



Component of Statistics Canada
Catalogue no. 82-003-S Health Reports

Article

Canadian Health Measures Survey: Sampling strategy overview

by Suzelle Giroux

December 5, 2007



Canada

Canadian Health Measures Survey: Sampling strategy overview

Suzelle Giroux

Abstract

This article presents an overview of the sampling strategy developed for the Canadian Health Measures Survey (CHMS). The CHMS was designed to collect key health information using a computer-assisted personal interview and a physical health examination. It began in March 2007 and will be conducted to 2009. A nationally representative sample of approximately 5,000 respondents aged 6 to 79 will be interviewed at home and asked to visit a mobile clinic where health care professionals measure several aspects of their physical health. The CHMS presents several challenges including: the need to have respondents who live within a reasonable commuting distance of the clinics; the difficulty of reaching the desired sample size for youths; and sub-sampling of measures related to exposure to environmental pollutants. The sampling strategy described in this article will address these challenges.

Keywords

area-frame, collection sites, cross-sectional studies, direct measures, health surveys, multi-stage sample, rejection method, sub-sampling, vector of selection, probabilities

Author

Suzelle Giroux (613-951-1481; Suzelle.Giroux@statcan.ca) is with the Household Survey Methods Division at Statistics Canada, Ottawa, Ontario, K1A 0T6.

Through personal interviews and physical health examinations, the Canadian Health Measures Survey (CHMS) will collect information from Canadians about their general health and lifestyle. The results of this voluntary survey will be used to estimate the number of people with selected diseases and risk factors.

The CHMS questionnaire will be used to gather information about medical history, diet, smoking habits, alcohol use, current health status and sexual behaviour, as well as demographic and socioeconomic characteristics. In addition, information will be collected in the form of direct physical measurements such as height and weight, blood pressure, blood and urine samples, and physical fitness testing. The results will create national baseline data on the prevalence of health concerns such as obesity, hypertension, cardiovascular disease and exposure to infectious diseases and environmental contaminants. The CHMS is one of the few Statistics Canada surveys to collect direct measurements.

This overview describes the sampling strategy used to meet the collection and estimation requirements of the CHMS. Other articles in this publication describe the survey's background,¹ the ethical, legal and social issues surrounding the CHMS,² the logistical and operational challenges,³ and the pre-test results.⁴

Survey design

Data collection for the CHMS will take place over two years and is done in two steps. First, an interviewer visits the respondent's home to administer the household questionnaire. Second, respondents are asked to attend an appointment at a nearby CHMS mobile clinic. At the clinic, trained health professionals take some physical health measurements. Respondents are also asked to wear a physical activity monitor (accelerometer) for a week to measure their activity levels.

The CHMS mobile clinic is stationed at each site for six to seven weeks. In determining collection sites, consideration must be given to the location of respondents' dwellings to ensure that selected dwellings are within a reasonable travelling distance from the clinic.³

The CHMS targets individuals aged 6 to 79 years who live in private households. People living on Indian reserves or Crown lands, residents of institutions, full-time members of the Canadian Forces and residents of certain remote regions are excluded.

Sample size

To meet the objective of providing national baseline estimates on a variety of health indicators, the CHMS requires 5,000 respondents equally distributed by age group (6 to 11, 12 to 19, 20 to 39, 40 to 59, and 60 to 79 years) and sex, for a total of 10 groups. Before the survey moves into the field, this number is inflated to take out-of-scope dwellings and non-response into account. This sample size should yield national level estimates by sex for each of the 5 age groups for conditions that have a prevalence of 10% or higher, with a coefficient of variation of the estimate of 16.5%.

Sub-samples of the survey's respondents are also required for laboratory analyses of environmental chemicals in the blood or urine.

Area frame—creation, allocation and selection of sites

Because the CHMS requires that respondents report to a clinic, they should be able to travel to that clinic within a reasonable period of time. The Labour Force Survey (LFS) area frame was used to design and control the size of collection sites in order to accommodate these requirements.

Using the LFS frame clusters,⁵ a total of 257 collection sites were created.⁶ Clusters are small geographic units that contain approximately 200 dwellings. A collection site is a geographic area with a population of at least 10,000 and a maximum respondent travel distance of 100 kilometres (km) (50 km in an urban area and 100 km in rural areas). Areas not meeting these criteria were excluded. The sites cover 96.3 % of the Canadian population.⁷

Although only national estimates were required, the collection sites were stratified in five regions to ensure a representative distribution of the sample across the country. Statistics Canada's standard regional boundaries were used: British Columbia, the Prairies (Alberta, Manitoba and Saskatchewan), Ontario, Quebec, and the Atlantic provinces (Newfoundland and Labrador, Prince Edward Island, Nova Scotia and New Brunswick).

A large number of collection sites with few respondents is recommended because it helps optimize the precision of the estimates.⁸ However, the logistical and cost constraints associated with the use of mobile clinics restricted the number of collection sites to 15. Each site will collect data on approximately 350 respondents, on average, for a total of roughly 5,000 respondents. The 15 collection sites were allocated to the regions in proportion to the size of the population (Table 1).

