CCHS and NPHS — An improved Health Survey Program at Statistics Canada

Yves Béland, Lorna Bailie, Gary Catlin, M.P. Singh
Yves Béland, Statistics Canada, Tunney's Pasture, Ottawa, Ontario, Canada, K1A 0T6, yves.beland@statcan.ca

Key words: Cross-sectional, longitudinal, regional survey, dual frame.

1. Introduction

The need for health information at all levels is more critical than ever before. Policy makers and health care professionals are concerned with the increased demand anticipated for health care services due to the ageing population. With continual advancements in the effectiveness of treatments, balancing costs and potential enhanced health outcomes will be a challenge. Planners of health-care delivery need basic information in order to better anticipate demands and evaluate services provided.

The current National Population Health Survey (NPHS) is a longitudinal household survey conducted by Statistics Canada, following a panel of approximately 17,000 people every two years for up to twenty years. The objective of this multipurpose survey, which was launched in 1994, is to measure the health of Canadians and the determinants of health over time. General health and sociodemographic information are collected for all members in the household while detailed health information is collected for the selected longitudinal panel member. The main goal of the NPHS has been not only to produce longitudinal health data for the panel members but also to provide users with reliable cross-sectional health data for all household members. (For complete details on the NPHS design, see Tambay and Catlin, 1995.)

Although the NPHS is an excellent source of information, it does not sample enough individuals to signal differences in health status between smaller subprovincial areas. Moreover, the studying of special populations such as seniors, visible minorities, single mothers, users of home care, shift workers, individuals with specific chronic conditions, etc. is very limited without sufficient sample. Thus, the Canadian Community Health Survey (CCHS) is currently being designed to produce cross-sectional estimates to address priority health data gaps at national, provincial and regional levels. As a result of extensive consultations with provincial and regional authorities, it is proposed that the CCHS comprise two cross-sectional surveys conducted over a two-year cycle and that the NPHS be strictly a longitudinal survey.

This paper describes the two-year cycle plan for the CCHS and its integration with the existing NPHS in order to build a more robust health survey program. Section 2 summarizes the high-level plan that initiated these changes. Section 3 outlines the results of the Canada-wide consultation process with various stakeholders which established the main requirements for the methodological developments of the CCHS. Section 4 describes the integrated NPHS/CCHS health survey program. Details on the sample design of the CCHS are provided in section 5. Finally, areas for future developments of the Statistics Canada health survey program are given in section 6.

2. Health Information Roadmap Initiative

In response to a need for better information on health, in 1998 the Federal Health Minister's Advisory Council on Health Infostructure, the Canadian Institute for Health Information (CIHI) and Statistics Canada brought together over 500 people including health administrators, researchers, caregivers, government officials, health advocacy groups, and consumers to discuss issues and elaborate a strategy for health information. Consultations led to the creation of the *Health Information Roadmap* Initiative (CIHI; 1999a, 1999b).

The Roadmap outlines a shared vision for modernizing health information in Canada. Priorities have been identified for immediate investment. There are areas where initiatives across the provinces and territories have the potential to yield significant benefits for individuals and organizations across the country were highlighted.

The Federal government's response was to announce a substantial investment in health information in its 1999 budget. The investment included a commitment to fund the first three to four years of the *Roadmap*. Funding will support a range of projects that will be led by CIHI and Statistics Canada. Of the *Roadmap* projects, the largest is the new CCHS.

3. Addressing Needs through Consultation

The key to success in providing relevant information to clients is consultation. Clearly, the importance of determining and meeting the information needs of clients is directly related to the success of this initiative. Identifying priority items for inclusion for such a major survey program is not, however, an easy

task. Where the CCHS is concerned, consultation has been extensive. As for the development of the NPHS in the early '90s, the CCHS is currently being designed with the advice and expertise of an external advisory committee, consisting of representatives from Health Canada, and each provincial and territorial ministry of health as well as several representatives of subprovincial health regions.

In addition to help provided by the advisory committee members, a cross-Canada consultation process with users of health information was undertaken in June 1999, in the form of workshops. Consultation participants included data users at the municipal, health-region, provincial health ministry and federal levels, as well as university and hospital researchers and special interest or agency groups. Among the key objectives identified in the consultations were:

- measure health factors at many levels of geography and for key subpopulations;
- collect data on issues relevant to health regions;
- respond quickly to emerging issues; and,
- explore key health issues in-depth.

