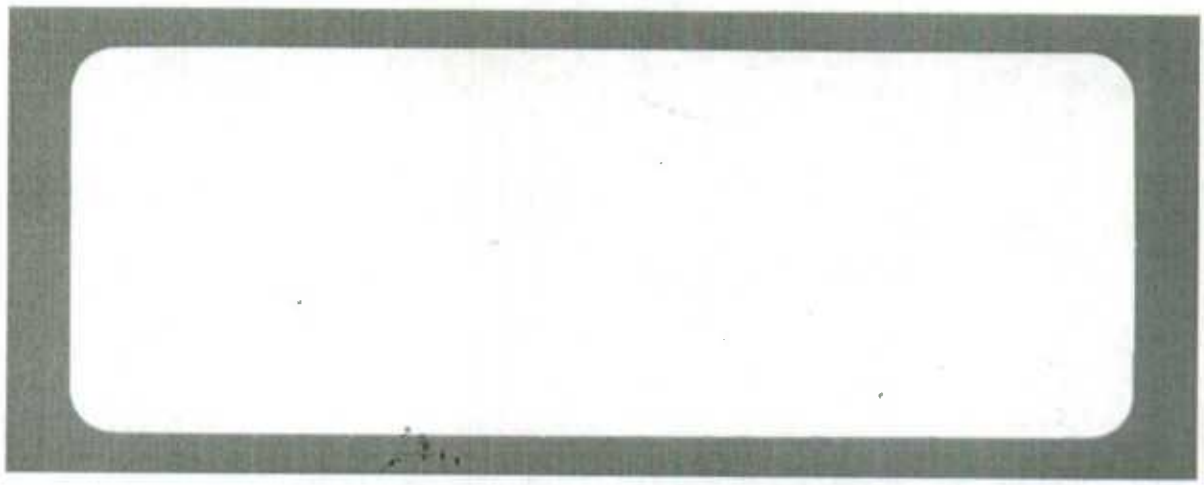
11-619
no. 2013-002

c.3

Methodology Branch

Direction de la méthodologie



Household Survey Methods Division

Division des méthodes d'enquêtes auprès des ménages

Statistics Canada Statistique Canada

Canada

WORKING PAPER
METHODOLOGY BRANCH

Sampling documentation for cycle 2 of the Canadian Health Measures Survey

HSMD - 2013 – 002E/F

Suzelle Giroux, France Labrecque and Andrew Quigley

Household Survey Methods Division
Statistics Canada

April 2013

Statistics Canada Statistique Canada
AVR APR 19 2013
LIBRARY BIBLIOTHÈQUE

The work presented in this paper is the responsibility of the author and does not necessarily represent the views or policies of Statistics Canada

Statistics Canada Statistique Canada
LIBRARY
BIBLIOTHÈQUE

Sampling documentation for cycle 2 of the Canadian Health Measures Survey

Suzelle Giroux, France Labrecque and Andrew Quigley
Household Survey Methods Division

ABSTRACT

Cycle 2 of the Canadian Health Measures Survey (CHMS) began collection in August 2009 and ended in November 2011. Data collection for the CHMS required interviewers to contact respondents in person to answer a health questionnaire, then the respondent was asked to travel to a nearby CHMS mobile examination center for physical measures and biospecimen collection. Finally, laboratory analyses of blood and urine were performed, making this survey's collection very costly and logistically challenging. It was therefore important to have a sampling strategy that considered the collection constraints and that could attain the objective of the survey without significantly exceeding the minimum total number of respondents required. In order to achieve this, a frame with the most up to date information about household composition and a flexible sampling strategy were necessary. This paper describes the frame and the sampling strategy used for cycle 2 to meet the survey's objective and provides results to determine if it was reached.

TABLE OF CONTENTS

ABSTRACT 3
Summary 7
1 Objective of the survey 8
1.1 Data requirements 8
1.2 Target population 8
1.3 Sample size 8
2 CHMS Sampling Design 8
2.1 First stage: Sampling of collection sites 9
2.1.1 Area frame creation for the collection sites 9
2.1.2 Collection sites sample size 9
2.1.3 Allocation of the collection sites 9
2.1.4 Selection of collection sites 10
2.1.5 Seasonality and order of collection sites 10
2.1.6 Addition of collection sites 10
2.2 Second stage: Sampling of dwellings 12
2.2.1 Frame creation for the dwellings 12
2.2.2 Frame updates – Adding new dwellings 12
2.2.3 Dwelling stratification 13
2.2.4 Dwelling stratification update 13
2.2.5 Quality of the match between the frame and the T1FF/T1 13
2.2.6 Dwelling sample size and allocation 14
2.2.7 Overlap control 15
2.2.8 Additional sample waves 15
2.3 Third stage: Sampling of persons 16
2.3.1 Person sampling 16
2.3.2 Subsampling for contaminants in blood and urine 17
2.3.3 Subsampling for fasted clinic appointments 18
3 Survey results 18
3.1 Frame quality 18
3.1.1 Accuracy of contact information – Addresses 18
3.1.2 Accuracy of contact information – Telephone numbers 19

3.1.3 Efficiency of stratification 20
3.2 Number of respondents and response rates by collection site 22
3.3 Comparing cycle 1 and cycle 2 respondent counts and response rates 24
3.4 Number of respondents for the subsamples 26
3.5 Response rates at the household-level by approximate driving distance 26
4 Conclusion 28
5 References 29
Appendix A 30

Summary

The objective of cycle 2 of the Canadian Health Measures Survey (CHMS) was to provide national baseline estimates on a variety of health indicators for individuals aged 3 to 79 years living in private dwellings. The collection strategy of the CHMS required that selected participants first answer a health questionnaire at their home and then travel to a nearby Mobile Examination Centre (MEC), often referred to as the clinic, where health measure specialists performed multiple physical measurements. The respondents were also required to provide blood and urine samples at the MEC on which laboratory analyses were performed to measure various biomarkers, environmental contaminants and infectious diseases. With the complex logistics and high costs of collection, it was important to have a sample design that met the objective of the survey with a sample size that was as small as possible.

A three-stage survey design was chosen to meet the collection logistics and the survey's objective. At the first stage, geographic areas called collection sites were constructed, since the CHMS required that respondents travel to the MEC and it was important to limit the distance a respondent had to travel. During the period of collection from August 2009 to November 2011, 18 collection sites were visited in five Canadian regions. At the second stage, dwellings were selected. The household composition of dwellings within a selected collection site was necessary to efficiently target all age groups in the target population. The most up to date information available was used: the frame of dwellings from the 2006 Census was updated with new dwellings found since Census Day and then all of these dwellings were updated with administrative files containing more recent information about the household composition. The dwellings were next stratified sequentially according to the presence of the following age groups: 3 to 5 year olds, 6 to 11, 12 to 19, 60 to 79, 20 to 39 and 40 to 59. The number of dwellings selected per collection site (around 500) was determined based on previous response rates. The allocation of dwellings per stratum was done using simulations, taking into account the person selection strategy and the previous response rates. This ensured that the target number of people per age group and sex was reached. At the third stage, an algorithm was programmed within the collection application to select persons. Once an interviewer contacted a dwelling and entered the household composition, the algorithm determined the household members that were in-scope and used a sampling factor vector and a random number to select one or two persons within the household, depending on household composition.

Measures were put in place during collection to ensure that the objective of the survey was met without significantly exceeding the minimum total number of respondents required. Monitoring of response rates by age group and sex provided information that was used to update the sample size and allocation for each subsequent collection site. Waves of supplementary sample were selected for each site and were released when the response rates for a given site were problematic.

Several evaluations were done to determine how efficient the sample design was in reaching the CHMS objective. The use of the 2006 Census with other administrative files improved the contact information as well as the coverage of the population. The efficiency of the method of stratification was also evaluated. The results demonstrated the necessity of updating the household composition using more recent administrative data file sources. It was found that the target number of respondents by age group and sex were also reached, although the number of females was noticeably higher than males. Changes to the person selection algorithm will be made to the future cycles of CHMS in order to decrease this gap. Response rates according to distance to the MEC were also examined. The clinic response rate at the household-level dropped after the approximate driving distance reached 50km and even more so after 75km. This was also taken into account in the redesign of collection sites for cycle 3.

The sample design met the objective of cycle 2 of the CHMS. Results from cycle 2 have shown the importance of some of the key elements of the sample design, such as having the most up to date sources of information for use in stratification, performing constant monitoring and having a design that allowed for quick adjustments.

1 Objective of the survey

The Canadian Health Measures Survey (CHMS) collects key information relevant to the health of Canadians in the form of direct physical measurements. Its objective is to produce national baseline estimates on a variety of health indicators.

1.1 Data requirements

Data collection for cycle 2 of the CHMS took place from August 2009 to November 2011 and was done in two steps. First, an interviewer visited the selected dwellings to administer the household questionnaire, where information related to nutrition, smoking habits, alcohol use, medical history, current health status, sexual behaviour, lifestyle, physical fitness and demographic and socioeconomic variables was gathered. Second, respondents were asked to attend an appointment at a nearby CHMS Mobile Examination Centre (MEC), often referred to as the clinic, where trained health professionals collected blood and urine samples, physical health measurements such as blood pressure, height and weight, as well as conducted physical fitness tests. Respondents were also asked to wear a physical activity monitor (accelerometer) for a week to measure their activity levels and an air quality sampler was provided to be installed at the dwelling for a week to measure many volatile organic compounds levels.

1.2 Target population

Cycle 2 of the CHMS covers the population 3 to 79 years of age living in the ten provinces and the three territories. Excluded from the survey's coverage are: persons living on reserves and other Aboriginal settlements in the provinces; full-time members of the Canadian Forces; the institutionalized population and residents of certain remote regions. Altogether these exclusions represent less than 4% of the target population.

1.3 Sample size

To meet the objective of providing national baseline estimates on a variety of health indicators, the CHMS required that 5,700 respondents visited the MEC, equally distributed by the following 11 groups: age group 3 to 5 (both sexes combined) and age groups 6 to 11, 12 to 19, 20 to 39, 40 to 59, and 60 to 79 years old by sex. Before the survey moved into the field, this number was inflated to take into account out-of-scope dwellings and household and person-level non-response. This sample size would yield national level estimates for each group for conditions that have a prevalence of 10% or higher, with a coefficient of variation (CV) of the estimate of 16.5%.

Subsamples of the respondents were also required. Two of the subsamples were for laboratory analyses on environmental contaminants: one set of measures on urine and one set on blood. A third subsample was also collected for measures required on fasted respondents. The indoor air sampler and the activity monitor were not designed as subsamples and therefore will not be included in this sampling document. Both measures were reweighted as subsamples since adjustments were necessary to correct for the bias due to the high non-response observed.

2 CHMS Sampling Design

The CHMS used a three-stage sampling design. The first stage was the sampling of the collection sites, the second stage was the sampling of dwellings in those collection sites and the third stage was the sampling of one or two persons in each of the selected dwellings. The three stages are now explained in detail through sections 2.1 to 2.3.

2.1 First stage: Sampling of collection sites

2.1.1 Area frame creation for the collection sites

Because the CHMS required that respondents report to the MEC, they should have been able to travel to that MEC within a reasonable period of time. The area frame developed for the Labour Force Survey (LFS) was used to design and control the size of collection sites in order to accommodate these requirements. Using the LFS frame clusters (Statistics Canada no 71-256-XPB, 1998), a total of 257 collection sites were created (Laflamme and Dochitoiu, 2005). LFS clusters are small geographic units that contain approximately 200 dwellings. A collection site was a geographic area with a population of at least 10,000 people and a maximum distance from the centre of the site to its farthest boundary of 50 km in an urban area and 100 km in a rural area. Areas not meeting these criteria were excluded. The collection sites covered 96.3% of the Canadian population (Giroux and Lavigne, 2005).

2.1.2 Collection sites sample size

In the early stage of the design of cycle 1 of the CHMS, it was recommended to select a minimum of 30 collection sites or primary sampling units (PSUs) (Giroux, 2004). However, due to cost and logistical constraints, the cycle 1 sample was selected in 15 collection sites. With such a limited number of PSUs, the priority became to have as much of a nationally representative sample as possible. Therefore the 15 sites were allocated proportionally to the population size of the five standard regional boundaries used by Statistics Canada: British Columbia, the Prairies (Alberta, Manitoba and Saskatchewan), Ontario, Quebec and the Atlantic region (Newfoundland and Labrador, Prince Edward Island, Nova Scotia and New Brunswick). Only one site was allocated to the Atlantic region, which made it problematic for variance calculations since the first stage variance could not be calculated for that region. For cycle 1, the solution was to collapse the Atlantic region with Quebec for variance estimation purposes.

For cycle 2, it was necessary to add an additional site to the Atlantic region for the variance to be calculated within this region but this had to be achieved without adding costs to the survey. One option would have been to remove one of the sites allocated in Ontario or Quebec and allocate it in the Atlantic region instead but this was not the preferred option. The second option was to see if some gain in efficiency could be achieved by simply adding a site in the Atlantic region (for a total of 16 sites instead of 15) but with a reduced sample size per site in the cycle and keeping the same overall sample size. In other words, could we reduce the design effect of the survey by adding a site?

A simulation study was done using the sites from cycle 1 for which collection was completed at the time to evaluate empirically the impact of adding a site by keeping the overall sample size constant. Although the analysis could not be exhaustively examined more because of the computer intensive steps required, early results showed that for several variables examined, the CVs were reduced with the addition of one collection site. The addition of one site to the survey was accepted and was selected in the Atlantic region. Further work on the number of sites and intracluster correlation was done in 2010 after cycle 1 data collection was finalized (MacIsaac, 2011) and is discussed further in section 2.1.6.

2.1.3 Allocation of the collection sites

As mentioned previously, although only national estimates were required, the collection sites were stratified in the standard five regional boundaries to ensure a representative distribution of the sample across the country. Yellowknife, Northwest Territories was included with the Prairies and Whitehorse, Yukon was included with British Columbia. A large number of collection sites with few respondents were recommended because it would help optimize the precision of the estimates (Giroux 2007). However, the logistical and cost constraints associated with the use of the MEC restricted the number of collection sites to 16. Each site was to collect data on approximately 350 respondents, for a total of roughly 5,700 respondents. The collection sites were allocated to the regions in proportion to the size of the population (Table 2.1.3-1) and to ensure that there would be two sites in the Atlantic region.

Table 2.1.3-1: Allocation of collection sites for cycle 2 of the CHMS, by region

Region	Estimated target population, aged 3 to 79, based on 2006 Census	Collection sites by region	Collection sites in sample
Atlantic	2,130,890	36	2
Quebec	7,044,190	50	4
Ontario	11,317,075	61	6
Prairies and Yellowknife	5,034,985	77	2
British Columbia and Whitehorse	3,848,985	33	2
TOTAL	29,431,725	257	16

2.1.4 Selection of collection sites

Within each region, the list of collection sites was sorted according to whether they belonged to a census metropolitan area (CMA) and by the size of the population. A CMA is an area consisting of one or more adjacent municipalities centered on a large urban area (known as the urban core). The CMA must have a population of at least 100,000 of which 50,000 or more must live in the core. 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 also among areas with larger and smaller populations. While not every province/territory would have a collection site, the CHMS sites were chosen to represent the Canadian population, east to west, with larger and smaller population densities (Giroux and Lavigne, 2005).

2.1.5 Seasonality and order of collection sites

Data collection for CHMS is done sequentially, one collection site after another. The sites were ordered considering the temporal effect between the two years of collection such that the sites in each region were distributed equally between the two years. The logistical and cost constraints related to the MEC travelling between collection sites were considered. The seasonality by region was also considered, including the fact that the data from cycles 1 and 2 could eventually be combined to produce estimates.

2.1.6 Addition of collection sites

Part-way through collection for cycle 2, it was decided that the CHMS would make additions to its original sampling design. The two options for cycle 2 were: increase the sample in each site (thereby extending collection) or add one or two sites to the end of collection. There were many reasons for the addition or extension of collection sites. First, there was pressure to move to a calendar year for collection and analysis, as following a calendar year is easier to interpret for data users. Under the original sampling plan for cycle 2, data collection for the 16th and final site was to be completed in September 2011. With the objective to start cycle 3 in January 2012, the addition of one or two extra sites or extra sample would push the completion of cycle 2 to November or December 2011. Thus, another advantage of extending the collection would be to keep the household interviewers and clinic staff continuously working on the survey and not have any issues with a lengthy interruption in collection. The CHMS uses a group of household interviewers and clinic staff that work solely on the CHMS due to the uniqueness of the survey. Finally, an obvious advantage of an increased sample is that we would do a better job of reaching the cycle 2 clinic targets for each of the age and sex groups.

To determine if and how many extra collection sites would be added and in which geographic region these additional sites would be most beneficial, a study was performed using the CHMS data from cycle 1 (MacIsaac, 2011). In the study, two modifications were examined to determine what would have the greatest impact on the sampling variances: the effect of increasing the number of collection sites per region while maintaining the same cycle 1 sample size or the effect of increasing the number of respondents per site. The results of the study showed that increasing the number of respondents per site would only have minimal effects on the sampling variances of the variables studied and thus was not examined further.

In the study, increasing the number of collection sites was shown to have a noticeable improvement on the sampling variances, particularly in British Columbia. There was a large improvement with the addition of one site but the magnitude of the improvement was less pronounced with the addition of two or more sites. The improvement in sampling variance in British Columbia was partly explained due to the differences in the two collection sites that were originally sampled in British Columbia in cycle 1. This included the very urban area of Vancouver and the very rural area of Quesnel / Williams Lake, thus the collected data was quite different between the sites and therefore an additional site would have helped improve the variances in this region. A similar situation was present in cycle 2, with the region including a large urban area (Richmond) and a small rural area (Central and East Kootenay region). Because the population distribution in British Columbia often meant there would be one urban and one rural site in the scenario of two selected sites, it was decided that it would be very beneficial for one additional site to be added to the British Columbia region.

Because of the time available for collection at the end of 2011 with the original sampling plan, a second additional site was proposed. The results of the study did not produce a clear decision on which other region should receive an additional site. To help determine where it would be added, the original proportional allocation of the number of collection sites amongst the regions for cycle 2 was revisited. At this time, the Prairies were being under-represented in terms of the number of collection sites being allocated. Based on the population, they would have expected an allocation of 2.7 sites out of 16 but only ended up receiving 2. To properly represent this region, the second additional site was allocated to the Prairies. To allow for the collection of two additional sites during cycle 2, the collection time for each of the additional sites was to be slightly shorter in duration than for the original 16 sites, with the clinic being open for 4 to 5 weeks per site, instead of 5 to 6 in the original 16 sites.

To determine the two collection sites that would be added to cycle 2, a method of increasing the number of primary sampling units (PSUs) to a sample that had already been selected was used (Gray, 1973). Both the British Columbia and Prairie regions had originally selected two collection sites (PSUs). The method dictates that an unbiased way of selecting one additional collection site in a region is to first select two additional collection sites, as if the original intention of the sampling was to have a total of four. This is performed by using the same random number for the original sample, selected using probability proportional to size, and then modifying the sampling fraction to select a total of four sites. Once the two extra collection sites are selected in a region, one of them can be randomly removed from the sample. In British Columbia, this method resulted in the addition of Coquitlam. In the Prairies, it resulted in the addition of Calgary.

Table 2.1.6-1: List of the 18 collection sites for cycle 2 of the CHMS

Site #	Location	Prov.	Region	Collection start date (Household quest.)	Collection end date (Clinic)
1	Central and East Ottawa	ON	Ontario	August 28, 2009	October 17, 2009
2	Oakville	ON	Ontario	October 2, 2009	November 22, 2009
3	South of Brantford	ON	Ontario	November 12, 2009	January 20, 2010
4	Laval	QC	Quebec	January 5, 2010	March 3, 2010
5	South Montérégie	QC	Quebec	February 17, 2010	April 14, 2010
6	St. John's	NL	Atlantic Canada	April 1, 2010	May 26, 2010
7	Winnipeg	MB	Prairies	May 14, 2010	July 7, 2010
8	Richmond	BC	British Columbia	July 13, 2010	September 1, 2010
9	Edmonton	AB	Prairies	August 23, 2010	October 20, 2010
10	Central and East Kootenay	BC	British Columbia	October 7, 2010	December 1, 2010
11	Calgary	AB	Prairies	November 13, 2010	January 19, 2011
12	Southwest Toronto	ON	Ontario	January 6, 2011	March 9, 2011
13	Kingston	ON	Ontario	February 25, 2011	April 20, 2011
14	Colchester & Pictou Counties	NS	Atlantic Canada	April 8, 2011	June 1, 2011
15	East Toronto	ON	Ontario	May 20, 2011	July 20, 2011
16	Gaspésie	QC	Quebec	July 22, 2011	September 14, 2011
17	North Shore Montréal	QC	Quebec	September 2, 2011	October 26, 2011
18	Coquitlam	BC	British Columbia	October 14, 2011	November 30, 2011

Once the two additional sites were determined, they had to be slotted into the remaining collection schedule for cycle 2. The sites were selected when collection was taking place in Site 7 – Winnipeg. After discussions within the project team, it was decided that Calgary would become the new Site 11 and that Coquitlam would become Site 18, the final site for cycle 2. This method respected the seasonal and temporal effects within each of the regions for cycle 2, as well as for the combined cycle 1 and cycle 2. It also controlled for the extra distance that the clinic trailers would need to travel. The final list of the 18 collection sites for cycle 2 can be found in Table 2.1.6-1.

2.2 Second stage: Sampling of dwellings

2.2.1 Frame creation for the dwellings

The strategy employed in constructing the sample frame for cycle 2 of the CHMS was heavily influenced by the experience gained in cycle 1. Originally, cycle 1 was to begin in March 2006. The strategy was to first sample dwellings with respondents to cycle 3.1 of the Canadian Community Health Survey (CCHS 3.1) to reach the desired number of young people more effectively, and then to complement this with a sample of dwellings from the area frame of the Labour Force Survey (which is why the collection sites were based on this frame). Since the area frame does not contain any demographic information (date of birth and sex of individuals), a rejection technique would have been used to reduce the number of respondents in some adult age groups. This would have been a costly strategy because it would have increased the response burden of the portion of households previously interviewed for the CCHS 3.1 and it involved interviewing households that would later be rejected. Due to the postponement of cycle 1, it was possible to use the 2006 Census instead of this strategy. It is important to note here that agreements giving the CHMS access to census data were confined to the use of dwelling addresses and telephone numbers for contact purposes and to demographic information (date of birth and sex of individuals) for stratifying the sample frame; no one can be identified with the information (there was no access to the names of census respondents). Using census data lowered costs and reduced the response burden, compared with using the area frame, since the demographic information provided by the census was used to develop a sampling plan that met the sampling requirements in each age-sex group more effectively. In short, for cycle 2, the source of the dwelling frame was originally the same as for cycle 1. It was therefore constructed with 2006 Census data, updated as described in the following sections.

2.2.2 Frame updates – Adding new dwellings

As CHMS collection proceeds and we get further from Census Day, the dwelling coverage of the frame deteriorates. The Address Register (AR) was used to remedy this situation. In addition to census dwellings, the AR contains any new dwellings discovered since the census. For example, a new dwelling may be a new building, a new apartment added to an existing dwelling or an existing dwelling that was missed in the census. To add new dwellings back in cycle 1, the AR was used as soon as the newest version was available after the 2006 Census, i.e. for site 11 in cycle 1 of the CHMS. Since then, the AR has been updated quarterly.

For cycle 2, about 8.7% of the dwellings on the frame were new dwellings. The percentages of new dwellings for each of the collection sites are shown in Table 2.2.2-1. The collection sites are numbered in order of collection, i.e. in chronological order as presented in Table 2.1.6-1. It would seem logical to assume that, because of the degradation of the frame, the percentage of new dwellings would increase as the census receded in time. While that is somewhat the case, Table 2.2.2-1 shows that the percentage of new dwellings did vary substantially from site to site. Hence, the percentage of new dwellings depended not only on the degradation of the frame over time but also on the population growth at each collection site.

Table 2.2.2-1 : Percentage of new dwellings added to the census frame

Collection Site																	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
5%	11%	6%	6%	7%	10%	3%	6%	8%	10%	10%	9%	10%	14%	6%	13%	15%	13%

2.2.3 Dwelling stratification

A strategy was developed to stratify the dwellings by age group of the occupants. First, the age at the start of each site's collection period was calculated for all persons using the dates of birth in the survey frame. For the first 8 sites of cycle 2, the dwellings were stratified as follows: if a dwelling was home to at least one child aged 3 to 11, it was assigned to the "3 to 11" stratum; otherwise, if there was at least one person aged 12 to 19 in the dwelling, its stratum was "12 to 19"; and so on for strata "60 to 79", "20 to 39" and "40 to 59", in that order. This stratification procedure was used because, with their distribution in the Canadian population and the CHMS's sampling strategy, people aged 3 to 11 are more difficult to find than people aged 12 to 19, who in turn are more difficult to find than people aged 60 to 79, and so on.

