



Statistics Statistique  
Canada Canada

# **Survey on Living with Chronic Diseases in Canada**

## **USER GUIDE**

**December 2009**



## TABLE OF CONTENTS

<b>1.</b>	<b>INTRODUCTION.....</b>	<b>1</b>
<b>2.</b>	<b>BACKGROUND .....</b>	<b>2</b>
<b>3.</b>	<b>OBJECTIVE .....</b>	<b>3</b>
<b>4.</b>	<b>SURVEY CONTENT .....</b>	<b>4</b>
4.1	QUALITATIVE TESTING .....	4
4.2	FINAL QUESTIONNAIRE CONTENT .....	5
<b>5.</b>	<b>SAMPLE DESIGN.....</b>	<b>10</b>
5.1	TARGET POPULATION .....	10
5.2	DOMAINS OF INTEREST .....	10
5.3	SAMPLING FRAME.....	10
5.4	SAMPLE SIZE AND ALLOCATION.....	10
<b>6.</b>	<b>DATA COLLECTION .....</b>	<b>13</b>
6.1	COMPUTER-ASSISTED INTERVIEWING .....	13
6.2	SLCDC APPLICATION DEVELOPMENT .....	13
6.3	INTERVIEWER TRAINING .....	14
6.4	THE INTERVIEW .....	15
6.5	FIELD OPERATIONS .....	16
6.6	QUALITY CONTROL AND COLLECTION MANAGEMENT.....	16
<b>7.</b>	<b>DATA PROCESSING .....</b>	<b>18</b>
7.1	EDITING.....	18
7.2	CODING .....	18
7.3	CREATION OF DERIVED AND GROUPED VARIABLES .....	18
7.4	WEIGHTING .....	18
<b>8.</b>	<b>WEIGHTING.....</b>	<b>20</b>
8.1	SAMPLE WEIGHTING .....	20
8.1.1	PROXY-LINK-PHONE ADJUSTMENT .....	20
8.1.2	SLCDC SELECTION CRITERIA ADJUSTMENT .....	21
8.1.3	OUT-OF-SCOPE IN SLCDC ADJUSTMENT .....	22
8.1.4	NON-RESPONSE IN SLCDC ADJUSTMENT .....	22
8.1.5	SHARE-LINK (FINAL) ADJUSTMENT .....	23
8.2	BOOTSTRAP WEIGHTS .....	23
<b>9.</b>	<b>DATA QUALITY .....</b>	<b>24</b>
9.1	OUT-OF-SCOPE CASES .....	24
9.2	RESPONSE RATES.....	24
9.3	DATA INTERPRETATION .....	26
9.4	SURVEY ERRORS.....	26
9.4.1	THE FRAME.....	27
9.4.2	NON-RESPONSE .....	27
9.4.3	MEASUREMENT OF SAMPLING ERROR .....	28
<b>10.</b>	<b>GUIDELINES FOR TABULATION, ANALYSIS AND RELEASE .....</b>	<b>29</b>
10.1	ROUNDING GUIDELINES .....	29
10.2	SAMPLE WEIGHTING GUIDELINES FOR TABULATION .....	30

10.2.1	DEFINITIONS: CATEGORICAL ESTIMATES, QUANTITATIVE ESTIMATES .....	30
10.2.2	TABULATION OF CATEGORICAL ESTIMATES .....	31
10.2.3	TABULATION OF QUANTITATIVE ESTIMATES .....	31
10.3	GUIDELINES FOR STATISTICAL ANALYSIS.....	32
10.4	RELEASE GUIDELINES .....	32
10.5	VARIANCES AND COEFFICIENTS OF VARIATION .....	33
<b>11.</b>	<b>FILE USAGE.....</b>	<b>35</b>
11.1	DATA FILE .....	35
11.1.1	SHARE PARTNERS.....	35
11.1.2	RESEARCH DATA CENTRES .....	35
11.1.3	CUSTOM TABULATIONS .....	36
11.2	USE OF WEIGHT VARIABLE.....	36
11.3	VARIABLE NAMING CONVENTION .....	36
11.3.1	VARIABLE NAME COMPONENT STRUCTURE IN SLCDC .....	36
11.3.2	POSITIONS 1-3: VARIABLE / QUESTIONNAIRE SECTION NAME .....	37
11.3.3	POSITION 4: CYCLE AND SURVEY NAME.....	37
11.3.4	POSITION 5: VARIABLE TYPE.....	37
11.3.5	POSITIONS 6-8: VARIABLE NAME .....	38
11.4	VARIABLE NAME COMPONENT STRUCTURE IN CCHS .....	38

## **1. Introduction**

The Survey on Living with Chronic Diseases in Canada (SLCDC) is a cross-sectional survey sponsored by the Public Health Agency of Canada (PHAC) that collects information related to the experiences of Canadians with chronic health conditions. The SLCDC will take place every two years, with two chronic diseases covered in each survey cycle. The 2009 survey focused on arthritis and hypertension.

There are two questionnaires for this survey—one questionnaire for arthritis and one questionnaire for hypertension. The SLCDC asked respondents about a number of issues related to chronic health conditions, including diagnosis of a chronic health condition, care received from health professionals, medication use and self-management of their condition.

The target population for this survey is Canadians aged 20 years or older who have been diagnosed with arthritis or hypertension living in private dwellings in the ten provinces. Residents of the three territories, persons living on Indian Reserves, residents of institutions, and full-time members of the Canadian Armed Forces are excluded from this survey. Data collection for the SLCDC took place in February and March 2009.

The purpose of this document is to facilitate the manipulation of the SLCDC data file and to describe the methodology used.

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

For custom tabulations or general data support:  
Client Custom Services, Health Statistics Division:  
E-mail:

613-951-1746  
[hd-ds@statcan.gc.ca](mailto:hd-ds@statcan.gc.ca)

## **2. Background**

The central objective of the Survey on Living with Chronic Diseases in Canada (SLCDC) is to gather information related to the experiences of persons living with chronic diseases, including diagnosis of a chronic health condition, care received from health professionals, medication use and self-management of their condition. The survey was sponsored by the Public Health Agency of Canada. The SLCDC is a cross-sectional survey and was a follow-up to the 2008 Canadian Community Health Survey.

### **3. Objective**

The purpose of the SLCDC is to provide information on the impact of chronic disease on individuals, as well as how people with chronic disease manage their health condition. More specifically, the survey had the following objectives:

- To assess the impact of chronic health conditions on quality of life
- To provide more information on how people manage their chronic health conditions
- To identify health behaviors which influence disease outcomes
- To identify barriers to self-management of chronic health conditions

## **4. 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 SLCDC.

The SLCDC content was developed based on an ongoing consultation process between the Health Statistics Division at Statistics Canada and the Public Health Agency of Canada (PHAC), with significant input from members of expert advisory groups in the areas of arthritis and hypertension. Content selection was based on objectives and data requirements specified by PHAC. Members of the PHAC project team were consulted on a regular basis throughout development and testing of the SLCDC questionnaires. The end result of the consultation process was two SLCDC questionnaires: (1) an arthritis-specific questionnaire and (2) a hypertension-specific questionnaire.