4. Plan for an Improved Health Survey Program

To meet the stated needs of users and keeping in mind the existing infrastructure at Statistics Canada, implementing a biennial cycle of data collection for the new CCHS was a logical choice. In fact, satisfying the need to produce i) estimates for a wide range of topics for 136 health regions and ii) estimates for a thematic subject on a provincial scale requires a sample of such a size that the data collection would be very difficult to manage operationally over a shorter period of time.

Two CCHS survey design scenarios were seriously studied from among many considered. One scenario recommended a sample size of 160,000 respondents for a health region-level survey conducted over a period of two years, and a province-level supplement administered to a subsample of 30,000 of the 160,000 respondents. The other scenario called for the development of two separate survey components: a survey at the health region-level the first year with a total sample of more than 130,000 respondents and a province-level survey in the second year with a sample of 30,000 respondents. In both scenarios, the health region-level survey would have varied content adapted to regional needs whereas the provincial level survey would deal with a specific theme in-depth.

Criteria such as relevance, accuracy, coherence, accessibility and interpretability, and especially timeliness led to the current design. These are the elements of Statistics Canada guidelines regarding data quality (Statistics Canada, 1998a). The CCHS biennial cycle consists, therefore, of the following two survey components: a health region-level survey the first year called the *regional component* and a province-level survey in the second referred to as the *provincial component.*

With the implementation of the CCHS, the cross-sectional objective of the NPHS is no longer a necessity. Thus starting with the fourth wave of data collection, which is currently in the field until June 2001, the NPHS becomes strictly longitudinal. Only the selected panel members are contacted to collect detailed health information for longitudinal estimates.

5. Sample Design for the CCHS

The primary objective of the regional component, which started collection in September 2000, is to provide reliable cross-sectional estimates at the health-region level in respect of health determinants, health status and utilization of the health system for 136 health regions. (Note however that, for simplicity, this section refers only to the design used in the ten provinces. The design of the CCHS in the three territories is different.)

The provincial component of this first occurrence of the CCHS is scheduled to begin collection in January 2002. Its main purpose will be to produce cross-sectional estimates at the provincial level on both positive aspects of mental health and the prevalence, determinants and related demographic features of mental disorders in Canada. What follows is a detailed description of the sample design for the regional component of the CCHS as well as an overview of the provincial component.

5.1 Sample Design of the Regional Component

5.1.1 Creation of the Health Regions

The health regions (HR) are geographical regions that the provinces use for administrative purposes in the field of health. For practical purposes, Statistics Canada, in agreement with the provinces, have revised the boundaries of some of the HRs, according to the geography of the 1996 Census, to allow the production of demographic projections for different age/sex groups. For statistical purposes, 133 HRs distributed in the ten provinces are considered for the regional component.

5.1.2 Target Population

The regional component of the CCHS targets persons living in private occupied dwellings who are aged 12 or older. Persons living on Indian Reserves and on Crown lands, residents of institutions, full-time

members of the Canadian Armed Forces and residents of certain remote regions are excluded from this survey. The CCHS covers approximately 97% of the Canadian population that is 12 years old or older. It should be noted that the National Population Health Survey (NPHS) already has an institutional component that deals with long-term residents in hospitals and resident care institutions (Statistics Canada, 1996).

5.1.3 Sample Size and Allocation

To provide reliable estimates to the 133 HRs, and given the budget allocated to the CCHS regional component, a sample of 130,750 is desired. Although producing reliable estimates at the HR level is a primary objective, the quality of the estimates at the provincial level for certain key characteristics is also an objective. Various scenarios have been considered for distributing this sample between the provinces and the HRs. As the size and number of HRs vary considerably from one province to the other, it is difficult to establish an equilibrium between regional and provincial needs. An approach focused primarily on the HRs penalizes heavily populated provinces that have a relatively small number of HRs, whereas an approach focused on the provinces produces the opposite effect. The strategy that has been adopted consists of three steps, which gives relatively equal importance to the HRs and to the provinces. The two first steps allocate the sample between the provinces as a function of their respective populations and of the number of HRs they contain. At the third step, the sample is distributed among the HRs of each province.

At the first step, approximately half the total sample of 130,750 respondents is allocated among the provinces as a function of the number of HRs that they have. As the primary objective of the CCHS is to provide reliable regional (HR) estimates, some 500 sample units are assigned to each HR. The only exception are the HRs in which the resulting sampling fraction is too high, in which case the sampling rate is capped at one household in twenty. A total of 65,830 sampling units have thus been allocated in this first step. At the second step, the rest of the total sample (64,920 units) is distributed among the provinces in proportion to the size of their populations. There is one exception to this: a minimum of 1,000 units are assigned to Prince Edward Island. The total sample size of any given province is found by adding the sizes obtained in the first two steps. Table 1 provides a detailed distribution of the regional component sample by province.