From site 9 onward, the "3 to 11" stratum was further divided into a "3 to 5" stratum and a "6 to 11" stratum. This allowed for more flexibility on how the sample was allocated amongst these two age groups, thereby helping to attain the age group target sample sizes at the clinic for both. In addition, all 18 sites included an "other" stratum for other dwellings, i.e. dwellings that were out of scope because they were either vacant, had missing stratum information (missing dates of birth of the members living in the household) or because of the age of the residents (for example, a dwelling whose only residents were 80 and over).

2.2.4 Dwelling stratification update

Since the cycle 2 target population covered 3 to 5 year olds and the dwelling frame was based on the 2006 Census, a birth update was required. While the frame information prior to the stratification update was valid on May 16, 2006, CHMS collection took place much later, between August 2009 and November 2011. For cycle 2, children born after the census were 3 years old in or after May 2009 and the 2006 Census household composition could not be used through the end of collection in 2011 to identify households with children aged 3 to S.

The T1 Family File (T1FF)[1] was used to update the census household composition and to assign the new dwellings from the Address Register to a stratum. The T1FF is produced with T1 and T4 income tax documents, the Child Tax Benefit file and birth files. A new version of the T1FF is produced around mid-to-late summer of each year. For example, the 2007 T1FF, containing information dated December 31, 2007, became available in September 2009. As a result of the production cycle of the T1FF, children as young as 2 and a half could be identified when the file becomes available. Note that the T1FF is not used as a frame on its own for two reasons: its coverage is not as good as the coverage provided by the census and the AR combined and the T1FF is based on the family concept, while the CHMS uses the concept of the household.

With the T1FF, the stratification could be updated for moves, births and deaths. The T1FF was also used to assign a stratum to new dwellings and dwellings that were vacant on Census Day but were occupied at the time of CHMS collection. However, the quality of the stratification update depends on the quality of the information in the T1FF and the quality of the match between the frame and the T1FF.

Starting with Site 13, the T1 tax file was also used to update the demographic information on the frame. Based on the results of a study, it was determined that the T1 file would help improve stratification but not as well as the T1FF. As such, the first priority for frame updates was to use information from the T1FF for all dwellings that properly linked between the census frame and the tax file. For dwellings that did not link to the T1FF, an attempt was made to link the frame to the T1 file and use this information to update the frame.

2.2.5 Quality of the match between the frame and the T1FF/T1

The T1FF/T1 and the frame were matched using addresses, which were standardized beforehand using PCODE (Statistics Canada, 2011). Wherever possible, matching was done one collection site at a time. Consequently, a collection site must be

[1] The information from the administrative files such as the T1FF, T1 and AR are used to improve the efficiency of our frame (see section 3.1) resulting in better coverage and lower collection costs.

assigned to each T1FF/T1 record, which involved a complex geographic coding process (assignment of a census block to each T1FF/T1 record).

Table 2.2.5-1 presents the linkage rates, namely the percentage of dwellings in the survey frame (a frame based on the 2006 Census with the addition of new dwellings from the AR) that were linked with the T1FF and the T1 (T1 used only for Site 13 onward). The linkage rate varied considerably from site to site, as it depended heavily on the quality of the addresses. The results were generally not as good for rural sites (sites 3, 5, 10, 14 and 16). Note that when no link with the T1FF or T1 was possible, the stratum was not updated and the stratification of the dwelling was therefore based solely on census information.

Table 2.2.5-1 : Linkage percentage between the survey frame and the T1FF and the T1*

Collection Site																	
1	2	3	4	5	6	7	8	9	10	11	12	13*	14*	15*	16*	17*	18*
72%	75%	43%	71%	62%	56%	74%	57%	59%	53%	52%	49%	68%	42%	72%	57%	81%	63%

* Site 13 onward for T1

2.2.6 Dwelling sample size and allocation

The original target for cycle 2 was 5,700 respondents for the clinic component of the survey, which worked out to approximately 356 respondents per collection site. To determine the number of dwellings to sample in each collection site to reach this target, previous response rates were used from both the CHMS and the Canadian Community Health Survey (CCHS). The CHMS and CCHS were both used to calculate:

- The expected probability that a dwelling would be eligible for the CHMS (the eligibility rate)
- The expected probability that a roster of all occupants of the household would be completed (the roster rate)
- The expected probability that a selected person would respond to the household questionnaire (the questionnaire rate)

Finally, rates from the previous CHMS sites were used to calculate the expected probability that a questionnaire respondent would also be a respondent to the clinic (the clinic rate). Since rural, urban and urban core (downtown) collection sites each have distinct response rates, each site was classified into one of these three categories and the previously mentioned response rates were calculated and applied separately for each category.

Once all of the previous response rates were calculated, a simulation of 100 independent samples of dwellings was performed for the site being sampled. The goal of each of the 100 simulated samples was to use the previous response rates to predict if each sampled dwelling would result in 0, 1 or 2 people responding to the clinic. The final frame for the site was used for each simulation. The average expected number of clinic respondents for each age and sex group over the 100 independent samples was used to determine if the specified sample size and allocation were sufficient. The entire simulation of 100 independent samples was performed multiple times with varying sample sizes and allocations in order to settle on a final overall sample size and allocation strategy. The first iteration used approximate age/sex targets to come up with starting values for the overall sample size and the stratum allocation. Subsequent iterations required manually adjusting the overall sample size and stratum allocation, based on the previous iteration, to produce final values that would satisfy the clinic target counts.

A sample size of 525 dwellings was used for the first 8 sites of cycle 2, as this allowed for a reasonable workload for each of the household interviewers. The sample was allocated amongst the 6 age group strata (3-5, 6-11, 12-19, 20-39, 40-59 and 60-79), with a small portion of the sample going to an "other" stratum. The goal of selecting a sample within the "other" stratum was to ensure a more complete coverage of the target population, since dwellings on the frame can go from out-of-scope to in-scope as households move. A maximum number of 35 dwellings per site were selected in this stratum, with fewer being selected for sites that had fewer dwellings in the stratum. This stratum size helped to prevent extreme dwelling sampling weights. Starting at Site 9, the sample sizes for the remaining sites were decreased, since the average number of

clinic respondents for the first 8 sites (379) was much higher than the target of 356. This was done to ensure that the survey did not exceed the overall target at the clinic.

The goal during the allocation of sample to each of the age group strata was to allow for the best chance of meeting the age and sex clinic respondent targets for cycle 2 without going too far over. Where possible, the sample was allocated in a way that emphasized the strata where more sample was required to meet the targets. After the midway point of cycle, 2, this meant focusing on the younger age groups (3-5, 6-11 and 12-19) and also on the oldest age group (60-79).

2.2.7 Overlap control

To avoid increasing response burden, a sample overlap control process was carried out with other household surveys at Statistics Canada. Overlap control is designed to ensure that the same dwelling is not contacted by another survey within a two-year period. Overlapping dwellings were excluded from the CHMS sample and the households in those dwellings were treated as non-respondents. The sample size was increased accordingly. The following are a few details concerning the sample overlap control for the CHMS:

- For each CHMS collection site, the LFS team provided the addresses of dwellings selected for the Labour Force Survey (LFS) or any other survey that used the LFS area frame. These dwellings were therefore excluded from the CHMS sample.
- The telephone numbers of CHMS respondent households were excluded by surveys that use the random-digit-dialling frame or the telephone frame.
- Canadian Community Health Survey: Healthy Aging respondent households were excluded from the CHMS dwellings sample.

In cycle 2, no overlap control was necessary for postcensal surveys since their collection was completed two years or more before collection began for cycle 2 of the CHMS.

2.2.8 Additional sample waves

During the sampling of dwellings for each collection site, a number of smaller secondary samples were selected at the same time as the primary sample but these secondary samples were not initially sent to collection. These secondary samples were known as waves and could be released into the field when collection was going poorly in a site and more dwellings were needed to achieve the target number of clinic respondents. The waves were created to target a wide range of respondent needs by age and sex and the specific composition of the waves is shown in Table 2.2.8-1.

Table 2.2.8-1: Composition of the waves used for cycle 2 of the CHMS

Wave	Sites 1 to 16		Sites 17 and 18	
	# of dwellings	Composition	# of dwellings	Composition
0	50	All age groups	40	3 to 5, 6 to 11, 12 to 19 and 60 to 79
1	25	All age groups	24	3 to 5, 6 to 11, 12 to 19 and 60 to 79
2	/	Not used	15	6 to 11 year old males
3	/	Not used	15	6 to 11 year old females
4	25	3 to 5 year olds	25	3 to 5 year olds
5	25	6 to 11 year olds	25	6 to 11 year olds
6	25	12 to 19 year olds	25	12 to 19 year old females
7	25	20 to 39 year old males	15	12 to 19 year old females
8	25	40 to 59 year old males	15	60 to 79 year old males
9	25	60 to 79 year old males	25	60 to 79 year old males

The age and sex information used in the creation and selection of the waves was based on the final CHMS household frame. For the first 16 sites of cycle 2, Wave 0 and Wave 1 were composed of dwellings from each of the 6 age group strata, with

the only difference between the two waves being the number of dwellings in each (50 and 25, respectively). Wave 4 through Wave 9 each targeted a specific age/sex group that could require additional sample if their response rates were low and/or not enough of the group was being selected during the person-level selection stage. The wave compositions shown in Table 2.2.8-1 for Site 17 and Site 18 drastically changed so that they would focus more on the age and sex groups that were having difficulties in meeting the cycle 2 clinic targets. To allow for more options, Wave 2 and Wave 3 were also used for the first time in sites 17 and 18.

In cycle 2, waves were released to collection in 6 sites: South of Brantford (Site 3), Southwest Toronto (Site 12), Colchester & Pictou Counties (Site 14), East Toronto (Site 15), Montréal North Shore (Site 17) and Coquitlam (Site 18). The results of collection for the waves are shown in Table 2.2.8-2. The wave that focused on 60 to 79 males was the most common one released, in a total of 4 sites. Slightly less than 50% of the clinic respondents from these three waves (22/45) were males in the 60 to 79 age group.

Table 2.2.8-2: Summary of the waves released in cycle 2

Site	Site location	Age group focus	Number of dwellings		House-holds rostered	Number of persons			
			Selected	In-scope		Selected	Quest. responses	Clinic responses	Responses to clinic in target age group
3	South of Brantford	3 to 5 M+F	25	24	17	29	23	18	5
3	South of Brantford	6 to 11 M+F	25	25	20	40	34	32	13
12	Southwest Toronto	All ages	25	25	14	16	14	11	
12	Southwest Toronto	60 to 79 M	25	24	15	15	14	10	7
14	Colchester & Pictou	60 to 79 M	25	23	20	20	19	15	6
15	East Toronto	All ages	25	25	16	18	15	10	
15	East Toronto	60 to 79 M	25	23	14	16	11	10	4
17	Montréal North Shore	3 to 5 M+F	25	25	16	30	30	22	9
17	Montréal North Shore	60 to 79 M	15	15	11	11	11	10	5
18	Coquitlam	3 to 5 M+F	25	25	18	32	31	28	9
Total - 6 sites		Various	**240**	**234**	**161**	**227**	**202**	**166**	

2.3 Third stage: Sampling of persons

2.3.1 Person sampling

When the household interviewer made contact with a sampled dwelling, the goal was to create a roster for the household. A roster is a list of all persons residing in the household and includes pertinent information such as age, sex and whether the individual works full-time for the Canadian Forces. With this information, the computer application randomly selected one or two persons to take part in the remaining part of the survey. The number of persons selected depended on the composition of the household:

- If there was at least one child between the ages of 3 and 11, two people were selected: one child between the ages of 3 and 11 and one other person between the ages of 12 and 79.
- If there were no children between the ages of 3 and 11, only one person in the 12 to 79 age group was selected.
- If there was no one eligible for the survey, no one was selected. This included households where all in-scope persons were under the age of 3, over the age of 79 and / or were full-time members of the Canadian Forces.

When the roster was completed, the computer application assigned a sampling factor to each eligible member of the household and this information was used to determine the probability of selection. The sampling factor assigned to each individual was based on the stratum that the dwelling was sampled from and the age of the person. The sampling factors shown in Table 2.3.1-1 were used for the majority of sites in cycle 2. They varied between age groups in order to do a better job of reaching the clinic targets for each age group. Small increases were made to the sampling factors for the 12 to 19 and 60 to 79 age groups for the last three sites of cycle 2 and for the 3 to 5 age group for the final site in order to increase the number of people selected in these groups. In households where two people were selected, the selection of the child (aged 3-11) was done independently of the person aged 12 to 79. A full explanation of the process used for person-level selection and a detailed example can be found in Appendix A.

Table 2.3.1-1: Sampling factors used during person-level respondent sampling for cycle 2

Stratum on frame	Sampling factors for 3 to 11		Sampling factors for 12 to 79			
	3 to 5	6 to 11	12 to 19	20 to 39	40 to 59	60 to 79
3 to 5	3	2	5	1	1	1
6 to 11	3	2	5	1	1	1
12 to 19	3	2	5	1	1	1
20 to 39	3	2	5	2	1	1
40 to 59	3	2	5	1	2	1
60 to 79	3	2	5	1	1	2
Out-of-scope	3	2	5	1	1	1

Households with a large number of members can have very large survey weights and therefore be overly influential when calculating survey estimates. To prevent this, two reset values were put in place that would reset some of the sampling factors to 1 when certain conditions were met. In both cases described below, each member of the 12 to 79 age group would have had an equal chance of being selected:

1. If a household contained six or more people in the 12 to 79 age group, all sampling factors for the 12 to 79 members were reset to 1.
2. If a household contained three or more people in the 12 to 19 age group, all sampling factors for the 12 to 79 members were reset to 1.

2.3.2 Subsampling for contaminants in blood and urine

Subsamples of the respondents were selected for laboratory analysis of specific environmental chemicals in their urine or blood. The entire sample was not used because of the high cost of the laboratory tests. The two subsamples on environmental measures were selected independently; that is, without consideration of who was selected for testing for the other group of environmental chemicals. This means that a subsampled respondent could be selected for one or both subsamples.

The target sample sizes by age group and sex for the subsamples performed on the urine and blood samples are shown in Table 2.3.2-1. The urine subsample had an overall target of 2,500 and no changes were made to this. The initial target sample size for the blood subsample was 250 respondents in each age category by sex, for a total of 1,000 respondents. Prior to Site 11, an extra 500 was added to the target for the 20 to 79 age category, split evenly between the males and females, bringing the final total to 1,500.

Table 2.3.2-1: Target counts for the urine and blood subsamples for contaminants

Age	Sex	Urine subsample target	Blood subsample target
3 to 5	Male	250	
	Female	250	
6 to 11	Male	250	
	Female	250	
12 to 19	Male	250	250
	Female	250	250
20 to 79	Male	500	500
	Female	500	500
Total 3 to 79	Male	**1,250**	**750**
	Female	**1,250**	**750**
	Both	**2,500**	**1,500**

2.3.3 Subsampling for fasted clinic appointments

Subsampling was also performed for measures that required the respondent to be fasted before their clinic appointment. Each sampled dwelling was randomly flagged to indicate whether a respondent should be fasted for the clinic appointment. This required that respondents fast overnight, whereas shorter eating restrictions were imposed on those with a non-fasted appointment. Pregnant women, people with diabetes, 3 to 5 year olds and people with special situations were not asked to fast, even if they had been selected for a fasting appointment.

3 Survey results

3.1 Frame quality

3.1.1 Accuracy of contact information – Addresses

The CHMS frame for cycle 2 consisted of addresses, telephone numbers and other sampling information for all dwellings located in the 18 sampled collection sites. The final addresses came from one of three possible sources: the census questionnaire, the census contact information (known commonly as the Master Control System or MCS) or the Address Register (AR). The addresses from the census questionnaire were standardized using the software PCODE before an attempt was made to use them, in order to deal with scanning errors or any extraneous information that was included in the address field on the census. Census questionnaire addresses that were found to be invalid using PCODE were not used. As the census was the primary source of addresses for the CHMS frame, a valid address from either the census questionnaire itself or the MCS was used prior to attempting to use an address from the AR. A quick summary of the procedures used in determining which source of addresses to use is presented in Table 3.1.1-1. Priority was given to addresses from the census questionnaire.

To be kept on the frame and therefore be eligible for sampling, a dwelling required at least a civic number and a street name. Of the dwellings on the frame that had enough information to be kept on the frame and therefore be eligible for sampling, approximately 2.7% had a final address that contained only a civic number and a street name, as the city and postal code were not available. This occurred most frequently in the rural sites, such as Colchester and Pictou Counties, where 29.9% of the eligible frame contained only civic number and street name, and South of Brantford, where it was 18.6% of the eligible frame. Sampled addresses that were missing their city and postal code were manually imputed using external sources such as PCODE and Canada 411. Frequently the city and postal code were available from the non-standardized census questionnaire address on the frame that had not been used due to it having been deemed invalid during the standardization process. Overall for cycle 2, 48.2% of the sampled addresses were from the census questionnaire, 38.0% were from the MCS and 13.8% were from the AR.

Table 3.1.1-1: Source of addresses for the sampling frame for cycle 2 of the CHMS

Address Description	Source of Addresses
Vacant during 2006 census and has a complete address from MCS	MCS
Vacant during 2006 census and does not have a complete address from MCS	Address Register
Valid address from census questionnaire	Census questionnaire
Invalid address from census questionnaire; complete address from MCS	MCS
Invalid address from census questionnaire and MCS	Address Register
Growth on the frame since 2006 Census	Address Register

Sometimes during household collection, an interviewer would be unable to find the address of a sampled dwelling. These problematic addresses were transferred to the Head Office, where attempts were made to determine if there was any additional information available to help locate the address. For these problematic addresses in cycle 2, approximately 60% had a sampled address that came from the census (either the MCS or the census questionnaire) and approximately 40% came from the AR. The process of seeking additional information involved using multiple resources, such as other address sources from the frame that were not used for the final address, Canada 411, Google Maps and other relevant information found from municipal websites or other Internet searches. A similar process was used when an interviewer discovered a potential multiple dwelling. The information for the potential multiple was transferred to the Head Office to determine if the second dwelling was already on the frame. If it was not present on the frame, the dwelling was considered a multiple and both were contacted to take part in the survey.

3.1.2 Accuracy of contact information – Telephone numbers

A complete address was provided to interviewers for each sampled dwelling. When available, a telephone number was also provided, which could be used to help the interviewer contact the dwelling or to verify the address. Similar to addresses, the three potential sources for telephone numbers were the census questionnaire, the MCS and the AR. The telephone number from the AR took precedence, as the information on this file was the most up-to-date. Telephone numbers from the MCS and the census questionnaire were used for dwellings on the census that did not link with the AR or when there was a link but the AR did not contain a telephone number. Of the dwellings that had a civic number and a street name and therefore were eligible for sampling, a telephone number was available for 84.6% of the total frame and 89.1% of the sampled dwellings. The rate was almost 5% higher in the sample since there was always a low sampling fraction in the "other" stratum, where frame information such as telephone numbers was more frequently missing.

Table 3.1.2-1: Source and accuracy of frame telephone numbers for responding households

Source of telephone number	Total (Responding households)		Correctly matched	
	Count	%	Count	%
Census	1,603	27.4%	634	39.6%
Address Register	3,776	64.5%	2,966	78.5%
No telephone	475	8.1%	-	-
Total	**5,854**	**100.0%**	**3,600**	**61.5%**

Table 3.1.2-1 presents a breakdown of the source of telephone number for the households that responded to the household interview in cycle 2 and whether or not the telephone number from the frame was correct. A 2-person household was considered as a responding household if at least one interview was completed. When verifying if the telephone number was correct for 2-person households, both telephone numbers (if provided) were compared to the telephone number from the frame. If one of the provided telephone numbers matched the frame, then the frame telephone number was deemed to be correct. The majority of responding households had a telephone number from the AR (64.5%). The telephone numbers from the AR also had a much higher accuracy rate (78.5% correct), compared to the census (only 39.6% correct). 8.1% of the responding households did not have a telephone number available on the frame.

3.1.3 Efficiency of stratification

Because of the high costs of the MEC and the laboratory analyses, it was especially important to ensure that the required number of respondents was obtained in each age-sex group without increasing the sample size unnecessarily. Consequently, stratification efficiency was vital for the CHMS. Stratification was efficient if a dwelling's frame-derived stratum—i.e. the stratum used in the sampling of dwellings—was the same as its "actual" stratum. The more efficient the stratification, the easier it would be to reach the targeted sample size for each of the age groups in the population of interest.

Neither the census nor the AR could be used to assign a new or vacant dwelling to a stratum. Therefore it is clearly beneficial to assign them to a stratum using the T1FF or T1. Yet for other dwellings that can be assigned to a stratum using the census, the update strategy depended on the quality of the T1FF and T1.

Before the start of cycle 2, there was a preliminary study of respondents using the last few sites of cycle 1 of the CHMS to assess the quality of the T1FF in terms of stratification efficiency. To assess the quality of the T1FF, the stratification efficiency that would have been obtained with the 2006 T1FF was compared with the efficiency that was obtained using only the 2006 Census. The study showed that the 2006 T1FF was no better or worse than the 2006 Census for stratification efficiency. It was therefore decided to take a cautious approach for the first few sites in cycle 2 and confine the stratification update to dwellings that were vacant on the census, dwellings that were constructed after the census date or which were missed during the census, or dwellings that contained a 3 to 5 year old according to the T1FF.

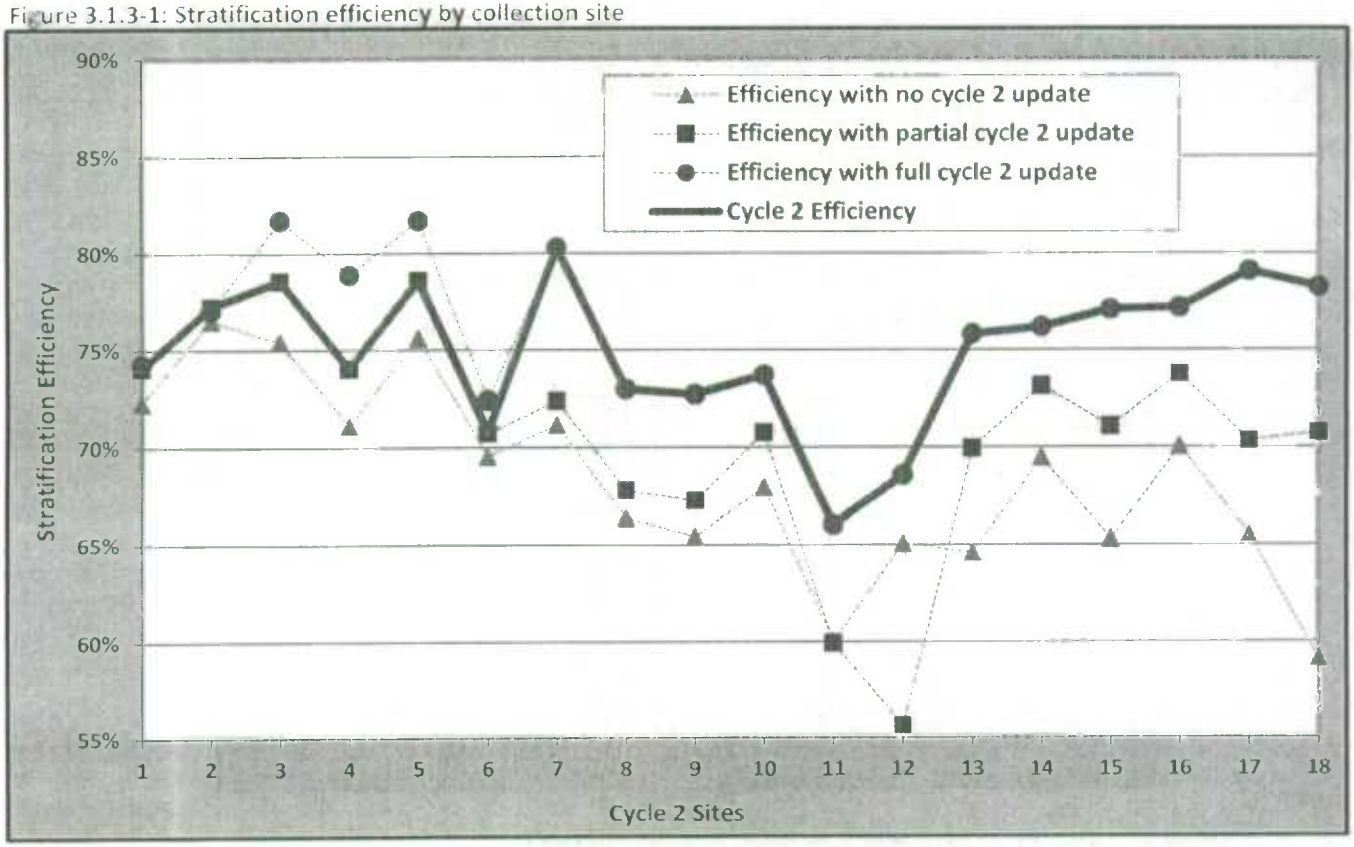
However, as the time gap between the census and data collection grew and more up-to-date versions of the T1FF became available, it was reasonable to assume that it would be more advantageous to use the T1FF. The 2007 T1FF became available at the time the dwellings were selected for site 3. The stratification efficiency was assessed again using the results of collection sites 3 and 4. This took until the end of collection for site 4, more than four months after the release of the 2007 T1FF. The results of the study showed that an update based on the 2007 T1FF would have improved the stratification efficiency for sites 3 and 4 not only in the "3 to 11" stratum, but also in all of the other strata. The decision was therefore made to carry out updates using the T1FF for all dwelling strata when the sample was selected for site 7. A similar study was performed using the T1 file just past the midway point of cycle 2 and a similar conclusion was made. Starting in Site 13, the primary source to update the frame was the T1FF. When dwellings did not link to the T1FF, an attempt was made to link to the T1 and information was used from it when a link was successful. Some results are presented in the stratification efficiency charts below and more details about the stratification updating process are provided in Giroux, Labrecque and Quigley (2010).

