Canadian Community Health Survey — Methodological overview

Yves Béland

● *Implementation of the Canadian Community Health Survey (CCHS) marks an important turning point in the collection of information on the health of Canadians. For the first time, it will be possible to compare health-related characteristics of the population at sub-provincial levels in all provinces and territories.*

Abstract

Objectives
This article describes the design, sampling strategy, interviewing procedures, data collection and processing of the Canadian Community Health Survey (CCHS).

Summary
Data collection for cycle 1.1 of the CCHS began in September 2000. This first cycle provides cross-sectional data at the regional level for 136 health regions; the first half of data collected for cycle 1.1 provides data for 133 health regions. In addition to the survey methods, this article reports the sample size and rates of proxy response and non-response for each province, for the first six months of cycle 1.1. A summary of methods used to impute values that were not provided by proxy respondents is provided. A discussion of survey errors and their sources follows.

Key words
health surveys, cross-sectional studies

Author
Yves Béland (613-951-1494; belayve@statcan.ca) is with the Household Survey Methods Division at Statistics Canada, Ottawa, Ontario, K1A 0T6.

The Canadian Community Health Survey (CCHS) is part of a recent federal initiative aimed at providing health information at the regional and provincial levels. The CCHS, for which data collection began in 2000, consists of two cross-sectional surveys conducted over a two-year, repeating cycle. The first survey, referred to as cycle 1.1, was designed to collect data from a sample large enough to provide information by health region. The second survey (cycle 1.2) will focus on a specific health topic, and will provide data at the provincial level.

This article describes the sampling strategy, data collection, data processing and sources of survey error in the CCHS. It supplements the other articles in this issue, each of which is based on analysis of data collected in the first half of cycle 1.1. Although the full cycle 1.1 data file will support analysis at the provincial and regional levels, the reports based on the first half of the data for cycle 1.1 provide preliminary findings at the national level only.

Target population

The CCHS targets individuals aged 12 or older who are living in private dwellings. People living on Indian reserves or Crown lands, residents of institutions, full-time members of the Canadian Armed Forces, and residents of certain remote regions are excluded. The CCHS covers approximately 98% of the Canadian population aged 12 or older. The three territories were not included in the analyses based on the preliminary file comprising the first half of data collected for cycle 1.1, because data collection in those areas began later than in the rest of the country.

Health regions

For administrative purposes, each province is divided into health regions (HR), and each territory is designated as a single HR (Table 1). Statistics Canada, in consultation with the provinces, has made minor changes to the boundaries of some of the HRs to correspond to the geography of the 1996 Census. Cycle 1.1 of the CCHS collects data in 133 HRs in the 10 provinces, in addition to one HR per territory, totalling 136 HRs.

Table 1
Number of health regions and targeted sample sizes, by province/territory, Canadian Community Health Survey, cycle 1.1

	Number of health regions	Sample size, first six months of data collection	Total sample size (projected)
Canada	**136**	**54,788**	**133,300**
Newfoundland	6	1,834	4,010
Prince Edward Island	2	909	2,000
Nova Scotia	6	2,158	5,040
New Brunswick	7	2,245	5,150
Québec	16	10,065	24,280
Ontario	37	16,508	42,260
Manitoba	11	3,823	8,000
Saskatchewan	11	3,702	7,720
Alberta	17	6,477	14,200
British Columbia	20	7,067	18,090
Yukon	1	0	850
Northwest Territories	1	0	900
Nunavut	1	0	800

Sample size and allocation

Although producing reliable estimates at the HR level is a primary objective of cycle 1.1, the quality of the estimates for certain key characteristics at the provincial level was also deemed important. Therefore, the sample allocation strategy, consisting of three steps, gives relatively equal importance to the HRs and the provinces. In the first two steps, the sample is allocated among the provinces and territories according to their respective populations and the number of HRs they contain (Table 1). In the third step, each province's sample is allocated among its HRs proportionally to the square root of the estimated population in each HR.

This three-step approach guarantees each HR sufficient sample with minimal disturbance to the provincial allocation of sample sizes. The sample sizes were enlarged before data collection to take into account out-of-scope and vacant dwellings and anticipated non-response. (For the complete list of updated HRs and projected sample sizes, see: http://www.statcan.ca/health_surveys.)