A summary describing each of the modules on the arthritis and hypertension questionnaires is provided in Section 4.2.

### **4.1 Qualitative testing**

As previously stated, the 2009 SLCDC consisted of two different questionnaires; an arthritis questionnaire and a hypertension questionnaire. The questionnaires were developed by Statistics Canada, in collaboration with PHAC. Hypertension and arthritis expert groups were also consulted during content development. The questionnaires were translated by the Official Languages and Translation Division of Statistics Canada. Both questionnaires (in English and French) were tested by Statistics Canada's Questionnaire Design and Review Centre (QDRC) using one-on-one interviews.

Qualitative testing was conducted to assess the content and flow of the SLCDC questionnaires. The questionnaires were administered face-to-face with respondents. The 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.

English testing was conducted in May 2008 in Ottawa, Ontario, followed by French testing in June 2008 in Montreal, Quebec. The frame used to select respondents for the interviews was the 2007 and 2008 CCHS. A total of 32 participants took part in the testing, representing a cross-section of persons who reported having either arthritis or hypertension diagnosed by a doctor or other health professional, or who had reported experiencing symptoms of arthritis (joint pain or stiffness) in their CCHS interview. All qualitative interviews were conducted by trained interviewers from QDRC and observed by members of the SLCDC project team, including personnel from STC's Health Statistics Division and PHAC. Some of the key findings from the qualitative testing are discussed below.



### Key findings from the arthritis questionnaire:

In general, the questions seemed to work well for participants with either arthritis or joint symptoms. However, for some participants, especially those with very minor symptoms, some of the questions did not seem applicable because symptoms were not severe enough to cause any problems. It was suggested that a lot of the questions were geared to people with more severe symptoms. A few participants felt that their condition was not serious enough to be included in the target population of the survey, however, these individuals were still able to answer most questions and said the survey was probably relevant.

Following qualitative testing, a decision was made to include only individuals who had received a diagnosis of arthritis from a health professional in the SLCDC target population. Individuals with joint symptoms but no diagnosis of arthritis would be excluded from the target population.

### Key findings from the hypertension questionnaire:

Overall, participants reported the hypertension questionnaire to be straightforward and easy to answer. In general, respondents did not find the questions to be overly sensitive, with the exception of some of the questions related to self-management (for example, losing weight, cutting down on alcohol intake, etc.). Some respondents also had difficulty recalling their diastolic and systolic blood pressure measurements the last time their blood pressure was measured by a health professional.

Following qualitative testing, questions were modified to reduce the impact of sensitive topics, for example, weight loss, alcohol use, and smoking, and to address instances where certain questions were not applicable for some respondents. To assist respondents in reporting their systolic and diastolic measurements, ranges were included as a response category option, in addition to allowing respondents to report their blood pressure measurement in exact values.

## **4.2 Final questionnaire content**

This section outlines the modules comprising the content of the SLCDC arthritis and hypertension questionnaires. The arthritis questionnaire was made up of 13 modules, and the hypertension questionnaire of 10 modules.

<b>Arthritis</b>	
<b>Module</b>	<b>Description</b>
<b>GENX</b>	<b>General health:</b> This module is the same as the module used in the CCHS. The intent was to make respondents feel comfortable before asking them specific questions about their arthritis.
<b>DHAX</b>	<b>Diagnosis and family history:</b> This module included questions that confirmed that the respondent had received a diagnosis of arthritis from a health professional, their age at the time of diagnosis, their age at the onset of symptoms, the type of arthritis they have been diagnosed with and whether they have any immediate family members who have ever been diagnosed with arthritis. The screening question is modified from the <i>Chronic conditions</i> module in the CCHS. The remaining questions are new.
<b>SSAX</b>	<b>Symptoms and severity:</b> Questions in this module asked about the frequency and severity of the main symptoms of arthritis—joint pain and fatigue. Respondents were also asked about the joints that cause them pain. Many of the questions in this module are based on questions in the Arthritis Conditions Health Effects Survey (ACHES, 2005), which was conducted by the Centers for Disease Control and Prevention in the United States.
<b>RAAX</b>	<b>Restriction of activities:</b> This module asked a series of questions about whether the respondent has been limited in doing a variety of activities in the past month because of their arthritis. This module is based on questions in ACHES but has been modified to use language similar to the SF-36 module in the CCHS.
<b>ADAX</b>	<b>Use of assistive devices:</b> This module asked a series of questions about whether the respondent currently uses any assistive devices to help with their usual activities, and for each device used, whether the device is used to help with problems caused by arthritis. The module was developed in cooperation with PHAC and uses many of the terms used by the Arthritis Society of Canada.
<b>RWAX</b>	<b>Restriction of work-related activities:</b> This module asked respondents whether they have had to modify their paid work or volunteering work because of their arthritis. Conceptually these questions are comparable to ACHES. The definition of volunteer work was taken from the Canada Survey on Giving, Volunteering and Participating (STC, 2007).
<b>MEAX</b>	<b>Medication use:</b> This is a short module which asked respondents whether they had taken any prescription, over-the-counter or natural health products for their arthritis in the past month. These questions are not comparable to the CCHS or ACHES questions as they are more general in nature.
<b>HUAX</b>	<b>Health care utilization:</b> The module asked respondents whether they have seen a variety of health professionals regarding their arthritis in the past 12 months. These questions are based on the format used by the CCHS, but have been modified to ask about arthritis-specific practitioners (i.e. orthopaedic surgeon).
<b>CLAX</b>	<b>Clinical recommendations:</b> This module asked respondents whether a health

	practitioner ever recommended they do certain activities to help manage their arthritis (i.e. lose weight, exercise). These questions are based on ACHES questions, which are based on best practice guidelines.
<b>SMAX</b>	<b>Self-management:</b> The module is a follow-up to the <i>Clinical recommendations</i> module and asks respondents whether they followed any of the recommendations given by their health professional, and, if not, what the barriers to doing so were. This module is somewhat based on the <i>Stages of Change</i> module in the CCHS. The barrier response categories are from the CCHS.
<b>SWAX</b>	<b>Support and well-being:</b> This module asked respondents about the availability of support from family and friends, and about emotional well-being related to managing their arthritis. The questions are a shortened form of content used by ACHES and uses similar wording to the CCHS.
<b>INAX</b>	<b>Information received:</b> This module asked respondents whether they have received information on a number of topics to help manage their arthritis. Questions in the module are new, and were developed for the SLCDC to address data priorities identified by PHAC to aid in the planning and development of programs for persons with arthritis.
<b>ADMX</b>	<b>Administration:</b> This module collected administrative information including permission to link information from the SLCDC with the respondent's previous CCHS information and to share the linked survey data with the SLCDC share partners: PHAC, Health Canada, provincial ministries of health, and the Institut de la Statistique du Québec (respondents living in the province of Quebec only). This module was asked of all respondents.