Stratification efficiency by collection site

Figure 3.1.3-1 shows the stratification efficiency by collection site for cycle 2 of the CHMS. The stratification efficiency rate is the percentage of dwellings (excluding dwellings in the "other" stratum) whose frame-based stratum was the same as their "actual" stratum. Since the sites are numbered chronologically, the chart shows how the efficiency changes over time as collection progresses. It is important to note, however, that the stratification efficiency also depended on the population growth at each site. Since each collection site was different, the data shown are not, strictly speaking, a time series.

The continuous blue line represents the efficiency rate for each collection site in cycle 2. The broken green line (joining the triangles) represents the census-based stratification efficiency rate that would have been obtained without updating the stratification with the T1FF / T1. For cycle 2, the overall efficiency rate was 75.2%; it would have been 68.6% without the T1FF and T1 file. Therefore, the increase in stratification efficiency associated with using the T1FF or T1 was 6.6 percentage points.

Figure 3.1.3-1: Stratification efficiency by collection site



The broken pink line (joining the squares) represents the efficiency rate with a partial update for dwellings in all of the strata. The stratification was updated only when there was a person aged 3 to 5 in the household, according to the T1FF or when the dwelling was vacant at the time of the census or was growth since the census. This type of update was performed for the first six collection sites, and as a result, the blue and pink lines coincide for the first six sites. The figures for site 7 show that carrying out a full stratification update with the T1FF was very beneficial, as the efficiency gain over the partial update was almost 8 percentage points.

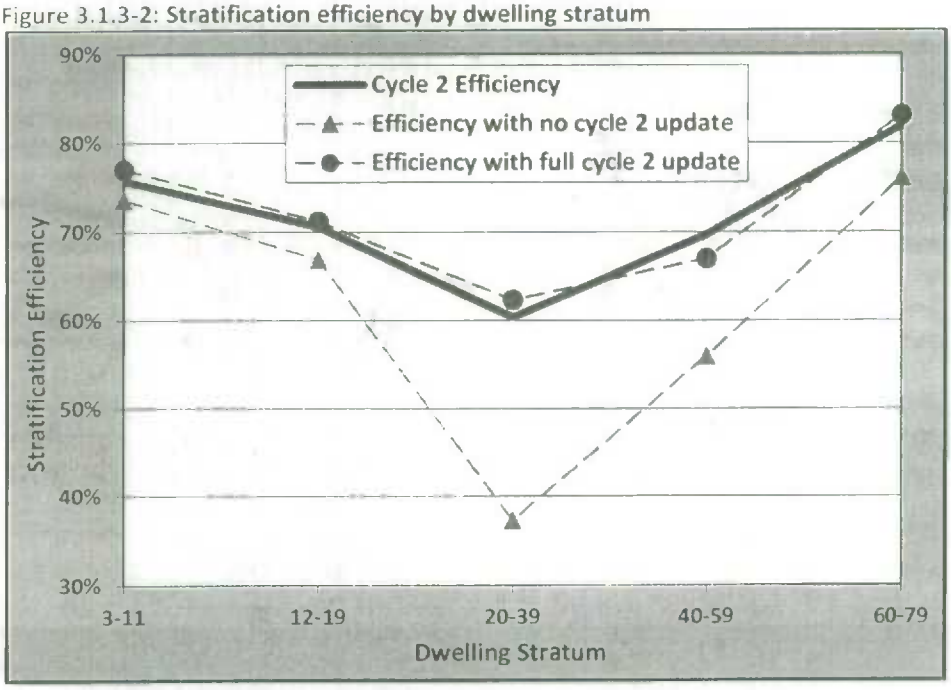
Finally, it is worth examining what the results would have been if a full update had been performed at the beginning of cycle 2. The broken orange line (joining the circles) shows the stratification efficiency rate with a full update based on the T1FF. Since this method was used starting in site 7, the blue and orange lines coincide from site 7 onward. The full update became worthwhile only after the 2007 T1FF came into play, starting with site 3. However, that observation could not be made until collection was completed for sites 3 and 4, just prior to the frame update for site 7. Having the ability to use the most recent T1FF seemed to make a significant difference, as the efficiency gain between the census with no update and the census with a full update improved with site 3 when the 2007 T1FF became available, with site 13 when the 2008 T1FF became available and with site 18 when the 2009 T1FF became available. Part of the increase observed with site 13 was due to the use of the T1 as a secondary source of demographic updates.

Stratification efficiency by dwelling stratum

Figure 3.1.3-2 shows the stratification efficiency by dwelling stratum for the collection sites in cycle 2. As in the previous figure, the continuous blue line represents cycle 2, the broken green line (joining the triangles) represents the census only, and the broken orange line (joining the circles) represents the full update. A comparison of the blue and green lines shows that updating with the T1FF or T1 resulted in an efficiency gain for all strata. The impact on the stratification efficiency was particularly large for dwellings in the "20 to 39" stratum, which had a 23 percentage points efficiency gain following the stratification update with the T1FF or T1 in cycle 2. The orange line in the chart shows that the full update at the beginning

of cycle 2 would have been beneficial for all strata except perhaps the "40 to 59" stratum (in part because of sites 1 and 2, for which the full update would not really have been beneficial).

Figure 3.1.3-2: **Stratification efficiency by dwelling stratum**



3.2 Number of respondents and response rates by collection site

The cycle 2 roster, questionnaire and clinic response rates are shown in Table 3.2-1. A rostered household was one in which the interviewer obtained the demographic composition of the household in the sampled dwelling and was able to select 1 or 2 people to take part in the survey. The household roster rate was calculated as the number of rostered households divided by the number of eligible dwellings in the sample. A dwelling was considered ineligible if it could not be located, it was vacant, it was not a primary residence or it contained only people that were not a part of the target population (such as 80+ years of age or full-time members of the Canadian Forces only). The overall roster rate for cycle 2 was 75.9%. The sites with the highest roster rates (Site 10 – Central and East Kootenay, Site 14 – Colchester and Pictou Counties and Site 16 - Gaspésie) were all smaller, rural sites. The sites with the lowest roster rates (Site 4 - Laval, Site 12 – Southwest Toronto, Site 15 – East Toronto and Site 18 - Coquitlam) were all major urban areas in Toronto, Montreal and Vancouver.

Of those that were selected to take part in the survey, the overall questionnaire response rate was 90.5%. The questionnaire rates were very consistent throughout cycle 2, with only four sites having a rate that deviated more than 3% from the average. These included two sites on the low side (Site 3 - South of Brantford at 86.8% and Site 15 – East Toronto at 85.2%) and two sites on the high side (Site 4 - Laval at 94.1% and Site 16 - Gaspésie at 93.6%).

For those that responded to the questionnaire, the clinic response rate was much more varied amongst the collection sites. The overall clinic rate was 81.7%. Six sites came in with a rate less than 80%, with Site 11 – Calgary coming in with the lowest rate at 75.6%. There were four sites with clinic response rates of 85% or higher, with Site 8 – Richmond coming in with the highest rate at 87.5%.

The blood and urine donation rates are shown in Table 3.2-2. The blood donation rate was fairly stable throughout the cycle, not straying too far from the overall cycle 2 proportion of 95.8%. Only Site 15 – East Toronto was more than 3% away from the average (at 91.5%). The urine donation rate was even more consistent throughout the cycle, with no sites being more than 1.4% away from the overall proportion of 98.9%.

Table 3.2-1: Household and person-level response rates by collection site

Site	Eligible dwellings	Household level		Person-level				
		Rostered	Roster rate	Selected	Responses to questionnaire	Quest. response rate	Responses to clinic	Clinic response rate
Canada	8,520	6,465	75.9%	8,652	7,830	90.5%	6,395	81.7%
1	490	392	80.0%	513	455	88.7%	386	84.8%
2	505	385	76.2%	523	460	88.0%	375	81.5%
3	521	386	74.1%	538	467	86.8%	360	77.1%
4	495	349	70.5%	471	443	94.1%	379	85.6%
5	480	348	72.5%	465	433	93.1%	338	78.1%
6	479	381	79.5%	505	457	90.5%	373	81.6%
7	493	397	80.5%	559	508	90.9%	409	80.5%
8	476	362	76.1%	513	473	92.2%	414	87.5%
9	484	370	76.4%	506	462	91.3%	376	81.4%
10	467	385	82.4%	506	464	91.7%	387	83.4%
11	427	319	74.7%	408	369	90.4%	279	75.6%
12	467	311	66.6%	393	353	89.8%	274	77.6%
13	439	338	77.0%	449	409	91.1%	344	84.1%
14	451	384	85.1%	500	453	90.6%	385	85.0%
15	538	360	66.9%	459	391	85.2%	319	81.6%
16	354	309	87.3%	407	381	93.6%	301	79.0%
17	452	340	75.2%	458	419	91.5%	356	85.0%
18	502	349	69.5%	479	433	90.4%	340	78.5%

Table 3.2-2: Blood and urine donation counts and rates by collection site

Site	Clinic responses	Blood donations	Blood donation % (out of clinic)	Urine donations	Urine donation % (out of clinic)
Canada	6,395	6,128	95.8%	6,327	98.9%
1	386	374	96.9%	384	99.5%
2	375	356	94.9%	373	99.5%
3	360	339	94.2%	358	99.4%
4	379	363	95.8%	378	99.7%
5	338	322	95.3%	334	98.8%
6	373	360	96.5%	369	98.9%
7	409	392	95.8%	402	98.3%
8	414	402	97.1%	412	99.5%
9	376	366	97.3%	372	98.9%
10	387	378	97.7%	379	97.9%
11	279	275	98.6%	275	98.6%
12	274	264	96.4%	273	99.6%
13	344	332	96.5%	339	98.5%
14	385	368	95.6%	379	98.4%
15	319	292	91.5%	315	98.7%
16	301	285	94.7%	300	99.7%
17	356	334	93.8%	347	97.5%
18	340	326	95.9%	338	99.4%

3.3 Comparing cycle 1 and cycle 2 respondent counts and response rates

The counts in Table 3.3-1 show a comparison of the final clinic and blood donation counts for cycle 1 and cycle 2. To produce quality estimates for each age by sex group, the primary target was approximately 500 respondents at the clinic. The lone exception was for the 3 to 5 age group, where the primary target was having 500 total blood donations (both sexes combined). This exception was due to not requiring final estimates broken down by sex for the 3 to 5 year olds. The secondary target for the 6 to 79 age groups was 500 blood donations by age group and sex, though cost constraints meant that this goal was not fully attained.

The primary target of 500 clinic respondents by age group and sex was achieved for all groups. The greatest challenge was for the 60 to 79 year old males. To reach the target for the 60 to 79 year old males, waves of additional sample that targeted this group were released in four of the last seven sites. The primary target of 500 blood donations for the 3 to 5 year olds was also achieved, due also in part to releasing waves during the final two sites. To increase the chance of reaching targets in these difficult groups, emphasis was also made during the initial sampling process to select more dwellings in their respective strata.

For the 6 to 79 year old age groups, the secondary target of 500 blood donations by age group and sex was achieved for seven out of the ten age / sex groups. The three age / sex groups that came in below the blood donation target were 12 to 19 females and both the male and female 6 to 11 year olds. These results were consistent with those from cycle 1, where all three of these groups finished with less than 500 blood donations.

The number of female clinic respondents in cycle 2 (3,319) was noticeably greater than the number of male clinic respondents (3,076). This was similar to what was collected in cycle 1 (2,898 females vs. 2,706 males). To help narrow the gap between the sexes, the person-level selection method has been changed for cycle 3 to take sex into account when assigning the sampling factors, which impact the final probabilities of selection. The sampling factors in the new method will help to decrease the number of 20 to 39 females selected and partition this extra sample to other groups. The new method will also allow for adjustments throughout cycle 3 if collection is having a difficult time in getting respondents for a specific age/sex group.

Table 3.3-1: Comparison of the number of respondents by age group and sex between cycles 1 and 2

Age	Sex	Clinic Counts		Blood Donation Counts	
		Cycle 1	Cycle 2	Cycle 1	Cycle 2
3 to 5	Male		308		264
	Female		293		254
6 to 11	Male	542	534	471	494
	Female	532	541	467	480
12 to 19	Male	514	550	490	529
	Female	499	510	464	479
20 to 39	Male	524	556	514	553
	Female	661	773	655	763
40 to 59	Male	582	619	578	619
	Female	654	613	648	608
60 to 79	Male	544	509	538	508
	Female	552	589	548	577
Total 3 to 79	Male		3,076		2,591
	Female		3,319		2,782
	Both		6,395		6,128
Total 6 to 79	Male	2,706	2,768	2,591	2,703
	Female	2,898	3,026	2,782	2,907
	Both	5,604	5,794	5,373	5,610

When compared to cycle 1, many of the age by sex groups obtained similar counts for cycle 2. Having samples from 18 collection sites in cycle 2, compared to the 15 sites in cycle 1, helped lead to an overall increase in counts at the clinic and for the blood donations. The addition of the 3 to 5 year olds in cycle 2 led to a notable increase in the number of 20 to 39 year old females taking part in the survey. The 40 to 59 year old females and the 60 to 79 year old males were the only age groups that showed noticeable decreases in clinic counts from cycle 1 to cycle 2.

A comparison of the person-level questionnaire, clinic, blood and urine response rates between cycle 1 and cycle 2 are shown in Table 3.3-2. The increase in questionnaire response rates was fairly consistent for the age / sex groups and led to an overall increase from 88.3% in cycle 1 (ages 6-79) to 90.5% in cycle 2 (ages 3-79). The opposite is true of the clinic response rates, as the decrease in rates was fairly consistent for the age / sex groups. When focusing only on the common age groups between the two cycles (6 to 79 year olds), the clinic response rate dropped from 84.9% in cycle 1 to 81.9% in cycle 2. The clinic response rate for the 3 to 5 year olds was quite low comparatively, at 79.5%, and caused the overall clinic response rate in cycle 2 to drop to 81.7%.

Table 3.3-2: Comparing questionnaire, clinic, blood and urine response rates between cycle 1 and cycle 2

Age	Sex	Questionnaire Response Rate		Clinic Response Rate		Blood Donation Rate		Urine Donation Rate	
		Cycle 1	Cycle 2	Cycle 1	Cycle 2	Cycle 1	Cycle 2	Cycle 1	Cycle 2
3 to 5	Male		91.6%		80.2%		85.7%		94.2%
	Female		93.7%		78.8%		86.7%		96.6%
	Both		92.6%		79.5%		86.2%		95.3%
6 to 11	Male	90.3%	92.7%	87.0%	84.0%	86.9%	92.5%	97.6%	99.6%
	Female	91.5%	94.2%	85.1%	84.9%	87.8%	88.7%	97.6%	98.3%
	Both	90.9%	93.5%	86.1%	84.4%	87.3%	90.6%	97.6%	99.0%
12 to 19	Male	87.5%	89.6%	88.3%	84.2%	95.3%	96.2%	98.4%	98.7%
	Female	86.9%	90.7%	86.5%	86.0%	93.0%	93.9%	98.2%	98.6%
	Both	87.2%	90.1%	87.4%	85.1%	94.2%	95.1%	98.3%	98.7%
20 to 39	Male	82.6%	85.1%	84.1%	79.1%	98.1%	99.5%	98.9%	99.5%
	Female	89.3%	88.8%	83.2%	81.3%	99.1%	98.7%	99.8%	99.9%
	Both	86.2%	87.2%	83.6%	80.4%	98.6%	99.0%	99.4%	99.7%
40 to 59	Male	84.1%	87.6%	82.7%	81.3%	99.3%	100.0%	99.5%	99.5%
	Female	89.4%	91.6%	85.3%	82.9%	99.1%	99.2%	99.1%	100.0%
	Both	86.8%	89.5%	84.0%	82.1%	99.2%	99.6%	99.3%	99.8%
60 to 79	Male	90.7%	92.3%	88.2%	81.2%	98.9%	99.8%	99.6%	99.6%
	Female	91.3%	92.3%	79.8%	76.1%	99.3%	98.0%	99.5%	99.0%
	Both	91.0%	92.3%	83.7%	78.4%	99.1%	98.8%	99.5%	99.3%
Total 3 to 79	**Male**		**89.4%**		**81.7%**		**96.5%**		**98.9%**
	Female		**91.5%**		**81.6%**		**95.2%**		**99.0%**
	Both		**90.5%**		**81.7%**		**95.8%**		**98.9%**
Total 6 to 79	**Male**	**86.8%**	**89.2%**	**85.9%**	**81.9%**	**95.8%**	**97.7%**	**98.8%**	**99.4%**
	Female	**89.7%**	**91.3%**	**83.9%**	**81.9%**	**96.0%**	**96.1%**	**98.9%**	**99.2%**
	Both	**88.3%**	**90.3%**	**84.9%**	**81.9%**	**95.9%**	**96.8%**	**98.9%**	**99.3%**

The blood donation and urine donation rates were consistently higher in cycle 2 when focusing only on the common age groups (6 to 79 year olds). The blood donation rates for the 6 to 79 year olds increased from 95.9% to 96.8%. The urine

donation rates for the 6 to 79 year olds increased from 98.9% to 99.3%. Similar to the clinic response rates, the blood and urine response rates for the 3 to 5 year olds were lower than the rates for the 6 to 79 year olds and led to bringing down the cycle 2 blood and urine response rates to 95.8% and 98.9%, respectively. However it is worth noting that the final blood donation rate for the 3 to 5 year olds (86.2%) was much higher than the anticipated value of 80% that was expected prior to the start of cycle 2. This high response rate helped in reaching the target of 500 blood donations.

3.4 Number of respondents for the subsamples

The results of the urine and blood subsamples are shown in Table 3.4-1. The goal for the urine subsample was to obtain 250 responses for each age group by sex, except for the 20 to 79 group, where the goal was 500 responses by sex. The target was achieved for each of the groups.

For the blood subsample, the original goal was to obtain 250 responses for both the 12 to 19 and 20 to 79 age groups by sex. Just after the mid-point of collection for cycle 2, an extra 500 sample was allocated to this subsample and this was evenly split between the males and females of the 20 to 79 age group, bringing the final goal for this age group to 500 by sex. Once again, the target was achieved in each of the groups.

Table 3.4-1: Number of respondents for the urine and blood subsamples

Age	Sex	Urine subsample		Blood subsample	
		Target	Actual	Target	Actual
3 to 5	Male	250	280		
	Female	250	264		
6 to 11	Male	250	262		
	Female	250	254		
12 to 19	Male	250	256	250	254
	Female	250	257	250	253
20 to 79	Male	500	504	500	511
	Female	500	506	500	506
Total 3 to 79	**Male**	**1,250**	**1,282**	**750**	**765**
	Female	**1,250**	**1,281**	**750**	**759**
	Both	**2,500**	**2,563**	**1,500**	**1,524**

The final subsample was for fasted respondents. The goal for this subsample was simply to have 50% of the eligible clinic respondents fasted before their clinic appointment. Those ineligible for fasting included 3 to 5 year olds, diabetics and pregnant women. Since it was known that not all respondents selected for a fasted appointment would actually fast, the percentage of households that were initially selected as part of the fasting subsample was closer to 55%. The results of the fasted subsample are shown in Table 3.4-2. Overall, 55.6% (3,223 / 5,794) of the eligible clinic respondents were selected to take part in the fasted subsample. Since the laboratory measures on the fasted subsample were done on blood, only fasted respondents that successfully donated blood and had at least one valid lab measure on the variables related to fasting status were included on the final subsample file. 86.7% of those selected to fast did fast and had at least one valid lab measure, resulting in an overall fasting rate of 48.2% (2,793 / 5,794). The reason why this final rate was below the 50% target was due to the fasted respondents that did not provide a blood sample or did not have at least one valid lab measure.

3.5 Response rates at the household-level by approximate driving distance

The CHMS requires each respondent to travel to the clinic. As such, it is important that respondents do not have an overly long distance to actually get to the clinic. Table 3.5-1 contains the questionnaire and clinic response rates at the household level broken down by the approximate distance required to attend the clinic.

Table 3.4-2: Number of respondents and fasting rate for the fasted subsample by collection site

Site	Clinic Respondents (6 to 79)	Not selected to fast	Selected to fast			
			Total	Actually fasted with at least 1 valid lab measure	% of those selected to fast that did fast	% fasted of total respondents
1	357	147	210	197	93.8%	55.2%
2	341	149	192	166	86.5%	48.7%
3	323	150	173	158	91.3%	48.9%
4	340	133	207	171	82.6%	50.3%
5	308	136	172	152	88.4%	49.4%
6	343	156	187	169	90.4%	49.3%
7	369	163	206	173	84.0%	46.9%
8	378	196	182	160	87.9%	42.3%
9	331	152	179	151	84.4%	45.6%
10	356	144	212	183	86.3%	51.4%
11	255	113	142	113	79.6%	44.3%
12	249	117	132	122	92.4%	49.0%
13	311	133	178	151	84.8%	48.6%
14	357	161	196	178	90.8%	49.9%
15	293	128	165	143	86.7%	48.8%
16	274	123	151	114	75.5%	41.6%
17	316	135	181	155	85.6%	49.1%
18	293	135	158	137	86.7%	46.8%
Total	5,794	2,571	3,223	2,793	86.7%	48.2%

For most dwellings, the approximate distance was calculated based on the distance between the spatial coordinates of the location of the clinic and the spatial coordinates of the LFS cluster that contained the selected address. This distance is the "as the crow flies" direct distance between the clinic and the selected dwelling and does an acceptable job of predicting driving distance for the more urban sites due to the greater number of roads available for travel. Using this approximation of driving distance for the more rural sites is more problematic, as fewer road options can lead to severely underestimating the driving distance. For the towns in rural sites where this issue existed, the approximate driving distances were determined based on a Google Map search between the actual address of the clinic and the listed city for the address. The response rates in Table 3.5-1 are at the household level, such that if there was at least one person responding to the questionnaire or clinic, the household was considered a respondent. This was used to avoid double-counting the 2-person households where both selected people responded. A graphical representation of the questionnaire and clinic response rates at the household-level is shown in Figure 3.5-1.

The questionnaire response rates at the household-level for people who had approximate driving distances less than 75 km were consistently close to the cycle 2 average of 68.9%. For dwellings that were between 75 km and 100 km, the questionnaire response rate at the household-level decreased to 63.0%. The rate surprisingly increased to 79.1% for the dwellings that were greater than or equal to 100 km away from the clinic and this group had the highest questionnaire rate. Since it was the interviewer that had to travel to the selected dwellings to perform the household interview, we would expect the probability of questionnaire response to not be significantly affected by how far away a dwelling was from the clinic.

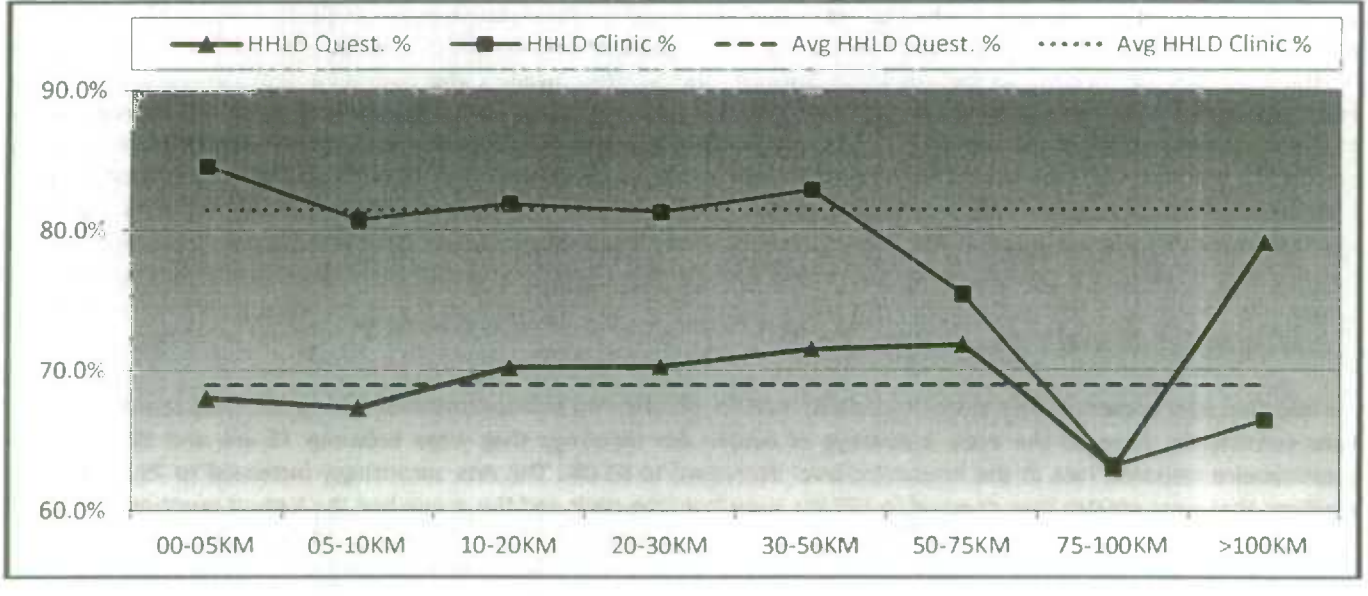
For the clinic response rates at the household-level, it was expected that greater driving distances would lead to lower response rates and this was observed in cycle 2. Dwellings with approximate driving distances less than 50 km had their clinic response rates at the household-level in the low 80%, close to the overall average of 81.4%. There was a decrease in the rate as the distance increased to between 50 km and 75 km (at 75.4%) and then an even greater decrease for the dwellings that were between 75 km and 100 km (at 63.2%) and greater than 100 km (at 66.3%). Due to these results, the

requirement of driving distance was modified for the collection sites for cycle 3. For cycle 1 and cycle 2, the maximum distance between the centre of a collection site and its farthest border was set at approximately 50 km for urban sites and approximately 100 km for rural sites. The reason why there are some respondents that had to travel more than 100 km in Table 3.5-1 is that some of the rural sites in cycle 2 had the clinic located on a boundary of the site (since this is where the largest town / city was located). As such, the respondents that lived on the opposite boundary had to travel more than the 100 km. During the creation of collection sites for cycle 3, the maximum driving distance for the rural sites was decreased to 75 km. In addition, more of an attempt was made to use the expected location of the clinic (where possible) when determining the approximate driving distances. For some sites, approximate driving routes were used to calculate distances, rather than the "as the crow flies" distances. These changes should make driving distances in the rural sites more reasonable and help with the clinic response rates.

