supervisors were responsible for planning CAPI interviewer assignments.

Completed CAPI interviews were transmitted daily from the interviewer’s home directly to Statistics Canada’s head office using a secure telephone transmission.

At the end of data collection, a national response rate of 74% was achieved. Complete details regarding the response rates can be found in Chapter 8, Section 8.1.

5.6 Quality control and collection management

During the CCHS – Healthy Aging data collection, several methods were used to ensure data quality and to optimize collection. These included using internal measures to verify interviewer performance, field interview observation, the use of interviewer newsletters as well as a series of ongoing reports to monitor various collection targets and data quality.

Head office team members observed interviews by accompanying interviewers during door to door interviews. The main goal of those observations was to ensure that proper interviewing techniques and procedures (reading the question as worded on the application, not prompting respondents for answers, etc.) were followed.

Three interviewer newsletters were produced and distributed to all interviewers. This communication tool provided direct feedback to interviewers on data quality issues. It also provided updates on collection, responses to frequent collection questions, additional information and the reinforcement of proper interviewing procedures and techniques.

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

Cumulative reports were generated at the end of each collection period, showing response, link, share and proxy rates for each collection region. Reports were also generated by the survey methodologists to monitor sample clusters and age groups that were below collection target levels, allowing the regional offices to focus efforts in these regions.

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

6.0 Data processing

6.1 Editing

Most editing of the data was performed at the time of the interview by the computer-assisted personal interviewing (CAPI) application. It was not possible for interviewers to enter out-of-range values and flow errors were controlled through programmed skip patterns. For example, the CAPI application 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".

6.1.1 Editing of physical activities module

In the physical activities 2 module, questions were asked about the respondent's level of physical activity. Activities were recorded for four levels (light, moderate, strenuous, muscle strength and endurance) and, for each level, the weekly frequency and daily duration of all activities was recorded. Weekly frequency and duration variables were categorical variables. Problems occurred when activities that should have been recorded under one level were recorded under another level (i.e. reporting bicycling as a light activity when it should be a strenuous activity). To correct this misclassification, these activities have been moved to the correct level.

For every activity that was moved, the weekly frequency and daily duration in the two affected levels were corrected. A proportional amount of the weekly frequency was removed from the previous activity level and transferred to the new activity level. With respect to the daily duration, the same duration was carried forward from the previous level to the new level. When activities already existed at the new activity level, the values were added to the existing weekly frequency and daily duration. In this manner, the total amount of activity (in terms of frequency times duration) is preserved, but the level of effort for any particular activity is changed to be consistent among all records.

6.2 Coding

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

In the event that a respondent’s answer could not be easily assigned to an existing category, several questions also allowed the interviewer to enter a long-answer text in the “Other-specify” category. Following collection, all such questions were reviewed in head office processing. For some of these questions, write-in responses were coded into one of the existing listed categories if the write-in information duplicated a listed category. For

all questions, the 'Other-specify' responses are taken into account when refining the answer categories for future cycles.

6.2.1 Coding of cognition module

The cognition module of the survey asked respondents to complete four tasks for which their responses (with their permission) were audio recorded. Those tasks were: naming as many animals as they could, counting from 1 to 20, reciting the letters of the alphabet, and alternating numbers and letters. Respondents were allowed a maximum of one minute to complete the animal naming task and 30 seconds for each of the remaining tasks.

Processing and coding of the recorded responses took place over several steps. First, the audio recordings (audio files) were returned to the Statistics Canada Head Office in blocks of audio files separate from the survey data; each with an unique name so that they could later be identified and associated with respondent's survey data. The blocks of audio files were unpacked and sorted into separate databases for each of the four tasks. Audio files from each cognition task were then moved to their separate database. An internally developed interface allowed the members of the data coding team to access the audio files in those databases, listen to the recordings and coding them. Coding rules were provided. Once all four tasks were coded for a respondent, the coded data were integrated with the respondent's survey data.

6.3 Creation of derived variables

To facilitate data analysis and to minimize the risk of error, a number of variables on the file have been derived using items found on the CCHS – Healthy Aging questionnaire. Derived variables generally have a "D", "G" or "F" in the fourth character of the variable name. In some cases, the derived variables are straightforward, involving the collapsing of response categories. In other cases, several variables have been combined to create a new variable. The *Derived Variables Documentation (DV)* provides details on how these more complex variables were derived. For more information on the naming convention, please go to Section 11.5.

6.4 Partials

In some cases, it was possible to carry out some part of the interview, but a complete interview was not obtained for a variety of reasons. Some respondents were willing to give only a certain amount of time to the completion of the survey. In other cases, an interviewer completed a portion of the survey with the respondent and made an appointment to continue at another time but was unable to re-contact the respondent. In such situations, it was necessary to come up with standard criteria for deciding what to do with these "partial" interviews.

For the CCHS – Healthy Aging, it was decided that to qualify as a “partial interview”, the respondent must have completed the Entry and Exit modules as well as a minimum part of the general health questionnaire up to and including the first question of the Social Participation (SPA) module. Anything less was considered a non-response – i.e. the household was dropped from the responding sample. The variable ADMD_STA indicates whether a case was partially or fully completed, indicated by a 71 and 70, respectively.

