Article

New Initiatives in Data Collection for the Canadian Community Health Survey

by Marie-Claude Duval

2009

Statistics Canada   Statistique Canada

Canada

New Initiatives in Data Collection for the Canadian Community Health Survey

Marie-Claude Duval[1]

Abstract

The Canadian Community Health Survey (CCHS) was redesigned in 2007 so that it could use the continuous data collection method. Since then, a new sample has been selected every two months, and the data have also been collected over a two-month period. The survey uses two collection techniques: computer-assisted personal interviewing (CAPI) for the sample drawn from an area frame, and computer-assisted telephone interviewing (CATI) for the sample selected from a telephone list frame. Statistics Canada has recently implemented some data collection initiatives to reduce the response burden and survey costs while maintaining or improving data quality. The new measures include the use of a call management tool in the CATI system and a limit on the number of calls. They help manage telephone calls and limit the number of attempts made to contact a respondent. In addition, with the paradata that became available very recently, reports are now being generated to assist in evaluating and monitoring collection procedures and efficiency in real time. The CCHS has also been selected to implement further collection initiatives in the future. This paper provides a brief description of the survey, explains the advantages of continuous collection and outlines the impact that the new initiatives have had on the survey.

Key Words: Continuous collection; Call scheduler, Paradata, Data collection monitoring.

1. Introduction

The Canadian Community Health Survey (CCHS) is a cross-sectional survey that Statistics Canada has been conducting since 2001. In January 2007, the CCHS became the first Statistics Canada survey to collect data continuously in the field. The sample is 65,000 respondents annually. Before 2007, the survey was administered over a one-year period every two years, with an overall sample of 130,000 respondents.

Aside from continuous collection, the Agency has implemented some new initiatives in the area of collection methods and tools to reduce the response burden and survey costs while maintaining or improving data quality. This paper describes those initiatives, the reasons for them and the impact on the CCHS. Section 2 provides an overview of the survey, and each of the subsequent sections covers one of the new initiatives: continuous collection, limit on the number of calls per unit, telephone call management, and collection monitoring. The final section deals with future initiatives.

2. Overview of the survey

The purpose of Statistics Canada's CCHS is to produce regional estimates concerning health determinants, health status and use of the health care system.   Its target population is all persons aged 12 or over living at home in the ten provinces and three territories. People living on Indian reserves or Crown lands, institutional residents, full-time members of the Canadian Forces and people living in selected remote areas are excluded. The CCHS covers about 98% of the Canadian population aged 12 and over. The areas of interest are health regions, which are defined within the ten Canadian provinces and three territories. In 2007, there were 122 health regions. Also of interest are estimates by age group and sex.

---

[1] Marie-Claude Duval, Statistics Canada, R. H. Coats Bldg, Ottawa, Ontario, Canada, K1A 0T6
(marie-claude.duval@statcan.gc.ca)

For the most part, the survey uses two sample frames: an area frame composed of dwelling clusters, and an administrative list frame composed of telephone numbers. A sample of dwellings is taken from the area frame, and a sample of telephone numbers is selected from the list frame. The two samples are of equal size, about 32,500 respondents annually. When a household is contacted, one member aged 12 or over is selected to answer a questionnaire. Different selection probabilities are used to select the one household member aged 12 or over. The probabilities vary with the age and number of people in the household. On average, the questionnaire takes 45 minutes to complete. The survey uses two collection methods: computer-assisted personal interviewing (CAPI) for the area frame sample, and computer-assisted telephone interviewing (CATI) for the list frame sample. For more details concerning the survey, see Statistics Canada, 2007.

The interviewing is managed at Statistics Canada's seven regional offices (ROs) across the country. Each RO receives a sample and is responsible for managing and monitoring data collection in the field. A head office is responsible for providing the sample as well as the tools and applications required for collection operations at the ROs. In conjunction with the ROs, the head office produces reports and conducts the studies needed to evaluate, monitor and improve collection.