Table 3.5-1: Response rates at the household-level by approximate driving distance to the clinic

Approximate Distance	Dwellings in-scope	Households with at least 1 questionnaire respondent	Questionnaire % at the household-level	Households with at least 1 clinic respondent	Clinic % at the household-level
≥0 and <5KM	2,813	1,912	68.0%	1,615	84.5%
≥5 and <10KM	2,259	1,520	67.3%	1,227	80.7%
≥10 and <20KM	1,788	1,255	70.2%	1,027	81.8%
≥20 and <30KM	501	352	70.3%	286	81.3%
≥30 and <50KM	415	297	71.6%	246	82.8%
≥50 and <75KM	391	281	71.9%	212	75.4%
≥75 and <100KM	138	87	63.0%	55	63.2%
≥100KM	215	170	79.1%	113	66.5%
Overall	**8,520**	**5,874**	**68.9%**	**4,782**	**81.4%**

Figure 3.5-1 - Response rates at the household-level by approximate driving distance to the clinic



4 Conclusion

With its survey design, cycle 2 of the CHMS was able to obtain the target number of respondents required to produce reliable national estimates by age group and sex. This was achieved with a three-stage sample design that used the 2006 Census frame updated with the most up to date administrative files. Through careful monitoring during data collection and additional measures such as adjusting sampling factors for person-level selection, modifying probabilities of selection of

subsamples and the release of sample waves, the sampling design allowed the target number of respondents to be reached without significantly exceeding the minimum total number of respondents required. Results have shown the importance of updating the household composition used for stratification as soon as an auxiliary data source is available, not only for the dwellings in the 3 to 5 year old stratum but for all age group strata. Results have also shown areas of improvement for the next cycle, namely modifying the person selection strategy to take into account age group and sex instead of only age group and also adapting the driving distance during the creation of the collection sites to try and further minimize the total driving distance required by the respondent.

5 References

Giroux S., *Study on the number of sites for the Canadian Health Measures Survey,* HSMD, March 2004, internal paper.

Giroux S., Lavigne N., *Sampling Strategy for the Canadian Health Measures Survey, Part I: Selection of Collection Sites,* HSMD, March 2005, internal paper.

Giroux S., *Canadian Health Measures Survey: Sampling Strategy Overview,* Health Reports (Statistics Canada, catalogue 82-003) 2007; 18 (supplement): 31-36.

Giroux S., Labrecque F., Quigley A., *Canadian Health Measures Survey, Cycle 2: Combining Census and Administrative Data to Improve Survey Frame Efficiency,* Proceedings of the Symposium 2010, Social Statistics: The Interplay among Censuses, Surveys and Administrative Data, Statistics Canada, Ottawa.

Gray, G.B., *On increasing the sample size (number of psu's).* Technical memorandum, 1973.

Laflamme G., Dochitoiu C., *Creating Sites for the Canadian Health Measures Survey,* April 2005, internal paper.

MacIsaac, K., *The Effect of Cluster Sampling in the Canadian Health Measures Survey,* 2011, internal paper.

Statistics Canada. *PCODE Centralized Service,* November 2011, internal document.

Statistics Canada. *Methodology of the Canadian Labour Force Survey,* (Catalogue no 71-526-X) Ottawa: Minister of Industry, 1998.

Appendix A

Example of person-level respondent sampling:

To determine which household members would be sampled for the survey, each eligible member was assigned a sampling factor from Table 2.3.1-1. The total of the sampling factors for the household and the cumulative value for each member were then calculated. These total and cumulative values were calculated independently for the 3 to 11 and 12 to 79 age groups, since the sampling was performed independently. The 3 to 11 and 12 to 79 groups were each assigned a random number, which was multiplied by their respective total to determine the selection point(s). For both the 3 to 11 and 12 to 79 age groups, the first eligible person on the roster with a cumulative value that was equal to or greater than their respective selection point was selected.

Here is the household member information obtained from a roster of a fictitious sampled dwelling (CF = full-time member of the Canadian Forces):

Name	Age	CF?
Mike	3	No
Molly	8	No
David	13	No
Mitch	34	No
Sheila	37	Yes
Mary	58	No
John	62	No

Relevant information:

- As at least one person aged 3 to 11 is present in this household, two persons will be selected
- Two random numbers are generated in the household application:
  - 0.5312 for 3 to 11
  - 0.4204 for 12 to 79
- On the survey frame, this household was sampled from the 3 to 5 stratum, so the corresponding sampling factors from Table 2.3.1-1 will be used
- Since Sheila works full-time for the Canadian Forces, she will not be eligible for the survey
- There are 4 eligible persons between the age of 12 and 79. Since this is less than the reset value for the number of 12 to 79 year olds in the household (6), the sampling factors from Table 2.3.1-1 remain as is.
- There is only 1 person between the age of 12 and 19. Since this is less than the reset value for the number of 12 to 19 year olds (3), the sampling factors from Table 2.3.1-1 remain as is.

The next table summarizes the eligibility of each member, their sampling factors and the cumulative sampling factors.

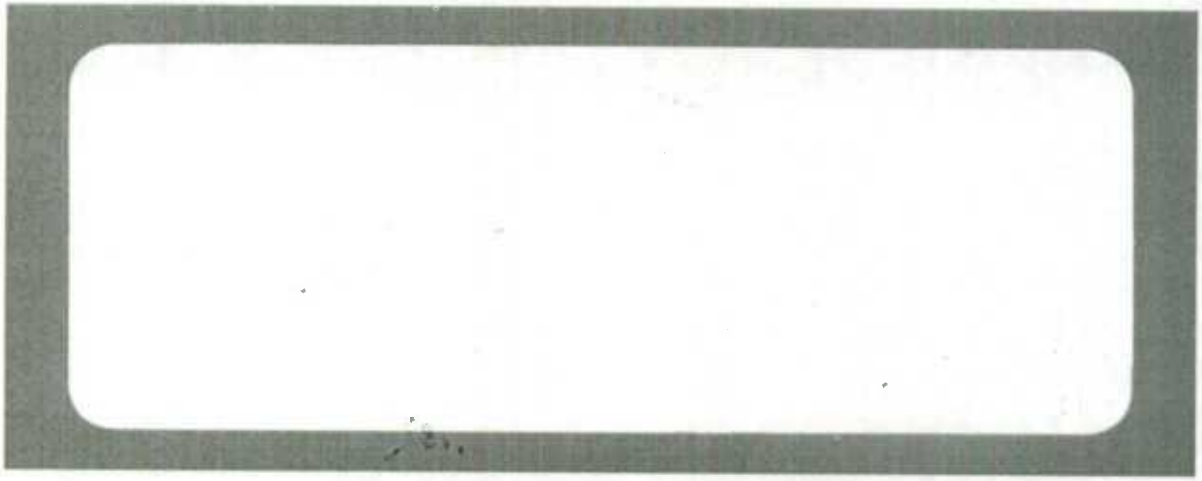
Name	Age	Canadian Forces?	Eligible?	Sampling Factor	Cumulative value 3 to 11	Cumulative value 12 to 79
Mike	3	No	Yes	3	3	-
Molly	8	No	Yes	2	5	-
David	13	No	Yes	5	-	5
Mitch	34	No	Yes	1	-	6
Sheila	37	Yes	No	-	-	-
Mary	58	No	Yes	1	-	7
John	62	No	Yes	1	-	8
Total 3 to 11					5	
Total 12 to 79						8

The total sampling factors are 5 for the 3 to 11 year olds and 8 for the 12 to 79 year olds. These total values are multiplied by their respective random numbers (0.5312 for 3 to 11 and 0.4204 for 12 to 79) to get the selection points.

- The selection point for 3 to 11 is 5*0.5312 = 2.6560. The first eligible member whose cumulative value 3 to 11 is greater than or equal to the random selection point is **Mike**.
- The selection point for 12 to 79 is 8*0.4204 = 3.3632. The first eligible member whose cumulative value is greater than or equal to the random selection point is **David**.

Methodology Branch

Direction de la méthodologie



Household Survey
Methods Division

Division des méthodes
d'enquêtes auprès des ménages

10105322 38

Statistics Canada Statistique Canada

Canada

DOCUMENT DE TRAVAIL
DIRECTION DE LA MÉTHODOLOGIE

Documentation du plan d'échantillonnage du cycle 2 de l'Enquête canadienne sur les mesures de la santé

HSMD - 2013 – 002F/A

Suzelle Giroux, France Labrecque et Andrew Quigley

Division des methods d'enquêtes auprès des ménages
Statistique Canada

Avril 2013

Le travail présenté dans ce document est la responsabilité des auteures et ne représente pas nécessairement le point de vue ou les politiques de Statistique Canada.

Documentation du plan d'échantillonnage du cycle 2 de l'Enquête canadienne sur les mesures de la santé

Suzelle Giroux, France Labrecque et Andrew Quigley
Division des méthodes d'enquêtes auprès des ménages

RÉSUMÉ

La collecte du cycle 2 de l'Enquête canadienne sur les mesures de la santé (ECMS) a commencé en août 2009 et a pris fin en novembre 2011. La collecte des données de l'ECMS débute par une interview en personne des répondants pour qu'ils répondent à un questionnaire sur la santé, puis les répondants devaient se rendre à un centre d'examen mobile de l'ECMS à proximité, pour la prise de mesures physiques et le prélèvement de spécimens biologiques. Enfin, des analyses en laboratoire ont été effectuées sur le sang et l'urine, ce qui a rendu la collecte de données de cette enquête très coûteuse et complexe au niveau logistique. Il était par conséquent important d'avoir une stratégie d'échantillonnage tenant compte des contraintes de la collecte et permettant d'atteindre l'objectif de l'enquête, sans dépasser de façon significative le nombre total minimum de répondants requis. À cette fin, une base de sondage comportant les données les plus à jour sur la composition du ménage et une stratégie d'échantillonnage souple étaient nécessaires. Le présent document décrit la base de sondage et la stratégie d'échantillonnage utilisées pour le cycle 2 afin d'atteindre l'objectif de l'enquête, et fournit des résultats pour déterminer si celui-ci a été atteint.

TABLE DES MATIÈRES

RÉSUMÉ 3
Sommaire 6
1 Objectif de l'enquête 7
1.1 Besoin en données 7
1.2 Population cible 7
1.3 Taille d'échantillon 7
2 Plan d'échantillonnage de l'ECMS 7
2.1 Premier degré: échantillonnage des emplacements de collecte 8
2.1.1 Création de la base aréolaire des emplacements de collecte 8
2.1.2 Taille d'échantillon des emplacements de collecte 8
2.1.3 Répartition des emplacements de collecte 8
2.1.4 Sélection des emplacements de collecte 9
2.1.5 Saisonnalité et ordre des emplacements de collecte 9
2.1.6 Ajout d'emplacements de collecte 9
2.2 Deuxième degré : échantillonnage des logements 11
2.2.1 Création de la base de sondage des logements 11
2.2.2 Mises à jour de la base de sondage – ajout de nouveaux logements 12
2.2.3 Stratification des logements 12
2.2.4 Mise à jour de la stratification des logements 13
2.2.5 Qualité de l'appariement entre la base de sondage et le FFT1 / T1 13
2.2.6 Taille et répartition de l'échantillon des logements 14
2.2.7 Contrôle du chevauchement 15
2.2.8 Vagues d'échantillon supplémentaire 15
2.3 Troisième degré: échantillonnage des personnes 17
2.3.1 Échantillonnage des personnes 17
2.3.2 Sous-échantillonnage pour les contaminants dans le sang et l'urine 18
2.3.3 Sous-échantillonnage des personnes qui doivent être à jeun lors de leur visite à la clinique 18
3 Résultats de l'enquête 19
3.1 Qualité de la base de sondage 19
3.1.1 Exactitude de l'information de contact – Adresses 19
3.1.2 Exactitude de l'information de contact – Numéros de téléphone 20

3.1.3 Efficacité de la stratification 20
3.2 Nombre de répondants et taux de réponse par emplacement de collecte 22
3.3 Comparaison du nombre de répondants et des taux de réponse des cycles 1 et 2 25
3.4 Nombre de répondants pour les sous-échantillons 27
3.5 Taux de réponse au niveau des ménages par distance approximative à parcourir 28
4 Conclusion 30
5 Références 30
Annexe A 32

Sommaire

L'objectif du cycle 2 de l'Enquête canadienne sur les mesures de la santé (ECMS) était de fournir des estimations nationales de référence pour divers indicateurs de la santé, pour les personnes de 3 à 79 ans vivant dans des logements privés. La stratégie de collecte de l'ECMS prévoyait que les participants sélectionnés répondent d'abord à un questionnaire sur la santé à leur domicile, puis se rendent à un centre d'examen mobile (CEM) à proximité, souvent appelé la clinique, où des spécialistes des mesures de la santé devaient procéder à plusieurs mesures physiques. Les répondants devaient aussi fournir des échantillons de sang et d'urine au CEM, échantillons qui ont été soumis à des analyses en laboratoire pour mesurer différents biomarqueurs, contaminants environnementaux et maladies infectieuses. Du fait de la logistique complexe et des coûts élevés de la collecte, il était important d'avoir un plan d'échantillonnage répondant à l'objectif de l'enquête avec une taille d'échantillon aussi petite que possible.

On a retenu un plan de sondage à trois degrés pour respecter la logistique de collecte et l'objectif de l'enquête. Au premier degré, des régions géographiques appelées emplacements de collecte ont été établies, étant donné qu'aux fins de l'ECMS, les répondants devaient se rendre au CEM et qu'il était important de limiter la distance que ceux-ci devaient parcourir. Au cours de la période de collecte d'août 2009 à novembre 2011, 18 emplacements de collecte ont été visités dans cinq régions du Canada. Au deuxième degré, les logements ont été sélectionnés. Il était nécessaire d'avoir la composition des ménages occupant les logements d'un emplacement de collecte donné, afin de cibler efficacement tous les groupes d'âge de la population cible. On a utilisé les données disponibles les plus à jour : la base de sondage des logements du Recensement de 2006 a été mise à jour au moyen des nouveaux logements trouvés depuis le jour du recensement, puis les informations de tous ces logements ont été mises à jour au moyen des fichiers administratifs comprenant des renseignements plus récents sur la composition des ménages. Les logements ont par la suite été stratifiés de façon séquentielle selon la présence des groupes d'âge suivants : 3 à 5 ans, 6 à 11 ans, 12 à 19 ans, 60 à 79 ans, 20 à 39 ans et 40 à 59 ans. Le nombre de logements sélectionnés par emplacement de collecte (environ 500) a été déterminé sur la base des taux de réponse précédents. La répartition des logements par strate a été effectuée au moyen de simulations, en tenant compte de la stratégie de sélection des personnes et des taux de réponse précédents. Cela a fait en sorte que le nombre cible de personnes par groupe d'âge et par sexe a été atteint. Au troisième degré, un algorithme a été programmé dans l'application de collecte, afin de sélectionner les personnes. Une fois qu'un intervieweur avait pris contact avec un logement et entré la composition du ménage, l'algorithme déterminait si les membres du ménage étaient compris dans le champ de l'enquête et utilisait un vecteur de facteur d'échantillonnage et un nombre aléatoire pour sélectionner une ou deux personnes au sein du ménage, dépendamment de la composition du ménage.

Des mesures ont été mises en place au cours de la collecte pour assurer l'atteinte de l'objectif de l'enquête, sans dépasser de façon significative le nombre total minimum de répondants requis. La surveillance des taux de réponse, selon le groupe d'âge et le sexe, a fourni des renseignements qui ont servi à mettre à jour la taille et la répartition de l'échantillon pour chaque emplacement de collecte subséquent. Des vagues d'échantillon supplémentaire ont été sélectionnées pour chaque emplacement et ont été utilisées lorsque les taux de réponse pour un emplacement donné posaient un problème.

On a effectué plusieurs évaluations afin de déterminer l'efficacité du plan de sondage en vue de l'atteinte de l'objectif de l'ECMS. L'utilisation des données du Recensement de 2006 avec d'autres fichiers administratifs a amélioré l'information de contact ainsi que la couverture de la population. On a aussi évalué l'efficacité de la méthode de stratification. Les résultats ont fait ressortir la nécessité de mettre à jour la composition des ménages au moyen de sources de données administratives plus récentes. On a constaté que le nombre visé de répondants par groupe d'âge et par sexe avait aussi été atteint, même si le nombre de répondants parmi les femmes était remarquablement plus élevé que celui parmi les hommes. On apportera des changements à l'algorithme de sélection des personnes pour les cycles futurs de l'ECMS afin de réduire cet écart. Les taux de réponse en fonction de la distance par rapport au CEM ont aussi été examinés. Le taux de réponse au volet clinique au niveau du ménage diminuait lorsque la distance approximative à parcourir dépassait 50 km, et encore davantage après 75 km. On a aussi tenu compte de cela dans le remaniement des emplacements de collecte pour le cycle 3.

Le plan de sondage a respecté l'objectif du cycle 2 de l'ECMS. Les résultats du cycle 2 ont fait ressortir l'importance de certains des éléments clés du plan de sondage, par exemple l'utilisation des sources d'information les plus à jour pour la stratification, la surveillance constante et l'utilisation d'un plan de sondage permettant des réajustements rapides.

1 Objectif de l'enquête

L'Enquête canadienne sur les mesures de la santé (ECMS) permet de recueillir de l'information clé relativement à la santé des Canadiens, sous forme de mesures physiques directes. Son objectif est de produire des estimations nationales de base relativement à une gamme variée d'indicateurs de la santé.

1.1 Besoin en données

La collecte des données pour le cycle 2 de l'ECMS s'est déroulée d'août 2009 à novembre 2011, en deux étapes. Tout d'abord, un intervieweur a visité les logements sélectionnés, afin d'administrer le questionnaire du ménage, grâce auquel des données concernant la nutrition, l'usage du tabac, la consommation d'alcool, les antécédents médicaux, l'état de santé actuel, le comportement sexuel, le mode de vie, la condition physique et des variables démographiques et socioéconomiques ont été recueillies. En deuxième lieu, on a demandé aux répondants de se rendre, sur rendez-vous, à un centre d'examen mobile (CEM) de l'ECMS à proximité, souvent appelé la clinique, où des professionnels de la santé qualifiés ont recueilli des spécimens de sang et d'urine, ainsi que des mesures physiques, comme la tension artérielle, la taille et le poids, et ont procédé à des tests de condition physique. On a aussi demandé aux répondants de porter un moniteur d'activité physique (accéléromètre) pendant une semaine, afin de mesurer leur niveau d'activité, et un échantillonneur de la qualité de l'air a été fourni, en vue d'être installé dans le logement pendant une semaine, afin de mesurer de nombreux niveaux de composés organiques volatils (COV).

1.2 Population cible

Le cycle 2 de l'ECMS vise la population âgée de 3 à 79 ans vivant dans les 10 provinces et les trois territoires. Sont exclus du champ de l'enquête les personnes vivant dans les réserves et autres peuplements autochtones des provinces, les membres à temps plein des Forces canadiennes, la population vivant en établissement et les habitants de certaines régions éloignées. En tout, ces exclusions représentent moins de 4% de la population cible.

1.3 Taille d'échantillon

Afin de répondre à l'objectif de l'ECMS de produire des estimations nationales de base pour une gamme variée d'indicateurs de la santé, 5 700 répondants devaient se rendre dans les CEM et être répartis également entre les 11 groupes suivants : groupe de 3 à 5 ans (deux sexes combinés) et groupes de 6 à 11 ans, de 12 à 19 ans, de 20 à 39 ans, de 40 à 59 ans et de 60 à 79 ans, selon le sexe. Avant que les opérations de l'enquête ne débutent sur le terrain, ce nombre a été augmenté, afin de tenir compte des logements hors du champ de l'enquête et de la non-réponse au niveau du ménage et de la personne. Cette taille d'échantillon devait produire des estimations au niveau national pour chaque groupe pour des problèmes de santé ayant une prévalence de 10 % ou plus, avec un coefficient de variation (c.v.) de l'estimation de 16,5 %.

On avait aussi besoin de sous-échantillons de répondants. Deux des sous-échantillons devaient servir aux analyses en laboratoire des contaminants environnementaux : un ensemble de mesures à partir de prélèvements d'urine et un à partir de prélèvements de sang. Des données ont aussi été recueillies auprès d'un troisième sous-échantillon pour des mesures prises à jeun. Les composantes des échantillonneurs d'air intérieur et des moniteurs d'activité n'ont pas été conçues comme des sous-échantillons, c'est pourquoi elles ne sont pas incluses dans ce document. Ces mesures ont été pondérées de la même manière que les sous-échantillons puisqu'un ajustement était nécessaire pour corriger le biais résultant de la non-réponse élevée observée.

2 Plan d'échantillonnage de l'ECMS

Dans le cadre de l'ECMS, on a utilisé un plan de sondage à trois degrés. Le premier degré consistait à échantillonner les emplacements de collecte, le deuxième, à échantillonner les logements dans ces emplacements de collecte, et le troisième, à échantillonner une ou deux personnes dans chacun des logements sélectionnés. Les trois degrés sont expliqués de façon détaillée dans les sections 2.1 à 2.3.

2.1 Premier degré: échantillonnage des emplacements de collecte

2.1.1 Création de la base aréolaire des emplacements de collecte

Comme les répondants à l'ECMS devaient se rendre au CEM, ils devaient pouvoir le faire en un temps raisonnable. La base aréolaire élaborée pour l'Enquête sur la population active (EPA) a servi à concevoir les emplacements de collecte et à en contrôler la taille, afin de répondre à ces exigences. À partir des grappes de la base de sondage de l'EPA (Statistique Canada, n° 71-256-XPB, 1998), on a créé 257 emplacements de collecte au total (Laflamme et Dochitoiu, 2005). Les grappes de l'EPA sont des petites unités géographiques qui comprennent environ 200 logements. Un emplacement de collecte est une région géographique comptant au moins 10 000 habitants, et la distance maximale entre le centre de l'emplacement et la limite la plus lointaine doit être de 50 km en région urbaine et de 100 km en région rurale. Les régions ne respectant pas ces critères ont été exclues. Les emplacements de collecte englobaient 96,3 % de la population canadienne (Giroux et Lavigne, 2005).

2.1.2 Taille d'échantillon des emplacements de collecte

Aux premières étapes de la conception du cycle 1 de l'ECMS, il a été recommandé de sélectionner au moins 30 emplacements de collecte ou unités primaires d'échantillonnage (UPE) (Giroux, 2004). Toutefois, en raison des contraintes de coûts et des contraintes logistiques, l'échantillon du cycle 1 a été sélectionné dans 15 emplacements de collecte. Avec un nombre aussi limité d'UPE, la priorité consistait à obtenir l'échantillon le plus représentatif au niveau national. Par conséquent, les 15 emplacements ont été répartis en proportion de la taille de la population des cinq limites régionales types de Statistique Canada : Colombie-Britannique, Prairies (Alberta, Manitoba et Saskatchewan), Ontario, Québec et région de l'Atlantique (Terre-Neuve-et-Labrador, Île-du-Prince-Édouard, Nouvelle-Écosse et Nouveau-Brunswick). Seulement un emplacement a été affecté à la région de l'Atlantique, ce qui a compliqué les calculs de la variance, la composante de premier degré de la variance n'ayant pu être calculée pour cette région. Au cycle 1, la solution a consisté à regrouper la région de l'Atlantique avec celle du Québec aux fins de l'estimation de la variance.

