Achieving the Unique Objectives of the Canadian Health Measures Survey

Sarah Maude Dion[1], Suzelle Giroux[2]
[1]Statistics Canada, R.H. Coats building, Tunney's Pasture Driveway, Ottawa, Ontario K1A 0T6, Canada
[2]Statistics Canada, R.H. Coats building, Tunney's Pasture Driveway, Ottawa, Ontario K1A 0T6, Canada

Abstract

The Canadian Health Measures Survey at Statistics Canada consists of an in-home health questionnaire and a visit to a clinic where direct measures of health indicators are taken from a nationally representative sample of Canadians. The sample was drawn from a multi-stage complex design that used a stratification based on auxiliary variables obtained from the most recent Census. Many challenges needed to be addressed such as deterioration of the frame over time as well as age groups of significantly different sizes and the specific objective of obtaining equal numbers of respondents in each age group. This paper will provide an overview of the CHMS with emphasis on the efficiency of the adopted strategy to obtain sufficient sample size in the target groups. Other issues will be discussed, such as the distance a respondent has to travel to get to the clinic and its effect on the response rates.

1. Introduction

Statistics Canada, in partnership with Health Canada and the Public Health Agency of Canada, initiated the Canadian Health Measures Survey (CHMS) beginning 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 Canadians. The collection period for Cycle 1 started in March 2007 and ended in February 2009.

CHMS collects information about the general health and health habits of Canadians. It will help 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 had 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 collects information that cannot be captured or could be inaccurately reported by Canadians. It is 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. The CHMS sample for Cycle 1 was drawn from a multi-stage complex design that was developed to meet the objectives of the survey. This special design used a stratification based on auxiliary variables obtained from the frame. Challenges such as deterioration of the frame, movers and new construction needed to be addressed. In particular, these factors made it difficult to attain the specific objective of obtaining an equal number of respondents in each age group which was further complicated by having survey defined groups of significantly different sizes.

This paper will provide an overview of the CHMS and its complex design with emphasis on the effectiveness of the stratification using Cycle 1 data and results.

2. Overview of the Survey

2.1 Overview of the pre-test

In the fall of 2004 a pre-test for the survey was conducted in Calgary, Alberta. (Morrison, Giroux, 2005) 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.

2.2 Target population

For Cycle 1, CHMS interviewed people of age 6 to 79 at the time of the survey who were living in private dwellings. Statistics Canada recognizes the importance of interviewing youths under the age of 6, but the logistics of interviewing this age group and performing physical measurements are more complicated than for other age groups. Therefore, it was not included in the first cycle of the survey. 80-year-olds and over were not included either as a high proportion of these persons are in institutions. This type of dwelling is not sampled for this survey. For those who are not in institutions, the pre-test showed that most refused to go to the clinic and this was especially true for women.

The target population of Cycle 1 was divided into 5 age groups.

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

Within these age groups, the survey excluded some people:

- Institutional residents
- Full time members of the Canadian Forces.

2.3 Health questionnaire and clinic visit

Respondents are requested to complete two components of the survey. First, a health questionnaire is administered at the household by a Statistics Canada interviewer using a computer-assisted method. The health questionnaire covers topics such as nutrition, physical activity habits, family medical history, sexual behavior, alcohol and drug use, current and past medical conditions and many more. Once the health questionnaire portion is completed, the

respondents are asked to travel to a mobile clinic where health measure specialists take physical measurements. The measures include, but are not limited to: blood pressure and heart rate; height and weight; skinfold measurements; fitness tests; an oral health examination and collection of blood and urine samples. Respondents fill out a screening questionnaire prior to any physical test or measurement to assess their health condition and the appropriateness of each measure. They are then tested accordingly and any measure that involves a risk for the respondent are not performed. For example, respondents aged 70 to 79 years old are all screened out of most fitness tests for safety reasons.

2.4 Objectives and constraints

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)

It was determined that the minimum number of respondents in each age-sex group should be 500. With 5,000 respondents, the survey should provide national level estimates for conditions that have a prevalence rate of 10% or higher, with a coefficient of variation of 16.5% or better (Giroux, S., 2007).

