Sample Design of the Canadian Health Measures Survey

Sarah Maude Dion
Statistics Canada, R.H. Coats building, Tunney's Pasture Driveway, Ottawa, Ontario, K1A OT6, Canada

1. The Canadian Health Measures Survey

Abstract

Statistics Canada, in partnership with Health Canada and the Public Health Agency of Canada, initiated the Canadian Health Measures Survey (CHMS) that started collection in the spring of 2007. The CHMS aims to overcome the limitations of existing health-monitoring information by directly measuring health indicators from a nationally representative sample of 5,000 Canadians aged 6 to 79. The respondents are asked to complete an in-home health questionnaire and then travel to a clinic to have physical health measurements taken by health professionals. A multistage design was developed to meet the objectives and logistics of the survey. This paper provides an overview of the CHMS with emphasis on the sample design. The effectiveness of the design is also being examined, using preliminary data.

KEY WORDS: Health Survey, Multistage Sample, Direct Measures

1.1 Introduction

The Canadian Health Measures Survey (CHMS) is a national survey collecting information about the general health and health habits of Canadians. The survey aims to overcome the limitations of existing health-monitoring information by directly measuring health indicators from a nationally representative sample of Canadians. It will help to explore emerging public health issues and to evaluate new measurement technologies. Although many countries have a long history of surveys that include direct physical measurements, it has been almost 30 years since such a survey was conducted in Canada. Direct physical measures are crucial for ascertaining relationships among risk factors, health protection practices and health status of the population. The CHMS will collect information that cannot be captured or could be inaccurately reported by Canadians. It will be an invaluable resource of benefit to individuals living in Canada, researchers and decision makers who will now have a comprehensive source of nationally representative measured health data to address the needs of Canadians.

Designing the survey was a challenge at different levels. It required exploring sampling strategies that are different from those traditionally used for large Statistics Canada surveys. This paper describes these challenges as well as the final sample design implemented. Section 1 presents an overview of the survey history and the pre-test, section 2 explains the sampling design put in place and section 3 shows some preliminary collection results on the efficiency of the sampling strategy.

1.2 Questionnaire and clinic visit

There are two components to the survey. First there is a health questionnaire, completed at the household using computer-assisted interviewing method and the help of regular interviewers. The health questionnaire covers topics such as nutrition, physical activity habits, family medical history, sexual behavior, alcohol and drug use, current and past medical condition and many more. Once the health questionnaire portion is completed, the respondents are asked to travel to a mobile clinic where specialists take physical measurements. The measures include blood pressure and heart rate, height and weight, skinfold measurements, fitness tests, an oral health examination and collection of blood and urine samples. The suitability of each measure is assessed for each respondent and they are tested accordingly. The CHMS is using 2 mobile clinics to conduct the clinic portion of the survey. The mobile clinics consist of a set of two trailers that once set up are linked by an enclosed pedestrian walkway. One of the trailers is used mostly as an administration and reception area and the other contains the clinic rooms and the hematology laboratory. The clinics travel from site to site and when one of the clinics is welcoming participants, the other one is traveling to the next site and setting up its examination rooms and equipments. The respondents receive a preliminary report of their health results upon completing their clinic visit. A few weeks after

their visit, a detailed report of the laboratory results is sent to them.

1.3 Overview of the pre-test

In the fall of 2004 a pre-test for the survey was conducted in Calgary, Alberta. (Morrison, Giroux, 2006) The objectives were mainly to determine the logistics of operating the clinic, taking physical measurements, booking appointments and assessing the costs and the time required to conduct all aspects of the survey. It also served to assess the participation rate and interest of the population in such a survey. The participation rate was considered satisfactory with most respondents agreeing to provide both the physical measurements and the blood and urine samples. The idea of taking direct physical measurements was well received by the population and proved necessary after the pretest results showed considerable differences between self-reported and directly measured data. In many cases the self-reported data in the pre-test were not as accurate as the direct physical measurements as respondents were not always aware of all of their medical conditions and therefore these were not always accurately reported at the time of interview.

Having a pre test a couple of years before collection started was a great helped with most of the logistical and planning issues. It helped identifying what can be problematic once in the field. The decision of using mobile clinics was also based on the pre test experience.