Au cycle 2, il a fallu ajouter un emplacement additionnel à la région de l'Atlantique pour que la variance puisse être calculée dans cette région, mais cela devait se faire sans augmenter les coûts de l'enquête. Une option aurait consisté à supprimer l'un des emplacements affectés en Ontario ou au Québec et à l'affecter plutôt à la région de l'Atlantique, mais ce n'est pas l'option qui a été privilégiée. La deuxième option consistait à déterminer si on pouvait améliorer l'efficacité en ajoutant simplement un emplacement dans la région de l'Atlantique (pour un total de 16 emplacements plutôt que de 15), mais avec une taille d'échantillon réduite par emplacement afin de ne pas augmenter la taille totale de l'échantillon pour le cycle. Autrement dit, il fallait déterminer si on pouvait réduire l'effet de plan de sondage de l'enquête en ajoutant un emplacement.

Une étude de simulation a été effectuée à partir des emplacements du cycle 1 pour lesquels la collecte était terminée, afin d'évaluer empiriquement les répercussions de l'ajout d'un emplacement, en conservant la taille de l'échantillon global constante. Même si l'analyse n'a pu être examinée davantage de façon plus exhaustive en raison des étapes nécessitant d'importantes ressources informatiques, les résultats préliminaires ont montré que, pour plusieurs variables examinées, les coefficients de variation étaient réduits par suite de l'ajout d'un emplacement de collecte. L'ajout d'un emplacement à l'enquête a été approuvé et effectué dans la région de l'Atlantique. D'autres travaux sur le nombre d'emplacements et la corrélation intra-grappe ont été menés en 2010, une fois la collecte des données du cycle 1 terminée (MacIsaac, 2011), et seront présentés plus en détails dans la section 2.1.6.

2.1.3 Répartition des emplacements de collecte

Comme il est mentionné précédemment, même si on avait uniquement besoin d'estimations nationales, les emplacements de collecte ont été stratifiés selon les cinq limites régionales types, afin d'assurer une répartition représentative au pays. Yellowknife, dans les Territoires du Nord-Ouest, a été incluse dans les Prairies, et Whitehorse, au Yukon, en Colombie-Britannique. Il a été recommandé d'utiliser un nombre plus important d'emplacements de collecte comportant peu de

répondants, parce que cela devait contribuer à optimiser la précision des estimations (Giroux, 2007). Toutefois, les contraintes logistiques et les contraintes de coûts associées à l'utilisation du CEM ont limité le nombre d'emplacements de collecte à 16. Dans chaque emplacement, on devait recueillir des données au sujet d'environ 350 répondants, pour un total d'environ 5 700 répondants. Les emplacements de collecte ont été répartis entre les régions en proportion de la taille de la population (tableau 2.1.3-1) et pour s'assurer d'avoir deux emplacements dans la région de l'Atlantique.

Tableau 2.1.3-1: Répartition des emplacements de collecte pour le cycle 2 de l'ECMS, par région

Région	Estimation de la population cible, âgées de 3 à 79 ans, selon le Recensement 2006	Emplacements de collecte par région	Emplacements de collecte échantillonnés
Atlantique	2 130 890	36	2
Québec	7 044 190	50	4
Ontario	11 317 075	61	6
Prairies et Yellowknife	5 034 985	77	2
Colombie-Britannique et Whitehorse	3 848 985	33	2
TOTAL	29 431 725	257	16

2.1.4 Sélection des emplacements de collecte

À l'intérieur de chaque région, la liste des emplacements de collecte a été triée selon que ces emplacements appartenaient à une région métropolitaine de recensement (RMR) et en fonction de la taille de la population. Une RMR est une région constituée d'une ou de plusieurs municipalités contiguës entourant une grande région urbaine (appelée noyau urbain). Une RMR doit avoir une population totale d'au moins 100 000 habitants et son noyau doit compter au moins 50 000 habitants. Les emplacements de collecte ont ensuite été échantillonnés systématiquement, avec une probabilité de sélection proportionnelle à la taille de leur population. Cette méthode de sélection, conjuguée au tri des emplacements selon l'appartenance ou non à une RMR et selon la taille de la population, a permis de s'assurer que les emplacements sélectionnés étaient répartis entre les RMR et les non-RMR, ainsi qu'entre les régions très peuplées et peu peuplées. Bien qu'il n'y en ait pas dans toutes les provinces et tous les territoires, les emplacements de l'ECMS ont été choisis pour représenter la population canadienne, d'est en ouest, compte tenu des grandes et des faibles densités de population (Giroux et Lavigne, 2005).

2.1.5 Saisonnalité et ordre des emplacements de collecte

La collecte des données pour l'ECMS est effectuée de façon séquentielle, un emplacement de collecte après l'autre. Les emplacements ont été ordonnés selon l'effet temporel entre les deux années de collecte, afin que les emplacements de chaque région soient répartis également entre les deux années. Les contraintes logistiques et de coûts liés au déplacement des CEM entre les emplacements de collecte ont été prises en compte. On a aussi tenu compte du caractère saisonnier selon la région, y compris le fait que les données des cycles 1 et 2 pourraient éventuellement être combinées pour produire des estimations.

2.1.6 Ajout d'emplacements de collecte

À mi-chemin de la collecte du cycle 2, il a été décidé de procéder à des ajouts au plan d'échantillonnage original de l'ECMS. Les deux options pour le cycle 2 étaient les suivantes : augmenter l'échantillon dans chaque emplacement (et par conséquent prolonger la collecte), ou ajouter un ou deux emplacements à la fin de la collecte. De nombreuses raisons justifiaient l'ajout d'emplacements de collecte ou leur prolongement. Tout d'abord, des pressions ont été exercées pour adopter l'année civile pour la collecte et l'analyse, les données pour une année civile étant plus faciles à interpréter par les utilisateurs. Selon le plan d'échantillonnage original pour le cycle 2, la collecte des données pour le 16^{e} et dernier emplacement devait être terminée en septembre 2011. Compte tenu de l'objectif d'entreprendre le cycle 3 en janvier 2012, l'ajout d'un ou de deux emplacements additionnels ou d'un échantillon supplémentaire aurait reporté la fin de la collecte du cycle 2 vers novembre ou décembre 2011. Le prolongement de la collecte avait donc aussi comme avantage l'absence

d'interruption dans le travail des intervieweurs et du personnel de la clinique et la suppression des problèmes liés à une longue interruption de la collecte. Dans le cadre de l'ECMS, on a recours à un groupe d'intervieweurs à domicile et d'employés à la clinique qui se consacrent uniquement à cette enquête, en raison de son caractère particulier. Enfin, un avantage évident de l'augmentation de l'échantillon était qu'il serait plus facile d'atteindre les objectifs du volet clinique du cycle 2 pour chacun des groupes d'âge et de sexe.

Afin de déterminer si et combien d'emplacements de collecte supplémentaires s'ajouteraient et dans quelle région géographique il serait le plus avantageux de les ajouter, une étude a été menée à partir des données du cycle 1 de l'ECMS (MacIsaac, 2011). Dans le cadre de l'étude, deux modifications ont été examinées, afin de déterminer ce qui aurait le plus de répercussions sur les variances d'échantillonnage : l'effet de l'augmentation du nombre d'emplacements de collecte par région, tout en maintenant la taille de l'échantillon du cycle 1, ou l'effet de l'augmentation du nombre de répondants par emplacement. Les résultats de l'étude ont montré que l'augmentation du nombre de répondants par emplacement n'aurait que des effets minimes sur les variances d'échantillonnage des variables étudiées, ce qui fait que cette option n'a pas été examinée de façon plus poussée.

Dans l'étude, il a été démontré que l'augmentation du nombre d'emplacements de collecte entraînerait une amélioration digne de mention des variances d'échantillonnage, particulièrement en Colombie-Britannique. L'ajout d'un emplacement a donné lieu à une amélioration importante, mais l'amélioration était moins marquée avec l'ajout de deux emplacements ou plus. L'amélioration de la variance d'échantillonnage en Colombie-Britannique s'explique en partie par les différences entre les deux emplacements de collecte échantillonnés au départ au cycle 1 en Colombie-Britannique. Il s'agissait d'une zone très urbaine de Vancouver et d'une zone très rurale de Quesnel/Williams Lake, ce qui fait que les données recueillies étaient assez différentes entre les emplacements, et qu'un emplacement additionnel aurait contribué à améliorer les variances dans cette région. Une situation similaire s'est présentée au cycle 2, la région incluant un grand centre urbain (Richmond) et une petite région rurale (région du centre et de l'est de Kootenay). Comme la répartition de la population en Colombie-Britannique signifiait souvent qu'il y aurait un emplacement urbain et un emplacement rural dans le scénario de deux emplacements sélectionnés, il a été décidé qu'il serait très profitable d'ajouter un emplacement dans la région de la Colombie-Britannique.

Étant donné le temps disponible pour la collecte à la fin de 2011 selon le plan d'échantillonnage original, un deuxième emplacement additionnel a été proposé. Les résultats de l'étude n'ont pas permis de décider clairement quelles autres régions devraient comporter un emplacement additionnel. À cette fin, la répartition proportionnelle originale du nombre d'emplacements de collecte entre les régions pour le cycle 2 a été revue. À ce moment là, les Prairies étaient sous-représentées du point de vue du nombre d'emplacements de collecte affectés. Compte tenu de leur population, ils auraient dû avoir 2,7 emplacements de collecte sur 16, mais n'en avaient que deux. Afin de représenter de façon appropriée cette région, le deuxième emplacement additionnel a été affecté aux Prairies. Pour permettre la collecte des données dans deux emplacements additionnels au cycle 2, la durée de la collecte de chacun de ces emplacements additionnels devait être légèrement plus courte que pour les 16 emplacements originaux, la clinique étant ouverte pendant 4 à 5 semaines par emplacement, plutôt que 5 ou 6 dans les 16 emplacements originaux.

Afin de déterminer les deux emplacements de collecte à ajouter au cycle 2, on a eu recours à une méthode d'augmentation du nombre d'unités primaires d'échantillonnage (UPE) d'un échantillon déjà sélectionné (Gray, 1973). Tant en Colombie-Britannique que dans la région des Prairies, on avait sélectionné au départ deux emplacements de collecte (UPE). Selon la méthode établie, pour sélectionner de façon non biaisée un emplacement de collecte additionnel dans une région, il faut d'abord sélectionner deux emplacements de collecte additionnels, comme si le but visé au départ dans l'échantillonnage était d'avoir un total de quatre emplacements. À cette fin, on utilise le même nombre aléatoire pour l'échantillon original, sélectionné au moyen de la probabilité proportionnelle à la taille, puis on modifie la fraction d'échantillonnage pour sélectionner quatre emplacements au total. Une fois que les deux emplacements de collecte additionnels sont sélectionnés dans une région, l'un d'entre eux peut être supprimé de façon aléatoire de l'échantillon. En Colombie-Britannique, cette méthode a donné lieu à l'ajout de Coquitlam. Dans les Prairies, elle a donné lieu à l'ajout de Calgary.

Une fois les deux emplacements additionnels déterminés, ils ont dû être intégrés dans le calendrier de la collecte restante pour le cycle 2. Les emplacements ont été sélectionnés lorsque la collecte se déroulait dans l'emplacement 7 à Winnipeg.

Après discussions avec l'équipe de projet, il a été décidé que Calgary deviendrait le nouvel emplacement 11 et que Coquitlam deviendrait l'emplacement 18, le dernier pour le cycle 2. Cette méthode respectait les effets saisonniers et temporels de chacune des régions pour le cycle 2, ainsi que pour le cycle 1 et le cycle 2 combinés. Elle permettait en outre de contrôler la distance additionnelle que les roulottes du CEM devaient parcourir. La liste finale des 18 emplacements de collecte pour le cycle 2 se trouve dans le tableau 2.1.6-1.

Tableau 2.1.6-1: Liste des 18 emplacements de collecte du cycle 2 de l'ECMS

Emp. #	Emplacement	Prov.	Région	Date début collecte (Quest. ménage)	Date fin collecte (Clinique)
1	Centre et est d'Ottawa	ON	Ontario	28 août 2009	12 octobre 2009
2	Oakville	ON	Ontario	2 octobre 2009	22 novembre 2009
3	Sud de Brantford	ON	Ontario	12 novembre 2009	20 janvier 2010
4	Laval	QC	Québec	5 janvier 2010	3 mars 2010
5	Sud de la Montérégie	QC	Québec	17 février 2010	14 avril 2010
6	St. John's	NL	Atlantique	1 avril 2010	26 mai 2010
7	Winnipeg	MB	Prairies	14 mai 2010	7 juillet 2010
8	Richmond	BC	Colombie-Britannique	13 juillet 2010	1 septembre 2010
9	Edmonton	AB	Prairies	23 août 2010	20 octobre 2010
10	Centre et est de Kootenay	BC	Colombie-Britannique	7 octobre 2010	1 décembre 2010
11	Calgary	AB	Prairies	13 novembre 2010	19 janvier 2011
12	Sud-ouest de Toronto	ON	Ontario	6 janvier 2011	9 mars 2011
13	Kingston	ON	Ontario	25 février 2011	20 avril 2011
14	Comtés de Colchester & Pictou	NS	Atlantique	8 avril 2011	1 juin 2011
15	Est de Toronto	ON	Ontario	20 mai 2011	20 juillet 2011
16	Gaspésie	QC	Québec	22 juillet 2011	14 septembre 2011
17	Couronne Nord de Montréal	QC	Québec	2 septembre 2011	26 octobre 2011
18	Coquitlam	BC	Colombie-Britannique	14 octobre 2011	30 novembre 2011

2.2 Deuxième degré : échantillonnage des logements

2.2.1 Création de la base de sondage des logements

La stratégie utilisée pour la construction de la base de sondage pour le cycle 2 de l'ECMS a été fortement influencée par l'expérience du cycle 1. Initialement, le cycle 1 de l'enquête devait débuter en mars 2006. La stratégie envisagée au cycle 1 était tout d'abord d'échantillonner des logements où il y a eu des répondants au cycle 3.1 de l'Enquête sur la santé des collectivités canadiennes (ESCC 3.1) afin d'atteindre plus efficacement le nombre désiré chez les jeunes et de compléter avec un échantillon de logements provenant de la base de sondage aréolaire, conçue pour l'Enquête sur la population active (c'est la raison pour laquelle les emplacements de collecte ont été construits à partir de cette base). Étant donné que la base de sondage aréolaire ne contient pas d'information démographique (date de naissance et sexe des personnes), une méthode de rejet aurait alors dû être utilisée pour réduire le nombre de répondants dans certains groupes d'âge chez les adultes. Cette stratégie aurait été coûteuse, car elle aurait augmenté le fardeau de réponse de la portion des ménages ayant déjà été interviewés pour l'ESCC 3.1 et impliqué l'interview de ménages qui sont ensuite rejetés. Or, le début du cycle 1 de l'enquête ayant été reporté, il a été possible d'utiliser le Recensement de 2006. Il est important de noter ici que les ententes reliées à l'accès aux données du recensement pour l'ECMS se limitaient à l'utilisation de l'adresse des logements et du numéro de téléphone aux fins de contact et à l'information démographique (date de naissance et sexe des personnes) aux fins de stratification de la base de sondage; aucune information ne permet d'identifier les personnes (il n'y avait aucun accès aux noms des répondants au recensement). L'utilisation du recensement a permis de réduire les coûts et le fardeau des répondants, par rapport à l'utilisation de la base de sondage aréolaire, puisque l'information démographique fournie par le recensement a permis d'établir un plan de sondage satisfaisant plus efficacement les exigences d'échantillonnage dans chaque groupe d'âge et de sexe. Bref, pour le cycle 2, la source de la base de sondage des logements est, à l'origine, la

même que celle du cycle 1. Elle a donc été construite à l'aide des données du Recensement de 2006, mises à jour comme il est décrit dans les prochaines sections de ce document.

2.2.2 Mises à jour de la base de sondage – ajout de nouveaux logements

Au fur et à mesure que la collecte de l'ECMS se déroule et que nous nous éloignons de la date du recensement, la couverture des logements provenant de cette base de sondage se détériore. Pour pallier cette situation, le Registre des adresses (RA) a été utilisé. Le registre contient, en plus des logements du recensement, tout nouveau logement qui a été découvert depuis. Par exemple, un nouveau logement peut être un nouvel immeuble, un nouvel appartement s'ajoutant à un logement existant ou un logement existant qui avait été manqué au moment du recensement. Pour ajouter de nouveaux logements au cycle 1, on a utilisé le Registre des adresses (RA), dès que la version la plus récente a été disponible après le Recensement de 2006, c'est-à-dire pour l'emplacement 11 du cycle 1 de l'ECMS. Depuis, le RA est mis à jour sur une base trimestrielle.

Pour le cycle 2, environ 8,7 % des logements de la base de sondage étaient des nouveaux logements. Les pourcentages de nouveaux logements pour chacun des emplacements de collecte figurent dans le tableau 2.2.2-1. Les emplacements de collecte sont numérotés selon l'ordre de collecte, c'est-à-dire l'ordre chronologique présenté dans le tableau 2.1.6-1. Logiquement, en raison de la dégradation de la base de sondage, le pourcentage de nouveaux logements devrait augmenter à mesure que l'on s'éloigne du recensement dans le temps. Même si cela est le cas dans une certaine mesure, le tableau 2.2.2-1 montre que le pourcentage de nouveaux logements a varié considérablement d'un emplacement à l'autre. Ainsi, le pourcentage de nouveaux logements dépendait non seulement de la dégradation de la base de sondage dans le temps, mais aussi de la croissance de la population dans chaque emplacement de collecte.

Tableau 2.2.2-1 : Pourcentage de nouveaux logements ajoutés à la base de sondage du recensement

Emplacement de collecte																	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
5 %	11 %	6 %	6 %	7 %	10 %	3 %	6 %	8 %	10 %	10 %	9 %	10 %	14 %	6 %	13 %	15 %	13 %

2.2.3 Stratification des logements

Une stratégie a été élaborée pour stratifier les logements selon le groupe d'âge des occupants. Tout d'abord, l'âge au début de la collecte pour chaque emplacement a été calculé pour toutes les personnes à partir des dates de naissance selon la base de sondage. Pour les huit premiers emplacements du cycle 2, les logements ont été stratifiés de la façon suivante : si un logement comptait au moins un enfant de 3 à 11 ans, il a été affecté à la strate « 3 à 11 »; autrement, si au moins une personne était âgée de 12 à 19 ans dans le logement, sa strate était « 12 à 19 », et ainsi de suite pour les strates « 60 à 79 », « 20 à 39 » et « 40 à 59 », dans cet ordre. On a utilisé cette méthode de stratification parce que, compte tenu de leur répartition dans la population canadienne et de la stratégie d'échantillonnage de l'ECMS, les enfants de 3 à 11 ans sont plus difficiles à cibler que les personnes de 12 à 19 ans qui, quant à elles, sont plus difficiles à cibler que les personnes de 60 à 79 ans, etc.

À partir de l'emplacement 9, la strate « 3-11 » a été séparée en une strate « 3 à 5 » et une strate « 6 à 11 ». Cela a permis plus de souplesse dans la répartition de l'échantillon entre ces deux groupes d'âge et a, ainsi, contribué à l'atteinte des tailles d'échantillon ciblées selon le groupe d'âge à la clinique dans ces deux groupes d'âge. Par ailleurs, les 18 emplacements incluaient une strate « autre » pour les autres logements, c'est-à-dire les logements qui étaient hors du champ de l'enquête parce qu'ils étaient vacants, comportaient des données manquantes sur la strate (dates de naissance manquantes des personnes vivant dans le ménage), ou en raison de l'âge de leurs résidents (par exemple, un logement dont tous les résidents avaient plus de 80 ans).

2.2.4 Mise à jour de la stratification des logements

Puisque la population cible du cycle 2 couvre les 3 à 5 ans et que la base de sondage des logements était construite à partir du Recensement de 2006, une mise à jour des naissances était nécessaire. En effet, l'information sur la base de sondage avant la mise à jour de la stratification datait du 16 mai 2006 alors que la collecte de l'ECMS s'est déroulée beaucoup plus tard des mois d'août 2009 à novembre 2011. Pour le cycle 2, les enfants nés après le recensement avaient 3 ans en mai 2009 ou après et il n'était donc pas possible d'utiliser, jusqu'à la fin de la période de collecte en 2011, la composition des ménages du Recensement de 2006 pour identifier les ménages ayant désormais des jeunes de 3 à 5 ans.

Le fichier FFT1 (Fichier sur la famille T1)[1] a été utilisé afin de mettre à jour la composition des ménages du recensement et d'assigner une strate aux nouveaux logements provenant du Registre des adresses. Le fichier FFT1 est produit à partir des déclarations de revenus T1 et T4, du fichier de prestation fiscale pour enfants et du fichier des naissances. Une nouvelle version du FFT1 est produite chaque année vers le milieu ou la fin de l'été. Par exemple, le FFT1 de 2007, contenant de l'information en date du 31 décembre 2007, est devenu disponible en septembre 2009. Le cycle de production du FFT1 a donc permis d'identifier des enfants aussi jeunes que 2 et demi ans au moment de sa disponibilité. Il est à noter que le FFT1 n'est pas utilisé seul comme base de sondage pour deux raisons : sa couverture n'est pas aussi bonne que celle obtenue par la combinaison du recensement et du RA et, le FFT1 utilise le concept de famille alors que la notion de ménage est utilisée par l'ECMS.

Grâce au FFT1, il a été possible de mettre à jour la stratification pour les déménagements, naissances et décès. Le FFT1 a aussi permis d'assigner une strate aux nouveaux logements et aux logements vacants selon le recensement, mais qui ne l'étaient plus au moment de la collecte de l'ECMS. Mais, la qualité de cette mise à jour de la stratification dépend de la qualité de l'information sur le FFT1 ainsi que de la qualité de l'appariement entre la base de sondage et le FFT1.

À partir de l'emplacement 13, le fichier de données fiscales T1 a aussi servi à mettre à jour les données démographiques de la base de sondage. Selon les résultats d'une étude, il a été déterminé que le fichier T1 contribuerait à améliorer la stratification, mais pas aussi bien que le FFT1. Ainsi, la première priorité de la mise à jour de la base de sondage a consisté à utiliser les données du FFT1 pour tous les logements pour lesquels un appariement approprié entre la base du recensement et le fichier de données fiscales était possible. Pour les logements pour lesquels un appariement n'était pas possible avec le FFT1, on a tenté d'apparier la base de sondage et le fichier T1 et on a utilisé ces données pour mettre à jour la base de sondage.

2.2.5 Qualité de l'appariement entre la base de sondage et le FFT1 / T1

L'appariement entre le FFT1 (ou le T1) et la base de sondage s'est fait à l'aide des adresses qui ont été d'abord standardisées à l'aide du logiciel PCODE (Statistique Canada, 2011). L'appariement s'est, en autant que possible, limité à un seul emplacement de collecte à la fois. Il fallait donc attribuer un emplacement de collecte à chaque enregistrement du FFT1 et T1, ce qui impliquait un processus complexe de codage géographique (attribution d'un îlot de recensement à chaque enregistrement du FFT1 et du T1).

Le tableau 2.2.5-1 présente les taux d'appariement, à savoir le pourcentage de logements de la base de sondage (une base fondée sur le Recensement de 2006 avec l'ajout de nouveaux logements du RA) qu'il était possible d'apparier avec le FFT1 et le T1 (le T1 a été utilisé uniquement à partir de l'emplacement 13). Le taux d'appariement variait considérablement d'un emplacement à l'autre, étant donné qu'il dépendait dans une large mesure de la qualité des adresses. Les résultats ont été généralement moins bons que pour les emplacements ruraux (emplacements 3, 5, 10, 14 et 16). À noter que lorsqu'aucun appariement avec le FFT1 ou le TI n'était possible, la strate n'a pas été mise à jour et la stratification du logement a été fondée uniquement sur les données du recensement.

Tableau 2.2.5-1 : Pourcentage d'appariement entre la base de sondage et le FFT1 ou le T1*

[1] Les données des fichiers administratifs, comme le FFT1, le fichier T1 et le RA, servent à améliorer l'efficacité de notre base de sondage (voir la section 3.1), ce qui donne lieu à une meilleure couverture et à des coûts de collecte plus faibles.

Emplacement de collecte																	
1	2	3	4	5	6	7	8	9	10	11	12	13*	14*	15*	16*	17*	18*
72 %	75 %	43 %	71 %	62 %	56%	74 %	57 %	59 %	53 %	52 %	49 %	68 %	42 %	72 %	57 %	81 %	63 %

* Fichier T1 utilisé pour l'emplacement 13 et les suivants

2.2.6 Taille et répartition de l'échantillon des logements

L'objectif original pour le cycle 2 avait été fixé à 5 700 répondants pour la composante de la clinique de l'enquête, soit environ 356 répondants par emplacement de collecte. Afin de déterminer le nombre de logements à échantillonner dans chaque emplacement de collecte pour atteindre cet objectif, on a utilisé les taux de réponse antérieurs de l'ECMS et de l'Enquête sur la santé dans les collectivités canadiennes (ESCC). L'ECMS et l'ESCC ont servi toutes les deux à calculer :

- La probabilité prévue qu'un logement soit éligible à l'ECMS (taux d'éligibilité)
- La probabilité prévue qu'une liste de tous les occupants du ménage soit obtenue (taux de listage)
- La probabilité prévue qu'une personne sélectionnée réponde au questionnaire du ménage (taux de réponse au questionnaire).