After reviewing the logistics, constraints and requirements, Statistics Canada was able to allocate a nationally representative sample of 5,600 respondents for Cycle 1. This implied that the sample had to be distributed as equally as possible between the groups as it contained a number of respondents barely higher than the minimum needed.

One of the main challenges with a sample equally distributed in each age group was reaching the targeted number. 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. The 6 to 11 age group has a span of 6 years, the 12 to 19 age group has a span of 8 years and all the other groups have a span of 20 years. Furthermore, children are harder to find in the population as they make up a smaller proportion. The difficulty was to obtain enough respondents aged 6 to 11 and 12 to 19 while keeping the number of respondents aged 20 to 39 and 40 to 59 at a reasonable level.

Table 1 shows that the 6 to 11 and 12 to 19 age groups had much smaller spans than the other groups. Also, it shows that the 60 to 79 group had the lowest density of population of all groups with just fewer than 240,000 persons per year covered. They were followed by the 6 to 11 group which had close to 380,000 persons per year covered.

Table 1: Average Distribution of Population Per Year Covered For Each Age Group (Density)

Age group	2006 Census population	Span	Average per year covered
6-11	2,275,425	6	379,238
12-19	3,410,095	8	426,262
20-39	8,294,450	20	414,723
40-59	9,672,985	20	483,649
60-79	4,757,800	20	237,890
Total	28,410,755	---	---

Overall, as shown in table 2, the 6- to 11-year-olds were going to be the hardest to reach as they made up only 8% of the target population. They were followed by the 12- to 19- and 60- to 79-year-olds.

Table 2: Distribution of Population by Age Group

Age group	2006 Census population	% population	Survey target %
6-11	2,275,425	8.0%	20%
12-19	3,410,095	12.0%	20%
20-39	8,294,450	29.2%	20%
40-59	9,672,985	34.1%	20%
60-79	4,757,800	16.7%	20%
Total	28,410,755	100.0%	100%

It is important to mention that rejection of certain respondents when targets were met was not an option. Quota sampling was not considered to be a viable solution. The design had to respect classic statistical theory such as giving everyone in the target population a probability of selection and having a random selection of respondents.

The CHMS had a small sample and the survey design also needed to help reduce the variance as much as possible. CHMS will be able to yield quality data on the health status of Canadians if all these constraints are respected.

Also, because respondents are asked to travel to a clinic, it was important to minimize their travel distance. It was expected that this could have a huge impact on the response rates and had to be taken into consideration. The deterioration of the sampling frame also has a considerable effect on the sampling strategy. All of these constraints were addressed by the survey design as explained in the next section.

3. Survey Design

The CHMS Cycle 1 survey design is a 3-stage design. Each stage was designed to reduce the variance and achieve the objectives using a small sample. The sample was drawn from a multi-stage complex design that used a stratification based on auxiliary variables obtained from the most recent Canadian Census of Population. The different steps of the sample selection are as explained in the next few sections.

3.1 First stage – Sample of collection sites

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 within a reasonable distance from the clinic. Because respondents were traveling to the clinic by their own means it was highly recommended to keep the sites under a maximum size to reduce respondent burden and increase the chances of achieving a high response rate. 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

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

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

In the end, a total of 257 sites were created across Canada.

The total number of sites selected was determined based on the following two constraints. First, 5,600 respondents were needed across the country, as mentioned at the beginning of section 2.4. Second, a minimum of 350 respondents per site were 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 were required, it was necessary to ensure all regions across Canada were 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. Nunavut did not have any area that met the sites constraints and was excluded from the frame in the end.

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 (the 2006 Census was not available at the time this was completed). 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 3.

Table 3: 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	1,217	4
Ontario	10,248,500	1,901	6
Prairies	4,539,000	842	2
B.-C.	3,540,000	657	2
Total	26,949,300	5,000	15

From east to west, the 15 sites selected were Moncton (Atlantic), Quebec City (Quebec), Montreal Downtown (Quebec), Montreal South Shore (Quebec), Mauricie South (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.). These sites were to be visited one at a time over a two year collection period. The order of the sites was determined prior to the start of the survey and dealt, to a certain extent, with seasonality and temporal effects.