6.5 Weighting

The principle behind estimation in a probability sample such as the CCHS – Healthy Aging 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 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 7.

7.0 Weighting

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

7.1 Sample weighting

Seven separate adjustments are part of the weighting strategy. The following sections describe the weighting process for the survey. Section 7.1.1 gives the initial weight, while Section 7.1.2 provides details on the removal of out-of-scope units. The treatment of non-responding households is covered in Section 7.1.3. This is followed by adjustments for a person-level weight in Section 7.1.4 and for person-level non-response in 7.1.5. Lastly, the final weighting steps of winsorization and calibration are discussed in Sections 7.1.6 and 7.1.7, respectively.

7.1.1 Initial weight

The weighting of the sample begins with an initial household weight (L1). The initial weight covers the first two stages of the sample design covered in Section 4.3 under the selection of clusters and the selection of households within each cluster.

The cluster weight is the inverse of the probability of selecting the cluster. As mentioned in Section 4.3.1, clusters in each province were stratified into urban and rural clusters. The desired number of clusters is then sampled proportional to the number of persons aged 45 and older in each cluster. According to the Hanurav-Vijayan algorithm[4], the selection probability for each cluster is:

$$\textbf{ClusterWeight}_i = n\frac{M_i}{M_{total}},$$

where i = the selected cluster, n = the number of clusters sampled by province and urban/rural status, M_i = the number of persons aged 45 and older in the selected cluster and M_{total} = the total number of persons aged 45 and older in all clusters with the same province and urban/rural status.

The dwelling weight is the inverse of the probability of selecting the dwelling within each cluster. Section 4.3.2 describes this procedure, in which clusters are divided into three strata and a specified number of dwellings are selected from each one. Using simple random sampling, the selection probability for each dwelling is calculated as follows:

[4] Vijayan, K. (1968). An Exact πps Sampling Scheme – Generalization of a Method of Hanurav. *Journal of the Royal Statistical Society*, Series B, 30, 556-566.

$$\textbf{DwellingWeight}_j = \frac{N_j}{n_j},$$

where j = the selected dwelling, N_j = the total number of dwellings in the cluster by stratum and n_j = the number of dwellings sampled in the cluster by stratum.

The initial weight is simply the product of these two components:

$$\textbf{L1} = \text{ClusterWeight}_i \times \text{DwellingWeight}_j.$$

7.1.2 Removal of out-of-scope units

Among all dwellings sampled, a certain proportion is identified during collection as being out of scope. Dwellings that are outside of the target population as specified in Section 4.1 (including those with no persons aged 45 years and older) as well as those that are demolished or under construction, vacant, seasonal or secondary are examples of out-of-scope cases for the CCHS – Healthy Aging. These dwellings and their associated weight are simply removed from the sample. This leaves a sample that consists of, and is representative of, in-scope dwellings. These remaining dwellings maintain the same weight as in the previous step, which is now called L2:

$$\textbf{L2} = \begin{cases} \text{L1} & \text{if unit is in scope} \\ 0 & \text{if unit is out of scope} \end{cases}.$$

7.1.3 Household non-response

During collection, a certain proportion of sampled households inevitably results in non-response. This usually occurs when a household refuses to participate in the survey, provides unusable data, or cannot be reached for an interview. Weights of the non-responding households are redistributed to responding households within response homogeneity groups (RHGs). In order to create the response groups, a scoring method based on logistic regression models is used to determine the propensity to respond, and these response probabilities are used to divide the sample into groups with similar response properties. The information available for non-respondents is limited; as a result, the regression model uses characteristics such as the collection period and geographic information, as well as paradata including the number of contact attempts, the time/day of each attempt, and whether the household was contacted on a weekend or weekday. An adjustment factor is calculated within each RHG as follows:

$$\frac{\text{Sum of weight L2 for all households}}{\text{Sum of weight L2 for all responding households}}.$$

Weight L2 is multiplied by this factor to produce weight L3 for the responding households. Non-responding households are dropped from the process at this point.

7.1.4 Creation of person level weight

Since persons are the desired sampling units, the household level weights computed to this point need to be converted to the person level. This weight is obtained by multiplying the weight L3 by the inverse of the probability of selection of the person selected in the household. This gives the weight I1. As mentioned earlier, the probability of selection for an individual changes depending on the number of persons aged 45 and older in the household and the ages of those individuals (see Section 4.3.3 for more details).

7.1.5 Person non-response

A CCHS – Healthy Aging interview can be seen as a two-part process. First, the interviewer gets the complete roster of the persons within the household. Second, (s)he interviews the selected person. In some cases, interviewers can only get through the first part, either because they cannot get in touch with the selected person, or because the selected person refuses to be interviewed. Such individuals are defined as person non-respondents and an adjustment factor must be applied to the weights of respondents to account for this non-response. Using the same methodology that was used in the treatment of household non-response, the adjustment was applied within response homogeneity groups based on the characteristics available for both respondents and non-respondents. All characteristics collected when creating the roster of household members were in fact available for the creation of the classes, as were geographic information and some paradata. A scoring method was used to define the classes. In the end, the following adjustment factor was calculated:

$$\frac{\text{Sum of weight I1 for all selected persons}}{\text{Sum of weight I1 for all responding selected persons}}.$$

Weight I1 for responding persons was multiplied by the above adjustment factor to produce weight I2. Non-responding persons were dropped from the weighting process from this point onward.