Enfin, les taux de réponse antérieurs de l'ECMS ont servi à calculer la probabilité prévue qu'un répondant au questionnaire soit aussi un répondant à la clinique (taux de réponse au volet clinique). Comme les emplacements de collecte ruraux, urbains et de noyau urbain (centre-ville) comportent chacun des taux de réponse distincts, chaque emplacement a été classé dans un de ces trois groupes, et les taux de réponse mentionnés précédemment ont été calculés et appliqués séparément à chaque groupe.

Une fois que tous les taux de réponse prévus ont été calculés, une simulation de 100 échantillons indépendants de logements a été effectuée pour l'emplacement échantillonné. L'objectif de chacun des 100 échantillons simulés était d'utiliser les taux de réponse antérieurs pour prévoir si, dans chaque logement échantillonné, 0, 1 ou 2 personnes répondraient au volet clinique. La base de sondage finale de l'emplacement a servi pour chaque simulation. Le nombre moyen prévu de répondants au volet clinique pour chaque groupe d'âge et de sexe pour les 100 échantillons indépendants a servi à déterminer si la taille et la répartition de l'échantillon choisies suffisaient. La simulation complète des 100 échantillons indépendants a été effectuée plusieurs fois, avec diverses tailles et répartitions d'échantillon, afin de déterminer une taille d'échantillon globale et une stratégie de répartition définitives. Dans le cadre de la première itération, on a utilisé des tailles d'échantillon approximatives de logements pour chaque groupe d'âge et de sexe en guise de valeurs de départ pour la taille globale de l'échantillon et la répartition entre les strates. Dans le cadre des itérations subséquentes, on a procédé à un ajustement manuel de la taille globale de l'échantillon et de la répartition entre les strates, selon les résultats de l'itération précédente, en vue de trouver les valeurs finales répondant aux objectifs pour la clinique.

Une taille d'échantillon de 525 logements a servi pour les huit premiers remplacements du cycle 2, parce qu'elle signifiait une charge de travail raisonnable pour chacun des intervieweurs des ménages. L'échantillon a été réparti entre six strates de groupe d'âge (3 à 5, 6 à 11, 12 à 19, 20 à 39, 40 à 59 et 60 à 79), une petite fraction de l'échantillon allant à une strate « autre ». L'objectif de sélection d'un échantillon à l'intérieur de l' strate « autre » était d'assurer une couverture plus complète de la population cible, les logements de la base de sondage hors du champ de l'enquête pouvant faire partie du champ de l'enquête à la suite d'un déménagement. Un nombre maximum de 35 logements par emplacement a été sélectionné dans cette strate, et un nombre moindre a été sélectionné pour les emplacements comportant un moins grand nombre de logements dans cette strate. Cette taille d'échantillon a contribué à prévenir les poids extrêmes d'échantillonnage pour ce type de logements. À partir de l'emplacement 9, les tailles d'échantillon des emplacements ont diminué, le nombre moyen observé de répondants au volet clinique pour les 8 premiers emplacements (379) étant beaucoup plus élevé que l'objectif de 356. On a procédé ainsi pour s'assurer de ne pas dépasser l'objectif global pour le volet clinique de l'enquête.

L'objectif pendant la répartition de l'échantillon entre chacune des strates de groupe d'âge était d'augmenter la probabilité de respecter les objectifs d'âge et de sexe pour les répondants au volet clinique du cycle 2, sans trop de dépassement. Dans

la mesure du possible, l'échantillon a été réparti de façon à mettre l'accent sur les strates où un échantillon plus important était requis pour atteindre les objectifs. À mi-chemin du cycle 2, cela a signifié de mettre l'accent sur les groupes plus jeunes (3 à 5, 6 à 11 et 12 à 19), ainsi que sur le groupe le plus âgé (60 à 79).

2.2.7 Contrôle du chevauchement

Afin d'éviter d'augmenter le fardeau des répondants, un contrôle du chevauchement de l'échantillon a été effectué avec les autres enquêtes-ménages de Statistique Canada. Le contrôle de chevauchement vise à éviter de contacter le même logement qu'une autre enquête dans un laps de deux ans. Les logements chevauchants ont été exclus de l'échantillon de l'ECMS et les ménages de ces logements ont été traités comme non-répondants. La taille d'échantillon a été augmentée en conséquence. Voici plus de détails sur le contrôle du chevauchement de l'échantillon de l'ECMS :

- Les adresses des logements sélectionnés pour l'Enquête sur la population active (EPA) ou toute autre enquête utilisant la base aréolaire de l'EPA, dans chacun des emplacements de collecte de l'ECMS, ont été envoyées par l'équipe de l'EPA. Ces logements ont donc été exclus de l'échantillon de l'ECMS.
- Les numéros de téléphone des ménages répondants à l'ECMS ont été exclus par les enquêtes utilisant la base de composition aléatoire du numéro de téléphone ou la base téléphonique.
- Les ménages répondants pour *l'Enquête sur la santé des communautés canadiennes – Composante thématique : Vieillissement en santé* ont été exclus de l'échantillon des logements de l'ECMS.

Au cycle 2, aucun contrôle de chevauchement n'a été nécessaire pour les enquêtes postcensitaires puisque leur collecte était terminée deux ans ou plus avant le début de la collecte du cycle 2 de l'ECMS.

2.2.8 Vagues d'échantillon supplémentaire

Au cours de l'échantillonnage des logements pour chaque emplacement de collecte, un certain nombre d'échantillons supplémentaires plus petits ont été sélectionnés en même temps que l'échantillon primaire, mais ces échantillons supplémentaires n'ont pas été envoyés initialement à la collecte. Ces échantillons supplémentaires ont été appelés « vagues » et pouvaient être utilisées lorsque la collecte allait moins bien dans un emplacement et qu'un plus grand nombre de logements était nécessaire pour atteindre le nombre visé de répondants au volet clinique. Les vagues ont été créées pour cibler une vaste gamme de besoins des répondants selon l'âge et le sexe, et la composition des vagues figure dans le tableau 2.2.8-1.

Les données sur l'âge et le sexe servant à la création et à la sélection des vagues ont été fondées sur la base de sondage finale des ménages de l'ECMS. Pour les 16 premiers emplacements du cycle 2, les vagues 0 et 1 étaient composées de

Tableau 2.2.8-1: Composition des vagues utilisées pour cycle 2 de l'ECMS

Vague	Emplacements 1 à 16		Emplacements 17 et 18	
	Nombre de logements	Composition	Nombre de logements	Composition
0	50	Tous les groupes d'âge	40	3 à 5, 6 à 11, 12 à 19 et 60 à 79
1	25	Tous les groupes d'âge	24	3 à 5, 6 à 11, 12 à 19 et 60 à 79
2	/	Non utilisé	15	6 à 11 ans, masculin
3	/	Non utilisé	15	6 à 11 ans, féminin
4	25	3 à 5 ans	25	3 à 5 ans
5	25	6 à 11 ans	25	6 à 11 ans
6	25	12 à 19 ans	25	12 à 19 ans, féminin
7	25	20 à 39 ans, masculin	15	12 à 19 ans, féminin
8	25	40 à 59 ans, masculin	15	60 à 79 ans, masculin
9	25	60 à 79 ans, masculin	25	60 à 79 ans, masculin

logements de chacune des 6 strates de groupe d'âge, la seule différence entre les deux vagues étant le nombre de logements dans chacune (50 et 25 respectivement). De la vague 4 à la vague 9, on a ciblé chaque groupe d'âge/de sexe particulier pouvant nécessiter un échantillon additionnel si les taux de réponse étaient trop faibles et/ou si une proportion insuffisante du groupe était sélectionnée à l'étape de la sélection au niveau de la personne. Les compositions des vagues figurant dans le tableau 2.2.8-1 pour les emplacements 17 et 18 ont changé radicalement, afin d'être axées davantage sur les groupes d'âge et de sexe pour lesquels il était difficile d'atteindre les objectifs du volet clinique au cycle 2. Pour permettre un plus grand nombre d'options, les vagues 2 et 3 ont aussi été utilisées pour la première fois dans les emplacements 17 et 18.

Au cycle 2, les vagues ont été envoyées à la collecte dans six emplacements : Sud de Brantford (emplacement 3), Sud-ouest de Toronto (emplacement 12), Comtés de Colchester et Pictou (emplacement 14), Est de Toronto (emplacement 15), Couronne Nord de Montréal (emplacement 17) et Coquitlam (emplacement 18). Les résultats de la collecte des vagues figurent dans le tableau 2.2.8-2.

Tableau 2.2.8-2: Sommaire des résultats de la collecte des vagues relâchées au cycle 2

Emp. #	Emplacement	Groupe d'âge ciblé	Nombre de logements		Ménages listés	Nombre de personnes			
			Sélec-tionnés	Dans le champ de l'ECMS		Sélec-tionnées	Répond. Quest.	Répond. Clinique	Répondants Clinique dans le groupe d'âge ciblé
3	Sud de Brantford	3 à 5 M+F	25	24	17	29	23	18	5
3	Sud de Brantford	6 à 11 M+F	25	25	20	40	34	32	13
12	Sud-ouest de Toronto	Tous	25	25	14	16	14	11	
12	Sud-ouest de Toronto	60 à 79 M	25	24	15	15	14	10	7
14	Comtés de Colchester & Pictou	60 à 79 M	25	23	20	20	19	15	6
15	Est de Toronto	Tous	25	25	16	18	15	10	
15	Est de Toronto	60 à 79 M	25	23	14	16	11	10	4
17	Couronne Nord de Montréal	3 à 5 M+F	25	25	16	30	30	22	9
17	Couronne Nord de Montréal	60 à 79 M	15	15	11	11	11	10	5
18	Coquitlam	3 à 5 M+F	25	25	18	32	31	28	9
Total – les six emplacements		Variés	**240**	**234**	**161**	**227**	**202**	**166**	

La vague axée sur les hommes de 60 à 79 ans est celle qui a été le plus couramment utilisée, dans quatre emplacements au total. Un peu moins de 50 % des répondants au volet clinique de ces trois vagues (22/45) étaient des hommes du groupe des 60 à 79 ans.

2.3 Troisième degré: échantillonnage des personnes

2.3.1 Échantillonnage des personnes

Lorsque l'intervieweur des ménages entrait en contact avec le logement échantillonné, l'objectif était de créer une liste des membres du ménage. Il s'agit d'une liste de toutes les personnes résidant dans le ménage, qui comprend des renseignements pertinents comme l'âge, le sexe et le fait que la personne travaille à temps plein pour les Forces canadiennes. À partir de ces données, l'application informatique a sélectionné de façon aléatoire une ou deux personnes pour participer au reste de l'enquête. Le nombre de personnes sélectionnées dépendait de la composition du ménage :

- S'il y avait au moins un enfant âgé entre 3 et 11 ans, deux personnes étaient sélectionnées : un enfant âgé entre 3 et 11 ans et une autre personne âgée entre 12 et 79 ans.
- S'il n'y avait pas d'enfant âgé entre 3 et 11 ans, seulement une personne du groupe des 12 à 79 ans était sélectionnée.
- Si aucune personne n'était admissible à l'enquête, personne n'était sélectionné. Cela comprenait les ménages dans lesquels toutes les personnes ciblées par l'enquête avaient moins de 3 ans ou plus de 79 ans et/ou étaient des membres à temps plein des Forces canadiennes.

Une fois la liste terminée, l'application informatique attribuait un facteur d'échantillonnage à chaque membre admissible du ménage, lequel servait à déterminer la probabilité de sélection. Le facteur d'échantillonnage attribué à chaque personne, l'était en fonction de la strate à partir de laquelle le logement avait été échantillonné et de l'âge de la personne. Les facteurs d'échantillonnage indiqués dans le tableau 2.3.1-1 ont servi à la majorité des emplacements du cycle 2. Ils variaient entre les groupes d'âge, afin de mieux permettre d'atteindre les objectifs du volet clinique pour chaque groupe d'âge. On a augmenté légèrement les facteurs d'échantillonnage pour les groupes des 12 à 19 ans et des 60 à 79 ans des trois derniers emplacements du cycle 2 et pour le groupe des 3 à 5 ans du dernier emplacement, afin d'augmenter le nombre de personnes sélectionnées dans ces groupes. Dans les ménages où deux personnes étaient sélectionnées, la sélection de l'enfant (entre 3 et 11 ans) s'est faite indépendamment de celle de la personne de 12 à 79 ans. Une explication complète du processus de sélection au niveau de la personne et un exemple détaillé se trouvent à l'annexe A.

Tableau 2.3.1-1: Facteurs d'échantillonnage utilisés lors de la sélection des personnes au cycle 2

Strate selon la base de sondage	Facteurs d'échantillonnage pour les 3 à 11 ans		Facteurs d'échantillonnage pour les 12 à 79 ans			
	3 à 5	6 à 11	12 à 19	20 à 39	40 à 59	60 à 79
3 à 5	3	2	5	1	1	1
6 à 11	3	2	5	1	1	1
12 à 19	3	2	5	1	1	1
20 à 39	3	2	5	2	1	1
40 à 59	3	2	5	1	2	1
60 à 79	3	2	5	1	1	2
Hors champ	3	2	5	1	1	1

Les ménages comptant un nombre important de membres peuvent avoir des poids d'enquête très élevés et, par conséquent, exercer une influence trop grande au moment du calcul des estimations. Afin de prévenir cela, deux valeurs de réinitialisation ont été mises en place, afin de ramener à 1 certains des facteurs d'échantillonnage dans certaines conditions. Dans les deux cas décrits ci-après, chaque membre du groupe des 12 à 79 ans avait une probabilité égale d'être sélectionné :

1. Si un ménage comprenait six personnes ou plus dans le groupe des 12 à 79 ans, tous les facteurs d'échantillonnage pour les membres de 12 à 79 ans ont été réinitialisés à 1.
2. Si un ménage comprenait trois personnes ou plus dans le groupe des 12 à 19 ans, tous les facteurs d'échantillonnage pour les membres de 12 à 79 ans ont été réinitialisés à 1.

2.3.2 Sous-échantillonnage pour les contaminants dans le sang et l'urine

Des sous-échantillons de répondants à l'enquête ont été sélectionnés pour des prélèvements de sang et d'urine sur lesquels on devait procéder à l'analyse de substances chimique particulières présentes dans l'environnement. On ne s'est pas servi de l'échantillon complet pour ces analyses, à cause du coût élevé de ces dernières. Les deux sous-échantillons pour les mesures des contaminants environnementaux ont été sélectionnés de façon indépendante, c'est-à-dire sans tenir compte des personnes sélectionnées pour le dépistage de l'autre groupe de produits chimiques présents dans l'environnement. Cela signifie qu'un répondant sous-échantillonné peut avoir été sélectionné pour l'un des sous-échantillons ou les deux.

Les tailles d'échantillon ciblées par groupe d'âge et par sexe pour les sous-échantillons des prélèvements d'urine et de sang figurent dans le tableau 2.3.2-1. Le sous-échantillon des prélèvements d'urine comportait un objectif global de 2 500 répondants, et aucun changement n'a été apporté. La taille de l'échantillon cible initial pour le sous-échantillon des prélèvements de sang était de 250 répondants dans chaque catégorie d'âge selon le sexe, pour un total de 1 000 répondants. Avant l'emplacement 11, 500 autres personnes se sont ajoutées à l'objectif pour la catégorie des 20 à 79 ans, réparties de façon égale entre les hommes et les femmes, portant le total final à 1 500.

Tableau 2.3.2-1: Nombre ciblé de répondants ayant fourni de l'urine / du sang pour les sous-échantillons sur les contaminants environnementaux

Âge	Sexe	Cible pour les sous-échantillons sur les contaminants mesurés dans l'urine	Cible pour les sous-échantillons sur les contaminants mesurés dans le sang
3 à 5	Masculin	250	
	Féminin	250	
6 à 11	Masculin	250	
	Féminin	250	
12 à 19	Masculin	250	250
	Féminin	250	250
20 à 79	Masculin	500	500
	Féminin	500	500
Total 3 à 79	Masculin	**1 250**	**750**
	Féminin	**1 250**	**750**
	Total (M+F)	**2 500**	**1 500**

2.3.3 Sous-échantillonnage des personnes qui doivent être à jeun lors de leur visite à la clinique

On a aussi procédé à un sous-échantillonnage pour les mesures nécessitant que le répondant soit à jeun avant son rendez-vous à la clinique. Chaque logement échantillonné a été désigné de façon aléatoire, afin d'indiquer si un répondant devait être à jeun pour le rendez-vous à la clinique. Dans ce cas, les répondants devaient être à jeun depuis la veille, alors que des restrictions alimentaires de moins longue durée ont été imposées à ceux qui n'avaient pas à être à jeun pour leur rendez-vous. Les femmes enceintes, les personnes souffrant de diabète, les enfants de 3 à 5 ans et les personnes dans des situations spéciales n'avaient pas à être à jeun, même si elles avaient été sélectionnées pour un rendez-vous à jeun.

3 Résultats de l'enquête

3.1 Qualité de la base de sondage

3.1.1 Exactitude de l'information de contact – Adresses

La base de sondage de l'ECMS pour le cycle 2 était constituée d'adresses, de numéros de téléphone et d'autres données d'échantillonnage pour tous les logements situés dans les 18 emplacements de collecte sélectionnés. Les adresses finales provenaient de l'une de trois sources possibles : le questionnaire du recensement, les données de contact du recensement (couramment appelé Système de contrôle principal ou SCP) ou le Registre des adresses (RA). Les adresses provenant du questionnaire du recensement ont été uniformisées au moyen du logiciel PCODE, avant que l'on tente de les utiliser, afin de résoudre les erreurs de lecture du scanneur optique ou de supprimer tout renseignement superflu compris dans le champ d'adresse du recensement. Les adresses du questionnaire du recensement considérées comme non valides au moyen de PCODE n'ont pas été utilisées. Étant donné que le recensement constituait la principale source d'adresses pour la base de sondage de l'ECMS, une adresse valide du questionnaire du recensement ou du SCP a été utilisée avant que l'on tente d'utiliser une adresse du RA. Un bref résumé des procédures servant à déterminer la source d'adresses à utiliser figure dans le tableau 3.1.1-1. La priorité a été donnée aux adresses provenant du questionnaire du recensement.

Pour être conservé dans la base de sondage et, par conséquent, être admissible à l'échantillonnage, un logement devait comporter au moins un numéro de voirie et un nom de rue. Parmi les logements de la base de sondage qui comportaient suffisamment de renseignements pour être conservés et, par conséquent, pour être admissibles à l'échantillonnage, environ 2,7 % comportaient une adresse finale comprenant uniquement un numéro de voirie et un nom de rue, la ville et le code postal n'étant pas disponibles. Cela s'est produit plus fréquemment dans les emplacements ruraux, comme les Comtés de Colchester et de Pictou (29,9 % de la base de sondage) et le Sud de Brantford (18,6 % de la base de sondage). Les adresses échantillonnées pour lesquelles il manquait la ville et le code postal ont été imputées manuellement à partir de sources externes comme PCODE et Canada 411. À l'occasion, la ville et le code postal étaient disponibles dans l'adresse non uniformisée du questionnaire du recensement de la base de sondage, qui n'avait pas été utilisée parce qu'elle avait été considérée comme invalide au cours du processus d'uniformisation. Dans l'ensemble, pour le cycle 2, 48,2 % des adresses échantillonnées provenaient du questionnaire du recensement, 38,0 % du SCP et 13,8 % du RA.

Tableau 3.1.1-1: Source des adresses pour la base de sondage du cycle 2 de l'ECMS

Description	Source des adresses
Vacant selon le Recensement 2006 et adresse complète sur le SCP	SCP
Vacant selon le Recensement 2006 et pas d'adresse complète sur le SCP	Registre des adresses
Adresse valide sur le questionnaire du Recensement	Questionnaire du Recensement
Adresse invalide sur le questionnaire du Recensement; adresse complète sur SCP	SCP
Adresse invalide sur le questionnaire du Recensement et sur le SCP	Registre des adresses
Nouveau logement sur la base de sondage depuis le Recensement 2006	Registre des adresses

Parfois, pendant la collecte auprès des ménages, un intervieweur n'a pas été en mesure de trouver l'adresse du logement échantillonné. Ces cas d'adresses problématiques ont été transférés au bureau central, où des tentatives ont été faites pour déterminer si des renseignements additionnels étaient disponibles pour aider à les localiser. Environ 60 % de ces adresses problématiques du cycle 2 comportaient une adresse échantillonnée provenant du recensement (soit du SCP ou du questionnaire), et environ 40 %, du RA. Le processus d'obtention de données additionnelles a nécessité l'utilisation de ressources multiples, comme d'autres sources d'adresses de la base de sondage qui n'avaient été utilisées pour l'adresse finale, Canada 411, Google Maps et d'autres données pertinentes trouvées dans des sites Web municipaux ou par suite d'autres recherches sur Internet. On a utilisé un processus similaire lorsqu'un intervieweur découvrait un logement multiple possible. Les données sur les logements multiples possibles ont été transférées au bureau central, afin de déterminer si le deuxième logement figurait déjà dans la base de sondage. S'il n'était pas déjà présent dans la base, le logement a été considéré comme un logement multiple, et les deux logements ont été contactés pour participer à l'enquête.

3.1.2 Exactitude de l'information de contact – Numéros de téléphone

Une adresse complète a été fournie aux intervieweurs pour chaque logement échantillonné. Lorsqu'il était disponible, le numéro de téléphone a aussi été fourni et pouvait être utilisé par l'intervieweur pour communiquer avec le logement pour vérifier l'adresse. Tout comme pour les adresses, les trois sources possibles de numéros de téléphone étaient le questionnaire du recensement, le SCP et le RA. Le numéro de téléphone du RA a eu préséance, les données de ce fichier étant les plus à jour. Les numéros de téléphone du SCP et du questionnaire du recensement ont été utilisés pour les logements du recensement qui n'avaient pas été appariés avec le RA ou lorsqu'il y avait un appariement, mais que le RA ne comprenait pas de numéro de téléphone. Dans le cas des logements comportant un numéro de voirie et un nom de rue, qui étaient par conséquent admissibles à l'échantillonnage, un numéro de téléphone était disponible pour 84,6 % de la base de sondage totale et 89,1 % des logements échantillonnés. Ce taux était presque 5 % plus élevé dans l'échantillon, étant donné qu'il y avait toujours une faible fraction d'échantillonnage dans la strate « autre », où les données de la base de sondage, comme les numéros de téléphone, étaient plus fréquemment manquantes.

Tableau 3.1.2-1: Source et exactitude des numéros de téléphone sur la base de sondage, ménages répondants

Source du numéro de téléphone	Total (Ménages répondants)		Numéro de téléphone exact	
	Nombre	%	Nombre	%
Recensement	1 603	27,4 %	634	39,6 %
Registre des adresses	3 776	64,5 %	2 966	78,5 %
Aucun numéro de téléphone	475	8,1 %	-	-
Total	**5 854**	**100,0 %**	**3 600**	**61,5 %**

Le tableau 3.1.2-1 présente une ventilation des sources de numéros de téléphone pour les ménages avec au moins un répondant au questionnaire au cycle 2, ainsi qu'une indication de l'exactitude du numéro de téléphone de la base de sondage. Un ménage de deux personnes a été considéré comme répondant si au moins une interview a été effectuée. Au moment de la vérification de l'exactitude du numéro de téléphone pour les ménages à deux personnes, les deux numéros de téléphone (le cas échéant) ont été comparés aux numéros de téléphone compris dans la base de sondage. Si un des numéros de téléphone fournis correspondait à celui provenant de la base de sondage, le numéro de téléphone de la base de sondage a été considéré comme exact. La majorité des ménages répondants avaient un numéro de téléphone du RA (64,5 %). Les numéros de téléphone du RA comportaient aussi un taux d'exactitude beaucoup plus élevé (78,5 %), par rapport à ceux du recensement (seulement 39,6 %). Pour 8,1 % des ménages répondants au total, il n'y avait pas de numéro de téléphone disponible dans la base de sondage.

3.1.3 Efficacité de la stratification

Étant donné les coûts élevés du CEM et des analyses en laboratoire, il était particulièrement important de veiller à ce que le nombre requis de répondants soit obtenu dans chaque groupe d'âge et de sexe, sans augmenter inutilement la taille de l'échantillon. Par conséquent, dans le cas de l'ECMS, l'efficacité de la stratification était primordiale. La stratification a été considérée comme efficace si la strate du logement selon la base de sondage, c'est-à-dire la strate utilisée pour l'échantillonnage des logements, était la même que la « vraie » strate. Plus la stratification était efficace, plus il était facile d'atteindre la taille d'échantillon cible pour chacun des groupes d'âge de la population visée.