Hypertension	
Module	Content
<b>GENX</b>	<b>General health:</b> This module is the same as the module used in the CCHS. The intent was to make respondents feel comfortable before asking them specific questions about their hypertension.
<b>CNHX</b>	<b>Confirmation of high blood pressure diagnosis:</b> This module included questions to confirm that the respondent had received a diagnosis of high blood pressure from a health professional, and their age at the time of diagnosis. Additionally, the module included questions from the CCHS that were used to screen out women whose only diagnosis of high blood pressure occurred during pregnancy. Questions were taken from the CCHS <i>Chronic conditions</i> module.
<b>BMHX</b>	<b>Blood pressure measurement:</b> This module asked about blood pressure measurement by a health professional. Questions assessed whether respondents were told details of their blood pressure reading the last time they had it measured by a health professional, their systolic and diastolic measurements, and whether a health professional had discussed a target rate for their blood pressure. These are new questions based on data requirements identified by PHAC in consultation with their expert working group on hypertension.
<b>MEHX</b>	<b>Medication use:</b> This module asked respondents whether they were currently taking any prescription medications, including medications for their high blood pressure. Respondents were then asked about the number of medications they take and the frequency that they take their medications. Respondents were also asked about their compliance with their medication plan and reasons for not taking their hypertension medication as prescribed. These questions are not comparable to the <i>Medication use</i> questions on the CCHS.
<b>HUHX</b>	<b>Health care utilization:</b> The module asked respondents to identify the health professional most responsible for treating their high blood pressure, and whether they have seen or talked to various health professionals regarding their high blood pressure in the past 12 months. These questions are based on the <i>Health care utilization</i> module on the CCHS, but have been modified for specific practitioners most likely to be involved in the treatment of hypertension.
<b>CLHX</b>	<b>Clinical recommendations:</b> This module asked respondents whether a health professional ever suggested they do certain activities to help control their blood pressure (i.e. diet modification, weight loss, exercise). These questions are based on the format of the Arthritis Conditions Health Effects Survey (ACHES, 2005), which was conducted by the Centers for Disease Control and Prevention in the United States.
<b>SMHX</b>	<b>Self-management:</b> The <i>Self-management</i> module follows the <i>Clinical recommendations</i> module and asks respondents about the degree to which they have made changes concerning various diet and lifestyle factors as a

	<p>result of being diagnosed with high blood pressure. The module covers a range of activities identified as modifiable risk factors for hypertension (i.e. diet modification, weight loss, exercise). Follow-up questions measuring barriers to change are asked of respondents who have never adopted diet and/or lifestyle changes, or who are not currently doing so. This module is based on the <i>Stages of change</i> module in the CCHS. The barrier response categories are from the CCHS.</p>
<b>MOHX</b>	<p><b>Self-monitoring of blood pressure:</b> This module asked questions about blood pressure monitoring respondents do themselves outside of a health professional's office. Topics include frequency of blood pressure self-monitoring, whether a respondent has received training on how to use a blood pressure measurement device and how respondents use information gained from blood pressure self-monitoring. Module content is based on a set of preliminary questions developed by PHAC in consultation with their expert working group on hypertension.</p>
<b>INHX</b>	<p><b>Information and training:</b> This module includes questions modified from the 1996 National Population Health Survey (NPHS) Asthma Supplement, with several additional questions developed for this survey. This module was developed for the SLCDC to address data priorities identified by PHAC to aid in the planning and development of programs for persons with hypertension.</p>
<b>ADMX</b>	<p><b>Administration:</b> This module collected administrative information including permission to link information from the SLCDC with the respondent's previous CCHS information and to share the linked survey data with the SLCDC share partners: PHAC, Health Canada, provincial ministries of health, and the Institut de la Statistique du Québec (respondents living in the province of Quebec only). This module was asked of all respondents.</p>

## **5. Sample Design**

### **5.1 Target population**

The 2009 Survey on Living with Chronic Disease in Canada (SLCDC) targets Canadians who have arthritis or hypertension diagnosed by a health professional, aged 20 years or older, living in private dwellings in the ten provinces. Residents of the three territories, persons living on Indian Reserves or Crown lands, in institutions, full-time members of the Canadian Forces and residents of certain regions are excluded from this survey. These exclusions represent 2% of the overall Canadian population.

### **5.2 Domains of interest**

The SLCDC is designed to produce reliable estimates at the national level by age group and sex. The targeted age groups are 20 to 44 years old, 45 to 64 years old, 65 to 74 years old and 75 years old and older.

### **5.3 Sampling frame**

The SLCDC used the 2008 Canadian Community Health Survey (CCHS) to select its sample. The SLCDC is built under a two-phase design where the first phase is the CCHS sample and the second phase, the SLCDC sample.

The CCHS sample (first-phase) is selected from multiple frames. The first frame is an area frame designed for the Canadian Labour Force Survey (LFS). The second is a list frame of telephone numbers. About half of the CCHS sample is selected from the area frame and the other half is selected from the list frame. For more detailed information on the CCHS sampling process, refer to the 2008 CCHS User Guide.

The SLCDC sample was selected from respondents to the 2008 CCHS who were aged 20 years or older as of December 2008 and who reported having been diagnosed by a health professional with arthritis or hypertension. Note that the CCHS respondents who reported only being diagnosed with high blood pressure during pregnancy were excluded from the hypertension sample.

### **5.4 Sample size and allocation**

In order to produce reliable estimates at the national level by age group and sex, a stratification of CCHS respondents by age group and sex was performed based on the age groups of interest: 20 to 44 years old (20-44), 45 to 64 years old (45-64), 65 to 74 years old (65-74) and 75 years and older (75+) for each of the chronic conditions. Therefore, the CCHS respondents were classified into eight categories—four age groups and two sex groups (female and male).

An overall sample of 7,062 persons for arthritis and 9,055 for hypertension were selected from 13,549 and 17,437 CCHS respondents respectively. The sample was inflated by 1.4 to take into account the out-of-scope rate (for example, respondents who were deceased or moved outside Canada) estimated at 10% and the non-response rate estimated at 20%. The distribution of the

first phase sample and the SLCDC sample by stratum is given in Tables 5.1 and 5.2 for each chronic disease.

**Table 5.1: SLCDC sample size by stratum for arthritis**

Stratum (Sex, Age group)		2008 CCHS respondents with arthritis (first phase)	2009 SLCDC arthritis sample size (second phase) raw sample	2009 SLCDC arthritis sample size (second phase) expected respondents
Female	20-44 years old	687	503	352
Female	45-64 years old	3,269	1,324	927
Female	65-74 years old	2,163	1,224	857
Female	75+ years old	2,682	1,311	918
Total Female		8,801	4,362	3,054
Male	20-44 years old	465	343	240
Male	45-64 years old	1,914	1,167	817
Male	65-74 years old	1,173	616	431
Male	75+ years old	1,196	574	402
Total Male		4,748	2,700	1,890
Total		13,549	7,062	4,944

**Table 5.2: SLCDC sample size by stratum for hypertension**

Stratum (Sex, Age group)		2008 CCHS respondents with hypertension (first phase)	2009 SLCDC hypertension sample size (second phase) raw sample	2009 SLCDC hypertension sample size (second phase) expected respondents
Female	20-44 years old	1,050	844	591
Female	45-64 years old	3,490	1,324	927
Female	65-74 years old	2,507	1,324	927
Female	75+ years old	3,042	1,324	927
Total Female		10,089	4,816	3,371
Male	20-44 years old	761	619	433
Male	45-64 years old	3,114	1,324	927
Male	65-74 years old	1,856	1,289	902
Male	75+ years old	1,617	1,007	705
Total Male		7,348	4,239	2,967
Total		17,437	9,055	6,338

To reduce response burden, it was decided that all sampled respondents could receive only one questionnaire, either the hypertension questionnaire or the arthritis questionnaire. For people

having the two conditions, they were split in two groups: one group assigned to the hypertension questionnaire and the other group to the arthritis questionnaire. The sample allocation by questionnaire was done proportionally to the size of the number of 2008 CCHS respondents for each condition.

