Statistics Canada Statistique Canada

Survey on Living with Chronic Diseases in Canada

USER GUIDE

November 2011

TABLE OF CONTENTS

1. Introduction

The Survey on Living with Chronic Diseases in Canada (SLCDC) is a cross-sectional survey that collects information related to the experiences of Canadians with chronic health conditions. Sponsored by the Public Health Agency of Canada (PHAC), the SLCDC takes place every two years, with two chronic diseases covered in each survey cycle. The 2011 survey focused on diabetes and respiratory conditions: asthma and chronic obstructive pulmonary disease (COPD).

The target population for the SLCDC is Canadians 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. The 2011 SLCDC included persons aged 20 years or older with diabetes, persons aged 12 years or older with asthma, and persons aged 35 years or older with COPD.

There were two data collection periods for the 2011 SLCDC: October and November of 2010 and March and April of 2011.

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

Any questions about the data sets and their use or about ordering of custom tabulations should be directed to:

Client Services, Health Statistics Division: 613-951-1746
E-mail: hd-ds@statcan.gc.ca

2. Background

The Survey on Living with Chronic Diseases in Canada (SLCDC) is a follow-up to the Canadian Community Health Survey (CCHS), an annual cross-sectional survey that collects information related to health status, health care utilization and health determinants for the Canadian population. Since the CCHS relies upon a large sample and identifies several diagnosed chronic conditions, it serves both as sample frame for the SLCDC and a source of additional health and socio-demographic information.

The central objective of the 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.

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 diabetes and respiratory conditions. 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) a diabetes-specific questionnaire and (2) an asthma and COPD - specific questionnaire.

A summary describing each of the modules on the diabetes and respiratory conditions questionnaires is provided in Section 4.2.

4.1 Qualitative testing

As previously stated, the 2011 SLCDC consisted of two different questionnaires: a diabetes questionnaire and a respiratory conditions questionnaire. The questionnaires were developed by Statistics Canada, in collaboration with PHAC. Diabetes and respiratory conditions 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.

Qualitative testing was conducted in February 2010. English testing took place in Toronto and French testing in Montreal. The frame used to select respondents for the interviews was the 2009 CCHS. A total of 36 participants took part in the testing, representing a cross-section of persons who reported in their CCHS interview having either diabetes, asthma, chronic bronchitis or chronic obstructive pulmonary disease (COPD) diagnosed by a doctor or other health professional. 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 testing the diabetes questionnaire:

In general, participants found the questionnaire straightforward and easy to answer. There were some questions where respondents were unsure of the terminology used in the questionnaire. For example, many respondents did not understand the term "A1C test" until a clarifying note was read to them. Certain questions in the *Coping and Support* module were poorly understood by respondents, especially in French.

Following qualitative testing, it was decided to remove the first five questions in the *Coping and Support* module and re-name the module to *Support and Well-being*. It was also decided to remove several questions related to difficulties experienced accessing health care and diabetes healthcare teams.

Key findings from testing the respiratory questionnaire:

Overall, participants judged the respiratory questionnaire to be clear-cut and easy to answer. Some respondents felt that the questionnaire was geared towards persons with more severe respiratory problems. One important observation made was that respondents were better able to report the colour of their inhaler rather than its type (rescue vs. controller), so this was emphasized in interviewer training.

The main change following qualitative testing was the decision to use the name of the respondent's respiratory condition in the dynamic text of the questionnaire rather than the term "breathing problems". The *Smoking History* module was expanded and the *Medication Cost* module was removed.

4.2 Final questionnaire content

This section outlines the modules comprising the content of the SLCDC diabetes and respiratory questionnaires. The diabetes questionnaire was made up of 15 modules, and the respiratory questionnaire of 17 modules.

Diabetes	
Module	**Description**
GENX	**General health:** The general health module is used to collect data on self-perceived health, satisfaction with life, self-perceived mental health and self-perceived stress. It 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 diabetes.
CNDX	**Confirmation of diabetes diagnosis:** This module is used to confirm that the respondent has received a diagnosis of diabetes from a health professional and the type of diabetes that they have. The module is also used to screen out women whose only diagnosis of diabetes occurred during pregnancy. The screening questions are modified from the *Chronic conditions* module in the CCHS. The remaining questions are new. A follow-up question is asked if the respondent says that they do not have diabetes to help determine why there is a discrepancy between what was reported in the CCHS and in the SLCDC. Even if the respondent says that they do not have diabetes because their condition is controlled by medication or lifestyle changes, they will continue with the survey. The experiences of persons who can control their condition through medication or lifestyle changes are of interest to the survey.
XHUX	**Health care utilization:** The *Health care utilization* module asks respondents about the health professional that is most responsible for treating their diabetes and whether in the past 12 months they have seen a variety of health professionals, as well as whether they have had difficulties receiving ongoing care for their diabetes.
CODX	**Clinical monitoring:** This module asks respondents about tests and exams that are commonly given by health professionals to monitor diabetes and its complications. The module asks about the frequency of the tests as well as their results. Test results and measurements can be used to determine how well-controlled the respondent's diabetes is.
MEDX	**Medication use:** This module asks respondents about prescription medications, including insulin injections, taken to control blood sugar, blood pressure and blood cholesterol. Respondents are also asked about compliance with their prescriptions as well as whether they take herbal or naturopathic remedies for their diabetes. Diabetes has a number of associated modifiable risk factors (e.g., diet, weight control, alcohol consumption) and researchers are interested in knowing whether changes to these modifiable risk factors are associated with a reduced rate of prescription medication use in people with diabetes.