1.4 Logistics

Putting the survey in place was a challenging task for everyone involved. The logistics of the CHMS are complicated and completely new for Statistics Canada. (Tremblay, 2005) Many different steps of the collection process have to be coordinated. First the in-home questionnaire takes approximately an hour to complete. Most of the time, due to the length of this survey, the interviewers have to make an appointment with the respondents to complete the questionnaire. Once the questionnaire is completed, another appointment is booked for the clinic visit. Each respondent is randomly assigned an AM or PM appointment and must be accommodated during opening hours of the clinic and before the end of collection. Before physical measurements are taken, the instruments have to be calibrated. This has to be done on an ongoing basis. Within the clinic, different physical measurements are taken at different stations and the respondents have to be coordinated between these measurement stations. Blood and urine samples are taken at the clinic and are sent to a laboratory for testing. The shipping has to be done in a secure and confidential environment and a bar coding system was put in place for the identification of the samples by the laboratory. Finally, once the collection in a site is over and analyses have been performed on blood, urine and other data, a confidential health report is sent to the respondents.

Because there was only one site in the pretest, one of the only things that was not tested was the transition between two sites. The scheduling of the sites and the transition between two consecutive ones was completely new in the planning process. Collection of the health questionnaires starts a week or two before the clinic opens to allow interviewers to complete a certain number of interviews. It would be a waste of resources to have the clinic opened on the first day with no appointment booked yet. Then the household questionnaire collection ends before the end of the clinic collection. This allows interviewers to travel to the next site and start the collection while the clinic is finishing the appointments with the last remaining respondents. When one clinic is set up and opened to accommodate respondents, the other one is on the road and getting ready for the next collection site. Interviewers and clinic staff have to travel from one site to another and the shipment of material, laptops, respondent information brochures and other materials have to be carefully planned to arrive on time at each site.

1.5 Target Population

The CHMS is interviewing people of age 6 to 79 at the time of the survey who are living in private dwellings. Statistics Canada recognizes the importance of interviewing youths under the age of 6, but the logistic of interviewing this age group and performing physical measurements is more complicated than for other age groups. Furthermore, the participation rate for youngster under the age of six predicted based on focus groups was too low. It was not implemented for the first cycle of the survey.

The target population was divided into 5 age groups.

- 6-11
- 12-19
- 20-39
- 40-59
- 60-79

Within these age groups, the survey excludes some people:

- Institutional residents
- Full time members of Canadian Forces.

1.6 Requirements and Challenges

With this survey, Statistics Canada wants to be able to produce baseline estimates at a national level for the 10 age and sex groups. (Haines and Kearney, 2001)

Targeting a coefficient of variation (CV) not higher than 16.5 % the minimum number of respondents required per age and sex groups is 500 to estimate a prevalence of 10 % for most measures at the national level. This was calculated assuming a design effect of 1.5.

One of the main challenges of the design is reaching the targeted number in each age groups. The age groups are not all of the same size; the range of age of the younger age groups is much smaller than for the older age groups. Furthermore, children are harder to find in the population as they make up a smaller portion. It will be difficult to obtain enough respondents aged 6 to 11 and 12 to 19 while keeping the number of respondents aged 40 to 59 at a reasonable level.

The CHMS has a small sample and the survey design must help with reducing the variance as much as possible. CHMS will be able to yield quality data on the health status of Canadians if all those constraints are respected.

2. Survey Design

The CHMS survey design is a complex 3 stage design. In this section the survey frame selected for the CHMS will be presented as well as the reasons it was chosen. This is followed by the three stages of the sample design which are: selection of the sites; selection of the dwellings; and selection of the respondents.

2.1 Survey Frame

2.1.1 Selection of the frame

The Canadian Labor Force Survey (LFS) area frame was used to create the CHMS sites. The LFS is a survey interviewing approximately 54,000 dwellings monthly. (Gambino, Singh, Dufour, Kennedy and Lindeyer, 1998) It produces the unemployment rate in Canada and is one of the largest surveys conducted by Statistics Canada. The LFS frame was selected because it provides an accurate and up to date source of information about Canadian dwellings. It would allow CHMS to obtain a list of dwellings that could later be used in the second stage of the design for the sample selection. In fact, once it was time to do the dwelling selection, a more convenient source of information became available and the idea of using the LFS address lists was abandoned.

2.1.2 Creation of the sites

Using the LFS clusters, the sites were created respecting the boundaries of Economic Regions and Employment Insurance Regions and when possible the LFS strata. This was done both for operational reasons and to reduce the variance.