3. Continuous collection

Despite a number of adjustments since its inception, the CCHS remains in growing demand to meet new requirements for health data from various health partners in government, the academic community and the private sector. The demands vary and relate to different dimensions, such as data on a specific health topic, or health data for a specific population or a more detailed geography than health regions. The data required to meet these emerging needs also have to be available in a reasonable period of time. The volume of demands exceeded the survey's capacity. Therefore, to address these needs more effectively, consultations were held with the various health partners to discuss a plan to restructure the survey and improve its effectiveness and flexibility. One of the suggestions made following the consultations was to run the survey in continuous collection mode, i.e. collect data continually in the field instead of once every two years (Béland, 2005).

As a result, the survey has been in continuous collection mode since 2007, using an approach that combines data from several collection periods and even several years, which definitely makes it more efficient and flexible at meeting the emerging needs described above. The sample of 130,000 respondents over two years is now split evenly among non-overlapping two-month collection periods (12 in a two-year period). The samples allocated to the collection periods are also non-overlapping and independent. There are about 10,500 respondents (5,250 from the area frame and 5,250 from the list frame) for each two-month collection period.

Continuous collection has a number of advantages. First, it offers substantial benefits in terms of more efficient use of collection infrastructure and resources. It stabilizes interviewer workloads over time by decreasing busy periods and slack periods both during the year and between years. This stability should also help retain interviewers and thus improve their skills through the accumulation of regular experience. Continuous collection is a good way to provide on-going training, as staff members deal with changes in the survey and difficulties encountered during collection. In addition, since it is continually administered in the field, the survey provides a framework for studying and testing new data collection initiatives. The CCHS is particularly attractive for new initiatives for two reasons aside from providing a continuous collection framework: it uses two collection methods (personal and telephone interviewing), and its sample is large.

Second, continuous collection as implemented by the CCHS – non-overlapping collection periods with independent, also non-overlapping samples of equal size for each collection period – helps to meet emerging needs for health data. As noted previously, it allows the accumulation of data over more than one year so that the number of units can be increased. This is especially useful for producing estimates for small domains, such as health characteristics that are rare in the population, and geographic areas that have small populations. Continuous collection also makes it possible to add new content for a single collection period so that a specific health-related topic that is not covered by the current survey can be addressed quickly. This feature, known as "Rapid Response", is available on request. With Rapid Response, content requiring up to two minutes of interviewing time can be added to the current survey

for a period of two months. The feature is intended primarily for specific health-related topics that are not covered by the survey, where there is a need for national data by age group or sex. The additional content is confined to two minutes in order the limit the response burden. In addition, continuous collection makes it possible to provide samples to other surveys. The CCHS may serve as an initial phase for selecting a sample of households or individuals with a specific characteristic over one or more collection periods.

Third, since the survey is continually in the field, changes can be made at various points in the year to improve collection. Changes can be made in terms of sampling, data collection applications, data collection procedures, questionnaire content and many other components of the survey process.

4. Call limits – CATI system

In 2007, Statistics Canada issued a new directive to reduce the response burden and survey costs, or at least to improve collection efficiency without increasing costs. The directive placed a limit on the number of calls per selected unit. It applied to the majority of telephone interviewing surveys. For cross-sectional telephone surveys using a telephone number list frame, such as the CCHS, the maximum number of calls was 25. The limit was the same regardless of the response rate, the length of the collection period or the type of respondent.

Before the cap was implemented in 2007, CCHS managers conducted a study to assess its potential impact on the response rates and estimates. They were concerned about some hard-to-reach groups, such as males and the 25-34 age group. The study used CCHS data from 2005. At that time, the survey was administered over a one-year period every two years. Thus, the study covered nearly 130,000 respondents. CATI respondents contacted after 25 calls were assigned non-respondent status, and new estimates were produced. The study revealed that, overall, 7% of respondents answered after 25 calls, and that the rate reached almost 13% among men aged 25 to 34. The distribution of the results by age group and sex is presented in Figure 4-1.