The sample for each chronic condition was selected using systematic sampling after the units were sorted by province, CCHS collection period and age. Exclusions were performed before the sample selection. Units were excluded if they did not have sufficient information to conduct the survey, such as missing telephone numbers on the CCHS, or to perform analysis, such as not agreeing to share or link the CCHS data. These exclusions represented 18% of the first phase sample and were taken into account at the estimation stage since they are part of the population of interest.



## **6. Data collection**

Collection for the SLCDC took place in February and March 2009. Over the collection period, a total of 10,707 valid interviews were conducted using computer assisted telephone interviewing (CATI).

### **6.1 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. CAI also provides an automated call scheduler, i.e. a central system to optimize the timing of call-backs and the scheduling of appointments used to support CATI collection.

The case management system routes the questionnaire applications and sample files from Statistics Canada's main office to regional collection offices (in the case of CATI). 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.
- Question text, including reference periods and pronouns, is customised automatically based on factors such as the age and sex of the respondent, the date of the interview and answers to previous questions.

### **6.2 SLCDC application development**

For the SLCDC, a CATI application was utilized. The application consisted of entry, survey content, and exit components.

Entry and exit components contain standard sets of questions designed to guide the interviewer through contact initiation, respondent confirmation, tracing (if necessary) and determination of case status. The survey content component consisted of the SLCDC arthritis and hypertension questionnaire modules, which made up the bulk of the application. Testing and development of the CATI application began in September 2008. This consisted of three stages of internal 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 the 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 sample file to the entry and exit and survey content components of the application. It also ensures that variables affecting skip patterns and logic flows are correctly passed between modules within the survey content component. 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, confirming that the correct respondent has been found, determining whether a case is in scope and creating appointments.

End-to-end testing occurs when the fully integrated application is placed in a simulated collection environment. The application is loaded onto computers that are connected to a test server. Data are then collected, transmitted and extracted in real time, exactly as 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 SLCDC application. It also provides a final chance of finding errors within the entry, survey content and exit components.

### **6.3 Interviewer training**

In late January and early February 2009, representatives from Statistics Canada’s Collection Planning and Management Division visited four of the five regional offices participating in the collection of the SLCDC data (Halifax, Sherbrooke, Toronto and Edmonton). The purpose of the visits was to train the regional office project managers and teams of interviewers for the SLCDC arthritis and hypertension surveys. Members of the SLCDC project team from Health Statistics Division also attended the training sessions to present information about the background and development of the SLCDC, and to offer additional support and clarify any questions or concerns that may have arisen. The project manager from the Winnipeg regional office was trained in Edmonton and then trained the Winnipeg interviewers.

The focus of these sessions was to make interviewers comfortable using the SLCDC application and familiarise interviewers with survey content. The training sessions focused on:

- goals and objectives of the survey
- survey methodology
- application functionality
- review of the questionnaire content and exercises
- use of mock interviews to simulate difficult situations and practise potential non-response

- situations
- survey management

One of the key aspects of the training was a focus on minimizing non-response. Exercises to minimise non-response were prepared for interviewers. The purpose of these exercises was to have the interviewers practice convincing reluctant respondents to participate in the survey.

## **6.4 The interview**

Sample units selected from the frame were interviewed from centralised call centres using the CATI application. The CATI interviewers were supervised by a senior interviewer located in the same call centre.

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

### **Introductory letters**

Before the start of the collection period, introductory letters explaining the purpose of the survey were sent to the targeted respondents. Mailing address information was not available for all respondents from the 2008 CCHS. For cases where mailing addresses were not available, an introductory letter was not sent out. The introductory letters explained the importance of the survey and provided examples of how the SLCDC data would be used.

### **Initiating contact**

Interviewers were instructed to make all reasonable attempts to obtain interviews. When the timing of the interviewer's call was inconvenient, an appointment was made to call back at a more convenient time. Numerous call-backs were made at different times on different days.

When a respondent was no longer available at the phone number provided on the 2008 CCHS, tracing of the respondent was initiated. In order to trace respondents, alternate contacts provided by the respondent on the 2008 CCHS were contacted to obtain the respondent's new telephone number.

### **Refusal conversion**

For individuals who at first refused to participate in the survey, a letter was sent from the regional office to the respondent, stressing the importance of the survey and the targeted respondent's participation. This was followed by a second call from a senior interviewer, a project supervisor or another interviewer to try to convince the respondent of the importance of participating in the survey.

## **Language barriers**

To remove language as a barrier to conducting interviews, the regional offices recruit 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.

## **Proxy interviews**

Proxy interviews were not permitted for the SLCDC.

## **6.5 Field operations**

The SLCDC consisted of one two-month collection period. However, the SLCDC sample was divided into two, with half of the sample starting collection in February and the other half starting collection in March. The regional collection offices were instructed to use the first two weeks of each month to resolve the majority of the sample, with the next two weeks being used to finalize 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 month.

Transmission of cases from the regional offices to head office was the responsibility of the regional office project supervisor, senior interviewer and the technical support team. These transmissions were performed nightly and all completed cases were sent to Statistics Canada's head office.

## **6.6 Quality control and collection management**

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

CATI interviewers were randomly chosen for validation. Validation during CATI collection consisted of senior interviewers monitoring interviews to ensure proper techniques and procedures (reading the questions as worded in the application, not prompting respondents for answers, etc.) were followed by the interviewers. In addition, members of the survey team from head office visited a number of regional offices to observe collection at various times during the collection period.

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 daily showing response rates, refusal rates and out-of-scope rates. The link and share rates were calculated weekly. Customised reports were also created and used to examine specific data quality issues that arose during collection.

One issue that arose during data collection was a higher than expected out-of-scope rate. It was found that many respondents who indicated that they had hypertension or arthritis on the 2008 CCHS said that this was not the case on the SLCDC. Because of this issue, specific instructions

were sent out to the regional offices explaining procedures that should be followed when interviewers encountered an out-of-scope case. In addition, out-of-scope cases from the beginning of the collection period (which did not follow the new out-of-scope procedures) were sent back to the field for follow-up by senior interviewers. Upon follow-up of the out-of-scope cases, these cases were then classified in one of three ways:

- Out-of-scope: these cases included respondents who reported having arthritis or hypertension on the 2008 CCHS, but when called for the SLCDC said that they did not have the condition. Part of this could be due to the wording of the questions on the CCHS and the SLCDC as the SLCDC placed more emphasis on the condition being diagnosed by a health professional. Many respondents reported that they never had the condition diagnosed by a health professional. Rather, they had symptoms and self-diagnosed their condition. Other respondents thought that they had arthritis but actually had a condition which wasn't arthritis, such as osteoporosis. These respondents were true out-of-scope cases and were not included in the survey.
- In-scope: These cases included respondents who said that they did not have the condition. However, upon follow-up, it was determined that these respondents had been diagnosed with arthritis or hypertension by a health professional but their symptoms were controlled either by medication or life-style modifications. These respondents should not be out-of-scope and attempts were made to interview them.
- Refusals: Some respondents did not want to participate in the survey and said that they did not have the condition in order to be screened out. Attempts were made to determine whether an out-of-scope case was actually a refusal.