3.2 Second stage – Sample of dwellings

At the second selection stage, the 2006 Census was available, providing a list of all dwellings in the 15 sites. The Census as a frame was favored for three main reasons. First, it provided the date of birth of every household member present at the time the Census was conducted. This information was helpful for targeting the age groups required for the CHMS. Second, the other frame considered would have introduced clusters in the design and avoiding it helped reduce the variance. Finally, the Census had an exhaustive list of dwellings and contact information.

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

For each site, the number of dwellings to be selected in each of the strata had to be determined. The number was based on estimated rates of responses and refusals at the household level, at the person level and at the clinic. Also taken into account was the vacancy rate of the sites. The

number of dwellings selected needed to yield a number of respondents at the clinic as equal as possible in each age group. This was not always possible within a site as the respondents are selected randomly by the household questionnaire application and all household members between the age of 6 and 79 have a chance of being selected. By closely monitoring each site it was possible to adjust the number of dwellings selected per stratum to compensate for high or low numbers of respondents. The flexibility of this technique was limited as a minimum number of dwellings had to be selected in any given stratum to avoid large sampling weights. Once the number of dwellings to select per stratum was determined, a simple random selection was performed within the stratum.

3.3 Third stage – Sample of persons

When the interviewers went to a dwelling, the up-to-date household composition was obtained. If the dwelling was in scope (see section 2.2 Target Population), one or two persons per household were selected. In the dwellings with people aged 12 to 79, only one person per household was selected. The decision was made to select only one person to reduce the cluster effect. In dwellings with 6- to 11-year-olds present, two persons were selected. First, one of the 6- to 11-year-olds was automatically selected by the application. Then, another person from the 12 to 79 age group was selected. The 6- to 11-year-olds being the most difficult to find in the population, it was important to try to increase their opportunities for participation. Experiences from other surveys indicated that the response rates tended to be higher when two persons were selected, meaning that the 6- to 11-year-olds were more likely to show up at the clinic if another person in the household was also participating in the survey. This can be due either to the fact that parents are going not only for their child but also for themselves or because they are taking another one of their children (age 12 and above) at the same time. Both respondents received a full health assessment report which can be an incentive for some and the appointments were usually booked at the same time to keep parents' burden to a minimum.

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 was determined by the stratum in which the dwelling was initially classified. Because one 6- to 11-year-old was always selected when one was present, this age group did not need to be favored in the selection process. The next hard to reach age group was the 12- to 19-year-olds. Therefore, this group had a higher chance of selection in three of the strata: 6 to 11, 12 to 19 and out-of-scope/vacant. The other 3 age groups, 20 to 39, 40 to 59 and 60 to 79 had a higher probability of selection in their respective strata. Table 4 shows which group was favored in each of the strata.

Table 4: 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 were part of a favored age group were given a probability of selection two to five times higher than all other household members. To avoid having extreme

sample weights in households where a large number of people reside, the probability of selection was set equal for everyone, in all age groups, when six or more persons lived in the same household.

When the distribution by age group became difficult to control with the stratification alone, mostly due to the out-dated information from the 2006 Census, it was possible to change the probability of selection of certain age groups. This was first examined and simulations were performed to assess the effectiveness of this possible solution and the impact it would have on the weights. Finally it was implemented and the age group with the lowest responses received a higher chance of selection than it had at the beginning of the survey in its assigned strata.

4. Evaluation of the Design

This design for the CHMS, along with the special person selection strategy was unusual at Statistics Canada. It was hard to predict the results it would yield but it was at the time the best way to meet the challenging objectives and constraints of the CHMS. The creation of this design respected the statistical theory, avoided rejection of respondents and also allowed some control over the selection of specific age groups. Control over the selection of the different age groups was of utmost importance and is what most influenced the choice of design.

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 Statistics Canada to achieve the desired number of persons in each age group.

The results presented in this paper are based on observations from Cycle 1.