Ni le recensement ni le RA n'ont pu être utilisés pour attribuer une strate aux logements nouveaux ou vacants. Il est par conséquent clairement avantageux de leur attribuer une strate à l'aide du FFT1 ou du TI. Toutefois, pour les autres logements, c'est-à-dire ceux auxquels une strate peut être attribuée à l'aide du recensement, la stratégie de mise à jour dépendait de la qualité du FFT1 ou T1.

Avant le début du cycle 2, on a procédé à une étude préliminaire des répondants, à partir des derniers emplacements du cycle 1 de l'ECMS, afin d'évaluer la qualité du FFT1 du point de vue de l'efficacité de la stratification. Pour évaluer la qualité du FFT1, l'efficacité de la stratification qui aurait été obtenue avec le FFT1 de 2006 a été comparée à celle qui a été obtenue en utilisant uniquement le Recensement de 2006. L'étude a démontré que le FFT1 de 2006 n'était ni meilleur ni pire que le Recensement de 2006 pour ce qui est de l'efficacité de la stratification. Il a par conséquent été décidé d'adopter une

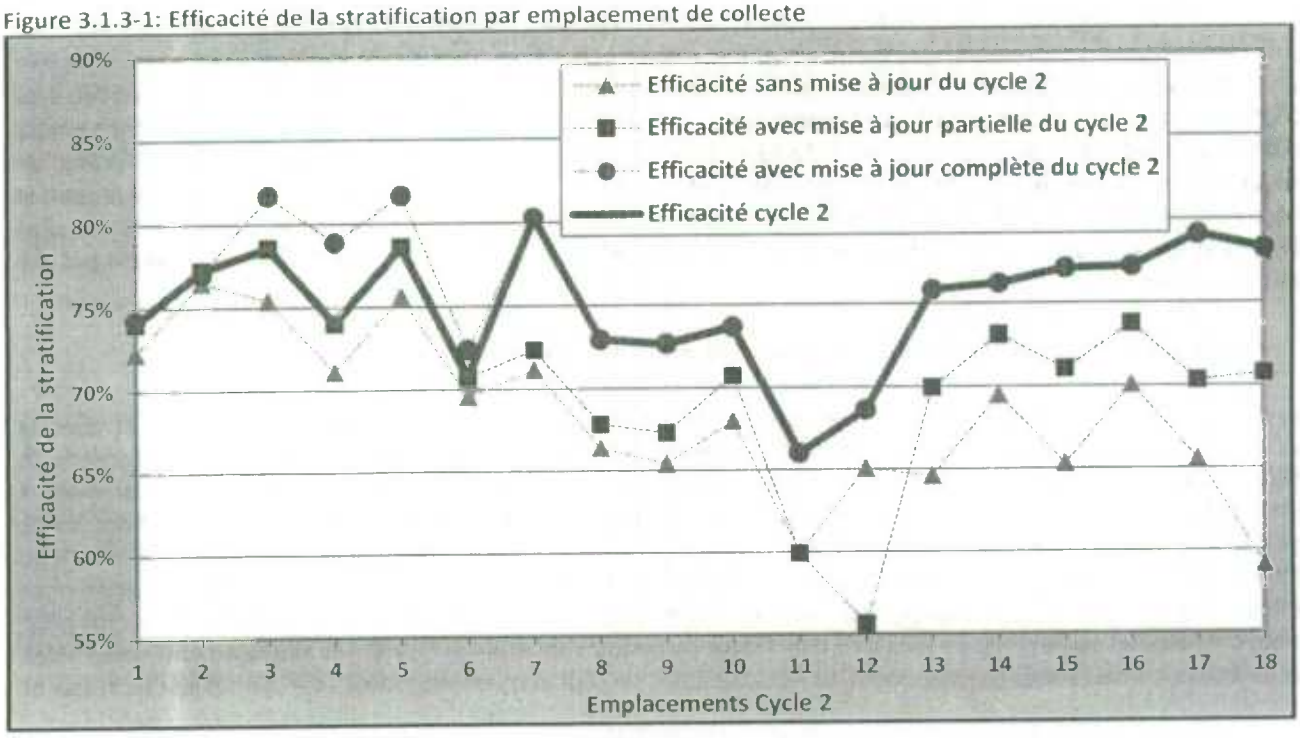
approche prudente pour les premiers emplacements du cycle 2 et de limiter la mise à jour de la stratification aux logements qui étaient vacants au moment du recensement, aux logements qui avaient été construits après la date du recensement ou qui avaient été laissés de côté durant le recensement, ou encore aux logements comptant une personne de 3 à 5 ans selon le FFT1.

Toutefois, à mesure que l'on s'éloignait du recensement dans le temps et que des versions plus à jour du FFT1 devenaient disponibles, il était plausible de présumer qu'il serait plus avantageux d'utiliser le FFT1. Le FFT1 de 2007 est devenu disponible au moment de la sélection des logements pour l'emplacement 3. L'efficacité de la stratification a de nouveau été évaluée à partir des résultats des deux emplacements de collecte, c'est-à-dire les emplacements 3 et 4. Cela s'est déroulé jusqu'à la fin de la collecte pour l'emplacement 4, plus de quatre mois après la diffusion du FFT1 de 2007. Les résultats de l'étude ont démontré qu'une mise à jour fondée sur le FFT1 de 2007 aurait amélioré l'efficacité de la stratification pour les emplacements 3 et 4, non seulement dans la strate « 3 à 11 », mais aussi dans toutes les autres strates. Il a donc été décidé de procéder aux mises à jour au moyen du FFT1 pour toutes les strates de logement au moment de la sélection de l'échantillon de l'emplacement 7. Une étude similaire a été effectuée au moyen du fichier T1, un peu après le milieu du cycle 2, et on est arrivé à une conclusion similaire. À partir de l'emplacement 13, la principale source de mise à jour de la base de sondage était le FFT1. Lorsque les logements n'étaient pas appariés au FFT1, on a tenté de les apparier au T1, et on a utilisé les données lorsque l'appariement était réussi. Certains résultats sont présentés dans les graphiques de l'efficacité de la stratification ci-après, et plus de détails sur le processus de mise à jour de la stratification figurent dans Giroux, Labrecque et Quigley (2010).

Efficacité de la stratification par emplacement de collecte

La figure 3.1.3-1 montre l'efficacité de la stratification par emplacement de collecte pour le cycle 2 de l'ECMS. Le taux d'efficacité de la stratification correspond au pourcentage de logements (sauf les logements de l'strate « autre ») pour lesquels la strate selon la base de sondage est la même que la strate « réelle ». Puisque les emplacements sont numérotés chronologiquement, le graphique présente l'évolution de l'efficacité dans le temps, à mesure que la collecte progresse. Il est important de noter, toutefois, que l'efficacité de la stratification dépendait aussi de l'augmentation de la population dans chaque emplacement. Comme chaque emplacement de collecte était différent, les données fournies ne représentent pas, à proprement parler, une série chronologique.

Figure 3.1.3-1: Efficacité de la stratification par emplacement de collecte



La ligne bleue continue représente le taux d'efficacité de chaque emplacement de collecte au cycle 2. La ligne verte pointillée (qui relie les triangles) représente le taux d'efficacité de la stratification selon le recensement qui aurait été obtenu sans la mise à jour de la stratification à l'aide du FFT1/T1. Pour le cycle 2, le taux d'efficacité global a été de 75,2 %; il aurait été de 68,6 % sans le FFT1 et le T1. Par conséquent, le gain d'efficacité de la stratification lié à l'utilisation du FFT1 ou du T1 a été de 6,6 points de pourcentage.

La ligne rose pointillée (qui relie les carrés) représente le taux d'efficacité avec une mise à jour partielle pour les logements de l'ensemble des strates. La stratification a été mise à jour uniquement lorsqu'il y avait une personne de 3 à 5 ans dans le ménage selon le FFT1, ou lorsque le logement était vacant au moment du recensement, ou qu'il avait pris de l'expansion depuis. C'est le type de mise à jour qui a été effectué pour les six premiers emplacements de collecte, ce qui fait que les lignes bleue et rose coïncident pour les six premiers emplacements. Les chiffres pour l'emplacement 7 montrent qu'il a été très avantageux de procéder à une mise à jour complète de la stratification au moyen du FFT1, le gain d'efficacité pour la mise à jour partielle ayant été de près de 8 points de pourcentage.

Enfin, il est intéressant de voir les résultats qui auraient été obtenus si une mise à jour avait été effectuée dès le début du cycle 2. La ligne orange en pointillé (qui relie les cercles) montre le taux d'efficacité de la stratification par suite d'une mise à jour complète fondée sur le FFT1. Étant donné que cette méthode a été utilisée à partir de l'emplacement 7, les lignes bleue et orange coïncident à partir de cet emplacement. La mise à jour complète est devenue intéressante uniquement après l'adoption du FFT1 de 2007, à partir de l'emplacement 3. Toutefois, cette observation n'a pu être faite tant que la collecte n'était pas terminée pour les emplacements 3 et 4, soit juste avant la mise à jour de la base de sondage pour l'emplacement 7. Le fait de pouvoir utiliser le FFT1 le plus récent a semblé faire une différence appréciable, les gains d'efficacité entre le recensement sans mise à jour et le recensement avec mise à jour complète ayant augmenté à partir de l'emplacement 3, lorsque le FFT1 de 2007 est devenu disponible, à partir de l'emplacement 13, lorsque le FFT1 de 2008 est devenu disponible, et à partir de l'emplacement 18, lorsque le FFT1 de 2009 est devenu disponible. Une partie de la hausse observée à l'emplacement 13 était attribuable à l'utilisation du T1 comme source secondaire pour les mises à jour démographiques.

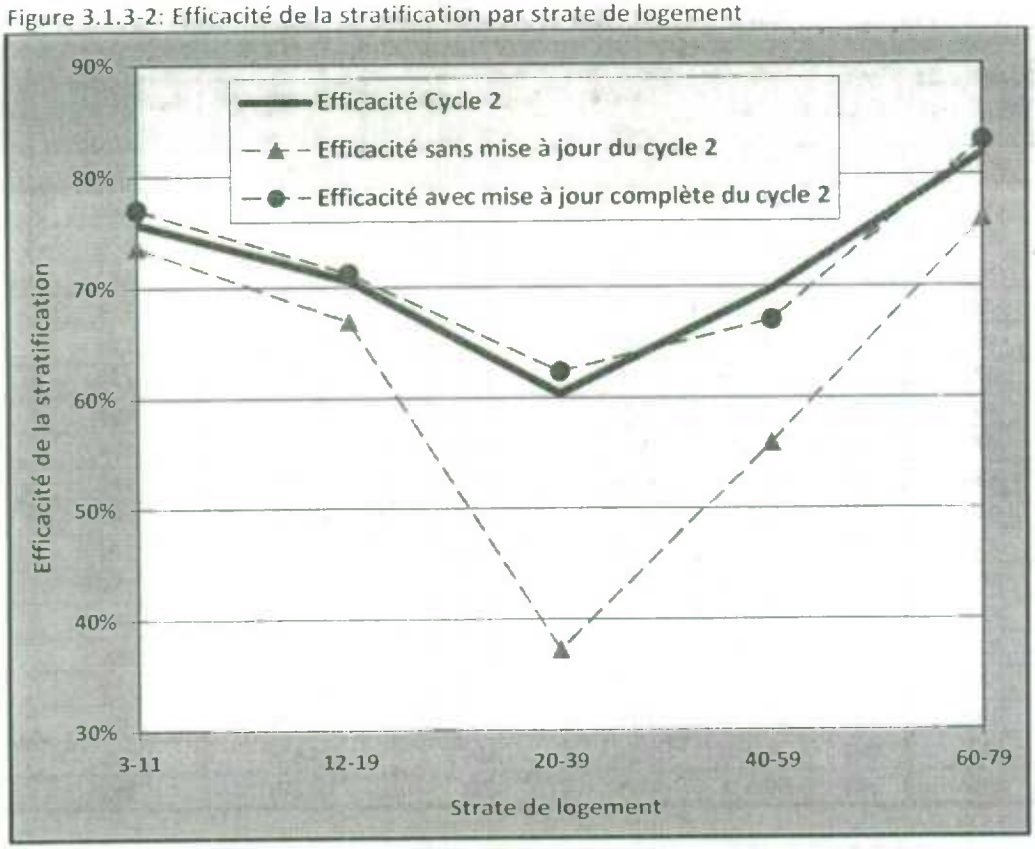
Efficacité de la stratification par strate de logement

La figure 3.1.3-2 montre l'efficacité de la stratification selon la strate de logement pour les emplacements de collecte du cycle 2. Comme dans la figure précédente, la ligne bleue continue représente le cycle 2, la ligne verte pointillée (reliant les triangles) représente le recensement seulement, et la ligne orange pointillée (reliant les cercles) représente la mise à jour complète. Une comparaison des lignes bleue et verte montre que la mise à jour à partir du FFT1 ou du T1 a donné lieu à un gain d'efficacité pour toutes les strates. Les répercussions sur l'efficacité de la stratification ont été particulièrement importantes pour les logements de la strate « 20-39 », pour laquelle on a noté un gain d'efficacité de 23 points de pourcentage par suite de la mise à jour de la stratification au moyen du FFT1 ou du T1 au cycle 2. La ligne orange dans le graphique montre que la mise à jour complète dès le début du cycle 2 aurait été avantageuse pour toutes les strates, sauf peut-être, la strate « 40-59 » (notamment à cause des emplacements 1 et 2, où la mise à jour complète n'aurait pas été vraiment avantageuse).

3.2 Nombre de répondants et taux de réponse par emplacement de collecte

Les taux finaux de listage, de réponse au questionnaire et de réponse au volet clinique du cycle 2 figurent dans le tableau 3.2-1. Un ménage listé était un ménage dont la composition démographique dans le logement échantillonné avait été obtenue par l'intervieweur et dans lequel celui-ci pouvait sélectionner une ou deux personnes pour participer à l'enquête. Le taux de listage des ménages a été calculé comme correspondant au nombre de ménages listés divisé par le nombre de logements éligibles dans l'échantillon. Un logement échantillonné a été considéré comme non éligible s'il n'a pu être localisé, s'il était vacant, s'il ne s'agissait pas d'une résidence principale ou s'il comprenait uniquement des personnes qui ne faisaient pas partie de la population cible (comme des personnes de 80 ans et plus ou des membres à temps plein des Forces canadiennes seulement). Le taux global de listage du cycle 2 s'est établi à 75,9 %. Les emplacements comportant les taux de listage les plus élevés (emplacement 10 – Centre et Est de Kootenay, emplacement 14 – Comtés de Colchester et

Figure 3.1.3-2: Efficacité de la stratification par strate de logement



de Pictou, et emplacement 16 – Gaspésie) étaient tous des emplacements ruraux. Les emplacements comportant les taux de listage les plus faibles (emplacement 4 – Laval, emplacement 12 – Sud-ouest de Toronto, emplacement 15 – Est de Toronto, et emplacement 18 – Coquitlam) étaient tous dans les grandes zones urbaines de Toronto, Montréal et Vancouver.

Parmi les personnes sélectionnées pour participer à l'enquête, le taux de réponse global au questionnaire a été de 90,5 %. Les taux pour le questionnaire ont été très uniformes tout au long du cycle 2, seulement quatre emplacements ayant un taux s'écartant de plus de 3 % de la moyenne : deux emplacements ayant de faibles taux de réponse (emplacement 3 – Sud de Brantford, à 86,8 %, et emplacement 15 – Est de Toronto, à 85,2 %) et deux emplacements ayant des taux élevés (emplacement 4 – Laval, à 94,1 %, et emplacement 16 – Gaspésie, à 93,6 %).

Dans le cas des personnes qui ont répondu au questionnaire, le taux de réponse au volet clinique a varié beaucoup plus d'un emplacement de collecte à l'autre. Le taux global pour le volet clinique s'est établi à 81,7 %. Six emplacements ont obtenu des taux inférieurs à 80 %, l'emplacement 11 – Calgary, affichant le taux le plus faible, soit 75,6 %. Au total, quatre emplacements ont eu des taux de réponse au volet clinique de 85 % ou plus, l'emplacement 8 – Richmond affichant le taux le plus élevé, soit 87,5 %.

Les taux de réponse aux prélèvements de sang et d'urine figurent dans le tableau 3.2-2. Le taux de prélèvement de sang a été relativement stable tout au long du cycle, ne s'écartant pas trop de la proportion globale de 95,8 % pour le cycle 2. Seul l'emplacement 15 – Est de Toronto s'écartait de plus de 3 % de la moyenne (à 91,5 %). Le taux de prélèvement d'urine a été encore plus uniforme tout au long du cycle, aucun emplacement ne s'écartant de plus de 1,4 % du taux global de 98,9 %.

Tableau 3.2-1: Taux de réponse au niveau des ménages et des personnes, par emplacement de collecte

Emp. #	Logements (éligibles)	Ménages		Personnes				
		Listés	%	Sélec-tionnées	Répondants Quest.	Taux réponse Quest.	Répondants Clinique	Taux réponse Clinique
Canada	8 520	6 465	75,9 %	8 652	7 830	90,5 %	6 395	81,7 %
1	490	392	80,0 %	513	455	88,7 %	386	84,8 %
2	505	385	76,2 %	523	460	88,0 %	375	81,5 %
3	521	386	74,1 %	538	467	86,8 %	360	77,1 %
4	495	349	70,5 %	471	443	94,1 %	379	85,6 %
5	480	348	72,5 %	465	433	93,1 %	338	78,1 %
6	479	381	79,5 %	505	457	90,5 %	373	81,6 %
7	493	397	80,5 %	559	508	90,9 %	409	80,5 %
8	476	362	76,1 %	513	473	92,2 %	414	87,5 %
9	484	370	76,4 %	506	462	91,3 %	376	81,4 %
10	467	385	82,4 %	506	464	91,7 %	387	83,4 %
11	427	319	74,7 %	408	369	90,4 %	279	75,6 %
12	467	311	66,6 %	393	353	89,8 %	274	77,6 %
13	439	338	77,0 %	449	409	91,1 %	344	84,1 %
14	451	384	85,1 %	500	453	90,6 %	385	85,0 %
15	538	360	66,9 %	459	391	85,2 %	319	81,6 %
16	354	309	87,3 %	407	381	93,6 %	301	79,0 %
17	452	340	75,2 %	458	419	91,5 %	356	85,0 %
18	502	349	69,5 %	479	433	90,4 %	340	78,5 %

Tableau 3.2-2: Taux de réponse aux prélèvements d'urine et de sang, par emplacement de collecte

Emp. #	Répondants Clinique	Répondants Sang	%	Répondants Urine	%
Canada	6 395	6 128	95,8 %	6 327	98,9 %
1	386	374	96,9 %	384	99,5 %
2	375	356	94,9 %	373	99,5 %
3	360	339	94,2 %	358	99,4 %
4	379	363	95,8 %	378	99,7 %
5	338	322	95,3 %	334	98,8 %
6	373	360	96,5 %	369	98,9 %
7	409	392	95,8 %	402	98,3 %
8	414	402	97,1 %	412	99,5 %
9	376	366	97,3 %	372	98,9 %
10	387	378	97,7 %	379	97,9 %
11	279	275	98,6 %	275	98,6 %
12	274	264	96,4 %	273	99,6 %
13	344	332	96,5 %	339	98,5 %
14	385	368	95,6 %	379	98,4 %
15	319	292	91,5 %	315	98,7 %
16	301	285	94,7 %	300	99,7 %
17	356	334	93,8 %	347	97,5 %
18	340	326	95,9 %	338	99,4 %

3.3 Comparaison du nombre de répondants et des taux de réponse des cycles 1 et 2

Les chiffres du tableau 3.3-1 présentent une comparaison des nombres de répondants obtenus pour le volet clinique et le prélèvement de sang pour le cycle 1 et le cycle 2. Pour produire des estimations de qualité pour chaque groupe d'âge selon le sexe, l'objectif principal a été fixé à environ 500 répondants pour le volet clinique. La seule exception était le groupe des 3 à 5 ans, pour lequel l'objectif principal était de recueillir 500 spécimens de sang au total (deux sexes combinés). Cette exception vient du fait qu'il n'était pas nécessaire de répartir les estimations finales selon le sexe pour les 3 à 5 ans. L'objectif secondaire pour le groupe des 6 à 79 ans était de 500 spécimens de sang selon le groupe d'âge et le sexe, même si les contraintes de coûts ont fait en sorte que cet objectif n'a pas été tout à fait atteint.

L'objectif initial de 500 répondants au volet clinique selon le groupe d'âge et le sexe a été atteint pour tous les groupes. Les hommes de 60 à 79 ans représentaient le défi le plus important. Pour atteindre l'objectif pour les hommes de 60 à 79 ans, des vagues d'échantillon additionnel axées sur ce groupe ont été utilisées dans quatre des sept derniers emplacements. L'objectif initial de 500 spécimens de sang pour les 3 à 5 ans a aussi été atteint en partie grâce à l'utilisation de vagues au cours des deux derniers emplacements. Pour augmenter la possibilité d'atteindre les objectifs pour ces groupes difficiles, l'accent a aussi été mis sur la sélection d'un plus grand nombre de logements dans leurs strates respectives au cours du processus d'échantillonnage initial.

Pour les groupes des 6 à 79 ans, l'objectif secondaire de 500 échantillons de sang selon le groupe d'âge et le sexe a été atteint pour sept des dix groupes d'âge/de sexe. Les trois groupes d'âge/de sexe se situant en dessous de l'objectif pour les spécimens de sang étaient les filles de 12 à 19 ans et les garçons et les filles de 6 à 11 ans. Ces résultats sont conformes à ceux du cycle 1, dans lequel ces trois groupes ont fourni au total moins de 500 spécimens de sang.

Tableau 3.3-1: Comparaison du nombre de répondants, par groupe d'âge et de sexe entre les cycles 1 et 2

Âge	Sexe	Répondants Clinique		Répondants Sang	
		Cycle 1	Cycle 2	Cycle 1	Cycle 2
3 à 5	Masculin		308		264
	Féminin		293		254
6 à 11	Masculin	542	534	471	494
	Féminin	532	541	467	480
12 à 19	Masculin	514	550	490	529
	Féminin	499	510	464	479
20 à 39	Masculin	524	556	514	553
	Féminin	661	773	655	763
40 à 59	Masculin	582	619	578	619
	Féminin	654	613	648	608
60 à 79	Masculin	544	509	538	508
	Féminin	552	589	548	577
Total 3 à 79	**Masculin**		**3 076**		**2 591**
	Féminin		**3 319**		**2 782**
	Total		**6 395**		**6 128**
Total 6 à 79	**Masculin**	**2 706**	**2 768**	2 591	**2 703**
	Féminin	**2 898**	**3 026**	**2 782**	**2 907**
	Total	**5 604**	**5 794**	**5 373**	**5 610**

Le nombre de répondants de sexe féminin au volet clinique du cycle 2 (3 319) était sensiblement plus élevé que le nombre de répondants masculins (3 076). Cela s'apparente aux résultats du cycle 1 (2 898 répondants de sexe féminin comparativement à 2 706 répondants de sexe masculin). Afin d'aider à rétrécir l'écart entre les sexes, la méthode de sélection au niveau de la personne a été modifiée pour le cycle 3, afin de tenir compte du sexe au moment de l'attribution des facteurs d'échantillonnage, ce qui a des répercussions sur les probabilités finales de sélection. Les facteurs

d'échantillonnage de la nouvelle méthode contribueront à faire diminuer le nombre de répondants de sexe féminin âgés de 20 à 39 ans qui sont sélectionnés et à répartir cet échantillon additionnel entre les autres groupes. La nouvelle méthode permettra aussi d'apporter des ajustements tout au long du cycle 3 s'il est difficile d'obtenir des répondants pour un groupe d'âge et de sexe en particulier.

Les résultats des cycles 1 et 2 sont similaires pour plusieurs groupes d'âge et de sexe. Le fait d'avoir 18 emplacements de collecte au cycle 2, comparativement aux 15 emplacements du cycle 1, a contribué à une hausse globale des chiffres pour le volet clinique et les spécimens de sang. L'ajout des 3 à 5 ans au cycle 2 a entraîné une augmentation appréciable du nombre des femmes de 20 à 39 ans participant à l'enquête. Les femmes de 40 à 59 ans et les hommes de 60 à 79 ans représentaient les seul groupes d'âge affichant une baisse appréciable des chiffres pour le volet clinique du cycle 1 au cycle 2.

Une comparaison des taux de réponse au niveau de la personne pour le questionnaire, le volet clinique et les prélèvements de sang et d'urine entre le cycle 1 et le cycle 2 figure dans le tableau 3.3-2. L'augmentation des taux de réponse au questionnaire a été relativement uniforme pour les groupes d'âge/de sexe et a entraîné une augmentation globale, la le taux de réponse étant passé de 88,3 % au cycle 1 (6 à 79 ans) à 90,5 % au cycle 2 (3 à 79 ans). Le contraire s'est produit pour le taux de réponse au volet clinique, la baisse des taux ayant été relativement uniforme pour les groupes d'âge/de sexe.