The impact of the out-of-scope cases on weighting and data quality will be discussed in Chapters 8 and 9, respectively.

## **7. Data processing**

### **7.1 Editing**

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

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

### **7.2 Coding**

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

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

### **7.3 Creation of derived and grouped variables**

To facilitate data analysis and to minimise the risk of error, a number of variables on the file have been derived using items found on the SLCDC questionnaire. Derived variables generally have a "D" or "G" in the fifth character of the variable name. In some cases, the derived variables are straightforward, involving collapsing of response categories. In other cases, several variables have been combined to create a new variable. 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 11.3.

### **7.4 Weighting**

The principle behind estimation in a probability sample such as the SLCDC 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.

The weighting phase is a step which calculates, for each record, what this number is. This weight appears on the microdata file, and **must** be used to derive meaningful estimates from the survey.

For example, if the number of individuals who have ever experienced joint symptoms of pain, aching, or stiffness related to their arthritis is to be estimated, this would be done by selecting the records referring to those individuals in the sample with that characteristic and summing the weights entered on those records.

Details of the method used to calculate these weights are presented in Chapter 8.

## 8. Weighting

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

As described in Chapter 5, the SLCDC survey frame is based on respondents to the 2008 CCHS. The starting point for the SLCDC weighting process is therefore the 2008 CCHS share weight. For more information on this weight see the 2008 CCHS User Guide.

### 8.1 Sample weighting

Table 8.1 presents an overview of the different adjustments that were part of the weighting strategy for the SLCDC, in the order in which they were applied.

**Table 8.1: Weighting steps for SLCDC**

CD 1 – Proxy-Link-Phone Adjustment
CD 2 – Selection Criteria Adjustment
CD 3 – Out-of-Scope in SLCDC Adjustment
CD 4 – Non-response in SLCDC Adjustment
CD 5 – Share-Link (Final) Adjustment

#### 8.1.1 Proxy-link-phone adjustment

The first step of the weighting for the SLCDC was to drop those CCHS respondents aged 12-19 and those in the Territories since they were not part of the target population. The next step was to adjust for the fact that some people were excluded from the SLCDC even if they were part of the target population mainly for practical reasons. The list of excluded people and the reasons for their exclusion are as follows:

- People with invalid phone numbers were excluded since it is impossible to contact them given that the SLCDC was a telephone survey.
- People who did not agree to share or link their 2008 CCHS information were excluded. This is because the intent of the SLCDC was to link the SLCDC survey responses with the 2008 CCHS. This linked survey file will then be provided to the survey share partners such as the survey sponsor, the Public Health Agency of Canada.
- People for whom their 2008 CCHS interview was done by proxy were excluded since the SLCDC questionnaire could not be answered by proxy.

Since these excluded people were in the population of interest, adjustments were made to allocate their weights to the remaining CCHS respondents. The adjustment process starts with the share weights from the 2008 CCHS. The probability of each sharer in the CCHS to be part of the



sampling population was predicted using a logistic regression and auxiliary variables from the CCHS. From the predicted probabilities, response homogeneous groups (RHG) were created.

The adjustment was calculated within each RHG as follow:

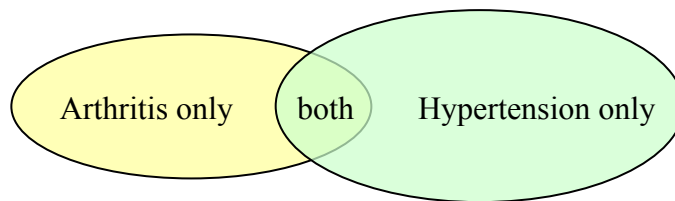
$$\text{adjCD1} = \frac{\text{Sum of CCHS share weights for all sharers}}{\text{Sum of share weights for all non - proxy, with valid phone numbers that agree to link CCHS data}}$$

The weight  $\text{wgtCD1}$  was calculated as  $\text{wgts3} * \text{adjCD1}$ , where  $\text{wgts3}$  is the final 2008 CCHS share weight. After the adjustment was calculated, the excluded units were dropped from the file.

### 8.1.2 SLCDC selection criteria adjustment

In the SLCDC sampling design, CCHS respondents were stratified by age groups and sex (see Chapter 5). In each stratum (age group by sex) a unit could have only one condition (arthritis or hypertension) or both conditions (see Graphic 8.1 below). For response burden purposes, every respondent sampled could receive only one questionnaire, either the hypertension questionnaire or the arthritis questionnaire. That led to a different adjustment for units with only one condition versus for units with both conditions.

**Graphic 8.1: Overlap of the two conditions**



The adjustment for units with only one condition was calculated within each stratum as follow:

$$\text{adjCD2a} = \frac{\text{Sum of weights of the units with only arthritis}}{\text{Sum of weights of the selected units with only arthritis}}$$

$$\text{adjCD2h} = \frac{\text{Sum of weights of the units with only hypertension}}{\text{Sum of weights of the selected units with only hypertension}}$$

The adjustment for units with both conditions was calculated within each stratum as follow:

$$\text{adjCD2a} = \frac{\text{Sum of weights of the units with both conditions (arthritis and hypertension)}}{\text{Sum of weights of the selected units with both condition whom received arthritis questionnaire}}$$

$$\text{adjCD2h} = \frac{\text{Sum of weights of the units with both conditions (arthritis and hypertension)}}{\text{Sum of weights of the selected units with both conditions whom received hypertension questionnaire}}$$

The weight for those who received the arthritis questionnaire was calculated as  $\text{wgtCD2a} = \text{wgtCD1} * \text{adjCD2a}$ . Similarly, the weight for those who received the hypertension questionnaire was calculated as  $\text{wgtCD2h} = \text{wgtCD1} * \text{adjCD2h}$ . After the adjustment was calculated, the non-selected units were dropped from the file.

### 8.1.3 Out-of-scope in SLCDC adjustment

After collection, the units were classified in two main groups. The first group is the resolved cases formed by the units confirmed through survey contact to be in-scope (i.e. in the target population) or out-of-scope (i.e. not in the target population). All the in-scope units were respondents. The second group was the unresolved units. For the units in the unresolved group, it was not possible to know if they were in our target population (in-scope) or not (out-of-scope) since no follow-up contact was made. Therefore, a logistic model was used to estimate the probability for an unresolved case to be in-scope. The resolved cases were used to determine the logistic model that predicted as well as possible the probability of being in-scope. The model was then applied to the unresolved cases to predict the probability of each unresolved unit to be in scope.