7.1.6 Winsorization

Following the series of adjustments applied to the respondents, some units may come out with extreme weights compared with other units of the same domain of interest (province by age group by sex). Some respondents could represent a large proportion of their domain or have a large impact on the variance. In order to prevent this, the weight of the outlier units that represent a large proportion of their province-age-sex group is adjusted downward using a "winsorization" trimming approach. For the majority of units that are not outliers, the weight I3 produced at this point will be the same as weight I2.

7.1.7 Calibration

The final step necessary to obtain the final CCHS – Healthy Aging weight is calibration (I4). Calibration is done using Calmar[5] to ensure that the sum of the final weights corresponds to the population estimates defined at the province level, for all 10 age–sex groups of interest. The 5 age groups are 45–54, 55–64, 65–74, 75–84 and 85+, for both males and females. A minimum domain size of 20 respondents is required to calibrate at the province by age by sex level. For domains that have fewer than 20 respondents, some collapsing is done within province and/or gender. Note that the calibration is done using the most up-to-date geography and may not match the geography used in sampling.

The population estimates are based on the most recent census counts and counts of births, deaths, immigration and emigration since that time. The average of these monthly estimates for each of the province-age-sex post-strata by collection period is used for calibration. The weight I3 was therefore adjusted using Calmar to obtain the final weight I4. Weight I4 corresponds to the **final CCHS – Healthy Aging weight** that can be found on the data file with the variable name WTS_M.

[5] Sautory, O. (2003). Calmar 2: A new version of the Calmar calibration adjustment program. *Proceedings of Statistics Canada's Symposium 2003: Challenges in Survey Taking for the Next Decade*, Catalogue no. 11-522-XIE.

8.0 Data quality

8.1 Response rates

In total, 41,496 of the selected units in the CCHS – Healthy Aging were in-scope for the survey[6]. Out of these, 33,517 households agreed to participate in the survey, resulting in an overall household-level response rate of 80.8%. Among these responding households, 33,517 individuals (one per household) were selected to participate in the survey, out of which a response was obtained for 30,865 individuals, resulting in an overall person-level response rate of 92.1%. At the Canada level, this yields a combined (household and person) response rate of 74.4% for the CCHS – Healthy Aging. Table 8.1 provides response rates by province and age group[7].

Table 8.1 CCHS – Healthy Aging response rates by province and age group

			Person level			
Province	**Provincial household response rate**	**Age**	**Number of selected people**	**Number of respondents**	**Response rate**	**Combined response rate**
10 - Newfoundland and Labrador	**86.8**	**Total**	**2,185**	**2,010**	**92.0**	**79.9**
		[45–54]	403	358	88.8	
		[55–64]	730	667	91.4	
		[65–74]	466	443	95.1	
		[75–84]	318	299	94.0	
		[85+]	268	243	90.7	
11 - Prince Edward Island	**83.3**	**Total**	**1,765**	**1,650**	**93.5**	**77.8**
		[45–54]	322	293	91.0	
		[55–64]	526	485	92.2	
		[65–74]	350	334	95.4	
		[75–84]	313	300	95.8	
		[85+]	254	238	93.7	
12 - Nova Scotia	**82.4**	**Total**	**2,536**	**2,282**	**90.0**	**74.1**
		[45–54]	399	353	88.5	
		[55–64]	794	706	88.9	

[6] Among the households selected, some are not in-scope for the survey. They are, for example, vacant or demolished dwellings or households that do not have any residents of the appropriate age. These households were identified during data collection; otherwise, they would have been excluded before the sample selection. These households are not considered in the calculation of response rates.

[7] Stratification for the CCHS – Healthy Aging was completed taking into consideration the projected ages of household members at the time of collection. Target response rates were tracked for the age groupings: [45–54], [55–64], [65–74], [75–84], and [85+] to ensure each group was covered sufficiently. The response rates are grouped similarly.

			Person level			
Province	**Provincial household response rate**	**Age**	**Number of selected people**	**Number of respondents**	**Response rate**	**Combined response rate**
		[65–74]	595	551	92.6	
		[75–84]	393	358	91.1	
		[85+]	355	314	88.5	
13 - New Brunswick	**85.2**	**Total**	**2,396**	**2,225**	**92.9**	**79.1**
		[45–54]	381	351	92.1	
		[55–64]	730	676	92.6	
		[65–74]	526	495	94.1	
		[75–84]	400	370	92.5	
		[85+]	359	333	92.8	
24 - Quebec	**80.0**	**Total**	**5,649**	**5,217**	**92.4**	**73.8**
		[45–54]	964	857	88.9	
		[55–64]	1,749	1,623	92.8	
		[65–74]	1,278	1,187	92.9	
		[75–84]	1,046	988	94.5	
		[85+]	612	562	91.8	
35 - Ontario	**79.1**	**Total**	**7,159**	**6,525**	**91.1**	**72.1**
		[45–54]	1,152	1,023	88.8	
		[55–64]	2,187	1,984	90.7	
		[65–74]	1,624	1,496	92.1	
		[75–84]	1,245	1,150	92.4	
		[85+]	951	872	91.7	
46 - Manitoba	**79.9**	**Total**	**2,386**	**2,177**	**91.2**	**72.9**
		[45–54]	429	378	88.1	
		[55–64]	678	622	91.7	
		[65–74]	528	486	92.0	
		[75–84]	399	369	92.5	
		[85+]	352	322	91.5	
47 - Saskatchewan	**80.7**	**Total**	**2,304**	**2,184**	**94.8**	**76.5**
		[45–54]	371	336	90.6	
		[55–64]	672	639	95.1	
		[65–74]	499	476	95.4	
		[75–84]	429	417	97.2	