ICDX	**Insurance coverage:** This module asks respondents whether they have insurance that covers the partial or full cost of prescription medications, glucose monitoring equipment, dental care and vision care. The cost of prescription drugs, glucose monitoring supplies, dental care and vision care can prevent some people from taking their prescription drugs, self-monitoring blood sugar levels or seeking preventative care.
CLDX	**Clinical recommendations:** This module is made up of a series of questions that ask about things that a doctor or other health professional may have suggested to a respondent to help them manage their diabetes. The purpose of these questions is to measure the extent to which doctors are following best-practice guidelines for the treatment of diabetes by discussing with their diabetic patients the importance of diet, exercise, weight control, etc.
SMDX	**Self-management:** This module asks respondents about things they may be doing to help manage their diabetes. The module covers activities that are recognized as being important for the management of diabetes (diet, exercise, weight control, etc.). Follow-up questions measuring barriers to change are asked of respondents who are not currently engaging in activities that are considered modifiable risk factors for diabetes, specifically diet, exercise and weight control. Questions are similar to those asked in the *Clinical recommendations* module. The purpose of these questions is to measure the extent to which respondents are applying best-practice guidelines to the treatment of their diabetes.
MODX	**Self-monitoring**: This module asks questions about diabetes monitoring respondents do themselves outside of a health professional's office. Topics include frequency of blood sugar self-monitoring, frequency of blood pressure monitoring and frequency of foot checks for sores or irritations.
DCDX	**Diabetes complications:** This module asks about health conditions and complications that are associated with diabetes. Conditions include cataracts, kidney failure, high blood pressure, etc. Respondents must have had the condition diagnosed by a health professional. The number of complications suffered by a respondent can be an indication of how severe or well-controlled their diabetes is.
RADX	**Restriction of activities:** This module asks a series of questions about being limited in daily or usual activities in the past 12 months because of the diabetes. When combined with information from the rest of the interview, this module can be used to compare persons who report having activity limitations due to diabetes with those who do not. In conjunction with the module on diabetes complications, this

	module will help identify persons with more severe symptoms of diabetes. Persons with more complications and more activity limitations will likely differ across variables such as health care utilization and medication use, compared to persons with fewer complications and limitations.
RWDX	**Restriction of work-related activities:** The purpose of this module is to identify respondents who have had to modify their paid work because of their diabetes. The module includes questions on the respondent's work life, including current and past employment status, and changes made to work activities due to their diabetes.
SWDX	**Support and well-being:** Having the support of family and friends, and being able to deal with emotions such as stress have been shown to be beneficial to people with diabetes. This module asks about self-perceived social support available to the respondent, and whether they have ever needed help for their emotions or mental health in order to cope with their diabetes.
PADX	**Patient activation:** This module asks a series of questions to determine how involved respondents are in their diabetes treatment and whether they feel confident to handle problems and complications that may develop.
ADMX	**Administration**: The module asks respondents' permission to link their information from the SLCDC to their responses from the 2010 CCHS. Respondents are then asked if this information can be shared with Statistics Canada's share partners.

Respiratory	
Module	**Content**
GENX	**General health:** The general health module is used to collect data on self-perceived health, satisfaction with life, self-perceived mental health and self-perceived stress.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 respiratory condition.
DHRX	**Diagnosis and family history:** This module is used to confirm that the respondent has received a diagnosis of asthma or chronic bronchitis/emphysema/COPD from a health professional. For respondents who indicate that they do not have a breathing problem, there is another question to determine why there is a discrepancy between what was reported in the CCHS and in the SLCDC. Respondents who say they feel better or they take medication to control their breathing problem continue with the survey. The experiences of persons who can control their condition through medication or lifestyle changes are of interest to the survey. Other questions in this module ask age at diagnosis, the type of respiratory condition, and whether the respondent has any blood relatives who have ever been diagnosed with any of asthma, chronic bronchitis, emphysema, or COPD.
SSRX	**Symptoms and severity:** Questions in this module ask about the frequency and severity of the main symptoms of asthma and COPD. In conjunction with the module on restriction of activities, this module will help identify persons with more severe respiratory symptoms. Persons with more severe symptoms and more activity limitations will likely differ across variables such as health care utilization and medication use, compared to persons with less severe symptoms.
TRRX	**Triggers:** Questions in this module ask respondents about things that bring on the symptoms of their asthma or COPD, or make them worse. A list of common triggers is read to the respondents and they are asked to indicate which ones affect them.
HURX	**Health care utilization:** This module refers to contacts respondents had with various health professionals regarding their asthma or COPD in the past 12 months. Information is also sought about the total number of visits to a doctor, to the emergency room and nights spent in hospital in the previous 12 months because of respondent's asthma or COPD.
MERX	**Medication use:** This module covers the current use of prescription medications for asthma and COPD. It collects detailed information on the two main types of inhalers used to treat these respiratory conditions (reliever and controller), as well as the use of corticosteroids and antibiotics, which are treatments used in more severe cases of the