Frames, household sampling strategies

The CCHS uses the area frame designed for the Canadian Labour Force Survey as its primary sampling frame. A multistage stratified cluster design was used to sample dwellings within the area frame.[1] In the first stage of the design, a list of the dwellings was prepared. At the second stage, a sample of dwellings was selected from this list. The households in the selected dwellings then formed the sample of households. The majority (88%) of the targeted sample was selected from the area frame, and face-to-face interviews were held with respondents randomly selected from households in this frame.

In some HRs, a random digit dialling (RDD) sampling frame was also used.[2] The sampling of households from the RDD frame used the Elimination of Non-Working Banks method, a procedure adopted by Statistics Canada's General Social Survey.[3] A telephone bank (the area code plus the first five digits of a seven-digit telephone number) was considered as "working" for the purposes of sampling if it included at least one residential telephone number. The working banks were regrouped to create RDD strata to encompass, as closely as possible, the HR areas. Within each

RDD stratum, a bank was randomly chosen, and a number between 00 and 99 was generated at random to create a complete 10-digit telephone number. This was repeated until the required number of telephone numbers within the RDD stratum was reached. Respondents in the RDD frame, who accounted for the remaining 12% of the targeted sample, were interviewed by telephone.

Respondent sampling

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

In approximately 82% of the households selected from the area frame, one person aged 12 or older was randomly selected; two people (12 or older) were randomly chosen in the remaining households. Selection of respondents from households in the area frame depended on the household composition, and was intended to increase the representation of the two age groups of special interest in the sample: youths and seniors. For households selected from the RDD frame, one person aged 12 or older was randomly chosen from all household members. The sample design of the CCHS is described in a previous report.[4]

Data collection

Data collection for cycle 1.1 began in September 2000 and was conducted over 12 months. This helped balance interviewer workload and minimize seasonal effects on certain health-related characteristics such as physical activity. The sample of households was allocated randomly over the 12-month period, and every HR was visited in each collection month. The data for the first half of cycle 1.1 were collected between September 5, 2000 and March 2, 2001.

Questionnaire design

The CCHS questionnaire was designed for computer-assisted interviewing (CAI). Thus, as the questions were developed, the associated logical flow into and out of the questions was programmed. This included specifying the type of answer required, minimum and maximum response values, on-line edits associated with the question, and procedures for handling item-non-response. With CAI, the interview can be directed based on the respondent's answers. On-screen prompts appear when an invalid entry is recorded. In this way, immediate feedback is given to the interviewer so that inconsistencies can be addressed.

In the developmental phase of the CCHS, consultation was undertaken with users of health information across Canada. Data needs were identified, and the questionnaire content was developed. An important goal of the CCHS is to collect data on issues of specific relevance to the HRs. To achieve this goal, the questionnaire was divided into two parts—a common content section 35 minutes in length, and a 10-minute optional content section containing questions selected to meet the particular needs of each HR. This resulted in 27 different versions of the questionnaire. For a summary of the common and optional content, see Appendix Table A. The complete CCHS questionnaire is available on Statistics Canada's website at: http://www.statcan.ca/health_surveys.

Interviewing

At the initial contact, Statistics Canada field interviewers visited all dwellings selected from the area frame. An inventory was made of everyone residing in the household, and one household member was randomly selected to be the survey respondent. Interviewers were instructed to administer the questionnaire directly to the selected respondent whenever possible. The procedure was similar for the sampling units from the RDD frame, where interviewers made as many as eight attempts to contact a householder.

When the selected respondent was unavailable at the time of the visit or telephone call, interviewers were instructed to return at a later date or to attempt to reach the respondent by telephone. When a selected respondent remained unavailable after repeated contact attempts, interviewers requested that another resident of the household complete a

Table 2
Percentage of interviews provided by proxy respondents in first six months of data collection, by province, Canadian Community Health Survey, cycle 1.1, September 2000 to February 2001

	%
Canada	**7.6**
Newfoundland	6.3
Prince Edward Island	7.9
Nova Scotia	4.9
New Brunswick	11.0
Québec	5.9
Ontario	6.8
Manitoba	10.0
Saskatchewan	7.2
Alberta	8.1
British Columbia	9.5

proxy interview on behalf of the designated respondent. After the first six months of data collection for cycle 1.1, nearly 8% of all interviews had been obtained by proxy; the rate varied somewhat among the provinces (Table 2).