			Person level			
Province	**Provincial household response rate**	**Age**	**Number of selected people**	**Number of respondents**	**Response rate**	**Combined response rate**
		[85+]	333	316	94.9	
48 - Alberta	**78.8**	**Total**	**3,012**	**2,735**	**90.8**	**71.6**
		[45–54]	633	558	88.2	
		[55–64]	853	775	90.9	
		[65–74]	648	597	92.1	
		[75–84]	473	436	92.2	
		[85+]	405	369	91.1	
59 - British Columbia	**79.4**	**Total**	**4,125**	**3,860**	**93.6**	**74.3**
		[45–54]	713	659	92.4	
		[55–64]	1,242	1,153	92.8	
		[65–74]	964	910	94.4	
		[75–84]	675	633	93.8	
		[85+]	531	505	95.1	
Canada	**80.8**	**Total**	**33,517**	**30,865**	**92.1**	**74.4**
		[45–54]	5,767	5,166	89.6	
		[55–64]	10,161	9,330	91.8	
		[65–74]	7,478	6,975	93.3	
		[75–84]	5,691	5,320	93.5	
		[85+]	4,420	4,074	92.2	

8.1.1 Response rate formulae and example

The response rate formulae are presented and a simple example is given to illustrate how the household- and person-level response rates are used to calculate the combined response rate.

Household-level response rate

$$HHRR = \frac{\text{Number of responding households}}{\text{Number of in - scope households}}$$

Person-level response rate

$$PPRR = \frac{\text{Number of responding persons}}{\text{Number of persons selected (one per responding household)}}$$

Combined response rate

$$CombRR = HHRR \times PPRR$$

Example 1 – Calculation of CombRR for Newfoundland and Labrador based on Table 8.1

$$HHRR = \frac{\text{Number of responding households}}{\text{Number of in-scope households}} = \frac{2{,}185}{2{,}516} = 86.8\%$$

$$PPRR = \frac{\text{Number of responding persons}}{\text{Number of persons selected (one per responding household)}} = \frac{2{,}010}{2{,}185} = 92.0\%$$

$CombRR = HHRR \times PPRR = 86.8\% \times 92.0\% =$ **79.9%**

8.1.2 Limitations of response rate calculations

Two variables of interest for CCHS – Healthy Aging, age group and sex, are person specific and, as such, response rates can only be reported for these variables at the person level.

8.2 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 census of the target population under similar conditions is called the *sampling error* of the estimate.

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

8.2.1 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, due to 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 it is estimated that 25% of Canadians aged 45 and over exercise at least once a week and that this estimate is found to have a standard deviation of 0.003. Then the CV of the estimate is calculated as:

(0.003/0.25) x 100% = 1.20%

Statistics Canada commonly uses CV results when analyzing data and urges users producing estimates from the CCHS – Healthy Aging data files to do so as well. For details on how to determine CVs, see Section 10. For guidelines on how to interpret CV results, see the table at the end of Section 9.4.

8.2.2 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 devoted to reducing non-sampling errors in the CCHS – Healthy Aging. 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 who were trained extensively with respect to the survey procedures and questionnaire and the observation of interviewers to detect problems. Testing of the CAPI application and field tests were also essential procedures for ensuring that data collection errors were minimized.

A major source of non-sampling errors in surveys is the effect of non-response on survey results. The extent of non-response varies from partial non-response (failure to answer one or a few questions) to total non-response. Partial non-response to the CCHS – Healthy Aging was minimal; once the questionnaire was started, it tended to be completed with very little non-response. Total non-response occurred either because a person refused to participate in the survey or because the interviewer was unable to contact the selected person. Total non-response was handled by adjusting the weight of persons who responded to the survey to compensate for those who did not respond. See Section 7 for details on the weight adjustment for non-response.

Non-response to any particular question (item non-response) is generally low, but may be higher in some modules for various reasons. Item non-response increases as the survey progresses due to respondent fatigue. For example, item non-response is 0.1% at the start of the survey (GEN_Q1), and 0.7% at the end (SDC_Q1). Item non-response is higher by 2.2% in modules which were not to be completed by proxy (for example, in Satisfaction with Life (SLS)). Questions in the Social Support Availability (SSA) module that deals with friends, feelings and hypothetical situations have higher non-response (e.g. 5.4% in SSA_Q2), mainly due to more respondents replying that they did not know. Questions on income are particularly sensitive and are generally known to have high item non-response: 16.8% non-response for household income (IN2DHH) and 21.8% for personal income (IN2DPER), similar to levels seen in the Canadian Community Health Survey Annual Survey. This trend also affects derived variables from other modules that use income variables in the derivation (e.g. Retirement status - objective (RETDRS) with 9.6% non-response). Users are cautioned to account for the larger variations during analysis, for example, by including a dummy variable for income non-response in regression analysis.