The decision was made to have the clinic set up as close as possible to the center of the sites to accommodate all respondents. The sites were created to ensure that respondents were at a reasonable distance from the clinic. It was decided to create the sites respecting these constraints:

- Maximum distance from the center of the site to the boundaries would be 100 km in rural areas and 50 km in urban areas
- Minimum of 10,000 people per site
- Respecting provincial and Census Metropolitan Area (CMA) boundaries.

Sites were first created respecting these constraints. The sites that did not meet all of the constraints were reviewed. When a low population count was found in a site it was collapsed with another site if the distance was still reasonable. A total of 3.7 % of the

Table 1 : Summary of population, exclusion and number of sites by region and Canada

	Atlantic	Quebec	Ontario	Prairies	British Columbia	CANADA
Census 2001 pop.	2,341,000	7,397,000	11,897,600	5,277,000	4,108,500	31,021,100
CHMS exclusions (% of Census 2001 pop)	110,965 (4.7%)	147,245 (2.0%)	239,131 (2.0%)	421,934 (8.0%)	222,991 (5.4%)	1,142,266 (3.7%)
# of sites	36	50	61	77	33	257

population was excluded from the survey for one of the following reasons:

- Low population density
- Remote, high vacancy and high cost area
- Indian reserves and crown lands.

In the end, a total of 257 sites were created all across Canada. Table 1 shows the distribution of population and sites in each of the regions.

2.2 First Stage – Sample of Collection Sites

The total number of sites selected was determined based on the following two constraints. First, 1,000 respondents required for each age group means 5,000 respondents needed across the country (see section *1.5.1 requirements*). Second, approximately 350 respondents per site are needed to justify the cost of moving and setting up the clinic. These two constraints led to the selection of 15 sites (Giroux and Lavigne, 2005).

Because national estimates are required, it was necessary to ensure all regions across Canada are represented. The idea of stratifying the sample of 15 sites by province was dismissed as there are 10 provinces and 3 territories. To obtain a good representation, the sample was stratified by region. The 5 Canadian regions commonly used at Statistics Canada are:

- Atlantic
- Quebec
- Ontario
- Prairies
- British Columbia

These regions are known to have similar geography and fairly homogeneous populations. The Yukon is included in the British Columbia region and the Northwest Territories and Nunavut are included in the Prairies region.

The sample of 15 sites was allocated proportionally to the size of the population in the regions. The allocation algorithm used the target population counts obtained from the 2001 Census. Within each region, the sites were sorted by population size and randomly selected using a systematic method with probability proportional to the size of the population. The number of sites selected by region is shown in table 2.

Table 2: Sample allocation of the 15 sites by region

Regions	Estimated target pop 6 - 79, using 2001 Census	Required respondents using proportional allocation	Allocated # of sites after adjustment
Atlantic	2,061,400	382	1
Quebec	6,560,400	1217	4
Ontario	10,248,500	1901	6
Prairies	4,539,000	842	2
B.-C.	3,540,000	657	2
Total	26,949,300	5000	15

From east to west, the 15 sites selected are Moncton (Atlantic), Quebec city (Quebec), Montreal Downtown (Quebec), Montreal South Shore (Quebec), Mauricie (Quebec), Clarington (Ontario), Toronto North (Ontario), Toronto Centre (Ontario), St-Catharines-Niagara (Ontario), Kitchener-Waterloo (Ontario), Northumberland County (Ontario), Edmonton (Prairies), Red Deer (Prairies), Vancouver (B.-C.) and Williams Lake-Quesnel (B.-C.).

2.3 Second Stage – Sample of Dwellings

2.3.1 List of dwellings

At the sample selection stage, the 2006 Census was available, providing a list of all the dwellings in the 15 sites. The initial idea was to use the LFS to obtain a list of the dwellings but the Census was favored for three main reasons. First, it provides the date of birth of every household member present at the time the Census was conducted. This information was helpful for targeting the five age groups required for the CHMS. Second, the LFS clusters would have been introduced in the design and avoiding it was helping with reducing the variance. Finally, the Census already had all the dwellings listed. Using LFS would have required more work as some of the areas are not yet listed.