	conditions. The module also has questions about reasons for not taking prescription medications for asthma or COPD and medication compliance.
HCRX	**Health conditions:** This module includes questions about health conditions that are associated with respiratory conditions, such as sleep apnea, osteoporosis, and heart failure. The age of the respondent determines which conditions they are asked about.
ALRX	**Allergies:** This module asks respondents with asthma about allergies they have and whether they have received treatment for them. We are only interested in allergies that have been diagnosed by a health professional, as the result of allergy testing.
RARX	**Restriction of activities:** This module includes a series of questions about whether the respondent has been limited in their daily or usual activities in the past 12 months because of their asthma or COPD. When combined with information from the rest of the interview, this module can be used to compare persons who report having activity limitations due to asthma or COPD with those who do not. In conjunction with the module on symptoms and severity, this module will help identify persons with more severe symptoms. Persons with more severe symptoms and more activity limitations will likely differ across variables such as health care utilization and medication use, compared to persons with less severe symptoms.
RWRX	**Restriction of work-related activities:** The purpose of this module is to identify respondents who have had to modify their paid work because of their respiratory condition. This module includes a number of questions on the respondent's work life, including current and past employment status, changes made to work activities due to their respiratory condition, and exposure to dust, fumes, or gases at work.
RERX	**Restriction of educational activities:** The purpose of this module is to identify respondents who have had to modify their educational activities because of their respiratory condition. This module is only asked of respondents under 50 years of age and focuses only on current school activities.
RVRX	**Restriction of volunteer activities:** The purpose of this module is to identify respondents who have had to modify their volunteer activities because of their respiratory condition. This module includes a number of questions on the respondent's current and past volunteer status and changes made to volunteer activities due to their respiratory condition.
SMRX	**Self-management:** The *Self-management* module asks respondents about things that a

	doctor or health professional may have suggested they do to help manage their breathing problems. Respondents are also asked whether they did any of these things to help manage their asthma or COPD. The module covers a number of activities and changes that are recognized as being important for the management of respiratory conditions (e.g. seeing an asthma or COPD educator or changing the home environment). These questions are of interest to researchers who want to know if health professionals are recommending certain activities and whether people with asthma or COPD are making changes to their behaviour and environment as a result of their diagnosis.
SWRX	**Support and well-being:** Having the support of family and friends, and being able to deal with emotions such as stress have been shown to be beneficial to people with asthma and COPD. This module asks about self-perceived social support available to the respondent, and whether they have ever needed help for their emotions or mental health in order to cope with their asthma or COPD. There are also additional questions for respondents with COPD asking whether they have discussed their wishes for care in case of hospitalization and as their condition progresses.
SHRX	**Smoking history:** This module asks respondents about cigarette smoking over the course of their lifetime. The module contains questions about the length of time they smoked daily and occasionally, and the number of cigarettes they smoked. The purpose of this module is to calculate "pack years", which is the number of cigarettes smoked daily multiplied by the number of years smoked divided by 20. This calculation is used by researchers and physicians to determine the intensity of an individual's tobacco exposure over the course of their lifetime.
SCRX	**Smoking cessation:** This module asks respondents who are current smokers about their intention to quit smoking in the near future, if they have received advice or information from their doctor about quitting, as well as if they have tried to quit in the previous 12 months and any methods they may have used to quit. Respondents are also asked questions about exposure to second-hand smoke in their home.
ADMX	**Administration:** This module asks respondents' permission to link their information from the SLCDC to their responses from the 2010 CCHS. Respondents are then asked if this information can be shared with Statistics Canada's share partners. Respondents aged 12 and 13 first answer the link and share questions for themselves. Then their parent or guardian are asked to give their permission to share and link the data. Both the respondent's and their parent's or guardian's permissions are required to link and share the data.

5. Sample Design

5.1 Target population

The 2011 Survey on Living with Chronic Disease in Canada (SLCDC) targets Canadians living in private dwellings in the ten provinces with one of the following conditions:

- Asthma (excluding people who also have COPD) and aged 12 years or older as of December 31st 2010
- Diabetes (excluding women who are only diabetic during pregnancy) and aged 20 years or older as of December 31st 2010
- COPD (including people who also have asthma) and aged 35 years or older as of December 31st 2010.

The conditions had to have been diagnosed by a health professional. Residents of the three territories; persons living on Indian Reserves, Crown lands, or in institutions; full-time members of the Canadian Forces and residents of certain remote regions were not in-scope for this survey and were excluded. These exclusions represent about 2% of the overall Canadian population.

5.2 Domains of interest

The SLCDC aims to produce reliable estimates at the national level by age group and sex. The targeted age groups differ between conditions:

- For asthma (excluding people who also have COPD), they are 12 – 24, 25 – 39, 40 – 54, and 55+
- For diabetes (excluding women who are only diabetic during pregnancy), they are 20 – 64 and 65+
- For COPD (including people who also have asthma), they are 35+.