Tableau 3.3-2: Comparaison des taux de réponse au questionnaire, à la clinique et au prélèvement d'urine et de sang entre les cycles 1 et 2

Âge	Sexe	Taux de réponse Questionnaire		Taux de réponse Clinique		Taux de réponse Sang		Taux de réponse Urine	
		Cycle 1	Cycle 2	Cycle 1	Cycle 2	Cycle 1	Cycle 2	Cycle 1	Cycle 2
3 à 5	Masculin		91,6 %		80,2 %		85,7 %		94,2 %
	Féminin		93,7 %		78,8 %		86,7 %		96,6 %
	Total		92,6 %		79,5 %		86,2 %		95,3 %
6 à 11	Masculin	90,3 %	92,7 %	87,0 %	84,0 %	86,9 %	92,5 %	97,6 %	99,6 %
	Féminin	91,5 %	94,2 %	85,1 %	84,9 %	87,8 %	88,7 %	97,6 %	98,3 %
	Total	90,9 %	93,5 %	86,1 %	84,4 %	87,3 %	90,6 %	97,6 %	99,0 %
12 à 19	Masculin	87,5 %	89,6 %	88,3 %	84,2 %	95,3 %	96,2 %	98,4 %	98,7 %
	Féminin	86,9 %	90,7 %	86,5 %	86,0 %	93,0 %	93,9 %	98,2 %	98,6 %
	Total	87,2 %	90,1 %	87,4 %	85,1 %	94,2 %	95,1 %	98,3 %	98,7 %
20 à 39	Masculin	82,6 %	85,1 %	84,1 %	79,1 %	98,1 %	99,5 %	98,9 %	99,5 %
	Féminin	89,3 %	88,8 %	83,2 %	81,3 %	99,1 %	98,7 %	99,8 %	99,9 %
	Total	86,2 %	87,2 %	83,6 %	80,4 %	98,6 %	99,0 %	99,4 %	99,7 %
40 à 59	Masculin	84,1 %	87,6 %	82,7 %	81,3 %	99,3 %	100,0 %	99,5 %	99,5 %
	Féminin	89,4 %	91,6 %	85,3 %	82,9 %	99,1 %	99,2 %	99,1 %	100,0 %
	Total	86,8 %	89,5 %	84,0 %	82,1 %	99,2 %	99,6 %	99,3 %	99,8 %
60 à 79	Masculin	90,7 %	92,3 %	88,2 %	81,2 %	98,9 %	99,8 %	99,6 %	99,6 %
	Féminin	91,3 %	92,3 %	79,8 %	76,1 %	99,3 %	98,0 %	99,5 %	99,0 %
	Total	91,0 %	92,3 %	83,7 %	78,4 %	99,1 %	98,8 %	99,5 %	99,3 %
Total 3 à 79	**Masculin**		**89,4 %**		**81,7 %**		**96,5 %**		**98,9 %**
	Féminin		**91,5 %**		**81,6 %**		**95,2 %**		**99,0 %**
	Total		**90,5 %**		**81,7 %**		**95,8 %**		**98,9 %**
Total 6 à 79	**Masculin**	**86,8 %**	**89,2 %**	**85,9 %**	**81,9 %**	**95,8 %**	**97,7 %**	**98,8 %**	**99,4 %**
	Féminin	**89,7 %**	**91,3 %**	**83,9 %**	**81,9 %**	**96,0 %**	**96,1 %**	**98,9 %**	**99,2 %**
	Total	**88,3 %**	**90,3 %**	**84,9 %**	**81,9 %**	**95,9 %**	**96,8 %**	**98,9 %**	**99,3 %**

Lorsqu'on met l'accent uniquement sur les groupes d'âge communs entre les deux cycles (6 à 79 ans), le taux de réponse au volet clinique a diminué pour passer de 84,9 % au cycle 1 à 81,9 % au cycle 2. Le taux de réponse au volet clinique pour les 3 à 5 ans a été relativement faible, à 79,5 %, et a fait en sorte que le taux de réponse global au volet clinique au cycle 2 a diminué, pour s'établir à 81,7 %.

On note que les taux de prélèvement de sang et d'urine ont été uniformément plus élevés au cycle 2 lorsqu'on met l'accent uniquement sur les groupes d'âge communs (6 à 79 ans). Les taux de prélèvement de sang pour les 6 à 79 ans ont augmenté, pour passer de 95,9 % à 96,8 %. Les taux de prélèvements d'urine pour les 6 à 79 ans ont augmenté, pour passer de 98,9 % à 99,3 %. Tout comme les taux de réponse au volet clinique, les taux de fourniture d'échantillons de sang et d'urine pour les 3 à 5 ans ont été plus faibles que les taux pour les 6 à 79 ans, ce qui a fait baisser les taux de prélèvement de sang et d'urine au cycle 2, ceux-ci s'établissant à 95,8 % et 98,9 %, respectivement. Toutefois, il convient de souligner que le taux final de prélèvement de sang pour les 3 à 5 ans (86,2 %) était beaucoup plus élevé que la valeur anticipée de 80 % prévue avant le début du cycle 2. Le taux de réponse élevé a contribué à l'atteinte de l'objectif de 500 spécimens de sang.

3.4 Nombre de répondants pour les sous-échantillons

Les résultats pour les sous-échantillons des prélèvements d'urine et de sang figurent dans le tableau 3.4-1. L'objectif pour le sous-échantillon des prélèvements d'urine était d'obtenir 250 prélèvements pour chaque groupe d'âge selon le sexe, sauf pour le groupe des 20 à 79 ans, dont l'objectif était de 500 prélèvements par sexe. L'objectif a été atteint pour chacun des groupes.

En ce qui a trait au sous-échantillon des prélèvements de sang, l'objectif original était d'en obtenir 250, tant pour le groupe des 12 à 19 ans que pour celui des 20 à 79 ans, selon le sexe. Juste après le milieu de la collecte du cycle 2, un échantillon supplémentaire de 500 participants a été affecté à ce sous-échantillon et a été réparti également entre les hommes et les femmes de 20 à 79 ans, ce qui a porté l'objectif final pour ce groupe d'âge à 500 par sexe. Encore une fois, l'objectif a été atteint dans chacun des groupes.

Tableau 3.4-1: Nombre de répondants des sous-échantillons (urine et sang)

Âge	Sexe	Sous-échantillon Urine		Sous-échantillon Sang	
		Cible	Obtenu	Cible	Obtenu
3 à 5	Masculin	250	260		
	Féminin	250	264		
6 à 11	Masculin	250	262		
	Féminin	250	254		
12 à 19	Masculin	250	256	250	254
	Féminin	250	257	250	253
20 à 79	Masculin	500	504	500	511
	Féminin	500	506	500	506
Total 3 à 79	**Masculin**	**1 250**	**1 282**	**750**	**765**
	Féminin	**1 250**	**1 281**	**750**	**759**
	Tous	**2 500**	**2 563**	**1 500**	**1 524**

Le dernier sous-échantillon était constitué de répondants à jeun. L'objectif pour ce sous-échantillon était simplement de faire en sorte que 50 % des répondants au volet clinique soient à jeun avant leur rendez-vous à la clinique. Parmi les répondants, les 3 à 5 ans, les personnes diabétiques et les femmes enceintes n'étaient pas admissibles pour être sélectionnés pour être à jeun. Comme on savait que ce ne sont pas tous les répondants sélectionnés pour un rendez-vous à jeun qui seraient à jeun dans les faits, le pourcentage de ménages sélectionnés initialement dans le cadre du sous-échantillon des participants à jeun se situait près de 55 %. Les résultats pour le sous-échantillon des participants à jeun figurent dans le tableau 3.4-2. Dans l'ensemble, 55,6 % (3 223/5 794) des répondants admissibles au volet clinique ont été

sélectionnés pour faire partie du sous-échantillon des répondants à jeun. Comme les mesures de laboratoire relatives au sous-échantillon des répondants à jeun touchaient les prélèvements de sang, seuls les répondants à jeun qui avaient donné des échantillons de sang et qui avaient au moins une mesure de laboratoire valide pour les variables liées au fait d'être à jeun ont été inclus dans le fichier de sous-échantillon final. Au total, 86,7 % des personnes sélectionnées pour les rendez-vous à jeun étaient à jeun dans les faits et avaient au moins une mesure de laboratoire valide, ce qui a donné lieu à un taux global de répondants à jeun de 48,2 % (2 793/5 794). La raison pour laquelle le taux final était inférieur à l'objectif de 50 % est que les répondants à jeun n'ont pas fourni d'échantillon de sang ou n'avaient pas au moins une mesure de laboratoire valide.

Tableau 3.4-2: Nombre de répondants et taux de répondants à jeun pour le sous-échantillon, par emplacement de collecte

Emp. #	Répondants Clinique (6 à 79)	Pas sélectionnés pour être à jeun	Sélectionnées pour être à jeun			
			Total	À jeun avec au moins une mesure de labo valide	% sélectionnés pour être à jeun et qui étaient à jeun	% à jeun
1	357	147	210	197	93,8 %	55,2 %
2	341	149	192	166	86,5 %	48,7 %
3	323	150	173	158	91,3 %	48,9 %
4	340	133	207	171	82,6 %	50,3 %
5	308	136	172	152	88,4 %	49,4 %
6	343	156	187	169	90,4 %	49,3 %
7	369	163	206	173	84,0 %	46,9 %
8	378	196	182	160	87,9 %	42,3 %
9	331	152	179	151	84,4 %	45,6 %
10	356	144	212	183	86,3 %	51,4 %
11	255	113	142	113	79,6 %	44,3 %
12	249	117	132	122	92,4 %	49,0 %
13	311	133	178	151	84,8 %	48,6 %
14	357	161	196	178	90,8 %	49,9 %
15	293	128	165	143	86,7 %	48,8 %
16	274	123	151	114	75,5 %	41,6 %
17	316	135	181	155	85,6 %	49,1 %
18	293	135	158	137	86,7 %	46,8 %
Total	**5 794**	**2 571**	**3 223**	**2 793**	**86,7 %**	**48,2 %**

3.5 Taux de réponse au niveau des ménages par distance approximative à parcourir

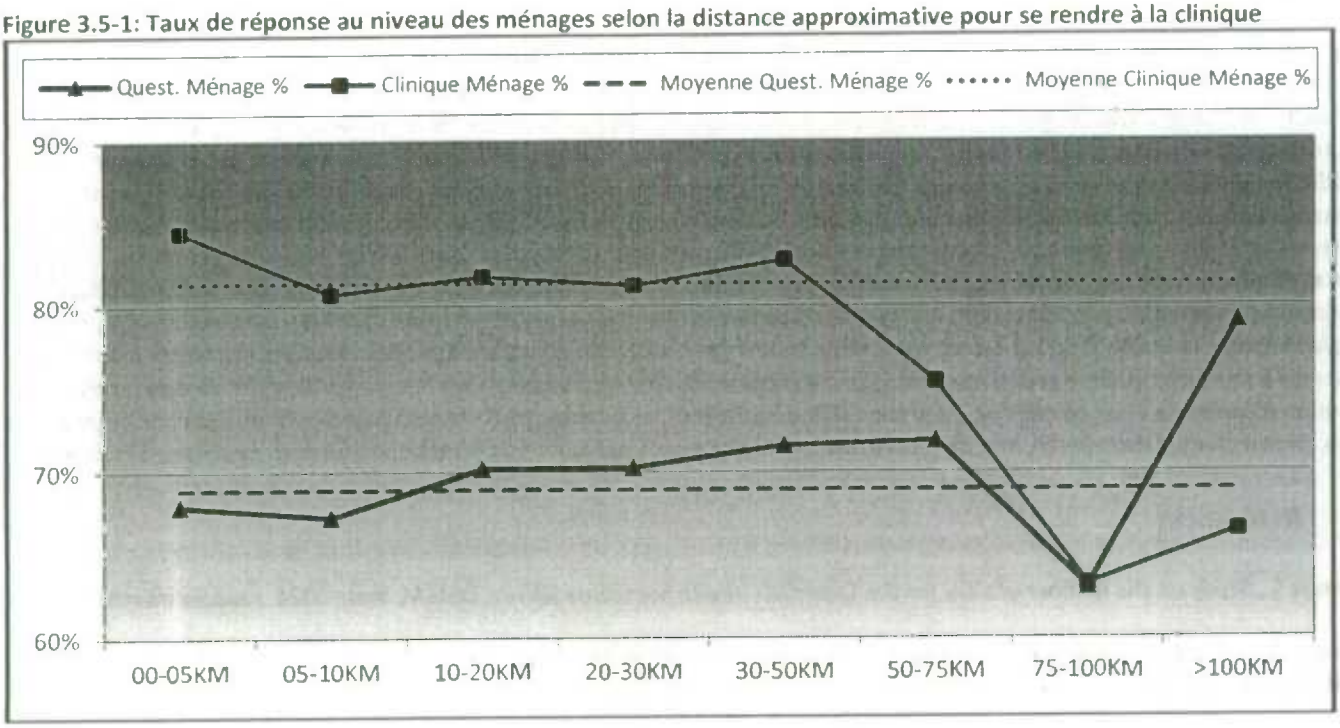
Dans le cadre de l'ECMS, chaque répondant doit se rendre à la clinique. Ainsi, il est important que les répondants n'aient pas à parcourir une distance excessivement longue pour se rendre à la clinique. Le tableau 3.5-1 comprend les taux de réponse au questionnaire et au volet clinique au niveau du ménage, ventilés selon la distance approximative à parcourir pour se rendre à la clinique.

Tableau 3.5-1: Taux de réponse au niveau des ménages selon la distance approximative pour se rendre à la clinique

Distance Approximative	Logements dans le champ de l'ECMS	Ménages avec au moins un répondant Questionnaire	Taux de réponse Questionnaire Niveau Ménage	Ménages avec au moins un répondant Clinique	Taux de réponse Clinique Niveau Ménage
≥0 et <5KM	2 813	1 912	68,0 %	1 616	84,5 %
≥5 et <10KM	2 259	1 520	67,3 %	1 227	80,7 %
≥10 et <20KM	1 788	1 255	70,2 %	1 027	81,8 %
≥20 et <30KM	501	352	70,3 %	286	81,3 %
≥30 et <50KM	415	297	71,6 %	246	82,8 %
≥50 et <75KM	391	281	71,9 %	212	75,4 %
≥75 et <100KM	138	87	63,0 %	55	63,2 %
≥100KM	215	170	79,1 %	113	66,5 %
Globalement	8 520	5 874	68,9 %	4 782	81,4 %

Pour la plupart des logements, la distance approximative a été calculée à partir de la distance entre les coordonnées spatiales de l'emplacement de la clinique et les coordonnées spatiales de la grappe de l'EPA comprenant l'adresse sélectionnée. La distance correspond à la distance directe à vol d'oiseau entre la clinique et le logement sélectionné et permet de prédire de façon acceptable la distance à parcourir en voiture pour les emplacements plus urbains, en raison du plus grand nombre de routes disponibles. L'utilisation de cette approximation de la distance à parcourir pour les emplacements plus ruraux présente davantage de problèmes, le moins grand nombre d'options de route pouvant mener à une sous-estimation importante de la distance à parcourir. Dans le cas des zones rurales pour lesquels ce problème se posait, les distances approximatives à parcourir en voiture ont été déterminées à partir d'une recherche sur Google Maps entre l'adresse de la clinique et la ville indiquée pour l'adresse. Les taux de réponse du tableau 3.5-1 se situent au niveau du ménage, ce qui fait que si au moins une personne a répondu au questionnaire ou au volet clinique, le ménage a été considéré comme répondant. On a procédé ainsi pour éviter le dénombrement en double des ménages à deux personnes qui ont toutes les deux répondu à l'enquête. Une représentation graphique des taux de réponse au questionnaire et au volet clinique au niveau du ménage figure dans la figure 3.5-1.

Figure 3.5-1: Taux de réponse au niveau des ménages selon la distance approximative pour se rendre à la clinique



Les taux de réponse au questionnaire au niveau du ménage pour les personnes qui avaient des distances approximatives à parcourir inférieures à 75 km ont été uniformément près de la moyenne de 68,9 % pour le cycle 2. Dans le cas des logements se situant entre 75 km et 100 km, le taux de réponse au questionnaire au niveau du ménage a diminué, pour s'établir à 63,0 %. Le taux a étonnamment augmenté pour atteindre 79,1 % dans le cas des logements qui se trouvaient à 100 km ou plus de la clinique, et ce groupe a obtenu le taux le plus élevé de réponse au questionnaire. Comme c'est l'intervieweur qui devait se déplacer pour procéder à l'interview du ménage dans le logement sélectionné, on ne s'attendait pas à ce que la probabilité de réponse au questionnaire soit significativement affectée par la distance entre le logement et la clinique.

Dans le cas des taux de réponse au volet clinique au niveau du ménage, on s'attendait à ce que de plus longues distances entraînent des taux de réponse plus faibles, et c'est ce qui a été observé au cycle 2. Les logements situés à des distances approximatives inférieures à 50 km ont affiché des taux de réponse au volet clinique de près de 80 % au niveau du ménage, ce qui se rapproche de la moyenne globale de 81,4 %. On a noté une baisse du taux lorsque la distance augmentait pour se situer entre 50 km et 75 km (soit 75.4 %) et à une baisse encore plus grande pour les logements se situant entre 75 km et 100 km de distance (soit 63,2 %) et à plus de 100 km (soit 66,3 %). Du fait de ces résultats, les exigences relatives à la distance à parcourir ont été modifiées pour les emplacements de collecte du cycle 3. Pour le cycle 1 et le cycle 2, la distance maximale entre le centre d'un emplacement de collecte et sa limite la plus éloignée avait été établie à environ 50 km pour les emplacements urbains et à environ 100 km pour les emplacements ruraux. La raison pour laquelle certains répondants ont dû parcourir plus de 100 km dans le tableau 3.5-1 est que la clinique de certains emplacements ruraux du cycle 2 se situait à la limite de l'emplacement (parce c'est là que se trouvait la ville/le village le plus important). De ce fait, les répondants qui vivaient à la limite opposée ont dû parcourir davantage que les 100 km prévus. Au cours de la création des emplacements de collecte pour le cycle 3, la distance maximale à parcourir pour les emplacements ruraux a été réduite à 75 km. En outre, on a effectué plus d'une tentative pour utiliser l'emplacement prévu de la clinique (dans la mesure du possible) au moment de la détermination des distances approximatives à parcourir. Pour certains emplacements, on a eu recours à des trajets approximatifs pour calculer les distances, plutôt qu'aux distances à vol d'oiseau. Ces changements devraient faire en sorte que les distances à parcourir dans le cas des emplacements ruraux soient plus raisonnables et devraient contribuer à améliorer les taux de réponse au volet clinique.

4 Conclusion

Du fait de son plan de sondage, le cycle 2 de l'ECMS a permis d'atteindre le nombre visé de répondants requis pour produire des estimations nationales fiables selon le groupe d'âge et le sexe. Il a été possible d'y arriver grâce à un plan d'échantillonnage à trois degrés utilisant la base de sondage du Recensement de 2006, mise à jour au moyen des fichiers administratifs les plus à jour. Grâce à une surveillance constante au cours de la collecte des données et à des mesures additionnelles, comme le réajustement des facteurs d'échantillonnage pour la sélection au niveau de la personne, la modification des probabilités de sélection des sous-échantillons et l'utilisation de vagues d'échantillon supplémentaire, le plan d'échantillonnage a permis d'atteindre le nombre visé de répondants, sans dépasser de façon significative le nombre total minimum de répondants requis. Les résultats ont fait ressortir l'importance de mettre à jour la composition du ménage utilisée pour la stratification, dès que des sources de données auxiliaires sont disponibles, non seulement pour les logements de la strate des 3 à 5 ans, mais pour toutes les strates de groupe d'âge. Les résultats ont aussi montré des aspects à améliorer pour le prochain cycle, à savoir modifier la stratégie de sélection des personnes afin de tenir compte du groupe d'âge et du sexe, plutôt que du groupe d'âge seulement, et adapter les distances à parcourir au cours de la création des emplacements de collecte, afin de réduire encore la distance à parcourir par les répondants pour se rendre à la clinique.

5 Références

Giroux S., *Study on the number of sites for the Canadian Health Measures Survey*, DMEM, mars 2004, rapport interne.

Giroux S., Lavigne N., *Sampling Strategy for the Canadian Health Measures Survey, Part I: Selection of Collection Sites*, DMEM, mars 2005, rapport interne.

Giroux S., *Enquête canadienne sur les mesures de la santé: aperçu de la stratégie d'échantillonnage*, Rapports sur la santé (Statistique Canada, catalogue 82-003) 2007; 18 (supplément): 35-40.

Giroux S., Labrecque F., Quigley A., *Le cycle 2 de l'Enquête canadienne sur les mesures de la santé: cambiner recensement et données administratives paur améliarer l'efficacité de la base de sandage*, Recueil du Symposium 2010, Statistiques sociales : interaction entre recensements, enquêtes et données administratives, Statistique Canada, Ottawa.

Gray, G.B., *On increasing the sample size (number af psu's)*, memorandum technique, 1973.

Laflamme G., Dochitoiu C., *Creating Sites far the Canadian Health Measures Survey*, avril 2005, rapport interne.

MacIsaac, K., *The Effect of Cluster Sampling in the Canadian Health Measures Survey*, 2011, rapport interne.

Statistiques Canada. *PCODE Centralized Service*, novembre 2011, document interne.

Statistique Canada., *Méthadalogie de l'enquête sur la papulation active du Canada*, (no 71-256-XPB au catalogue) Ottawa: Ministre de l'industrie, 1998.

Annexe A

Exemple de l'échantillonnage au niveau personne:

Afin de déterminer quels membres du ménage seraient échantillonnés pour l'enquête, on a attribué à chaque membre éligible un facteur d'échantillonnage tiré du tableau 2.3.1-1. Le total des facteurs d'échantillonnage pour le ménage et la valeur cumulative pour chaque membre ont par la suite été calculés. Ces totaux et valeurs cumulatives ont été calculés indépendamment pour les groupes des 3 à 11 ans et des 12 à 79 ans, parce que l'échantillonnage a été effectué de façon indépendante. Chacun des groupes de 3 à 11 ans et de 12 à 79 ans s'est vu attribuer un nombre aléatoire, qui a été multiplié par son total respectif, afin de déterminer le ou les points de sélection. Tant pour les 3 à 11 ans que pour les 12 à 79 ans, la première personne éligible dans la liste, dont la valeur cumulative était égale ou supérieure à son point de sélection respectif, a été sélectionnée.

Voici les données sur les membres du ménage tirées d'une liste de logements échantillonnés fictifs (FC = membre à temps plein des Forces canadiennes) :

Nom	Âge	FC?
Mike	3	Non
Molly	8	Non
David	13	Non
Mitch	34	Non
Sheila	37	Oui
Mary	58	Non
John	62	Non

Données pertinentes :

- Comme au moins une personne de 3 à 11 ans est présente dans le ménage, deux personnes seront sélectionnées.
- Deux nombres aléatoires sont produits dans l'application du ménage :
    - 0,5312 pour les 3 à 11 ans
    - 0,4204 pour les 12 à 79 ans
- Dans la base de sondage de l'enquête, ce ménage a été échantillonné à partir de la strate 3 à 5, ce qui fait que les facteurs d'échantillonnage correspondant du tableau 2.3.1-1 seront utilisés.
- Comme Sheila travaille à temps plein dans les Forces canadiennes, elle n'est pas éligible.
- Il y a quatre personnes âgées éligibles entre 12 et 79 ans. Étant donné que cela est inférieur à la valeur pour le nombre de 12 à 79 ans dans le ménage (6), les facteurs d'échantillonnage du tableau 2.3.1-1 demeurent tels quels.
- Il n'y a qu'une personne entre 12 et 19 ans. Étant donné que cela est inférieur à la valeur de réinitialisation pour le nombre de 12 à 79 ans (3), les facteurs d'échantillonnage du tableau 2.3.1-1 demeurent tels quels.

Le tableau qui suit résume l'éligibilité de chaque membre, les facteurs d'échantillonnage et les facteurs d'échantillonnage cumulatifs.

Nom	Âge	Forces armées canadiennes?	Éligible?	Facteur d'échantillonnage	Valeur cumulative 3 à 11	Valeur cumulative 12 à 79
Mike	3	Non	Oui	3	3	-
Molly	8	Non	Oui	2	5	-
David	13	Non	Oui	5	-	5
Mitch	34	Non	Oui	1	-	6
Sheila	37	Oui	Non	-	-	-
Mary	58	Non	Oui	1	-	7
John	62	Non	Oui	1	-	8
Total 3 à 11					5	
Total 12 à 79						8

Les facteurs d'échantillonnage totaux sont de 5 pour les 3 à 11 ans et de 8 pour les 12 à 79 ans. Ces valeurs totales sont multipliées par leurs nombres aléatoires respectifs (0,5312 pour les 3 à 11 ans et 0,4204 pour les 12 à 79 ans), afin d'obtenir les points de sélection.

- Le point de sélection pour les 3 à 11 ans est 5*0,5312 = 2,6560. Le premier membre admissible dont la valeur cumulative 3 à 11 est supérieure ou égale au point de sélection aléatoire est **Mike**.
- Le point de sélection pour les 12 à 79 ans est 8*0,4204 = 3,3632. Le premier membre admissible dont la valeur cumulative est supérieure ou égale au point de sélection aléatoire est **David**.