9.0 Guidelines for tabulation, analysis and release

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

9.1 Rounding guidelines

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

a) Estimates in the main body of a statistical table are to be rounded to the nearest hundred units using the normal rounding technique. In normal rounding, if the first or only digit to be dropped is from 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is from 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 subtotals and totals in statistical tables are to be derived from their corresponding unrounded components and are then 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 are then 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 from 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is from 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 are then 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.

9.2 Sample weighting guidelines for tabulation

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

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

9.2.1 Definitions: categorical estimates, quantitative estimates

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

Categorical estimates:

Categorical estimates are estimates of the number or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of individuals aged 65 or over who experienced falls in the past twelve months 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 a categorical question:

*In the past 12 months, did YOU have any falls? (**FAL_01**)*
__ Yes
__ No

Quantitative estimates:

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

An example of a quantitative estimate is the average number of falls experienced in the past 12 months by individuals aged 65 or over who fell at least once. The numerator is an estimate of the total number of falls experienced in the past 12 months, and its denominator is an estimate of the number of individuals who experienced at least one fall

in the past 12 months. The module on falls is asked only to persons aged 65 and over and focuses on falls where the persons hurt themselves enough to limit some of their normal activities.

Example of a quantitative question:

*How many times ^HAVE ^YOU1 fallen in the past 12 months? (**FAL_02**)*
|_|_| Times

9.2.2 Tabulation of categorical estimates

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

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

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

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

c) dividing the numerator estimate by the denominator estimate.

9.2.3 Tabulation of quantitative estimates

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

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

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

c) dividing the numerator estimate by the denominator estimate.

For example, to obtain the estimate of the average number of falls experienced in the past 12 months by individuals aged 65 or over who fell at least once, first compute the numerator ($\hat{X}$) by summing the product between the value of variable **FAL_02** and the weight **WTS_M** for those records with a value of "yes" for variable **FAL_01**. The denominator ($\hat{Y}$) is obtained by summing the final weight of those records with a value of "yes" for variable **FAL_01**. Divide ($\hat{X}$) by ($\hat{Y}$) to obtain the average number of falls experienced in the past 12 months by individuals aged 65 or over who fell at least once.

9.3 Guidelines for statistical analysis

The CCHS is based on a complex design involving stratification and multiple stages of selection as well as and unequal probabilities of selection of respondents. Using data from such complex surveys causes problems for 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. As a result, while in many cases the estimates produced by the packages are correct, the variances that are calculated are almost meaningless.

For many analysis techniques (e.g., 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 1, then the results produced by the standard packages will be more reasonable. Note that they will still not take into account the stratification and clustering of the sample's design, but they will take into account the unequal probabilities of selection. 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.

In order to provide a means of assessing the quality of tabulated estimates, Statistics Canada has produced a set of Approximate Sampling Variability 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 10 for more details.

9.4 Release guidelines

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

Table 9.1 Sampling variability guidelines

Type of Estimate	CV (in %)	Guidelines
Acceptable	$0.0 \leq CV \leq 16.5$	Estimates can be considered for general unrestricted release. Requires no special notation.
Marginal	$16.5 < CV \leq 33.3$	Estimates can be considered for general unrestricted release but should be accompanied by a warning cautioning subsequent users of the high sampling variability associated with the estimates. Such estimates should be identified by the letter E (or in some other similar fashion).
Unacceptable	$CV > 33.3$	Statistics Canada recommends not releasing estimates of unacceptable quality. However, if the user chooses to do so, then estimates must be flagged with the letter F (or in some other fashion) and the following warning must 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."

10.0 Approximate sampling variability tables

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

The coefficients of variation (CV) are derived using the variance formula for simple random sampling and by incorporating a factor that reflects the multi-stage, clustered nature of the sample design. This factor, known as the ***design effect,*** is 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 is then used to generate a table that applies to the entire set of characteristics.

The design effects, sample sizes and population counts used to produce the Approximate Sampling Variability Tables as well as the tables themselves are presented in a document, which is included on the PUMF CD. All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Options concerning the computation of exact coefficients of variation are discussed in Section 10.7.

Remember: As indicated in the Sampling Variability Guidelines in Section 9.4, if the number of observations on which an estimate is based is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation. Coefficients of variation based on small sample sizes are too unpredictable to be adequately represented in the tables.

10.1 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, as well as for ratios and differences between such estimates.

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

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

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

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion and the size of the numerator on 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 on a subgroup 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 the CVs in the tables decline in value reading across a row from left to right). For example, the estimated proportion of individuals 65 and over who experienced at least two falls in the past 12 months out of those who fell at all is more reliable than the estimated number who fell at least twice.

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

Rule 3: Estimates of differences between aggregates or percentages

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

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

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

Rule 4: Estimates of ratios

When the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. For example, this would apply to the case where the numerator is the number of individuals aged 65 or over who experienced at least one fall in the past 12 months and the denominator is the number of individuals aged 65 and over.