4.1 Accuracy of the information

As reported previously, the sampling strategy was based on the dwelling stratification which relied on the 2006 Census information. As collection moved further away from the Census collection date, the accuracy of the stratification decreased. The more people moved, the less accurate the stratification. Deaths also played a role in the deterioration of the stratification but births were not considered in the results as anyone born after the census date would not be included in any of the CHMS age groups. If the stratification became inaccurate, it would have led to a drop in the number of youths selected aged 6 to 19, and a significant increase in the number of 40- to 59-year-olds selected. This assumption was based on the known distribution of the population and observations made on collection results. This implied that the minimum number of respondents for the groups aged 6 to 11 and 12 to 19 would not be obtained and it would increase the variance.

After two years of collection, the number of dwellings selected for the survey with the exact same household members as the ones reported at time of Census was on average 47%. The collection of Cycle 1 took place approximately 11 to 32 months after the Census collection date. The details of the distribution of household changes are reported in Table 5.

Table 5: Changes in the Household Composition Since the 2006 Census

Changes in household members	Completely different household members	Some changes in the list of household members	Same household members
Site 1	9%	31%	60%
Site 2	14%	28%	58%
Site 3	16%	28%	57%
Site 4	23%	36%	41%
Site 5	29%	26%	45%
Site 6	22%	28%	51%
Site 7	34%	26%	40%
Site 8	34%	30%	36%
Site 9	28%	33%	39%
Site 10	24%	28%	49%
Site 11	25%	27%	48%
Site 12	23%	28%	49%
Site 13	21%	34%	45%
Site 14	29%	30%	41%
Site 15	42%	23%	35%
Average	25%	29%	47%

As would be expected, the percentage of households with the exact same composition decreased over time. The site had a considerable influence on the percentage. Sites 7, 8 and 9 had a much lower percentage than the following sites but overall, there was a constant decrease in the accuracy.

In major urban cores, the moving rate is usually higher than in rural areas. In big cities, the percentage of household with no changes in the composition since the 2006 Census was 44.2% compared to 47.5% for the households in rural areas. Furthermore, the response rates in the major urban cores are always a little lower than for rural areas adding to the challenge of attaining the targets in these regions.

4.2 Efficiency of the strata

Moving and deaths had an impact on the household composition but what was most important was the impact on the stratification. It is possible to observe many changes but still have a similar age composition afterward. It was important to ascertain whether the household stratification observed at the time of survey was the same as the one assigned based on the Census household age composition. In other words, if the strata were assigned based on survey data, would they have been the same as the ones assigned based on Census data? For the entire Cycle 1 collection period, it was observed that the strata were the same, on average, 79% of the time. Table 6 shows the accuracy of the stratification for each age group. A high percentage means that most of the strata planned with the Census were the same as the strata that would have been obtained based

on the up-to-date information acquired at time of the survey. It was observed that the most accurate stratum was the 60 to 79 and the least accurate one was 20 to 39.

Table 6: Percentage of Accuracy in Assessing the Strata

Strata	Percentage
6-11	80%
12-19	76%
20-39	72%
40-59	75%
60-79	86%
Average	79%

Sites are independent from one another and have different characteristics. They were interviewed one at a time over a period of two years. Collection of Site 1 started in April of 2007. Each site's collection period lasted approximately six to eight weeks and collection of Site 15 ended in February of 2009. Even though the stratification accuracy was influenced by the characteristics of each site, there seemed to be a constant decrease in the accuracy as we moved farther away from the census date.

Table 7: Percentage of Accuracy in Assessing the Strata by Site

Sites	Percentage
Site 1	90%
Site 2	85%
Site 3	83%
Site 4	81%
Site 5	81%
Site 6	80%
Site 7	80%
Site 8	67%
Site 9	74%
Site 10	80%
Site 11	79%
Site 12	83%
Site 13	77%
Site 14	75%
Site 15	71%
Average	79%

The accuracy of the stratification, as shown in Figure 1, was directly influenced by the accuracy of the household composition. However, the accuracy of the stratification is much higher and more stable which is one factor that positively contributed to the achievement of the targets.

