

PREVENTING NON-RESPONSE IN THE CANADIAN COMMUNITY HEALTH SURVEY

Yves Béland,¹ Johane Dufour and Marc Hamel

ABSTRACT

To fill statistical gaps in the areas of health determinants, health status and use of the health system by the Canadian population at the health region level, Statistics Canada established a new survey, the Canadian Community Health Survey (CCHS). The CCHS consists of two separate components: a regional survey in the first year and a provincial survey in the second year. The main purpose of the regional survey, for which collection took place between September 2000 and October 2001, was to produce cross-sectional estimates for 136 health regions in Canada based on a sample of more than 134,000 respondents. This article focuses on the various measures taken at the time of data collection to ensure a high level of quality for this large-scale survey.

KEYWORDS: Collection, CAI, call centre

1. INTRODUCTION

1.1 Rationale for the new survey

In connection with the Health Information Roadmap Initiative (Canadian Institute for Health Information, 1999), Statistics Canada held consultations across the country in the spring of 1999 with more than 225 key stakeholders in the health field to determine their data needs. These consultations gave rise to some requirements that dictated the methodological development of the Canadian Community Health Survey (CCHS). Statistics Canada had to develop a flexible survey instrument that would produce cross-sectional estimates for 136 health regions in Canada. In addition, development of the CCHS had to be consistent with Statistics Canada's data collection infrastructure, policies, and survey capacity. It is clear that the introduction of such a large-scale survey has led to major changes in the Agency's collection infrastructure.

1.2 Sample design

The CCHS's target population covers individuals aged 12 or over living in private dwellings. Persons living on Indian reserves and Crown lands, institutional residents, full-time members of the Canadian Forces and residents of certain remote areas were excluded. The CCHS covers about 98% of the Canadian population aged 12 and over.

To provide reliable estimates for the country's 136 health regions with the survey's budget, a net sample of about 134,000 respondents was needed. Although producing reliable estimates at the health region level was the primary objective, the quality of the provincial estimates for certain key characteristics was an important goal. A three-stage approach was taken to ensure that the health regions and the provinces were treated more or less equally. We also made sure that with a few exceptions, every health region had at least 500 respondents (Béland, Dufour, Lessard, Morano, Saint-Denis and Thivierge; 2000).

¹ Yves Béland, Household Survey Methods Division, Statistics Canada, 16th Floor, R.H. Coats Building, Tunney's Pasture, Ottawa, Ontario, Canada, K1A 0T6.

The CCHS used two overlapping frames for its sample of households: the Labour Force Survey (LFS) area frame and the random digit dialling (RDD) frame (Morano, Lessard and Béland; 2000). In the CCHS, the area frame was used as the primary frame, and the RDD frame served as a supplementary frame in some health regions. The basic design of the LFS is a multistage stratified sample in which the dwelling is the ultimate sampling unit (Statistics Canada, 1998). For each geographic area selected at the first stage, a list of dwellings was prepared and then updated in the field. At the second stage, a sample of dwellings was chosen from each list. Households living in the selected dwellings made up the household sample. Most of the sample (88%) came from the area frame, and personal interviews were to be conducted with all persons selected from the households in that frame.

The random digit dialling (RDD) sampling method uses the Elimination of Non-Working Banks technique (Norris and Paton; 1991). A bank of telephone numbers (area code plus the first five digits) was considered as “working” for the purposes of sampling if it contained at least one valid residential number. Working banks were regrouped to create RDD strata to encompass, as closely as possible, the health region areas. In each RDD stratum, one bank was selected at random, and a random number between 00 and 99 was generated to form a complete 10-digit telephone number. This process was repeated until the number of telephone numbers required for the sample was reached in that RDD stratum. The persons selected from the households in this frame, who made up 12% of the sample, were interviewed by telephone.