However, since the 2011 SLCDC sample was selected from the 2010 Canadian Community Health Survey (CCHS), the sample size of the SLCDC was limited by the number of people with the conditions in the CCHS. It is therefore possible that some age groups will end up having insufficient sample size and need to be collapsed for analytical purposes.

5.3 Sampling frame

The 2011 SLCDC used the 2010 CCHS to select its sample. The SLCDC employs a two-phase design in which the first phase is the CCHS sample and the second phase is the SLCDC sample.

The CCHS sample 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 2010 CCHS User Guide.

5.4 Sample size and allocation

In order to produce reliable estimates at the national level by age group and sex, CCHS respondents were stratified by condition, age group and sex. Because the age groups of interest differ between conditions and there were units with more than one condition (e.g. both asthma and diabetes), the age groups used in stratification and allocation consist of the intersection of the age groups listed under section 5.2.

To reduce response burden, it was decided that respondents could receive only one questionnaire. For people having two conditions[1], they were randomly assigned the questionnaire that corresponded to one of their conditions. The sample allocation by questionnaire was done by sex and age groups in proportion to the number of 2010 CCHS respondents for each condition.

Some in-scope units were excluded before sample selection. Units were excluded if they were a proxy respondent or if they did not agree to share or link their CCHS data. These exclusions represented 13% of the first phase sample and were taken into account at the estimation stage since they are part of the population of interest.

Since the number of CCHS respondents with each of the three conditions was not large, the decision was made to include all such units in the 2011 SLCDC sample. In the end, after exclusions, as well as assigning only one questionnaire to units with two conditions, there were 3,650 units in the asthma sample, 3,747 units in the diabetes sample and 1,733 units in the COPD sample. Tables 5.1, 5.2 and 5.3 show the sample sizes of each of the three conditions by age groups and sex:

[1] According to the definitions of the conditions, nobody in the SLCDC sample could be classified to both asthma and COPD. As a result, the maximum number of conditions to which anyone could be classified was two: COPD and diabetes or asthma and diabetes.

Table 5.1: SLCDC sample size by stratum for the asthma condition

Stratum (Sex, Age group)		2011 SLCDC sample size
Female	12-24 years old	512
Female	25-39 years old	544
Female	40-54 years old	427
Female	55+ years old	708
Total Female		2191
Male	12-24 years old	504
Male	25-39 years old	326
Male	40-54 years old	272
Male	55+ years old	357
Total Male		1459
Total		3650

Table 5.2: SLCDC sample size by stratum for the diabetes condition

Stratum (Sex, Age group)		2011 SLCDC sample size
Female	20-64 years old	769
Female	65+ years old	1086
Total Female		1855
Male	20-64 years old	875
Male	65+ years old	1017
Total Male		1892
Total		3747

Table 5.3: SLCDC sample size by stratum for the COPD condition

Stratum (Sex, Age group)		2011 SLCDC sample size
Female	35+ years old	1078
Male	35+ years old	655
Total		1733

6. Data collection

Collection for the SLCDC took place in October and November of 2010 and continued in March and April of 2011. Over the collection period, a total of 6,573 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 the following:

- 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 diabetes and respiratory questionnaire modules, which made up the bulk of the application. Development and testing of the CATI application began in April 2010. There were 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 October 2010 and March 2011, representatives from Statistics Canada's Collection Planning and Management Division visited the four regional offices participating in the collection of the SLCDC data (Halifax, Sherbrooke, Sturgeon Falls, and Edmonton). The purpose of the visits was to train the regional office project managers and teams of interviewers for the SLCDC diabetes and respiratory 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 focus of these sessions was to make interviewers comfortable using the SLCDC application and familiarise interviewers with survey content. The training sessions covered the following topics:

- goals and objectives of the survey
- survey methodology
- application functionality
- the questionnaire content and exercises

- mock interviews to simulate difficult situations and practise ways of dealing with non-response
- 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, several practices were used to minimise non-response:

Introductory letters

Before the start of the collection period, introductory letters explaining the purpose of the survey were sent to the targeted respondents. The letters described the importance of the survey and provided examples of how the SLCDC data would be used.

Mailing address information was not available for all respondents from the 2010 CCHS. For cases where mailing addresses were not available, an introductory letter was not sent out.

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 2010 CCHS, tracing of the respondent was initiated. In order to trace respondents, alternate contacts provided by the respondent on the 2010 CCHS were used 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 two six week collection periods. Half of the sample was collected in each collection period. The regional collection offices were instructed to use the first two weeks of each collection period to complete 40% of the cases, with the rest of the collection period being used to finalize the remaining sample and to follow up on outstanding non-response cases.

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 internal measures to verify interviewer performance and daily reports monitoring 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 the beginning of each 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 observed during collection.