The weight for the unresolved units who received the arthritis questionnaire was calculated as  $\text{wgtCD3a} = \text{wgtCD2a} * p_{\text{inscope}}$ , where  $p_{\text{inscope}}$  is the predicted probability of being in scope. Similarly, the weight for unresolved units who received the hypertension questionnaire was calculated as  $\text{wgtCD3h} = \text{wgtCD2h} * p_{\text{inscope}}$ . This adjustment had the effect of reducing the total weight of the unresolved units by the predicted number of out-of-scope units in the population that they represented. After the adjustment was calculated, the resolved out-of-scope units were dropped from the file.

### 8.1.4 Non-response in SLCDC adjustment

Note that all the unresolved cases are now considered as non-respondents since their weights have been adjusted for out-of-scope and all the resolved cases that remained (in-scope units) are respondents. A logistic model using mainly CCHS auxiliary variables was built to predict the probabilities of being a respondent. From the predicted probabilities, response homogeneous groups (RHGs) were created. To make sure that the non-response adjustment did not change the estimated number of people with a condition at the stratum level or at the regional level, the RHG were created within each stratum by region since these were important domains of interest. Sometimes in order to meet the quality criteria's (minimum of 20 units by RHG and at least 50% respondent in each RHG) of the RHG creation, RHG from different regions were grouped together. But two RHG from different stratum were not allowed to be grouped together.

The adjustment was calculated within each RHG as follow:

$$\text{adjCD4a} = \frac{\text{Sum of weights for all units that received arthritis questionnaire}}{\text{Sum of weights for the respondents that received arthritis questionnaire}}$$

$$\text{adjCD4h} = \frac{\text{Sum of weights for all units that received hypertension questionnaire}}{\text{Sum of weights for the respondents that received hypertension questionnaire}}$$

The arthritis weight  $\text{wgtCD4a}$  was calculated as  $\text{wgtCD3a} * \text{adjCD4a}$ . Similarly, the hypertension weight  $\text{wgtCD4h}$  was calculated as  $\text{wgtCD3h} * \text{adjCD4h}$ . After the adjustment, the non-responding units were dropped from the file.

### 8.1.5 Share-link (final) adjustment

Only the information for the people who agreed to share and link their SLCDC data will be released (98% of respondents agreed to share and link). Since the people who did not agree to share or link their information are in the population of interest, adjustments were made to allocate the weights of the non-sharers / non-linkers to the remaining units. The probability of a respondent agreeing to share and link their SLCDC data was predicted using a logistic regression. From the predicted probabilities, response homogeneous groups (RHGs) were created the same way as in 8.1.4.

The adjustment was calculated within each RHG as follow:

$$\text{adjCD5a} = \frac{\text{Sum of weights for the respondents that received arthritis questionnaire}}{\text{Sum of weights (wgtCD4a) that agree to share and link SLCDC data}}$$

$$\text{adjCD5h} = \frac{\text{Sum of weights for for the respondents that received hypertension questionnaire}}{\text{Sum of weights (wgtCD4h) that agree to share and link SLCDC data}}$$

The final arthritis weight  $\text{wgtCD5a}$  for the respondents who received the arthritis questionnaire and agreed to share and link their SLCDC data was calculated as  $\text{wgtCD4a} * \text{adjCD5a}$ . Similarly, the final hypertension weight  $\text{wgtCD5h}$  was calculated as  $\text{wgtCD4h} * \text{adjCD5h}$ . After the adjustment was calculated, the respondents who did not agree to share or link their 2009 SLCDC data were dropped from the file.

The weights  $\text{wgtCD5a}$  and  $\text{wgtCD5h}$  correspond to the final 2009 SLCDC weight that can be found in the arthritis and hypertension data files with the variable name  $\text{WTSX\_S}$ .

## 8.2 Bootstrap weights

Coordinated bootstrap weights were used for SLCDC because of its dependence on the CCHS 2008 sample. Hence, the starting point for the SLCDC bootstrap weights was the 500 replicates from the 2008 CCHS share bootstrap file. Each bootstrap replicate was adjusted using the five adjustments listed in Table 8.1.

## 9. Data quality

### 9.1 Out-of-scope cases

As mentioned earlier in the User Guide, a higher than expected number of cases were out-of-scope. While every effort was made to reduce the out-of-scope cases, 17% of the arthritis cases and 13% of hypertension cases were out-of-scope. Reasons for this include:

- Respondents were incorrectly classified as having arthritis on the CCHS. For example, a number of respondents reported that their arthritis had not been diagnosed by a health professional, a requirement to be included in the SLCDC. In addition, some respondents thought that they had arthritis but really did not (i.e. osteoporosis).
- Respondents said that they did not have the condition because they no longer have symptoms.
- Respondents were conditioned from the CCHS to answer no to certain questions knowing that they would then be screened out of the survey (these could be considered refusals).

As a result of the high out-of-scope rate, the total number of people having the condition differs between the CCHS and the SLCDC. The CCHS likely included some respondents who reported having the condition but really do not (false positives). However, the SLCDC likely excluded some respondents who really do have the condition but said that they did not to avoid completing the survey (false negatives). Which survey is used depends on the needs of the data user. The CCHS provides a time series of arthritis and hypertension prevalence rates. In addition, the CCHS data can be used when looking at co-morbidities with other conditions. However, the SLCDC is able to measure detailed information about the quality of life and health behaviours of the arthritis and hypertension population.

### 9.2 Response rates

A total of 16,117 people were selected to take part in the 2009 SLCDC: 7,062 for the arthritis questionnaire and 9,055 for the hypertension questionnaire.

**For the arthritis questionnaire**, of the resolved cases (those that could clearly be determined to be in- or out-of scope), 979 cases were no longer in the SLCDC target population (for example, due to death, moving outside of Canada, not having the chronic condition, etc). Of the 5,820 estimated eligible people, 4,565 responded to the survey and agreed to share their data with the share partners and link back to their CCHS responses. This resulted in an overall response rate of 78.4%. The table below contains a summary of the SLCDC response rates by age group and sex for arthritis.

**Table 9.1: SLCDC sample selected, hit-rate and response rate by sex and age group for arthritis.**

Sex	Age Group	Sample Selected	Potential Survey Respondents	Hit Rate (%)	Respondents	Response Rate (%)
Female	20-44 years old	503	393	78.1	284	72.3
Female	45-64 years old	1,324	1,152	87.0	924	80.2
Female	65-74 years old	1,224	1,068	87.3	877	82.1
Female	75+ years old	1,311	1,053	80.3	801	76.1
Total Female		4,362	3,666	84.0	2,886	78.7
Male	20-44 years old	343	252	73.6	181	71.8
Male	45-64 years old	1,167	965	82.7	759	78.7
Male	65-74 years old	616	491	79.7	385	78.4
Male	75+ years old	574	445	77.6	354	79.6
Total Male		2,700	2,153	79.7	1,679	78.0
Total		7,062	5,820	82.4	4,565	78.4

**For the hypertension questionnaire**, of the resolved cases (those that could clearly be determined to be in- or out-of scope), 903 were no longer in the SLCDC target population. Of the 7,862 estimated eligible people, 6,142 responded to the survey and agreed to share their data with the share partners and link back to their CCHS responses. This resulted in an overall response rate of 78.2%. The table below contains a summary of the SLCDC response rates by age group and sex for hypertension.