Figure 4-1
Percentage of CATI respondents contacted after 25 calls by age group (12-17,18-24, 35-44, 45-64, 65+) and sex (male, female) (CCHS 2005)
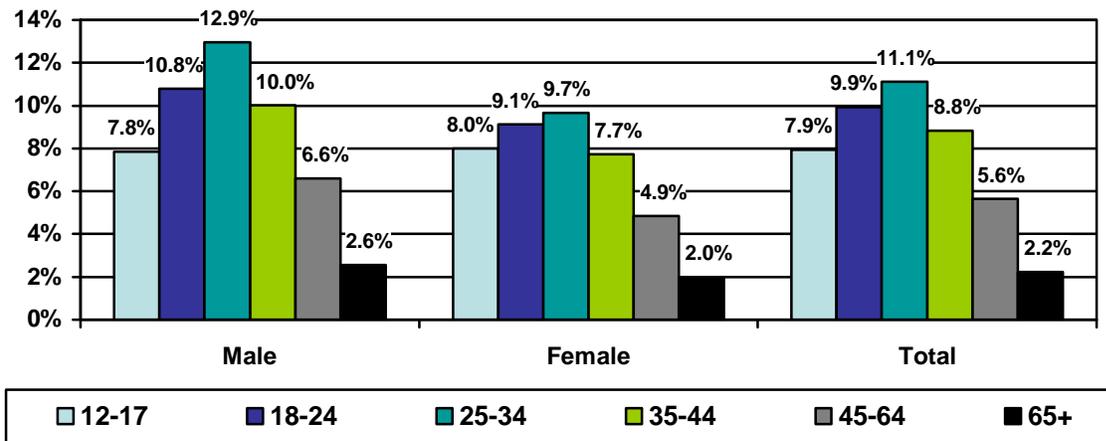
Some estimates were also affected: 2.5% of the 46,479 estimates calculated for key variables and areas of interest (health regions, age-sex groups) were significantly different at the 5% level when a limit of 25 calls was applied (Deleva, 2006). In response to these results, Statistics Canada raised the cap to 40 calls for CCHS CATI respondents for a specified period so that the impact could be analyzed further and the studies could be redone with the survey in continuous collection mode. The 40-call limit reduced the number of respondents by about 2%, which was deemed acceptable.

The imposition of a call limit forced a review of the collection procedures to improve call management so that the selected units could be contacted in as few calls as possible, which meant reaching them at more convenient times. Statistics Canada's BLAISE telephone data collection system[2] has a call scheduling utility that can be used to manage calls. The use and impact of the call scheduler in the CCHS are discussed in the next section.

5. Call scheduler or call management

The call scheduler or call management system available in BLAISE[1] can limit the number of calls during various periods of the day or week. It helps manage calls so that respondents are contacted during different periods and at more convenient times. The user can define settings such as the various time slices in a day and the maximum number of calls in each time slice. For example, the maximum number of calls could be set at two between 8:00 a.m. and noon and five between 6:00 and 9:00 p.m., since the chances of reaching a respondent in the evening are higher. The total number of calls in all time slices must equal the maximum number of calls per unit, which is 40 for the CCHS.

The calling strategy can be customized in the call scheduler, i.e. different time slices can be used depending on variables such as the respondent's age and the dwelling type. In the CCHS, only the hour and the day are used prior to initial contact, since no auxiliary information is available before that. However, once a household member is selected, his or her age is considered in determining the maximum number of calls per period: 12-17, 18-24, 25-64, 65 and over. The first group consists of young people who may be at school; the second, of young adults who may be at school or at work; the third, of adults who may be at work; and the fourth, of people who may be retired. In defining the settings for the CCHS, paradata from the 2005 survey were used to select the most effective periods of the day. To determine effectiveness, we developed a success rate by dividing the number of respondents by the total number of calls during the period. Results by age group and time of day are shown in Figure 5-1.