One issue that arose during data collection of the 2009 SLCDC was a higher than expected out-of-scope rate. As a result of this, a series of questions was developed and included in the 2011 questionnaires to follow-up with any respondents who reported that they had not been diagnosed

with diabetes, asthma, or COPD. These questions were aimed at identifying respondents who had been diagnosed, but were no longer experiencing symptoms or who were able to manage their condition through medication or changes to their lifestyle. This reduced the number of out-of-scope cases.

For the 2011 SLCDC only COPD produced a higher than expected out-of-scope rate. The question on the CCHS that is used to identify respondents with COPD is: *Do you have chronic bronchitis, emphysema or chronic obstructive pulmonary disease or COPD?* After the SLCDC collection the notes and remarks for the cases that were coded out-of-scope were examined to gain a better understanding of why respondents reported having COPD on the CCHS but did not on the SLCDC. There was an indication that there was confusion among some respondents about the term "chronic bronchitis". Many seemed to confuse it with the form of bronchitis that is a bacterial or viral infection and reported having "chronic bronchitis" on the CCHS and were screened into the SLCDC as a result.

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" category.

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 taken insulin injections for their diabetes 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

To ensure estimates produced from survey data are representative of the surveyed 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, which consists of all in-scope respondents to the survey. Intuitively, the weight corresponds to the number of persons in the population that are represented by the respondent.

As described in Chapter 5, the SLCDC survey frame is composed of respondents to the 2010 CCHS. The starting point for the SLCDC weighting process is therefore the 2010 CCHS share weight. For more information on this weight, please refer to the 2010 CCHS User Guide.

8.1 Weight adjustment for sample weight

Table 8.1 presents an overview of the different adjustments that are part of the weighting strategy for SLCDC 2011, in the order in which they are applied.

Table 8.1: Weighting steps for SLCDC

CD 1 – Proxy-Link 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
CD 6 – Winsorization
CD 7 – Post-Stratification

8.1.1 Proxy-link adjustment

The first step of weighting for SLCDC was to drop the CCHS units in the territories, since they were not part of the target population. The next step was to adjust for the fact that some CCHS respondents were excluded from the SLCDC for practical reasons, even though they were still part of the target population. The reasons for their exclusion are as follows:

- People who did not agree to link their 2010 CCHS information were excluded. One of the objectives of the SLCDC was to link the SLCDC survey responses with the 2010 CCHS and provide the linked files to survey share partners. Without their permission to link, there was no reason to survey these CCHS respondents.

- People for whom their 2010 CCHS interview was done by proxy were excluded since the SLCDC questionnaire could not be answered by proxy.

Since these CCHS respondents belonged to 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 2010 CCHS, and within each cell (defined as the intersection of condition, age group, sex, and region) the adjustment is calculated as

$$\text{adjCD1} = \frac{\text{Sum of CCHS share weights for all sharers}}{\text{Sum of share weights for all non - proxies who agreed to link CCHS data}}$$

The weight wgtCD1 is calculated as wgts3*adjCD1, where wgts3 is the final 2010 CCHS share weight. After the adjustment is calculated, the excluded units are dropped from the file.

8.1.2 SLCDC selection criteria adjustment

In the SLCDC sampling design, CCHS respondents were stratified by age group and sex (see Chapter 5). In each age group by sex stratum, a unit could have either one condition (asthma, COPD, or diabetes) or two conditions (COPD and diabetes or asthma and diabetes). Those with COPD and asthma were considered as having only COPD. All of those identified as having one condition were selected for the SLCDC and no weight adjustments were necessary. To lower response burden for those with two conditions, the sample unit could receive only one questionnaire. As a result of this selection, weight adjustments were necessary. The adjustment for units with two conditions were calculated as follows:

$$\text{adjCD2a} = \frac{\text{Sum of weights of the units with two conditions (asthma and diabetes)}}{\text{Sum of weights of units with two conditon who are selected to receive asthma questionnaire}}$$

$$\text{adjCD2c} = \frac{\text{Sum of weights of the units with two conditions (COPD and diabetes)}}{\text{Sum of weights of units with two conditon who are selected to receive COPD questionnaire}}$$

$$\text{adjCD2d} = \frac{\text{Sum of weights of the units with two conditions (COPD and diabetes or asthma and diabetes)}}{\text{Sum of weights of units with two conditon who are selected to receive diabetes questionnaire}}$$

The weight for those who received the asthma questionnaire was calculated as wgtCD2a = wgtCD1*adjCD2a. Similarly, the weight for those who received the COPD questionnaire was calculated as wgtCD2c = wgtCD1*adjCD2c and for those who received the diabetes questionnaire was wgtCD2d = wgtCD1*adjCD2d.

8.1.3 SLCDC out-of-scope adjustment

After collection, units were classified into two main groups: resolved cases and unresolved cases. Resolved cases are units where contact was made with the CCHS respondent and it was confirmed whether or not they still had the condition (in-scope or out-of-scope). The out-of-scope units were dropped from the file. The unresolved cases are units that cannot be contacted and so it is not possible to know if they are in-scope or out-of-scope. Therefore, logistic models were used (based on the resolved cases) to estimate the probability for an unresolved case to be in-scope. This probability was then used in adjusting the weights.