**Table 9.2: SLCDC sample selected, hit-rate and response rate by sex and age group for hypertension.**

Sex	Age Group	Sample Selected	Potential Survey Respondents	Hit Rate (%)	Respondents	Response Rate (%)
Female	20-44 years old	844	431	51.1	309	71.7
Female	45-64 years old	1,324	1,214	91.7	969	79.8
Female	65-74 years old	1,324	1,253	94.7	1,029	82.1
Female	75+ years old	1,324	1,207	91.1	951	78.8
TOTAL Female		4,816	4,105	85.2	3,258	79.4
Male	20-44 years old	619	465	75.2	305	65.6
Male	45-64 years old	1,324	1,202	90.8	953	79.3
Male	65-74 years old	1,289	1,200	93.1	956	79.7
Male	75+ years old	1,007	890	88.3	670	75.3
TOTAL Male		4,239	3,757	88.6	2,884	76.8
Total		9,055	7,862	86.8	6,142	78.2

### 9.3 Data interpretation

The SLCDC was a follow-up survey that collected additional data from targeted respondents from the 2008 CCHS. Therefore, the CCHS and the SLCDC share the same 2008 survey population. Data that were collected for the SLCDC during February and March 2009 appear on the data files along with data collected from the CCHS during 2008. The data collected by the SLCDC reflect the status of the 2008 survey population for the February and March 2009 reference period, while data collected through the CCHS reflects the same survey population but for the 2008 reference period. Interpretation of estimates from the SLCDC should consider the reference period if it is felt that this would affect the responses from respondents. For example, the proportion of people with arthritis who are taking prescribed medication can be measured at 38.6%. Information on medication use was collected as part of the SLCDC follow-up so this information reflects the 2008 population as of February or March 2009. In a similar analysis, the proportion of people with hypertension who consider themselves as overweight can be measured at 61%. Information on obesity was collected as part of the CCHS so this reflects the 2008 population as of 2008. This slight nuance should only be an issue if it is felt that the collection period would have an effect on the response.

### 9.4 Survey errors

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

Errors which are not related to sampling may occur at almost every phase of a survey operation. Interviewers may misunderstand instructions, respondents may make errors in answering questions, the answers may be incorrectly entered on the questionnaire and errors may be introduced in the processing and tabulation of the data. These are all examples of 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 were taken to reduce non-sampling errors in the survey. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures include the use of highly skilled interviewers, extensive training of interviewers with respect to the survey procedures and questionnaire, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions, procedures to ensure that data capture errors were minimized, and coding and edit quality checks to verify the processing logic.

#### **9.4.1 The frame**

The 2009 SLCDC was a supplement to the 2008 CCHS, which is based mainly on an area frame and a telephone frame. The coverage of the 2009 SLCDC should then be the same as the CCHS, which is estimated at 98% of the Canadian population. It is unlikely that the under-coverage introduces any significant bias into the survey data.

#### **9.4.2 Non-response**

A major source of non-sampling errors in surveys is the effect of non-response on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. In the case of the 2009 SLCDC, only complete responses were kept for the survey. It has to be mentioned that respondents tended to complete the questionnaire once they started the interview. Total non-response occurred because the interviewer was either unable to contact the respondent, or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of individuals who responded to the survey to compensate for those who did not respond.

It is important to note that the SLCDC interview took place between three and 14 months after the 2008 CCHS interviews. During the SLCDC some units were unable to be reached because they moved or changed phone number. For these unresolved cases, an estimated portion of in-scope units and out-of-scope units was derived based on respondents and out-of-scope units resolved. The out-of-scope portion was estimated at 17% for arthritis and 13% for hypertension. This was taken into account at the estimation stage.

See Chapter 8 for more details on weighting adjustments for non-response.

### 9.4.3 Measurement of sampling error

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. This section of the documentation outlines the measures of sampling error which Statistics Canada commonly uses and which it urges users producing estimates from this microdata file to use also.

The basis for measuring the potential size of sampling errors is the standard error of the estimates derived from survey results. However, because of the large variety of estimates that can be produced from a survey, the standard error 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 error of the estimate by the estimate itself and is expressed as a percentage of the estimate.

For example, suppose that, based on the survey results, 45.1% of Canadians visited a health care professional in the past twelve months for their hypertension or arthritis and this estimate is found to have a standard error of 0.009. Then the coefficient of variation of the estimate is calculated as:

$$\left( \frac{0.009}{0.451} \right) \times 100 \% = 2.0 \%$$

There is more information on the calculation of coefficients of variation in Chapter 10.



## **10. Guidelines for tabulation, analysis and release**

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

### **10.1 Rounding guidelines**

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

- a) Estimates in the main body of a statistical table are to be rounded to the nearest hundred units using the normal rounding technique. In normal rounding, if the first or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is raised by one. For example, in normal rounding to the nearest 100, if the last two digits are between 00 and 49, they are changed to 00 and the preceding digit (the hundreds digit) is left unchanged. If the last digits are between 50 and 99 they are changed to 00 and the preceding digit is incremented by 1;
- b) Marginal sub-totals and totals in statistical tables are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units using normal rounding;
- c) Averages, proportions, rates and percentages are to be computed from unrounded components (i.e., numerators and/or denominators) and then are to be rounded themselves to one decimal using normal rounding. In normal rounding to a single digit, if the final or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is increased by 1;
- d) Sums and differences of aggregates (or ratios) are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units (or the nearest one decimal) using normal rounding;
- e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released that differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s);
- f) Under no circumstances are unrounded estimates to be published or otherwise released by users. Unrounded estimates imply greater precision than actually exists.

## 10.2 Sample weighting guidelines for tabulation

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

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

### 10.2.1 Definitions: categorical estimates, quantitative estimates

Before discussing how 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. How often individuals experience joint pain is an example of such an estimate. An estimate of the number of persons possessing a certain characteristic or exhibiting certain behaviours may also be referred to as an estimate of an aggregate.

Example of categorical question:

*In the past month, how often have you experienced joint pain? (SSAX\_01)*

- Always
- Often
- Sometimes
- Rarely
- Never

#### Quantitative estimates:

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

An example of a quantitative estimate is the average age at which individuals are first diagnosed with high blood pressure. The numerator is an estimate of the age at which individuals with

hypertension were first diagnosed with high blood pressure, and its denominator is an estimate of the number of individuals who have high blood pressure.

Example of quantitative question:

*How old were you when you were first diagnosed with high blood pressure?*  
(CNHX\_05)  
|\_|\_|\_| Age of diagnosis

### 10.2.2 Tabulation of categorical estimates

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

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

- summing the final weights of records having the characteristic of interest for the numerator ( $\hat{X}$ );
- summing the final weights of records having the characteristic of interest for the denominator ( $\hat{Y}$ ); then
- dividing the numerator estimate by the denominator estimate.

### 10.2.3 Tabulation of quantitative estimates

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

- 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 ( $\hat{X}$ );
- summing the final weights of records having the characteristic of interest for the denominator ( $\hat{Y}$ ); then
- dividing the numerator estimate by the denominator estimate.