Figure 5-1
Success rate (respondents/number of calls) by age group (12-17, 18-24, 25-64 and 65+) and time slice (CCHS 2005)
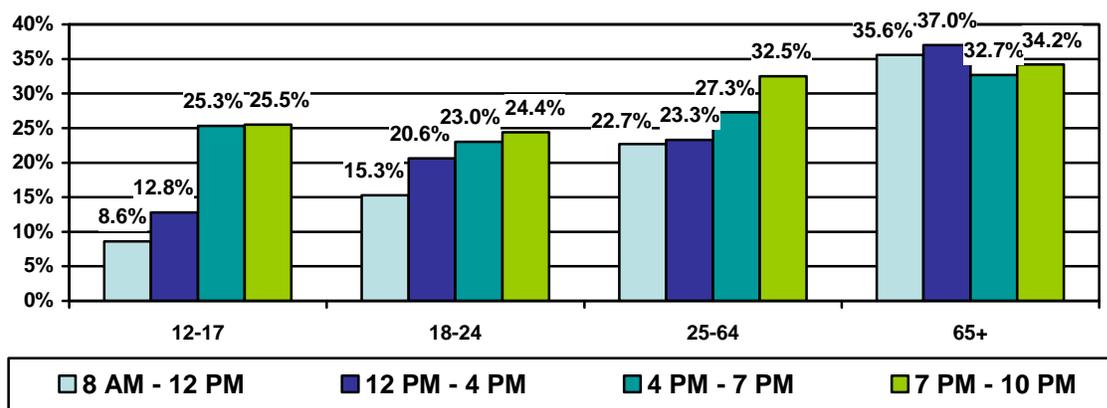
As the figure indicates, the success rate is generally higher in the evening. It is considerably better in the evening for the 12-17 age group, while for the 65 and over group, it is slightly higher during the day but also good in the evening. These results were used to program the settings so that evening calls would be given priority before initial contact, and then to adjust the settings when the respondent's age was known.

Although the call scheduler's purpose is to contact the selected units more efficiently, its use is limited. First, the initial call is random. The call scheduler does not come into play until the second call. Second, it is used only for numbers for which there was no answer. It plays no role in refusals or cases where a response is obtained. For

---

[2] BLAISE: A program that serves as a platform for computer-assisted telephone interviewing (CATI) applications.

respondent calls where the interview cannot be conducted immediately (the selected person is busy or not at home), the interviewer makes an appointment. An appointment is considered more effective than the call scheduler. Refusals are managed by more experienced interviewers who have specific follow-up procedures. For these reasons, the call scheduler is used in only 30% of the calls on average. Nevertheless, it is effective at organizing calls and should improve the chances of contacting respondents with fewer calls.