Consider the case where the numerator is not a subset of the denominator, such as the ratio of the number of individuals aged 65 or over who experienced one fall in the past 12 months to the number who experienced more than one. 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 α1 and α2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$, respectively.

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

Rule 5: Estimates of differences of ratios

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

10.2 Examples of using the CV tables for categorical estimates

The following "real life" examples are included to assist users in applying the foregoing rules[8].

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

Suppose that a user estimates that 865,484 individuals aged 65 or over experienced at least one fall in the past 12 months. How does the user determine the coefficient of variation of this estimate?

1) Refer to the CANADA level CV table.

2) The estimated aggregate (865,484) does not appear in the left-hand column (the "Numerator of Percentage" column), so it is necessary to use the smallest figure closest to it, namely 750,000.

[8] Values used in this section come from the CCHS – Healthy Aging master file. An estimate coming from different microdata files will differ slightly, but the same application of the CV tables still applies.

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, 4.9%.

4) The approximate coefficient of variation of the estimate is thus 4.9%. According to the Sampling Variability Guidelines presented in Section 9.4, the finding that 865,484 individuals aged 65 or over experienced at least one fall in the past 12 months is publishable with no qualifications.

Example 2: Estimates of proportions or percentages possessing a characteristic

Suppose that the user estimates that 865,484 / 4,366,101 = 19.8% of individuals aged 65 or over in Canada experienced at least one fall in the past 12 months. 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 that is based on a subset of the total population (i.e., individuals aged 65 or over), it is necessary to use both the percentage (19.8%) and the numerator portion of the percentage (865,484) in determining the coefficient of variation.

3) The numerator (865,484) does not appear in the left-hand column (the "Numerator of Percentage" column), so it is necessary to use the smallest figure closest to it, namely 750,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the smallest figure closest to it: 15.0%.

4) The figure at the intersection of the row and column used, 4.8%, is the coefficient of variation (expressed as a percentage) to be used.

5) The approximate coefficient of variation of the estimate is thus 4.8%. According to the Sampling Variability Guidelines presented in Section 9.4, the finding that 19.8% of individuals aged 65 or over in Canada experienced at least one fall in the past 12 months can be published with no qualifications.

Example 3: Estimates of differences between aggregates or percentages

Suppose that a user estimates that, among people aged 65 to 74, 409,499 / 2,407,219 = 17.0% experienced at least one fall in the past 12 months (estimate 1), while this percentage is estimated at 132,527 / 490,775 = 27.0% for people aged 85 and over (estimate 2). How does the user determine the coefficient of variation of the difference between these two estimates?

1) The CV for estimate 1 can be obtained using the Canada by age group CV table for 65 to 74 year olds. Since the percentage is based on the total population covered by the table, the CV of the percentage is the same as the CV of the numerator of the proportion, which is 409,499. Since this number doesn't appear in the left-hand column, it is necessary to use the smallest figure closest to it, namely 400,000. This gives a CV of

4.0% for estimate 1. Similarly for estimate 2, a CV of 5.0% is obtained by looking at the CV table for the 85 and over age group.

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 α1 and α2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$, respectively. The standard error of the difference $\hat{d}$ = (0.270 - 0.170) = 0.100 is

$$\sigma_{\hat{d}} = \sqrt{[(0.270)(0.050)]^2 + [(0.170)(0.040)]^2} = 0.015.$$

3) The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.015/0.100 = 0.150.

4) The approximate coefficient of variation of the difference between the estimates is thus 15.0% (expressed as a percentage). According to the Sampling Variability Guidelines presented in Section 9.4, this estimate can be published with no qualifications.

Example 4: Estimates of ratios

Suppose that the user estimates that 547,815 individuals aged 65 or over experienced one fall in the past 12 months, while 314,031 individuals in the same age group experienced more than one. The user is interested in comparing the estimate of individuals who fell once to those who fell more than once 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 aged 65 or over who experienced one fall in the past 12 months. The denominator of the estimate (= $\hat{X}_2$) is the number of individuals aged 65 or over who experienced more than one fall in the past 12 months.

2) Refer to the CANADA level CV table.

3) The numerator of this ratio estimate is 547,815. The smallest figure closest to it is 500,000. 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, 6.2%.

4) The denominator of this ratio estimate is 314,031. The figure closest to it is 300,000. 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, 8.0%.

The approximate coefficient of variation of the ratio estimate is given by rule 4, which is as follows:

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

that is,

$$\alpha_{\hat{R}} = \sqrt{(.062)^2 + (.08)^2}$$

$$= 0.101$$

where α1 and α2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$, respectively. The obtained ratio of people aged 65 or over who fell once to those who fell more than once is 547,815 / 314,031, which is 1.74:1. The coefficient of variation of this estimate is 10.1% (expressed as a percentage), which is releasable with no qualifications according to the Sampling Variability Guidelines presented in Section 9.4.