For example, to obtain the estimate of the average age at which individuals are diagnosed with hypertension, first compute the numerator ( $\hat{X}$ ) by summing the product between the value of variable CNHX\_05 and the weight WTSX\_S. The denominator ( $\hat{Y}$ ) is obtained by summing the final weight of those records with a value of “2” to the variable CONFLAG. Divide ( $\hat{X}$ ) by ( $\hat{Y}$ ) to

obtain the average age at which individuals are diagnosed with hypertension.

### **10.3 Guidelines for statistical analysis**

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

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

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

### **10.4 Release guidelines**

Before releasing and/or publishing any estimate from the data file, users must first determine the number of sampled respondents having the characteristic of interest (for example, the number of respondents with arthritis who experience joint pain). If this number is less than 30, the un-weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the rounded estimate and follow the guidelines below.

**Table 10.1 Sampling variability guidelines**

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

**10.5 Variances and coefficients of variation**

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

The computation of coefficients of variation (or any other measure of precision) with the use of the bootstrap method requires access to information that is considered confidential.

For the computation of coefficients of variation, the bootstrap method is advised. A macro program, called “Bootvar”, was developed in order to give users easy access to the bootstrap method. The Bootvar program is available in SAS and SPSS formats, and is made up of macros

that calculate the variances of totals, ratios, differences between ratios, and linear and logistic regressions.

Although some standard statistical packages allow sampling weights to be incorporated in the analyses, the variances that are produced often do not take into account the stratified and clustered nature of the design properly, whereas the exact variance program would do so.

## **11. File usage**

This section begins by describing the data files and how the data files can be accessed, the weight variable of the data files and an explanation of how it should be used when doing tabulations. This is followed by an explanation of the variable naming convention that is employed by the SLCDC.

### **11.1 Data file**

The SLCDC consists of two different data files: one for arthritis and one for hypertension. Both of these data files have been linked to the 2008 CCHS.

Since the variables from the two surveys are on the same data file, it is important that users are aware of the variables which they are using in their analysis. For example, some demographic variables (age, sex and province of residence) were collected on the CCHS and the SLCDC. Users should therefore be aware which variables they are using in order to ensure consistency in their estimates. More information on how to differentiate the variables from the SLCDC and CCHS are provided in Sections 11.3 and 11.4.

Unlike other CCHS data files, the SLCDC does not have a Master file separate from a Share file. Rather, the SLCDC data file contains only the respondents who agreed to link their SLCDC data to their 2008 CCHS data. Furthermore, only respondents who agreed to share the linked data with the share partners are included on the data file.

The data can be accessed in a number of ways and are described in the next sections.

#### **11.1.1 Share partners**

Share partners have access to the data under the terms of the data sharing agreements. These data files only contain information on respondents who agreed to share their data with Statistics Canada's partners. The share partners for the SLCDC are the Public Health Agency of Canada (the survey sponsor), Health Canada and the provincial health departments. 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.

#### **11.1.2 Research Data Centres**

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

### 11.1.3 Custom tabulations

One way to provide access to the data files is to offer 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 obtain 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.ca](mailto:hd-ds@statcan.ca).

### 11.2 Use of weight variable

The weight variable **WTSX\_S** represents the SLCDC sampling weight. For a given respondent, the sampling weight can be interpreted as the number of people the respondent represents in the population. This weight must always be used when computing statistical estimates in order to make inferences at the population level possible. The production of un-weighted estimates is not recommended. The sample allocation, as well as the survey design, can cause such results to not correctly represent the population. Refer to Chapter 8 on weighting for a more detailed explanation on the creation of this weight.

### 11.3 Variable naming convention

The SLCDC adopted a variable naming convention that allows data users to easily use and identify the data based on module and condition. The variable naming convention follows the mandatory requirement of restricting variable names to a maximum of eight characters for ease of use by analytical software products.

#### 11.3.1 Variable name component structure in SLCDC

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

- Positions 1-2:** Module reference (e.g. SS – Symptoms and severity, ME – Medication use, AD – Use of assistive devices and HU – Health care utilization)
- Position 3:** Questionnaire-specific reference (A – Arthritis, H – Hypertension)
- Position 4:** Reference to the Survey on Living with Chronic Diseases in Canada (X)
- Position 5:** Variable type ( \_ - question, D - derived variable)
- Positions 6-8:** Question number

For example: The variable corresponding to Question 1, Health care utilization module, Hypertension questionnaire, SLCDC (HUHX\_01):

- Position 1-2:** HU Comes from the Health care utilization module
- Position 3:** H Hypertension questionnaire component
- Position 4:** X SLCDC
- Position 5:** \_ underscore ( \_ = collected data)



**Position 6-8:**            01        question number (& answer option where applicable)

**11.3.2 Positions 1-3: variable / questionnaire section name**

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

GEN	General health (on both arthritis and hypertension questionnaires)
DH	Diagnosis and family history
SS	Symptoms and severity
RA	Restriction of activities
AD	Use of assistive devices
RW	Restriction of work related activities
ME	Medication use
HU	Health care utilization
CL	Clinical recommendations
SM	Self-management
SW	Support and well-being
IN	Information and training
CN	Confirmation of high blood pressure diagnosis
BM	Blood pressure measurement
MO	Self-monitoring of blood pressure
ADM	Administration (on both arthritis and hypertension questionnaires)

The third position of the variable name consists of either an A if the module is on the arthritis questionnaire or an H if the module is on the hypertension questionnaire. A number of modules are on both questionnaires but with different questions for arthritis and hypertension.

**11.3.3 Position 4: Cycle and survey name**

The X in position four of the variable name indicates that the variable is part of the SLCDC.

**11.3.4 Position 5: variable type**

_	Collected variable	A variable that appeared directly on the questionnaire
C	Coded variable	A variable coded from one or more collected variables (e.g., SIC, Standard Industrial Classification code)
D	Derived variable	A variable calculated from one or more collected or coded variables, usually calculated during head office processing (e.g., Health Utility Index)

<b>F</b>	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)
<b>G</b>	Grouped variable	Collected, coded, suppressed or derived variables collapsed into groups (e.g., age groups)

### 11.3.5 Positions 6-8: variable name

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

For questions which allow for more than one response option (also referred to as a “mark-all” question), 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” and “no” answer for each response option. For example, if Q2 had 4 response options, the new questions would be named Q2A for option 1, Q2B for option 2, Q2C for option 3, etc. If only options 2 and 3 were selected, then Q2A = No, Q2B = Yes, Q2C = Yes and Q2D = No.

### 11.4 Variable name component structure in CCHS

Since the SLCDC data files have been linked to the CCHS, it is important to be able to distinguish between the surveys from which the variables originate. The variable naming convention for the CCHS and SLCDC is very similar. The only exception is that the SLCDC uses an X in the fourth position to indicate that the variable comes from the SLCDC.

The example below shows the age variable from the SLCDC and CCHS:

SLCDC: DHHX\_AGE  
CCHS: DHH\_AGE

Users should therefore be aware which variables they are using in order to ensure consistency in their estimates.