The CCHS has been using the call scheduler since January 2007, and its experience suggests that the scheduler's settings follow the interviewers' work schedule, and vice versa. It is pointless to program the settings so that 90% of calls are made in the evening if only 60% of the interviewers work evenings. In that case, the interviewers would not have enough units available to contact during the day, which is not desirable. The system may also stop presenting cases to contact for a given time slice if the maximum number of calls has been reached or if there are too many interviewers in that time slice. Case shortages tend to occur at the end of a collection period, when there are fewer cases to contact. If that is true, the settings should be adjusted over the course of the collection period, which the call scheduler allows, or the interviewers' work schedules should be changed. Efficient management of resources and continuous monitoring of the scheduler's results with paradata during each collection period are required. Lastly, in the CCHS, the call scheduler is not used on the final three collection days (usually Friday, Saturday and Sunday), during which there is a final push to complete as many cases as possible and the scheduler's call-limit-per-time-slice feature is not needed. It is important to note that the 40-call limit per unit is still applied.
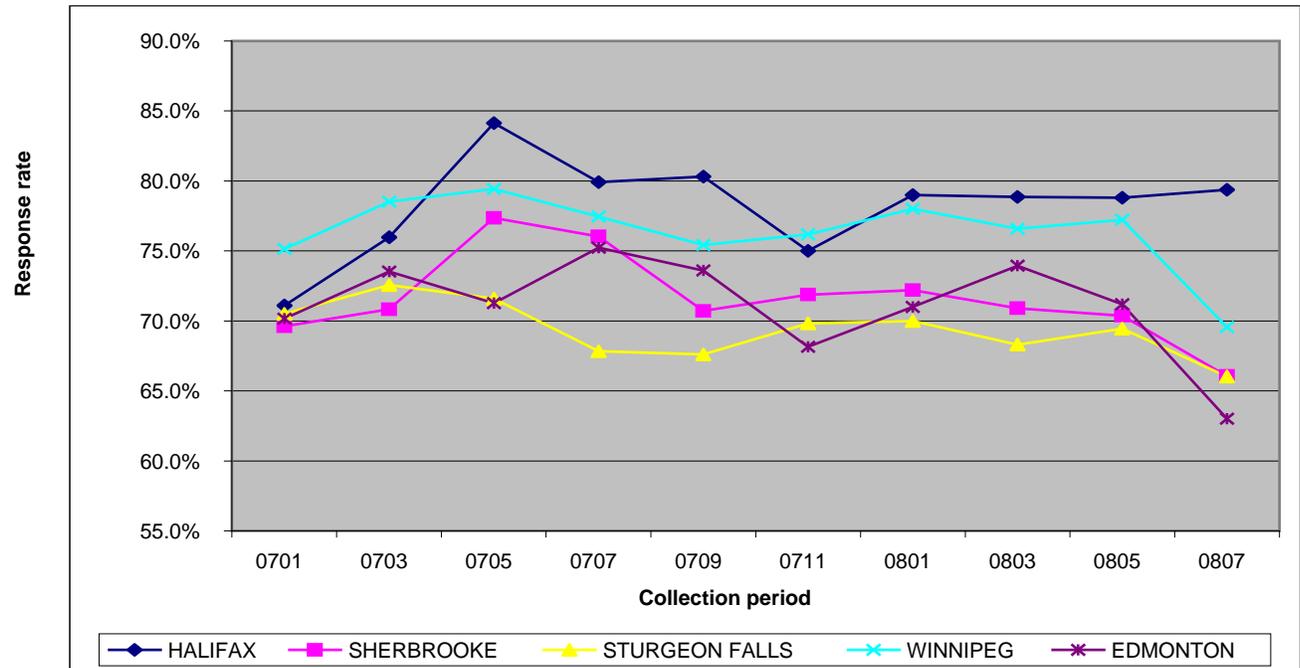
6. Monitoring data collection

Monitoring data collection involves observing and checking collection work and ensuring that it is proceeding well so that high-quality data are collected, with as many completed questionnaires as possible within the specified time frame. Collection monitoring is considered effective if it helps identify flaws and if collection can be adjusted in mid-stream to correct the flaws and/or make improvements. The flaws and/or improvements might affect any aspect of a survey, including sampling, computer applications, collection procedures and questionnaire content. With continuous collection, improvements can be made not only in the middle of the current collection period but also in subsequent collection periods. For example, in the sampling process, sample sizes can be adjusted on the basis of the latest response rates from previous collection periods; in the collection process, an effort can be made to reduce the number of refusals by mailing out follow-up letters as early as possible in the collection period or providing regular training. Likewise, in the area of questionnaire content, a question can be improved or reworded when interview monitoring shows that respondents are having trouble understanding it.

To identify flaws and make improvements, the CCHS uses paradata, which are available on a regular basis. The paradata are used to produce various reports: daily progress reports that show response rates per day, cases resolved and cases in progress; more detailed progress reports produced every two weeks during a collection period, showing the response rate for each health region, the number of refusals and refusal conversions, the average number of calls per unit, the number of units for which the call limit has been reached, and so on. In addition, many reports are generated after each collection period to assess the collection process for that period or over a longer time frame. The results are compared between collection periods and between regional offices. Other reports are generated to measure the impact of new collection initiatives.

The results are then shared at various points in the survey between managers at head office and those in regional offices. The results reveal the effectiveness of a collection period and of any new procedures that have been introduced, so that improvements can be made if necessary. Head office and regional office managers hold monthly conference calls to discuss problems and make suggestions and adjustments. Annual meetings are held at head office or at the regional offices to discuss broader issues concerning management, new tools, initiatives, best practices and results analysis. These discussions between the regional offices and Head Office are vital. Each regional office works independently and has operational expertise, while head office tends to focus on developing collection tools, procedures and guidelines. Each party has a specific role, and these roles complement one another. The parties need to hold discussions, exchange information and work together to improve collection, which has been very beneficial for the CCHS.