2.3.2 Stratification

The dwellings within the selected sites are first stratified based on the age composition of the household members. This stratification is based on the presence or absence of certain age groups in the households. The stratification is achieved by favoring the hardest to reach age group to the easiest to reach age group. If a 6-11 year old is present based on the Census information, the dwelling is stratified in the 6 to 11 stratum. If there is no 6-11 year old, but a 12-19 year old is present according to the Census, then the dwelling is stratified in the 12 to 19 stratum. If there is no 6-11 year old and no 12-19 year old but a 60-79 year old is present then the dwelling is stratified in the 60 to 79 stratum. Similarly, the other dwellings are stratified in the 20 to 39 stratum and then the 40 to 59 stratum. All the other dwellings, either out of scope or vacant, are stratified in another stratum. The out of scope and vacant dwellings are included in the sample as the composition of these dwellings could have changed since the 2006 Census.

2.3.3 Selection

A simulation program was created and for each site it is used to help determine the number of dwellings to be selected in each of the strata. The simulation is based on predicted rates of responses and refusals at the household level, at the person level and at the clinic. It also takes into account the vacancy rate of the sites. The simulation program randomly selects a large number of different samples and simulates the selection of respondents and the rates of response, refusal and vacancy. It then gives the average number of respondents that would complete the clinic portion of the survey. The number of dwellings selected must give a number of respondents at the clinic as equal as possible in each age group. This is not always possible within a site as the respondents are selected randomly. By closely monitoring each site it is possible to adjust the number of dwelling selected per strata to compensate for high or low numbers of respondents. The simulations and predictions are adjusted at each site based on the results observed for the previously completed sites. The flexibility of this technique is limited as a minimum number of dwellings have to be selected in any given strata to avoid extremely large weights. However, it is still considered a good indicator. Once the number of dwellings to select per strata is determined, a simple random selection is performed within the strata.

2.4 Third Stage – Sample of Persons

When the interviewers go to a dwelling, the up-to-date household composition is obtained. If the dwelling is in scope (see section 1.5 Target Population and sampling requirements), one or two persons per household are selected. In the dwellings with people aged 12 to 79, only one person per household is selected. The decision was made to select only one person to reduce the cluster effect. In dwellings with 6-11 year olds present, two persons are selected. First, one of the 6-11 year olds is automatically selected by the application. Then, another person from the 12 to 79 age groups is selected. The 6-11 year olds being the most difficult to find in the population, it is important to try to increase their opportunities for participation. It was concluded after the pre-test and based on information gathered form focus groups that the 6-11 year olds were more likely to show up at the clinic if another person in the household was also participating in the survey.

After the household members have been listed, the application gives each of them a probability of selection. The probability of selection of each age group is determined by the strata in which the dwelling was classified. Because the 6-11 year olds are always being selected when one is present, this age group does not need to be favored in the selection process. The next hard to reach age group is the 12-19 year olds.

Therefore, this group has a higher chance of selection in three of the strata: 6-11, 12-19 and out-of-scope/vacant. The other 3 age groups, 20-39, 40-59 and 60-79 have a higher probability of selection in their respective strata. Table 3 shows which group is favored in each of the strata.

Table 3: Selection of person

Strata	**Age group favored for the selection**
6-11	12-19 (for 2nd person)
12-19	12-19
20-39	20-39
40-59	40-59
60-79	60-79
Out of Scope/Vacant	12-19

All household members who are part of a favored age group are given a probability of selection two or three times higher than all other household members. To avoid having extreme sample weights in households where a lot of people reside, the probability of selection is equal for every one, in all age groups, when six or more persons live in the same household.

3. Preliminary Evaluation of the Design

Collection of the CHMS started in the winter of 2007. The results presented in this paper are based on observations of the first three sites.

As soon as the survey went into the field, one of the results of interest, after response rates and refusal rates, was the effectiveness of the stratification. It was important to determine if the sample design, using the 2006 Census information would allow us to get the desired number of persons in each age group.

3.1 Accuracy of the information

As reported previously, the sampling strategy is based on the dwelling stratification which relies on the 2006 Census information. As collection moves further away from the Census collection date, the accuracy of the stratification decreases. The more people move, the less accurate the stratification. This would lead to a drop in the number of youths, aged 6 to 19, selected and a significant increase in the number of 40 to 59 year olds selected. This assumption is based on the known distribution of the population and observations made so far. This implies that the minimum number of respondents for the groups aged 6 to 11 and 12 to 19 would not be obtained and it would also increase the variance. After 3 sites, the number of dwellings selected for the survey with the exact same household members as the ones reported at time of Census was about 58%. The collection of these three sites took place approximately 11 to 14 months after the Census collection date. The details of the moving distribution are reported in Table 4.