Non-response

CCHS interviewers were instructed to make all reasonable attempts to obtain interviews. Designated respondents who initially refused to be interviewed were contacted by a senior interviewer, who stressed the importance of the survey and the household's co-operation. Additional attempts were then made to schedule the interview at the respondent's convenience. To maximize response rates for the first half of cycle 1.1, many non-

Table 3
Non-response rate after six months of data collection, by province, Canadian Community Health Survey, cycle 1.1, September 2000 to February 2001

	%
Canada	**20.0**
Newfoundland	9.7
Prince Edward Island	13.2
Nova Scotia	18.2
New Brunswick	14.5
Québec	18.4
Ontario	23.9
Manitoba	13.6
Saskatchewan	15.3
Alberta	17.5
British Columbia	22.7

respondents were contacted again in March 2001 at the end of the six-month period of data collection and were encouraged to participate. Despite these efforts, some non-response remained (Table 3).

Data processing

Through the use of a computer-assisted interviewing (CAI) application, a great deal of editing is either precluded or performed as the data are collected. For example, the CAI application does not allow out-of-range values and controls flow errors. CAI ensures that questions that do not apply to a specific respondent are not asked. In other situations, warning messages are displayed. For instance, when contradictory information is entered, the interviewer is alerted. Then, depending on instructions specific to the question being asked, the interviewer may leave the response as given or request clarification. At the completion of data collection, some inconsistencies are removed during editing by Head Office staff.

Imputation for proxy respondents

Because of their private or sensitive nature, many CCHS questions are appropriate for self-response only, and are skipped when the questionnaire is answered by proxy respondents. During the first half of data collection for cycle 1.1, an unexpectedly high proportion of interviews were completed by proxy (Table 2). Consequently, important information was missing for the individuals represented in those interviews. Values for missing information in proxy interviews were imputed during data processing. This affected key variables in two articles in this issue: "Fruit and vegetable consumption" and "Community belonging and health."

Data for proxy interviews were imputed using the "nearest neighbour" imputation method. Within pre-defined imputation classes, another respondent with characteristics similar to the designated respondent was identified; this individual became the "donor" whose data were imputed to the designated respondent.

Imputation was used only to complete information for respondents whose interviews were provided by proxy respondents. Imputation was

not performed in cases of total non-response or when responses to individual questionnaire items were refused in non-proxy interviews. The issue of total non-response was addressed in the weighting strategy, and item-non-responses in non-proxy interviews were left as missing values on the data file.

Weighting

Each respondent was assigned a weight to represent his or her contribution to the total population. The weight was used to derive estimates for all characteristics surveyed. Taking into account the sample design, estimates were produced from the sample data by employing estimation techniques from survey sampling theory. Because the CCHS used two overlapping sampling frames with separate sample designs, two weighting strategies with various adjustments were processed side-by-side and integrated at a certain point with a dual-frame technique. The integrated weights were then calibrated to population projections using a one-dimensional post-stratification adjustment of 10 age/sex post-strata (that is, the age groups 12 to 19, 20 to 29, 30 to 44, 45 to 64 and 65 or older for each sex) within each province.

Before the integration and calibration adjustments, key factors determined the weighting strategy for the CCHS data collected during the first half of cycle 1.1. For the area frame sample units, these factors included:

- use of a stratified, multistage design, involving probability sampling proportional to size at all stages except the final stage, when systematic sampling of dwellings was used;
- monthly stabilization of sampled dwellings;
- use of only half of the full, pre-defined cycle 1.1 sample;
- household-level non-response;
- selection of one or two respondents, based on household composition;
- person-level non-response.

For the RDD frame sampling units, some of the determining factors were:

- use of simple random sampling of telephone numbers within working banks of each RDD stratum;

- use of only six monthly RDD samples;
- household-level non-response;
- households not included in the frame because of no telephone line;
- selection of only one person per household;
- person-level non-response.