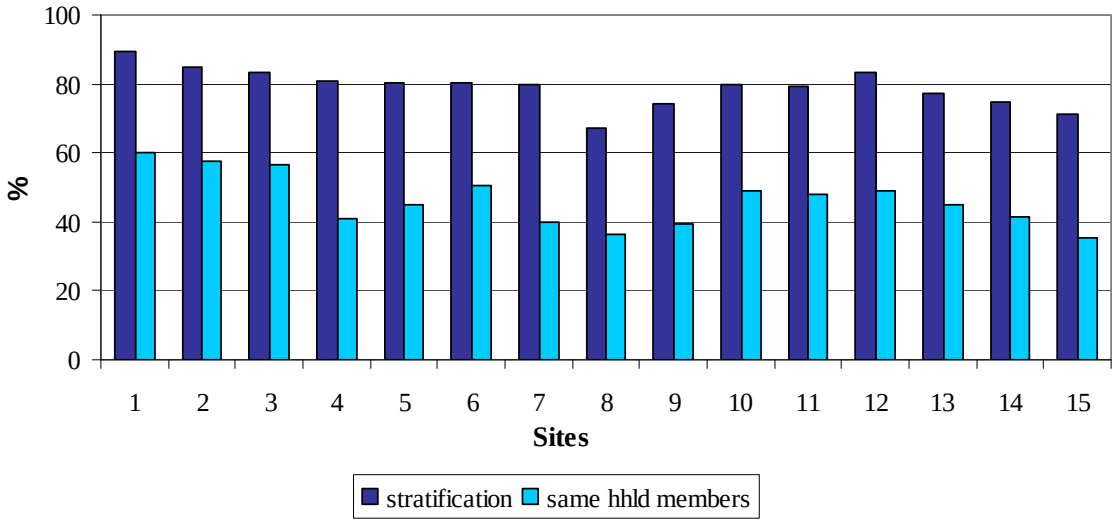

Figure 1: Accuracy of Stratification and Household Composition

5. Influence of the Distance between the Clinic and the Respondents' Household

The CHMS collection sites can easily be mapped to the clusters of the Canadian Labour Force Survey (Beaumont, Boyer, Laflamme, Lebrasseur, Lindeyer and Turmelle, 2008). Using the known coordinates of these LFS clusters, distances between the households and the clinic were approximated. The coordinates reflect approximately the center of each cluster. The distances were based on the differences between the coordinates of the cluster with a responding household and the cluster in which the clinic was located. Distances were then grouped in the following categories:

- 0 km to 5 km
- more than 5 km to 10 km
- more than 10 km to 15 km
- more than 15 km to 20 km
- more than 20 km to 30 km
- more than 30 km to 50 km
- more than 50 km to 100 km
- more than 100 km

Results showed that response rates decreased systematically as distance increased between the clinic and the respondent's home. Response rates varied from 88.2% in the 0km-5km class to 68.8% in the 100+km class. This trend was not observed for the household portion of the survey as the questionnaire was completed at the respondent's home. Since interviewers were the ones traveling to reach the respondents the distance did not influence the response rates.

These data were analyzed for all sites combined as the number of respondents per site was not high enough to lead to results of acceptable quality. However, it was possible to group data from major urban cores (Montreal Downtown (Quebec), Montreal South Shore (Quebec), Toronto North (Ontario), Toronto Centre (Ontario), Vancouver (B.C.)) and compare them to the grouped data of the remaining sites. The decreasing trend in the response rates was not significantly different for the two groups when data were divided this way. One can easily observe from Figure 2 that both curves are very similar for 0 to 30 km. There were no observations over 30 km for

major urban cores as the dwellings were always located within 30 km of the clinic. Based on these results, it was concluded that people who did not live in major urban cores were not more inclined to travel than the others. Furthermore, the response rates were much lower for the highest category of distances (30 km and over). When creating the sites, the assumption was that the non-urban sites could be larger because people living outside of major urban cores were more inclined to travel. This assumption was based on information gathered from focus groups held in February 2004 but did not appear to apply to CHMS once collection started. Thus, when the sites are remodeled for Cycle 3 this new information will be taken into account.