The technique used to choose the respondents in the selected households provided a representative number of young people (age 12-19) and elderly people (65 and over). User needs, costs, sample design efficiency, response burden and operational constraints were considered in its development (Thivierge; 1999). In about 83% of the households selected from the area frame, one person aged 12 or over was chosen at random; two persons aged 12 or over were selected at random in the remaining households. The technique used to select the persons was based on household composition, and its primary objective was to increase the representation of the two subgroups of interest. In RDD households, one person aged 12 or over was chosen at random from the members of the household.

2. INITIAL STRUCTURE OF COLLECTION OPERATIONS

2.1 Organizational restructuring

To conduct the CCHS, Statistics Canada made major changes to its collection infrastructure. This organizational restructuring took the form of a consolidation of collection resources (Statistics Canada; 1999), which changed the role of field interviewers substantially. Whereas in the past the field interviewer’s job could involve either personal or telephone interviewing, it now consisted solely of personal interviewing. In addition, the majority of telephone interviewing for Statistics Canada’s household surveys was shifted to collection centres (or call centres) in the Agency’s five regional offices (Halifax, Montréal, Toronto/Sturgeon Falls, Edmonton and Vancouver). The collection centres are actually satellite offices of the regional offices.

This restructuring was a necessary response to the huge workload generated by the number of personal interviews that the CCHS required. The infrastructure change was made gradually between May and September 2000, when CCHS collection began. Needless to say, it not only resulted in massive recruitment of new interviewers but also put considerable strain on all of Statistics Canada’s collection programs. For the CCHS, all interviewers on both the field team and the call centres team took part in collection operations.

2.2 Initial data collection plan for the CCHS

The initial plan called for data collection between September 2000 and early October 2001, a period of 13 months. This plan was carefully designed to ensure that the survey’s quality objectives would be met. A number of the plan’s operational aspects were borrowed from similar Statistics Canada surveys. For

example, the response rate targets were set at 90% for the area frame sample and 85% for the RDD frame. These targets were based on Statistics Canada's experience with other surveys of the same kind, such as the National Population Health Survey for area frame units and the General Social Survey for telephone frame units. The following strategies, common to a number of Agency surveys, were also employed: introductory letters describing the survey's importance were mailed to area frame units, and out-of-service telephone numbers were checked with the telephone companies.

In addition, a number of parameters needed in planning the CCHS were fine-tuned in various pilot tests prior to collection. The length of the interview, 45 minutes on average, was determined in qualitative tests conducted in early 2000. Various operational procedures such as case transfers between the regional offices and Head Office were also refined in two pilot tests. A pilot test for area frame units was held in April 2000 with a sample of 2,500 cases from across Canada. A similar test for telephone frame units was carried out in June 2000 on a nation-wide sample of 4,225 cases. On the basis of the results obtained in these tests, it was determined that the optimum feasible size of a monthly task was 18 personal interview cases for field team interviewers and 80 cases per computer workstation in the call centres. If the latter figure seems high, it is because two shifts – a day shift and an evening shift – were using the same computer each day.

To even out the interviewers' workload and eliminate any seasonal effects on health-related characteristics (such as physical activity), the CCHS's final sample was divided in 12 so that each month of the year would be properly represented for each health region. Hence, a sample of new cases was transferred to the regional offices at the beginning of each month and returned to Head Office at the end of the month. In addition to these monthly caseloads, follow-up had to be done for non-respondent units from previous months. This follow-up procedure is common to several of Statistics Canada's household surveys. After three consecutive months with no contact, non-response cases were set aside until the very end, when a 13th month of collection was planned. This additional month provided interviewers with an opportunity for a final attempt to convert the non-response cases.

For the CCHS, all the interviews, whether they were personal or telephone interviews, were computer-assisted (CAI), and all the interviewers were trained accordingly. In addition, to ensure that collection operations ran smoothly, a quality assurance program involving an interviewer supervision and monitoring process was in operation during the collection period.

Because of some technical constraints in the collection system, all the cases, whether complete or not, had to be transferred to Head Office at the end of each month so that the system would have space to download the next month's sample. It should be noted that this process completely shut down collection operations for two or three days. As a result, the interviewers had between 18 and 22 days each month to complete their tasks during the collection period.