Table 1 – Provincial Sample Sizes

Prov.	Size ('000)	# of HRs	1st step 500/HR	2nd step X-prop.	Total Sample
Nfld.	551	6	2,780	1,230	4,010
P.E.I.	135	2	1,000	1,000	2,000
N.S.	909	6	3,000	2,040	5,040
N.B.	738	7	3,500	1,650	5,150
Que.	7,139	16	8,000	16,280	24,280
Ont.	10,714	37	18,500	23,760	42,260
Man.	1,114	11	5,500	2,500	8,000
Sask.	990	11	5,400	2,320	7,720
Alta.	2,697	17	8,150	6,050	14,200
B.C.	3,725	20	10,000	8,090	18,090
Canada	29,000	133	65,830	64,920	130,750

The third and last step consists of allocating each provincial sample among the provincial HRs proportionally to the root of the size of the population in the regions. This three-step approach guarantees each HR sufficient sample with minimal disturbance to the provincial allocation of sample sizes. Table 2 summarizes the distribution of HRs by population size as well as the average sample size planned by HR category.

Table 2 – Mean Sample Sizes by Category of HRs

Category of HRs	Population Size	# of HRs	Mean Samp. Size
Small	Less than 75,000	41	530
Medium	75,000 - 240,000	60	900
Large	240,000 – 640,000	25	1,500
Very Large	640,000 and more	7	2,500

The final sample size for the regional component is 130,750 respondents, despite the opportunity that the HRs and the provincial authorities had to increase the sample sizes by buying supplementary units. None chose to do so. The allocated sample sizes were considered satisfactory. These sample sizes were boosted before data collection to take into account out-of-scope dwellings, vacant dwellings and anticipated household level non-response.

5.1.4 Sampling Frame

Statistics Canada has several sampling frames used by various household surveys that could be useful for a health survey. The area frame used for the Canadian Labour Force Survey (LFS) is the preferred frame for household surveys. With this frame, it is possible to select either a sample of new dwellings or a sample of dwellings expired from the LFS. (It is worth noting that, for the LFS, the dwellings selected remain in the

sample for six consecutive months and is then replaced. These rotated-out dwellings are then available for other surveys.) On the other hand, Statistics Canada's Address Register (AR), the Random Digit Dialling (RDD) frame of telephone numbers, and provincial health administration files are also potential bases.

The choice of a sampling frame depends on many factors but, first and foremost, the frame must correspond as much as possible to the target population of the survey. Moreover, the creation, utilization, updating and verification of the sampling frame must fall within the operational and budget constraints of the survey. For the regional component, it was decided to use two overlapping sampling frames, the area frame established for the LFS and the RDD frame of telephone numbers. Within the regional component, the area frame is used as the primary frame whereas the RDD frame is complementary to the primary frame in certain HRs. (Morano, Lessard and Béland, 2000)

Apart from the fact that the target population is the same as that of the LFS, the advantages of using the area frame set-up for the LFS for selecting the sample are undeniable. The infrastructure, which is already in place for updating new buildings, demolished buildings and excluded units, as well as the entire evaluation process of the frame coverage, are definite assets. Moreover, given that several other Statistics Canada household surveys also use this area frame, it is easier to control sample overlap between surveys. For various reasons, but primarily in consideration of response burden, it was decided to choose a sample of new dwellings and not a sample of rotated-out dwellings from the LFS.

The limitations to using the RDD frame are also evident: under-representation of households without telephones (~2%) or with cellular phones only (estimated between 1% and 2%), the generally lower response rate and the need to make several calls before contacting a valid household. A dual frame approach is necessary for the following reasons: i) the high cost of face-to-face data collection in certain areas; ii) the inability of the area frame to provide the required sample for certain HRs; and iii) the desire for a permanent and flexible infrastructure for collecting data by telephone.

5.1.5 Sampling Strategy

Among the targeted sample of 130,750 respondents, the majority (115,000 respondents) comes from the area frame in which personal face-to-face interviews are held, with the rest coming from the RDD frame for telephone interviews.

5.1.5.1 Sampling of Households

The area frame, as designed for the LFS, covers almost the entire country, from which a sample of dwellings is selected under a multistage stratified cluster design (Statistics Canada, 1998b). For those areas selected in the first stage of the design, a list of dwellings is prepared and maintained in the field. A sample of dwellings is then selected at the second stage from each list. The households in the selected dwellings then form the sample of households. To get a base sample of 97,000 households, 123,000 dwellings are selected from the area frame to account for vacant dwellings and non-response.