The weight for the unresolved units that received the asthma questionnaire was calculated as wgtCD3a = wgtCD2a*p_inscope, where p_inscope is the predicted probability of being in scope. Similarly, wgtCD3c = wgtCD2c*p_inscope for COPD units and wgtCD3d = wgtCD2d*p_inscope for diabetes units. This adjustment reduces the total weight of the unresolved units by the predicted number of out-of-scope units that they represent in the population. The units remain in the file to be treated in the non-response adjustment.

8.1.4 SLCDC non-response adjustment

Note that at this point all the unresolved cases were considered to be non-respondents since their weights were adjusted for out-of-scope. Similarly, all the resolved cases that remained (i.e. in-scope units) were respondents. Logistic models using mainly CCHS auxiliary variables were built to predict the probabilities of being a respondent. From the predicted probabilities, response homogeneous groups (RHGs) were created. To ensure 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 RHGs were created within each stratum by region.

The adjustment was calculated within each RHG as follows:

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

$$\text{adjCD4c} = \frac{\text{Sum of weights for all units that receive COPD questionnaire}}{\text{Sum of weights for the respondents that receive COPD questionnaire}}$$

$$\text{adjCD4d} = \frac{\text{Sum of weights for all units that receive diabetes questionnaire}}{\text{Sum of weights for the respondents that receive diabetes questionnaire}}$$

The asthma weight wgtCD4a was calculated as wgtCD3a*adjCD4a. Similarly, the COPD weight wgtCD4c was calculated as wgtCD3c*adjCD4c and the diabetes weight wgtCD4d was wgtCD3d*adjCD4d. After the adjustment, the non-responding units were dropped from the file.

8.1.5 Share-link adjustment

Only the information for the people who agreed to share and link their SLCDC data will be released. Of the 2011 SLCDC respondents, 99% agreed to share and link. Since the people who did not agree to share or link were still 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 is predicted using a logistic regression model. From the predicted probabilities, response homogeneous groups (RHGs) were created the same way as described in 8.1.4.

The adjustment was calculated within each RHG as follows:

$$\text{adjCD5a} = \frac{\text{Sum of weights for respondents who receive asthma questionnaire}}{\text{Sum of weights for respondents who agree to share and link SLCDC data}}$$

$$\text{adjCD5c} = \frac{\text{Sum of weights for respondents who receive COPD questionnaire}}{\text{Sum of weights for respondents who agree to share and link SLCDC data}}$$

$$\text{adjCD5d} = \frac{\text{Sum of weights for respondents who receive diabetes questionnaire}}{\text{Sum of weights for respondents who agree to share and link SLCDC data}}$$

At this point, the weight wgtCD5a for the asthma respondents who agreed to share and link their SLCDC data was calculated as wgtCD4a*adjCD5a. Similarly, wgtCD5c = wgtCD4c*adjCD5c and wgtCD5d = wgtCD4d*adjCD5d. After the adjustment, the respondents who did not agree to share or link their 2011 SLCDC data were dropped from the file.

8.1.6 Winsorization

Following the series of weight adjustments, some units may come out with extreme weights compared to other units of the same domain of interest. These units can represent a large proportion of their strata or have a large impact on the variance. To prevent this, the weight of these outlier units is adjusted downward using a "winsorization" trimming approach similar to the one used by CCHS. After winsorization, the weights for asthma, COPD, and diabetes became wgtCD6a, wgtCD6c, and wgtCD6d, respectively.

8.1.7 Post-stratification

To ensure the total numbers of estimated people with the conditions agree with the counts prior to winsorization by stratum and region, post-stratification is employed. Within each intersection of stratum and region, the following adjustment factors were calculated:

$$\text{adjCD7a} = \frac{\text{Sum of weights before winsorization}}{\text{Sum of weights after winsorization}}$$

$$\text{adjCD7c} = \frac{\text{Sum of weights before winsorization}}{\text{Sum of weights after winsorization}}$$

$$\text{adjCD7d} = \frac{\text{Sum of weights before winsorization}}{\text{Sum of weights after winsorization}}$$

The final asthma weight wgtCD7a was calculated as wgtCD6a*adjCD7a. Similarly, the final COPD weight wgtCD7c = wgtCD6c*adjCD7c and the final diabetes weight wgtCD7d = wgtCD6d*adjCD7d. The weights wgtCD7a, wgtCD7c, and wgtCD7d correspond to the final 2011 SLCDC weight that can be found under the variable name WTSX_S.

8.2 Bootstrap weights

Coordinated bootstrap weights are used for SLCDC because of its dependence on the 2010 CCHS sample. Hence, the starting point for the SLCDC bootstrap weights was the 500 replicates from the 2010 CCHS share bootstrap file. Each bootstrap replicate was adjusted using the seven adjustments listed in Table 8.1.

9. Data quality

9.1 Out-of-scope cases

The out-of-scope rates of SLCDC vary among the different conditions. The rates are 7% for asthma, 18% for COPD, and 3% for diabetes. There may be several reasons for units becoming out-of-scope:

- Respondents were incorrectly classified as having the condition according to the CCHS. For example, a number of respondents reported that their condition had not been diagnosed by a health professional, a requirement to be in-scope for the SLCDC.
- Respondents indicated that they did not have the condition because they no longer had symptoms.
- Respondents were conditioned by the CCHS to answer "no" to certain questions, knowing that they would then be screened out of the survey. In a certain sense, these units can be considered refusals.