10.3 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 with 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 around 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 will be less than one standard error, about 95 out of 100 that the difference will be less than two standard errors, and about 99 out of 100 that the difference 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, namely, $\left(\hat{X} - k, \hat{X} + k\right)$, where k is determined depending on 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 the coefficient of variation of the estimate $\hat{X}$ from the appropriate table and then using the following formula to convert it 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 the 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: The release guidelines presented in Section 9.4 that apply to the estimate also apply to the confidence interval. For example, if the estimate is not releasable, then neither will be the confidence interval.

10.4 Example of using the CV tables to obtain confidence limits

A 95% confidence interval for the estimated proportion of individuals aged 65 or over who experienced at least one fall in the past 12 months (from example 2, Section 10.2) would be calculated as follows:

$\hat{X}$ = 0.198

z = 2

$\alpha_{\hat{X}}$ = 0.046 (0.046 is the coefficient of variation of this estimate as determined from the tables)

$CI_{\hat{x}}$ = {0.198 - (2) (0.198) (0.046), 0. 198 + (2) (0.198) (0.046)}

$CI_{\hat{x}}$ = {0.180, 0.216}

10.5 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 they are in fact identical.

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

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

Suppose that we wish to test, at the 5% level of significance, the hypothesis that there is no difference between the proportion of the persons who experienced at least one fall in the past 12 months among people aged 65 to 74 AND those aged 85 or over. From example 3 in Section 10.2, the standard error of the difference between these two estimates is found to be = 0.015. Therefore,

$$z = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.17 - 0.27}{0.015} = \frac{-0.1}{0.015} = -6.7 .$$

Since z = -6.7 is lower than -2, it must be concluded that there is a significant difference between the two estimates at the 5% level of significance. Note that the two subgroups compared are considered to be independent, so the test is correct.

10.7 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 no simple mathematical formula can account for all CCHS sampling frame and weighting aspects. Therefore, other techniques such as resampling methods must be used in order to estimate measures of precision. Among these methods, the bootstrap method is recommended for analysing CCHS data.

The computation of coefficients of variation (or any other measure of precision) using the bootstrap method requires access to information that is considered confidential and thus not available on the PUMF. This computation must be done using the Master file. Access to the Master file is discussed in Section 11.1.

A macro program called Bootvar was developed 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 and differences between ratios, as well as linear and logistic regressions.

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

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

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

Third, for estimates of quantitative variables, separate tables are required to determine their sampling error. Since most of the variables for the CCHS are primarily categorical in nature, this has not been done. Consequently, 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, neither will be the quantitative estimate. For example, the coefficient of variation of the estimate of the total number of falls experienced in the past 12 months by individuals aged 65 or over who fell at least once would be greater than the coefficient of variation of the corresponding estimate of the number of individuals who fell at least once. Thus, 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 about 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 whether it is releasable without a qualifying note. The recalculation is useful because the coefficients of variation produced by the tables are based on a wide range of variables and are thus considered crude, whereas the exact variance program will give an exact coefficient of variation associated with the variable in question.

10.8 Release cut-offs for the CCHS

The Approximate Sampling Variability Tables document provided with the PUMF CD also presents tables giving the minimum cut-offs for estimates of totals at the Canada and provincial levels as well as those for various age groups at the Canada level. Estimates smaller than the value given in the "Marginal" column may not be released under any circumstances.

11.0 Microdata files: description, access and use

The CCHS – Healthy Aging produces three types of microdata files: master files, share files and public use microdata files (PUMF). The PUMF is expected to be released in 2011.

11.1 Master files

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

11.1.1 Research Data Centre

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

11.1.2 Custom tabulations

Another way to access the master files is to offer all users the option of having staff in Client Services of the Health Statistics Division prepare custom tabulations. This service is offered on a cost-recovery basis. It allows users who do not possess knowledge of tabulation software products to get custom results. The results are screened for confidentiality and reliability concerns before release. For more information, please contact Client Services at 613-951-1746 or by e-mail at hd-ds@statcan.gc.ca.

11.1.3 Remote access

Finally, the remote access service to the survey master file is another way to have access to these data if, for some reason, the user cannot access a Research Data Centre. Each purchaser of the microdata product can be supplied with a synthetic or 'dummy' master file and a corresponding record layout. The 'dummy' master file will be available starting in late 2010. With these tools, the researcher can develop his own set of analytical computer programs. The code for the custom tabulations is then sent via e-mail to cchs-escc@statcan.gc.ca. The code will then be transferred into Statistics Canada's internal secured network and processed using the appropriate master file of the CCHS – Healthy Aging data. Estimates generated will be released to the user, subject to meeting the guidelines for analysis and release outlined in Section 9 of this document. Results are screened for confidentiality and reliability concerns and then the output is returned to the client. There is no charge for this service.

11.2 Share files

The share files contain all processed variables but only those records for CCHS – Healthy Aging respondents who agreed to share their data with Statistic Canada's partners, which are the provincial health departments, Health Canada and the Public Health Agency of Canada. Statistics Canada also asks respondents living in Quebec for their permission to share their data with the Institut de la statistique du Québec. The share file is released only to these organizations. Personal identifiers are removed from the share files to respect respondent confidentiality. Users of these files must first certify that they will not disclose, at any time, any information that might identify a survey respondent.