Within each region, the collection sites were sorted according to whether they belonged to a census metropolitan area (CMA), and then by the size of the population before the selection. A CMA is an area consisting of one or more adjacent

Table 1
Selection of Canadian Health Measures Survey collection sites, by region

Region	Estimated target population, aged 6 to 79, 2001 Census	Sites in region	Allocated sites number	Sample
Total	26,949,315	257	15	4,995
Atlantic (Newfoundland and Labrador, Prince Edward Island, Nova Scotia, New Brunswick)	2,061,425	36	1	333
Quebec	6,560,375	50	4	1,332
Ontario	10,248,545	61	6	1,998
Prairies (Alberta, Saskatchewan, Manitoba; includes Yellowknife)	4,538,970	77	2	666
British Columbia (includes Whitehorse)	3,540,000	33	2	666

municipalities centered on a large urban area (known as the urban core). The urban core must have a population of at least 100,000 to form a CMA. The sites were then sampled systematically with a probability of selection proportional to the size of their population. This method of selection, combined with the sorting of sites by CMA and non-CMA and by population size, ensured that the selected sites would be distributed among CMA and non-CMA areas and among areas with larger and smaller populations.

While not every province/territory will have a collection site, the CHMS sites were chosen to represent the Canadian population, east to west, with larger and smaller population densities (see *Canadian Health Measures Survey collection sites*).

Dwelling sampling

Several options were examined to determine how best to obtain the required number of respondents by age group. One option proposed using dwellings with known household composition from the 2005 Canadian Community Health Survey (CCHS cycle 3.1) and a fresh sample of dwellings with unknown household composition from the Labour Force Survey frame.⁹ Although this approach met the objectives of reaching the target number of respondents by age group, a random rejection method¹⁰ for households was required to avoid interviewing too many people in the age groups that constitute relatively large shares of the population (20 to 59 years) before reaching the target number of respondents for the other age groups.

Another option was using the 2006 Census as a frame. The household composition of dwellings as of May 2006 was available and could be used to develop a design to meet the sample requirements more efficiently in each age group. This option was chosen.

Within each site, dwellings with known household composition at the time of the 2006 Census are stratified by age at the time of the survey. Five age-

Canadian Health Measures Survey collection sites

Atlantic

- Moncton area, New Brunswick

Quebec

- Québec City area
- Montréal area (two sites)
- South of Mauricie area

Ontario

- Oshawa area
- Toronto area (two sites)
- St. Catharine's-Niagara area
- Kitchener-Waterloo area
- Northumberland County area

Prairies

- Edmonton area, Alberta
- Red Deer area, Alberta

British Columbia

- Vancouver area
- Williams Lake and Quesnel area

group strata corresponding to the five CHMS age groups (6 to 11, 12 to 19, 20 to 39, 40 to 59, and 60 to 79 years) were created. Age is determined based on the starting date of data collection at each site.

- Stratum 1: dwellings where at least one 6- to 11-year-old is present, else,
- Stratum 2: dwellings where at least one 12- to 19-year-old is present, else,
- Stratum 3: dwellings where at least one 60- to 79-year-old is present, else,
- Stratum 4: dwellings where at least one 20- to 39-year-old is present, else,
- Stratum 5: dwellings where at least one 40- to 59-year-old is present, else,
- Stratum with dwellings not falling in the above strata, such as vacant dwellings at the time of the Census or dwellings with people outside the CHMS age range based on household composition at the time of Census.

Each stratum has a high probability of having dwellings with respondents in the desired age groups. Within each site, a simple random sample of dwellings is selected in each stratum. The sample is allocated to each stratum so that, combined with the respondent sampling strategy, an equal number of respondents by age group can be obtained. No random rejection of dwellings is expected at the beginning of the survey. Towards the end of the two-year collection, however, the 2006 Census strata may be less efficient in reaching specific age groups, and random rejection of households with predominant age groups in the 20- to 59-year-old population may be necessary.

Each selected dwelling is contacted and asked to provide a list of current household members, and this list is used to select survey participants.

Because CHMS data collection continues until 2009, other sources may be used to supplement the addresses provided at the time of the 2006 Census to compensate for new or missed dwellings and to reduce undercoverage. Possible alternative sources are the Address Register or the Labour Force Survey Frame.⁵

Respondent sampling

Different selection probabilities by age group within each stratum are used to ensure that the sampling targets are attained. The dwellings selected are contacted and asked to provide a list of current household members. One or two people are selected, depending on the household composition. Because children have to be accompanied to the clinic, two people are selected from households with children aged 6 to 11: one child randomly selected among those aged 6 to 11, and a second person aged 12 to 79. If no 6- to 11-year-olds live in the household, only one person is selected among the household members aged 12 to 79 years.