Due to out-of-scope units, the total number of people having the condition differs between the CCHS and the SLCDC. This is especially true for the COPD condition. The CCHS likely includes some respondents who report having the condition but really do not (false positives). However, the SLCDC likely excludes some respondents who really do have the condition but who indicate that they do not to avoid completing the survey (false negatives). The objectives of the analysis dictate which survey should be used. For example, the CCHS provides a time series of the prevalence rates by condition while the SLCDC can be considered a one-time survey. In addition, the CCHS data should be used when looking at co-morbidities with other conditions. However, the SLCDC is able to provide detailed information about the quality of life and health behaviours of persons living with the chronic diseases.

9.2 Response rates

A total of 9,130 people were selected to take part in the 2011 SLCDC: 3,650 for the asthma questionnaire, 1,733 for the COPD questionnaire and 3,747 for the diabetes questionnaire.

For the asthma questionnaire, there were 238 units deemed out-of-scope among the resolved cases (units that had been contacted and could be classified to be in- or out-of scope). Among the unresolved cases (units with no contact, so it was not possible to determine whether they were in- or out-of-scope), the logistic model predicted 68 units to be out-of-scope. For more detailed information on the use of the logistic model, please refer to section 8.1.3. Therefore, there were a total of 306 modelled out-of-scope units. Of the 3,344 modelled in-scope units, 2,507 cases responded to the survey and agreed to share their data with the share partners and to link back to their CCHS responses. This resulted in a response rate of 75.0%. Table 9.1 below contains a summary of the SLCDC response rates by age group and sex for asthma.

Table 9.1: SLCDC initial sample size, modelled in-scope rate and response rate by sex and age group for the asthma questionnaire.

Sex	Age Group	Sample Selected	Modelled No. of In-scope Units	Modelled In-scope rate (%)	Respondents	Response Rate (%)
Female	12-24 years old	512	483	94.3	321	66.5
Female	25-39 years old	544	511	93.9	360	70.5
Female	40-54 years old	427	397	93.0	289	72.8
Female	55+ years old	708	648	91.5	566	87.3
Total Female		2191	2039	93.1	1536	75.3
Male	12-24 years old	504	459	91.1	331	72.1
Male	25-39 years old	326	294	90.2	197	67.0
Male	40-54 years old	272	247	90.8	185	74.9
Male	55+ years old	357	305	85.4	258	84.6
Total Male		1459	1305	89.4	971	74.4
Total		3650	3344	91.6	2507	75.0

For the COPD questionnaire, there were 315 units deemed out-of-scope among the resolved cases (units that had been contacted and could be classified as in- or out-of scope). Among the unresolved cases (units with no contact, so it was not possible to determine whether they were in- or out-of-scope), the logistic model predicted 57 units to be out-of-scope. Therefore, there were a total of 372 modelled out-of-scope units. Of the 1,361 modelled in-scope units, 1,133 cases responded to the survey and agreed to share their data with the share partners and to link back to their CCHS responses. This resulted in a response rate of 83.2%. Table 9.2 below contains a summary of the SLCDC response rates by age group and sex for COPD.

Table 9.2: SLCDC initial sample size, modelled in-scope rate and response rate by sex and age group for the COPD questionnaire.

Sex	Age Group	Sample Selected	Modelled No. of In-scope Units	Modelled In-scope rate (%)	Respondents	Response Rate (%)
Female	35+ years old	1078	870	80.7	728	83.7
Male	35+ years old	655	491	75.0	405	82.5
Total		1733	1361	78.5	1133	83.2

For the diabetes questionnaire, there were 129 units deemed out-of-scope among the resolved cases (units that had been contacted and could be classified as in- or out-of scope). Among the

unresolved cases (units with no contact, so it was not possible to determine whether they were in- or out-of-scope), the logistic model predicted 28 units to be out-of-scope. Therefore, there were a total of 157 modelled out-of-scope units. Of the 3,590 modelled in-scope units, 2,933 cases responded to the survey and agreed to share their data with the share partners and to link back to their CCHS responses. This resulted in a response rate of 81.7%. Table 9.3 below contains a summary of the SLCDC response rates by age group and sex for diabetes.

Table 9.3: SLCDC initial sample size, modelled in-scope rate and response rate by sex and age group for the diabetes questionnaire.