11.3 Public use microdata files

The public use microdata files (PUMF) are developed from the master files using a technique that balances the need to ensure respondent confidentiality with the need to produce the most useful data possible at the provincial or regional level. The PUMF must meet stringent security and confidentiality standards required by the *Statistics Act* before they are released for public access. To ensure that these standards have been achieved, each PUMF goes through a formal review and approval process by an executive committee of Statistics Canada.

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

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

11.4 Use of weight variable

The weight variable **WTS_M** represents the sampling weight for key survey files. For a given respondent, the sampling weight can be interpreted as the number of people the respondent represents in the Canadian population. This weight must always be used when computing statistical estimates in order to make inference at the population level possible. The production of unweighted estimates is not recommended. The sample allocation, as well as the survey design specifics, can cause such results to not correctly represent the population. Refer to Section 7 on weighting for a more detailed explanation on the creation of this weight.

The **Cognition** module was administered in English and French to non-proxy respondents who consented to participate. As the response rate for the cognition module was significantly lower than the overall survey response rate, it was decided to calculate separate cognition weights to accompany the CCHS – Healthy Aging survey. The cognition weight variable **WTS_CM** is found in a separate file. This variable should be used in place of the variable **WTS_M** for any analyses involving the cognition module. The weight variable **WTS_M** must be used for all non-cognition analyses.

11.5 Variable naming convention

The CCHS – Healthy Aging variable naming convention allows data users to easily use and identify the data based on the module and variable type. The CCHS – Healthy Aging variable naming convention fulfils two requirements: to restrict variable names to a maximum of eight characters for ease of use by analytical software products and to identify easily conceptually identical variables between different CCHS surveys. Questions to which changes are made between CCHS surveys, and where the changes alter the concept measured by the question, are entirely renamed to avoid any confusion in the analysis.

The CCHS survey program variable naming convention was changed beginning with the data from the CCHS annual survey 2007 collection period. The letter corresponding to the survey version (e.g., A = cycle 1.1, C = cycle 2.1 and E = 3.1) is no longer used in the variable names. A new variable (REFPER, format = YYYYMM-YYYYMM) was added to the microdata files in order to identify the beginning and the end of the reference during which data included in the file were collected.

The variable names are structured as follows:

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

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

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

Example 2 shows the structure of the variable name for question 6 of the Health Care Utilization 2 Module (HC2_06), which is a multi-response question:

Positions 1 to 3:	HC2	Health care utilization module
Position 4:	_	(underscore = collected data)
Position 5 to 8:	06A	Corresponding question number and answer option

Positions 1 to 3 contain the acronyms for each of the modules. These acronyms appear beside the module names given in the table in Section 3.3.

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

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

Table 11.1 Designation of codes used in the 4th position of the CCHS – Healthy Aging variable names

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

For questions that have more than one response option, the final position in the variable naming sequence is represented by a letter. For this type of question, new variables were created to differentiate between a "yes" or "no" answer for each response option. For example, if Question 2 had 4 response options, the new questions would be named _02A for option 1, _02B for option 2, _02C for option 3, etc. If only options 2 and 3 were selected, then _02A = No, _02B = Yes, _02C = Yes and _02D = No.

11.6 Variable naming convention for previous CCHS Surveys

As mentioned earlier, the variable naming convention for CCHS surveys was changed in 2007. The flag for the cycle in which the variables were collected was removed. This flag was found in the 4th position for cycles 1.1 to 3.1.

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

Letter	Cycle and cycle name
A	Cycle 1.1: Canadian Community Health Survey

B Cycle 1.2: Canadian Community Health Survey - Mental Health and Well-Being

C Cycle 2.1: Canadian Community Health Survey

D Cycle 2.2: Canadian Community Health Survey - Nutrition

E Cycle 3.1: Canadian Community Health Survey

11.7 Weighting files

To encourage appropriate use of the CCHS – Healthy Aging data, the cognition weights described in Section 11.5 are provided in a separate file. They have their own sampling weights and a corresponding set of bootstrap weights that should be used to produce valid estimates for all variables on the file. The cognition sampling weights are calibrated so that the sum represents the total Canadian population aged 45 and over. Table 11.2 describes the two data files.

Table 11.2 Names and content of CCHS – Healthy Aging data files

Files	File name	Sampling weight	Bootstrap weights file	Variables included	Records included
Main master file	HS.txt	WTS_M	b5.txt	All modules except COG	All respondent records
Share file	HS.txt	WTS_S	b5.txt	All modules except COG and CLS	Records of all respondents who agreed to share their data
Cognition master file	HSC.txt	WTS_CM	b5_c.txt	All modules	Records of all respondents who participated in the cognition module
Cognition share file	HSC.txt	WTS_CS	b5_c.txt	All modules except CLS	Records of all respondents for who participated in the cognition module and who agreed to share their data

11.8 Data dictionary

A data dictionary report, including universe statements and frequencies, is provided for the main master file.

11.9 Differences in calculation of variables using different files

Variables can be estimated using the master file, the share file, or the PUMF. Depending on which of these files is used, very small differences will be observed.

All official Statistics Canada estimates of variables excluding the cognition variables are based on the main master file sampling weight. All official Statistics Canada estimates involving the cognition variables are based on the cognition master file sampling weight.