Table 4: Moving Since the 2006 Census

Changes in Household Members	% site 1	% site 2	% site 3
Completely different household members	9	14	15
Some changes in the list of household members	31	29	28
Same household members	60	57	57

3.2 Effectiveness of the Strata

The moving has an impact on the household composition but what is most important is the affect on the stratification. It is possible to observe a lot of moving but still have a similar household composition afterward. It is important to ascertain whether the household stratification observed at the time of survey is the same as the one assigned based on the Census household composition. In other words, if the strata were assigned based on survey data, would they be the same as the ones assigned based on Census data? It was observed after the first three sites that the stratification for the in-scope strata is the same, on average, 85% of the time. Table 5 shows the accuracy of the stratification for each age group. A high percentage on the grey diagonal is desired. This means that the strata planned with the Census are the same as the strata that would have been obtained based on the up-to-date information acquired at time of survey. It was observed that the most accurate stratum is the 60 to 79 and the least accurate one is 12 to 19. This can most likely be explained by the fact that most of the moving is observed in the 12 to 19 age group.

Table 5: Percentage of Accuracy in Assessing the Strata

Strata based on 2006 Census	Strata based on survey observations					
	6-11	12-19	20-39	40-59	60-79	All
6-11	86%					
12-19		85%			9%*	
20-39			78%			
40-59				83%		
60-79		6%*			91%	
All						85%

*Results are an average of all cells, below or above the diagonal, combined.

4. Future Work

The collection for the CHMS will end in the spring of 2009. Until then, a close monitoring of each site is performed. The number of dwellings selected in each stratum is adjusted at each site in order to obtain an equal number of respondents in each age group at the end of the survey. The efficiency of the stratification is assessed after every site. So far, the stratification has provided more than satisfactory results. We are currently able to obtain a number of respondents in each age group as close as possible to the targeted number.

Should the distribution by age group become difficult to control, mostly due to the out-dated information from the 2006 Census, other mechanisms could be used. For instance, a solution to adjust the selection would be to increase the probability of selection of certain age groups. This is currently being examined. Simulations are performed to assess the effectiveness of this possible solution. This means that the age group with low responses would get a higher chance of selection than it has now in its assigned strata (see section 2.4). If the difficulties in obtaining the targeted number of respondents per age group are not resolved with this method, another solution would be to use the rejection method that is already in place in the application. This method would randomly reject dwellings where household members are all part of the same high response age group. This method was proved unbiased but is not the preferred one. (Tambay and Mohl, 1995) A lot of time and efforts are spent on those dwellings before they get rejected and this is not desirable from a production point of view.

It is expected that before the end of collection, the age stratification alone might not be sufficient and at this point Statistics Canada want to be ready to make changes that are going to help with the sampling process without having a negative impact on the weighting, the imputation and the variance. Also, lists of dwellings from sources other than the 2006 Census will most likely be used in addition to the current frame. This will ensure all new dwellings are included in the survey.

Acknowledgements

The author wants to thank Suzelle Giroux and Rebecca Morrison for their excellent work on the CHMS project, and particularly Suzelle for her contribution to section 2 and the information provided for the related tables. Also, a thank you to Michelle Simard for all her valuable inputs and recommendations. A special thank you to Guy Laflamme and Sarah Connor Gorber for their insightful comments and suggestions.

References

Morrison, R. and Giroux, S., (2005), "Collecting Blood and Urine : the Experience of the Canadian Health Measures Survey", 2005 JSM Proceedings, Survey Methods Research Section, 3398-3401.

Tremblay, M.S. (2006). "Learning the Unique and Peculiar Challenges of Direct Health Measures Surbeys: The Canadian Experience", 2006 Statistics Canada International Symposium Series, to be published.

Haines, D. and Kearney, J., (2001), "Sample Size Estimations for the CCHS Physical Measures Survey". Statistics Canada, 2001.

Gambino, J.G., Singh, M.P., Dufour, J., Kennedy, B. and Lindeyer, J., (1998), "LFS Methodology", Household Survey Method Division.

Giroux, S. and Lavigne, N. (2005), "Sampling Strategy for the Canadian Health Measures Survey, Part 1: Selection of collection sites"

Tambay, J.-L. and Mohl, C. (1995), "Improving sample representativity through the use of a

rejective method" Proceedings of the Annual Meeting of the American Statistical Association: Survey Research Method Section, August 1995. Orlando, Florida: American Statistical Association, 1995.