Sex	Age Group	Sample Selected	Modelled No. of In-scope Units	Modelled In-scoperate (%)	Respondents	Response Rate (%)
Female	20-64 years old	769	735	95.6	589	80.1
Female	65+ years old	1086	1042	95.9	874	83.9
Total Female		1855	1777	95.8	1463	82.3
Male	20-64 years old	875	841	96.1	678	80.6
Male	65+ years old	1017	972	95.6	792	81.5
Total Male		1892	1813	95.8	1470	81.1
Total		3747	3590	95.8	2933	81.7

9.3 Data interpretation

Since the 2011 SLCDC is a follow-up survey that collected additional data from targeted respondents from the 2010 CCHS, the two surveys share the same survey population. However, their reference periods can be different. The reference period for the 2010 CCHS is the 2010 calendar year while the data collected by the SLCDC reflects *the status of the same survey population in October – November 2010 or March – April 2011*, depending on whether the unit belonged to the first or second collection period. Under most circumstances, this will not affect the comparison of data from the CCHS and SLCDC. However, interpretation of estimates from the SLCDC should consider the reference period if it is felt that this would affect the responses from respondents.

9.4 Survey errors

The estimates derived from this survey are based on a sample of persons. Somewhat different estimates may have been obtained if a complete census has been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. The differences between the estimates obtained from the sample and those resulting from a census taken under similar conditions are 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 are taken to reduce non-sampling errors in the survey. Quality assurance measures are 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 are minimized, and coding and edit quality checks to verify the processing logic.

9.4.1 The frame

The 2011 SLCDC was a supplement to the 2010 CCHS, which is based mainly on an area frame and a telephone frame. The coverage of the 2011 SLCDC should then be the same as the CCHS in the ten provinces; the SLCDC does not cover the territories. For the ten provinces, it is unlikely that the under-coverage of CCHS and SLCDC will introduce 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 2011 SLCDC, only complete responses were kept for the survey. It is worthwhile to note that respondents tend to complete the questionnaire once they start the interview so partial non-response tends to be rare. Total non-response occurs because the interviewer is either unable to contact the respondent, or the respondent refuses 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 2011 SLCDC interview took place several months after the 2010 CCHS interviews. As a result, some units could not be contacted because they moved or changed phone number. The strategy of breaking the collection into two waves (October – November 2010 and March – April 2011) reduced this risk. For these unresolved (non-contact) cases, logistic models were used to estimate the portion of in-scope and out-of-scope units (see section 8.1.3 for more details).

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. It urges users to provide similar measures when producing estimates from this microdata file.

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.

A fictitious example is used to illustrate. Suppose that, based on the survey results, 45.1% of Canadians visited a health care professional in the past twelve months for their chronic disease and this estimate has a standard error of 0.009. Then the coefficient of variation of the estimate is calculated as:

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

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 proceeding digit is incremented by 1;

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

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

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

e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released that differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s);

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

10.2 Sample weighting guidelines for tabulation

The sample design used for 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 health professional checks blood pressure of individuals at their diabetes related appointments 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:

How often does your health professional check your blood pressure at your diabetes related appointments? (CODX_04)

__ 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 asthma. The numerator is an estimate of the age at which individuals with asthma were first diagnosed with this condition, and its denominator is an estimate of the number of individuals who have been diagnosed with asthma.

Example of quantitative question:

How old were you when you were first diagnosed with asthma? (DHRX_07)

|_|_|_| 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:

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

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

c) dividing the numerator estimate by the denominator estimate.

10.2.3 Tabulation of quantitative estimates

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

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

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

c) dividing the numerator estimate by the denominator estimate.

For example, to obtain the estimate of the average age at which individuals are diagnosed with asthma , first compute the numerator ($\hat{\boldsymbol{X}}$) by summing the product between the value of variable DHRX_07 and the weight WTSX_S. The denominator ($\hat{\boldsymbol{Y}}$) is obtained by summing the final weight of those records with a value of "2" to the variable CONFLAG. Divide ($\hat{\boldsymbol{X}}$) by ($\hat{\boldsymbol{Y}}$) to obtain the average age at which individuals are diagnosed with asthma .

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, or 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 diabetes who have been diagnosed with glaucoma). 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 diabetes and one for respiratory conditions. Both of these data files have been linked to the 2010 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 is 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 2010 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 contain only 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 some 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.

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, SM –Self-management and HU – Health care utilization)
Position 3:	Questionnaire-specific reference (D – Diabetes, R – Respiratory)
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, Respiratory questionnaire, SLCDC (HURX_01):

Position 1-2:	HU	Comes from the Health care utilization module
Position 3:	R	Respiratory 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 diabetes and respiratory questionnaires)
CN	Confirmation of diabetes diagnosis
DH	Diagnosis and family history
SS	Symptoms and severity
TR	Triggers
CO	Clinical monitoring
HU	Health care utilization
ME	Medication use
IC	Insurance coverage
HC	Health conditions
AL	Allergies
CL	Clinical recommendations
RA	Restriction of activities
RW	Restriction of work-related activities
RE	Restriction of educational activities
RV	Restriction of volunteer activities
SM	Self-management
MO	Self-monitoring
SW	Support and well-being
SH	Smoking history
SC	Smoking cessation
DC	Diabetes complications
PA	Patient activation
ADM	Administration (on both diabetes and respiratory questionnaires)

The third position of the variable name consists of either a D if the module is on the diabetes questionnaire or an R if the module is on the respiratory questionnaire. A number of modules are on both questionnaires but with different questions for respondents with diabetes and respiratory conditions.

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)
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)

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.