For example, Figure 6-1 shows the response rate for each collection period from January 2007 to July 2008 for each regional office. A presentation on this report was given at the regional offices. The variations in response rates over time were discussed. The Sherbrooke Regional Office pointed out that refusal conversion training had been given prior to the May-June 2007 collection period, which had the highest response rate.

Figure 6-1
CATI response rates by collection period and regional office

The next collection monitoring challenge for the CCHS will be to develop active management during the collection process, with greater use of sampling methodology in the collection of data (Groves, 2006).

7. Future initiatives

As noted above, the CCHS is often selected as the testing ground for new collection initiatives because it uses continuous collection and two collection techniques and because it has a large sample. The CCHS is currently involved in two new intiatives: computer-assisted recorded interviewing (CARI) and the Survey Master Control System (SMCS).

Computer-assisted recorded interviewing (CARI) involves the recording of personal interviews for the purposes of monitoring personal interviewing collection. The recording makes it possible to assess interviewer performance and provide feedback, identify interviews that do not meet standards, and ensure consistency in collection within a survey and between surveys. The CCHS is the first survey to use CARI. A pilot test was conducted in the November-December 2008 collection period in three health regions. The CCHS sample for the three regions was used, consisting of about 300 respondents. A few questions (unknown to the interviewers) were selected for recording. The interviewers were informed of the recording and the recording was made only if the respondent gave consent. The main purpose of the pilot test was to assess recording quality and the feasibility of evaluation methods based on recordings in a production environment. The pilot test will also be used to analyze the potential impact on responses, based on separate questionnaires administered to the interviewers and the respondents, and the potential impact on the survey in general (the impact on refusals, for example). The analyses are currently in progress.

The Survey Master Control System (SMCS) can switch from one collection technique to another (for example, from CATI to CAPI or vice versa) or from one collection site to another (for example, from one regional office to another). This is also a first at Statistics Canada. The SMCS will lower the risk factor by allowing cases to be transferred from one site to another in the event of an overload or some other limitation (strike, natural disaster, interview in a non-official language spoken by an interviewer at a particular office, etc.). It will also make it possible to switch collection techniques to accommodate a respondent. For example, a respondent who is selected for a personal interview but is away during the collection period could be interviewed by telephone. Specific rules will be established for case transfers in order to limit the potential bias due to the collection method. The pilot test is scheduled for the summer of 2009.

8. Conclusion

The CCHS has been in continuous collection mode since January 2007. Since it is continually in the field, the CCHS is much more flexible and better prepared to address emerging needs for health data on a specific topic or population. It also allows the accumulation of data over time, enabling the volume of data concerning small domains to be increased. Continuous collection also means that the CCHS is available in the field at any time to test the Agency's new collection initiatives. It is doubly attractive since it uses two collection techniques (personal and telephone interviewing).

As in any survey, it is vital to have a continuous collection monitoring system to ensure that the survey is progressing well and that high quality data are being collected. The CCHS has monitoring procedures based on reports produced with paradata from the collection systems. The monitoring leads to improvements in sampling, collection system applications, data collection procedures and questionnaire content. With continuous collection, improvements can be made not only in subsequent collection periods but also in the middle of the current collection period. The next monitoring challenge for the CCHS will be to develop active management during the collection process, with greater use of sampling methodology in collection management.

9. Acknowledgements

References

Béland, Y., Dale, V., Dufour J. and Hamel, M. (2005). The Canadian Community Health Survey Building on the Success from the Past, *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Deleva I. (2007). CCHS Cap25 Study - Measuring the Impact on Estimation from Setting a Cap on Calls for Data Collection, Statistics Canada

Groves, R.M. and Heeringa, S.G. (2006). Responsive Design Responsive for Household Surveys: Tools for Actively Controlling Survey Errors and Costs, *Journal of Royal Statistics Society: Series A,* 169, 439-457.

Statistics Canada (2007). Canadian Community Health Survey (CCHS): User Guide – 2007 Microdata Files, Statistics Canada.