The sampling of households from the RDD frame uses the Elimination of Non-Working Banks (ENWB) method, a procedure adopted by the General Social Survey (Norris and Paton, 1991). A hundreds bank (the area code plus the first five digits of a seven-digit phone number) is considered as "working" for the purposes of sampling if it includes at least one residential telephone number. The working banks are regrouped to create RDD strata, in order to encompass as closely as possible the HR areas. Within a RDD stratum, a bank is randomly chosen at random and a number between 00 and 99 is generated at random to create a complete, ten-digit telephone number. This procedure is repeated until the required number of telephone numbers within the RDD stratum is reached. Frequently, the number generated is out of scope, that is, excluded from the survey or not in service. To reach the objectives of the required sample size, many additional telephone numbers must be generated. Within the framework of the regional component, the success rate varies between 15% and 61% in the HRs for which a sample of households is required from the RDD frame. Overall, a sample of more than 51,000 telephone numbers, therefore, is required (taking into account in addition non-responding households) for the purposes of the survey in order to get the required 15,750 responding households.

5.1.5.2 Sampling of Persons

During the pan-Canadian consultation process, regional and provincial representatives expressed their concern regarding the necessity to increase the representativeness of the sample for youth and older persons to allow for more in-depth analyses of these two important subgroups of the population. With that concern in mind, various options for defining the rules for selecting persons within a household have been studied.

Interviewing more than one person in a same household allows for economies in the cost of collection, since a large part of these costs are

attributable to the process required to reach the household. The incremental costs to interview additional persons in the household are small. However, strong similarities observed among members of the same household can lead to an undesired cluster effect for certain important survey characteristics (one of the key components of the survey is the utilization of the health system). Moreover, the response burden of the household is increased.

On the other hand, selecting only one person per household represents a significant increase in collection costs, since a greater number of households must be sampled. Also, as the chances of being part of a sample are inversely proportional to the number of persons in the household, certain age groups are either under- or over-represented. In particular, selecting only one person per household underrepresents persons coming from large households, typically parents and children, and overrepresents persons coming from small households, often single people and the elderly. Table 3 compares the age group distributions of the targeted population from the 1996 Census (column 2), the LFS sample where every household member is selected (column 3) and a simulated CCHS sample (regional component) where only one person per household is selected (column 4).

Table 3 – Percent Distribution by Age Group

Age group	1996 Census	LFS sample	* CCHS Sample
Col.1	Col. 2	Col. 3	Col. 4
12-19	13.2	13.7	8.5
20-29	16.4	14.4	14.3
30-44	30.8	28.7	29.1
45-64	25.8	28.0	27.9
65+	13.8	15.2	20.2

* Average distribution over 100 repetitions

The results clearly demonstrate that, by selecting only one person per household, the 12-to-19 age group is greatly under-represented while old persons are over-represented. For reasons of user needs, cost, the efficiency of the design, the response burden, and operational constraints, the rule for selecting persons implemented for this survey is a compromise.

For all households from the RDD frame, a single person aged 12 or older is chosen from among all members of the household. For the households from the area frame, one person aged 12 or older is selected from among 79,000 sampled households and two persons (12 or older) are chosen in the remaining 18,000 households. The rule for selecting persons from households in the area frame is defined as a function of the household composition, its purpose being to increase the representativeness of 12 to 19-year-olds in the sample. The table 4 describes the rule for selecting persons within the area frame sampled households.

Table 4 – Selection Strategy based on Household Composition – Area Frame Sample only

Number of 12-19	Number of persons aged 20 or over					
	0	1	2	3	4	5+
0	-	A	A	A	A	B
1	A	A	C	C	C	B
2	A	C	C	C	C	C
3+	A	C	C	C	C	C

A: random selection of one person aged 12 or over
B: random selection of two persons aged 12 or over
C: random selection of one person in the age group 12-19 **and** random selection of one person aged 20 or over

Table 5 shows the expected distribution of a simulated CCHS sample when selecting two persons (using the rule described in Table 4) in 18,000 households and one person, at random, in all the other households selected. The results show that the two subgroups of interest (12 to 19 and 65 or older) are now over-represented in the sample without overly penalizing the other age groups.