The weight vector for the selection of people aged 12 to 79 years has been designed to avoid large person sampling weights. Since some age groups have a weight that is twice that of the other age groups, it is possible that a selected person would have a very high sampling weight when there are many household members in a dwelling. Hence, when a specified minimum number of people aged 12 to 79 live in a household, the weight for each person is reset to 1. In such cases, each household member has an equal chance of being selected.

A careful balance of the parameters required for each of the measures put in place was obtained through studies and simulations. It is expected that the target number of respondents ($n=1,000$) by age group will be reached.

Sub-sampling

Sub-samples of the survey's respondents are also selected for laboratory analysis of specific environmental chemicals in their blood or urine.^{1,3} The entire sample population is not used because of the high cost of the laboratory tests.

The sub-samples are selected independently; that is, without consideration of who was selected for testing for the other groups of environmental chemicals. This means that a specific respondent can be selected for testing of one, two, or all environmental chemicals.

However, two people in the same household should not be selected for the same environmental chemical. It is believed that environmental chemicals will affect people within the same household in the same way. In order to avoid obtaining more or less the same information, two members of one household should not be selected for the same environmental chemical. Two people might be selected in households where sub-samples are required for both children aged 6 to 11 and other age groups. A collocated sampling method has been used to minimize or prevent this situation by using different selection intervals for 6- to 11-year-olds and the other age groups.

Other considerations

Each sampled dwelling is randomly flagged to indicate whether a respondent should be offered a clinic appointment in the morning or in the afternoon. A morning appointment requires that respondents fast overnight, whereas shorter eating restrictions are imposed on those with afternoon appointments. This random allocation reduces the potential for bias, which could occur if respondents were permitted to choose their appointment times. Pregnant women, people with diabetes and other special cases are not asked to fast, even if they are given a morning appointment.

Dwellings are randomly selected in order to monitor some of the physical measures taken of the respondents at the clinic as part of the quality control and quality assurance monitoring. For these cases, some physical measures are repeated on the same respondent and compared with the original measures taken.

During the two-year data collection period, Statistics Canada will ensure that the required number of respondents in the appropriate age and sex groups is reached. As a result, adjustments to the parameters will be made on the sample file and the number of dwellings as required in order to reach the expected number of respondents and to try to reduce potential non-response bias.

Data collection at the 15 sites will be done sequentially over the two years. The sites have been ordered to take into account seasonality by region and the temporal effect, subject to operational and logistical constraints. The temporal effect means that the number of sites by region is distributed evenly in year one and year two (with the exception of the Atlantic region which has only one site).

Concluding remarks

The sample design of the Canadian Health Measures Survey (CHMS) is unique. For the first time in decades, both household and physical measures data will be collected on a representative sample at the national level. The CHMS builds on the successes of previous health surveys, and will fill data gaps to ensure that emerging needs can be addressed ●

Acknowledgements

The author thanks Sarah Maude Dion, Nicolas Lavigne and Rebecca Morrison of Household Survey Methods Division for participating in the development of the Canadian Health Measures Survey sample design. She also thanks Johane Dufour and Yves Béland, Household Survey Methods Division, and Sarah Connor Gorber, Physical Health Measures Division, for their comments on this paper.

References

1. Tremblay MS, Wolfson M, Connor Gorber S. Canadian Health Measures Survey: Rationale, background and overview. *Health Reports* (Statistics Canada, Catalogue 82-003) 2007; 18(Suppl.): 7-20.
2. Day B, Langlois R, Tremblay M, et al. Canadian Health Measures Survey: Ethical, legal and social issues. *Health Reports* (Statistics Canada, Catalogue 82-003) 2007; 18(Suppl.): 37-51.
3. Bryan SN, St-Denis M, Wojtas D. Canadian Health Measures Survey: Operations and logistics. *Health Reports* (Statistics Canada, Catalogue 82-003) 2007; 18(Suppl.): 53-70.
4. Tremblay MS, Langlois R, Bryan SN, et al. Canadian Health Measures Survey pre-test: Design, methods, results. *Health Reports* (Statistics Canada, Catalogue 82-003) 2007; 18(Suppl.): 21-30.
5. Statistics Canada. *Methodology of the Canadian Labour Force Survey* (Catalogue 71-256-XPB) Ottawa: Minister of Industry, 1998.
6. Laflamme G, Dochitoui C. *Creating Sites for the Canadian Health Measures Survey*, April 2005, internal paper.
7. Giroux S, Lavigne N. *Sampling Strategy for the Canadian Health Measures Survey, Part I: Selection of Collection Sites*, HSMD, March 2005, Internal paper.
8. Giroux S. *Study on the Number of Sites for the Canadian Health Measures Survey*, March 26, 2004, Internal paper.
9. Lavigne N. *Stratégie de sélection de personnes*, HSMD, January 2006, Internal paper.
10. Tambay J-L, Mohl C. Improving sample representativity through the use of a rejective method. *Proceedings of the Annual Meeting of the American Statistical Association: Survey Research Methods Section, August 1995*. Orlando, Florida: American Statistical Association, 1995.