A detailed account of the cycle 1.1 weighting strategy has been reported previously.[5]

Sampling error

The survey produces estimates based on information collected from a sample of individuals. Sampling error is the error attributed to studying a fraction of the population rather than carrying out a complete census under the same general conditions (questionnaire, interviewers, processing methods, etc.). The extent of this error depends on factors such as sample size, the variability of the characteristic of interest, sample design and estimation method. Because of the complexity of the sample design, sampling error for CCHS estimates was calculated using the bootstrap resampling technique.

Non-sampling error

Errors not related to sampling are called non-sampling errors; these errors can arise during any survey activity. For example, interviewers may misunderstand instructions about questionnaire administration, respondents may give erroneous answers, responses may be incorrectly recorded, and errors may be introduced in data processing. Over a large number of observations, randomly occurring non-sampling errors will have little effect on overall estimates derived from a survey. However, errors that occur systematically will contribute to biased estimates.

Considerable time and effort have been expended to reduce non-sampling error in the CCHS. Extensive training of interviewers with respect to survey procedures and questionnaire content, use of skilled interviewers for follow-up of non-respondents, monitoring interviewers to detect problems, and quality assurance protocols were among the measures implemented to minimize non-sampling error.

Non-response

Non-response is a major source of non-sampling error. The extent of non-response ranges from item-non-response (failure to answer single questions) to total non-response. Partial non-response to the CCHS was rare; once an interview was started, the questionnaire was usually completed with very little item-non-response. Total non-response occurred either because a respondent refused to participate in the survey, or because the interviewer was unable to contact the selected respondent. After the first half of data collection for cycle 1.1, non-response varied considerably by province (Table 3).

Concluding remarks

The CCHS has been designed to study differences in health among sub-provincial units. Using the data from the CCHS, policy-makers and health care professionals will be able to identify benchmarks and track the progress of health programs within their regions. As well, the CCHS will allow the study of subgroups in the population with special needs, including seniors, home care recipients and single mothers. ●

References ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●

1 Statistics Canada. *Methodology of the Canadian Labour Force Survey* (Catalogue 71-526-XPB) Ottawa: Minister of Industry, 1998.

2 Morano M, Lessard S, Béland Y. Creation of a dual frame for the Canadian Community Health Survey. *Proceedings of the Statistical Society of Canada Annual Meeting, Survey Methods Section, 2000*. Ottawa: Statistical Society of Canada, 2001.

3 Norris DA, Paton DG. Canada's General Social Survey: Five years of experience. *Survey Methodology* (Statistics Canada, Catalogue 12-001) 1991; 17: 227-40.

4 Béland, Y, Bailie L, Catlin G, et al. CCHS and NPHS – An improved health survey program at Statistics Canada. *Proceedings of the American Statistical Association Meeting, Survey Research Methods, 2000*. Indianapolis: American Statistical Association, 2000.

5 Brisebois F, Thivierge S. The weighting strategy for the Canadian Community Health Survey, 2001. *Proceedings of the American Statistical Association Meeting, Survey Research Methods, 2001*. In press.

Appendix

Table A
Summary of common and optional content, by number of participating health regions, Canadian Community Health Survey, cycle 1.1

Common content	Optional content	
Item	**Item**	**Number of health regions (out of 136)**
Alcohol	Breast examinations (clinical)	53
Alcohol dependence/abuse	Breast self-examination	69
Blood pressure check	Changes made to improve health	100
Breastfeeding	Dental visits	56
Chronic conditions	Depression	134
Contacts with mental health professionals	Distress	24
Exposure to second-hand smoke	Driving under influence	103
Food insecurity	Drug use	29
Fruit and vegetable consumption	Eye examinations	2
General health	Flu shots	57
Health care utilization	Home care	41
Health Utilities Index (HUI)	Mastery	55
Height / Weight	Mood	14
Injuries	Physical check-up	4
Mammography	Sedentary activities	68
PAP test	Self-esteem	45
Physical activities	Sexual behaviours	57
PSA test	Smoking-cessation aids	56
Restriction of activities	Social support	86
Smoking	Spirituality	8
Tobacco alternatives	Suicidal thoughts and attempts	70
Two-week disability	Use of protective equipment	59
Administration	Work stress	97
Household record variables		