Table 5 – Expected CCHS Sample Distribution by Age Group

Age group	1996 Census	* CCHS Simulated Sample
12-19	13.2	14.9
20-29	16.4	13.1
30-44	30.8	28.1
45-64	25.8	26.3
65+	13.8	17.6

* Average distribution over 100 repetitions

5.1.6 Weighting and Estimation

The combination of the health region-level sample design and the data collection strategy will allow for the release of preliminary estimates over the course of the year. Sample weights will be computed after three, six, nine and twelve months of collection. Because the sample is obtained using a dual frame with separate sample designs, two separate weighting systems with various adjustments (including an adjustment for total non-response) will take place. The two weighting systems will be integrated using a dual-frame technique. Although a final decision has not been made, a dual-frame technique that will take into

account both the design effect and the effective sample size of each design is most likely to be adopted. The integrated weights will be calibrated using a one-dimensional poststratification adjustment of ten age/sex poststrata (i.e., 12-19, 20-29, 30-44, 45-64 and 65+ crossed with the two sexes) within each HR.

5.1.7 Questionnaire Design

As mentioned earlier, one key goal that came out of the consultation process was to collect data on issues unique to HRs. In order to achieve that goal within the maximum 45-minute interview length planned for this survey, a strategy that is innovative and unique to the CCHS was implemented (Bailie, Béland, Diverty and Ingram, 2000).

The development of the final questionnaire was divided into two parts – a common content section of 35 minutes in length and an optional content section of 10 minutes customized to the HR needs. Provinces and HRs were provided with a choice of 28 questionnaire modules to choose from for their HRs. This process has resulted in 27 different versions of the questionnaire. For a copy of all questions on the final questionnaires the reader is referred to http://www.statcan.ca/health_surveys.

5.2 Sample Design of the Provincial Component

The primary objective of this provincial component of the CCHS is to produce cross-sectional estimates on the different aspects of the mental health and well-being of Canadians at both provincial and national levels from a sample of 30,000 respondents aged 15 and over. Moreover, the survey will collect data on both positive and negative factors affecting mental health, the utilisation of mental health care services, and data on social impacts and the costs associated with mental health. All this will be rounded out with the collection of a number of social and demographic characteristics. The data collection is expected to begin in January 2002 and will extend over 12 months.

Both questionnaire content and sample design for this provincial component of the CCHS are still under development at the time of writing this paper.

6. Future Directions

The issues around health and health care are well known and are expected to remain a concern for years to come. The addition of the CCHS biennial cycle to the existing NPHS makes the Statistics Canada health survey program comprehensive and robust. It is an ambitious program where efforts from several parties are essential but the benefits for policy makers and health care professionals should be impressive.

Although the budget for future cycles of the CCHS is still under review, plans call for a repetition of the CCHS biennial cycle using a similar approach. A regional component with a large sample size will be conducted in the calendar year 2003 followed by a provincial component in 2004. The in-depth topics of interest proposed for the provincial component are nutrition and physical measures. In parallel, the NPHS will continue its data collection every two years by following the selected panel members and hence provide valuable longitudinal information for researchers. The NPHS will be enhanced in its sixth wave (2004/05) with a second panel of respondents and revised content to take into account the changing needs.

Using the data from this health survey program, policy makers and health care professionals will be able to identify benchmarks and track progress. This should hence help them taking appropriate action towards addressing the shortcomings of the health care system in Canada.

7. References

Bailie, L., Béland, Y., Diverty, B. and Ingram, S. (2000). Population health survey initiatives: Investing in health surveys, investing in health. *Health Reports,* Statistics Canada Cat. No. 82-003, Vol. 12 (3), To be published.

Canadian Institute for Health Information (1999a), *Health Information Roadmap: Beginning the Journey.* (1-895581-32-X).

Canadian Institute for Health Information (1999b), *Health Information Roadmap: Responding to Needs.* (1-895581-30-3).

Morano M., Lessard, S. and Béland, Y. (2000). Creation of a dual frame for the Canadian Community Health Survey, *2000 Proceedings of the Survey Methods* Section, Statistical Society of Canada. To be published.

Norris, D.A. and Paton, D.G. (1991). Canada's General Social Survey: Five years of experience, *Survey Methodology*, 17, 227-240.

Statistics Canada (1996). *National Population Health Survey – Institutional Component 1994-95, PUMF,* Statistics Canada Cat. No. 82M0010XDB.

Statistics Canada (1998a). *Quality Guidelines.* Third Edition, Statistics Canada. Cat. No. 12-539-X1E.

Statistics Canada (1998b). *Methodology of the Canadian Labour Force Survey.* Statistics Canada. Cat. No. 71-526-XPB.

Tambay, J.L. and Catlin, G. (1995). Sample design of the National Population Health Survey. *Health Reports,* Statistics Canada. Cat. No. 82-003, Vol. 7 (1), 33-42.