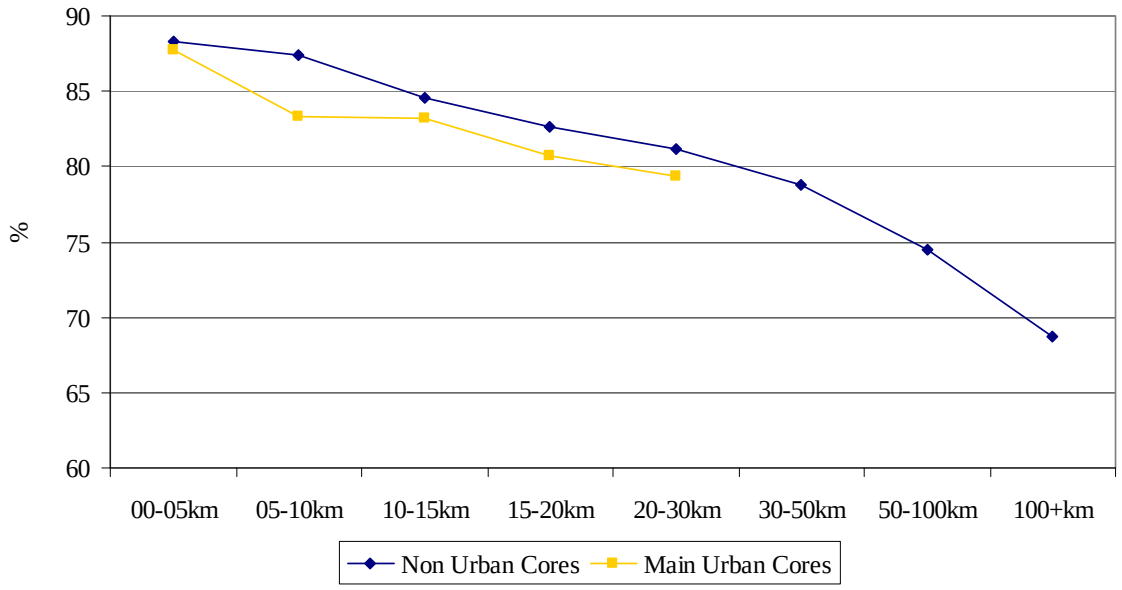

Figure 2: Response Rates by Distance to the Clinic
Note: The rates illustrated in Figure 2 are the clinic response rates of those who agreed to participate in the household health questionnaire. It is not representative of the complete survey response rates.

Data were also looked at by morning and afternoon appointment at the clinic. These appointments were distributed randomly and once a respondent was selected for one or the other, the time of the clinic visit could not be changed. It was a concern that perhaps morning appointments would be more difficult to attend by respondents. The analysis of these data showed that the response rates were the same for both morning and afternoon appointments.

6. Final Results

The collection for the CHMS ended in the spring of 2009. Throughout the first cycle, a close monitoring of each site was performed. The number of dwellings selected in each stratum was 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 was assessed after every site. The stratification has provided more than satisfactory results. It was possible, for Cycle 1, to obtain a number of respondents in each age group close to the targeted number.

The difficulty in obtaining the targeted number of respondents per age group came from the fact that a rejection method was to be avoided as much as possible. This rejection method was already in place in the application but it was available only as a last resort solution if all else failed. This method would have randomly rejected dwellings where household members were all part of the same high response age group. This method was proved unbiased (Tambay and Mohl, 1995) but was not the preferred one. Considerable time and effort would have been spent on those dwellings before they were rejected and this was not desirable from a production point of view. It is also not recommended from a weighting and estimation point of view as it adds an additional complication at the estimation stage.

In the end, the stratification of the dwellings based on age composition and the special person selection strategy were sufficient to meet the targets. It was possible to obtain the minimum of 500 respondents per age group without oversampling and without the use of the rejection strategy. The final results are shown in Table 8.

Table 8: Final Number of Respondents Per Age Group and Sex

	Sex	
Age Group	Male	Female
6-11	542	534
12-19	514	500
20-39	526	662
40-59	584	655
60-79	546	554
Sub Total	2712	2905
Total	5617	

The 12- to 19-year-old females were the hardest age-sex group to obtain. The target for this group was just met. The age-sex group with the highest number of respondents was the 20- to 39-year-old females. This is in part due to the fact that youths under the age of 12 were always selected along with another person and there are more single mothers in the population than there are single fathers. As expected, the number of 40- to 59-year-olds is also very high in particular for the female group. In fact, it is interesting to note that for the three older age groups, there was a higher number of female respondents than male respondents. This phenomenon could have become an issue if the sample size was smaller. It would have become almost impossible to obtain the minimum number of male respondents in some age groups without oversampling and going beyond the allocated total number of respondents for the survey. These observations will be useful when sampling strategies for subsequent cycles are put in place.

7. Cycle 1 Current Work

Work on non-response adjustment is currently being performed. There are three levels of non-response to consider for this survey. There is first the non-response at the household level when interviewers try to obtain the household roster. Then there is the non-response at the household questionnaire level once a person is selected for an interview and do not participate and finally there is the non-response at the clinic level when respondents who completed the household health questionnaire do not show up at the clinic.

There are also four different sub-samples that were randomly selected for different blood and urine measures. These four sub-samples will need the same kind of non-response adjustment with a fourth level of non-response. People selected for a measure who either refused or did not provide enough blood or urine are considered non-respondents. Adjustment models different than the one for the clinic level will have to be found for all four sub-samples and a total of five different set of weights are to be produced.

8. Work on Subsequent Cycles

Cycle 2 started in August of 2009 and will last approximately two years. As mentioned in section 2.2, Statistics Canada recognizes the importance of gathering health information on children under the age of 6 and the 3- to 5-year-olds were added to the survey for Cycle 2. The addition of this age group added a new challenge to the sampling strategy. By the end of Cycle 2 collection, these children will not be on the frame as they were not born at the time of Census. Updates from other sources are required in order to be able to target this specific age group. The 3-to 5-year-olds will be stratified along with the 6- to 11-year-olds in a 3 to 11 stratum. The person selection in this stratum will be the same as for the 6 to 11 stratum in Cycle 1, meaning that only one respondent aged 3 to 11 will be automatically selected along with another person aged 12 to 79. The number of respondents required for the 3- to 5-year-olds is half as much as for the 6- to 11-year-olds. Luckily, the natural distribution of the population for these two age groups is very similar to the survey requirements which will help reaching these targets without an overly complicated sampling strategy. To adjust the numbers of respondents in these two age groups, a probability of selection is assigned independently to the 3-to 5-year-olds and 6- to 11-year-olds and can be easily adjusted throughout the cycle.

The person selection strategy was slightly changed in Cycle 2. Instead of favoring the selection of the 12- to 19-year-olds in three of the six strata, they are now favored in all six. This change was implemented after Cycle 1 because the group of 12- to 19-year-olds was the most difficult to obtain. All other age groups are still favored only in their respective strata but with a probability of selection slightly lower than for the 12 to 19. Again, this does not affect the 3- to 11-year-olds that are automatically selected when one is present.

Finally, new sites will be created for Cycle 3. The information that was gathered from the analysis of the impact of the traveling distance to the clinic on the response rates will be useful for this purpose. The initial assumption that non-urban sites could be of considerably larger area without affecting the response rates was erroneous. When remodeling starts, the geographical areas of the non-urban sites will be kept to a smaller size. The person selection algorithm will also be improved for Cycle 3. Not only will age be taken into consideration as it was for Cycle 1 but sex will also be added. This will allow for different probabilities of selection between the two sexes. This improvement will balance out the total number of respondents per sex for each age group.

Acknowledgements

The authors want to thank, France Labrecque and Krista Collins for their excellent ongoing work on the CHMS project, and Rebecca Morrison and Nicolas Lavigne for their past involvement and dedication to the survey. A special thank you to Colin Babyak and Renée Langlois 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.

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

Giroux, S. (2007), "Canadian Health Measures Survey: Sampling Strategy Overview" Health Reports, No 82-003-SPE Supplement to vol. 18, 2007.

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

Beaumont, J.-F., Boyer, R., Laflamme, G., Lebrasseur, D., Lindeyer, J., and Turmelle, C., (2008), "Methodology of the Canadian Labour Force Survey", Household Survey Method Division.

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.