3. COLLECTION OPERATIONS

For most of Statistics Canada's household surveys, collection operations proceed smoothly and within the limits of established parameters. In this case, however, the total workload due to CCHS collection and the organizational restructuring proved to be a formidable challenge. To ensure the success of collection operations, a number of established procedures were altered, some more than others. For the reader's convenience, the observations and the changes made in collection operations are described separately for the field team and the call centres team.

3.1 Field team

For personal-interview units in the area frame, the response rates in the initial months of collection were below the targets. After three months of collection, the national average response rate was 81%. It was clear that the workload involved in CCHS collection was too much for the available human resources in some regions. In other words, the distribution of the CCHS sample across the health regions did not match the geographic distribution of available interviewers. The field team interviewers were distributed so as to best

meet the multiple requirements of Statistics Canada's many surveys. The CCHS's very specific sample requirements created a serious imbalance in interviewer workloads. Some interviewers had twice as much work as originally planned, while others had half. The imbalance was exacerbated by the day-to-day collection activities of other surveys and the organizational restructuring of the Agency's collection operations, which resulted in a high turnover among the interviewers. The impact of this problem is evident in the regional differences in response rates. After three months of collection, the response rates were over 90% in some regions and 70% in others. The recruitment of new interviewers and the economic conditions in some parts of the country also contributed to the problem. The imbalance that resulted in heavy workloads in some regions would remain a problem throughout CCHS collection.

An unexpected problem with the computerized collection application surfaced in the first few months of the survey operation. When collection operations were shut down for two or three days while the old month's cases were transferred to Head Office and the cases for the new month were downloaded, a number of outstanding cases had to be dropped. Specifically, cases in which an appointment had been made with the selected person at a date following the shutdown could not be returned the next month; this created a considerable number of cases of person-level non-response. For technical reasons, these cases could not be downloaded again the next month. (These cases included instances where a household member was contacted but the person selected for the interview was not home at the time; in such instances, an appointment was made to conduct the interview at a later date.) In the first three months of collection, the person-level non-response rate was about 10%, much higher than expected. The expected rates based on the National Population Health Survey were 3% to 4%. To remedy this problem, a special collection procedure was introduced so that these special cases could be downloaded. Over 700 cases were reinitialized at Head Office; fictitious households were created, and selection of the previously chosen respondent was forced. Once the fictitious selection was completed, the cases were downloaded separately in a distinct environment. Though costly in terms of human resources, this initiative was successful since over half of the cases were converted the following month.

Because of the initial technical problems and the imbalance in interviewer workloads, a number of non-response cases and outstanding cases did not receive the attention they required. That is, a number of field team interviewers were simply unable to properly carry out the non-response follow-up procedures. The result was that each month, a larger number of non-response cases had to be downloaded, which added to the interviewers' already excessive workloads.

This snowball effect, combined with the enormous pressure on interviewers to improve their response rates, led to a substantial increase in the number of proxy interviews. After six months of collection, about 8% of the interviews had been completed by proxy. In a health survey, which involves a number of personal characteristics that cannot be obtained from another person, the normal proxy rate is only 2% to 3%. Proxy interviews are supposed to be confined to cases where the person will be away for a long period or has a physical or mental disability. In the CCHS computerised collection application, about two thirds of the questionnaire can be completed during proxy interviews; the result is a significant amount of partial non-response. A special imputation model was added to data processing to address this partial non-response problem.

A number of steps were taken to remedy the various problems encountered in the field. Basic measures such as hiring more interviewers, redistributing workloads in some areas and assembling special teams to do non-response follow-up were developed and implemented from the regional offices.

Other measures were implemented from Head Office. The addresses of dwellings selected from the area frame were matched with the corresponding telephone numbers using various computer files available at the Agency. For about 65% of the addresses of selected dwellings, a telephone number was downloaded as well so that interviewers could make initial contact with the households. This saved time for the interviewers by reducing the number of no-contact field visits.

As soon as the computerized collection application had the capability, the collection period was lengthened to two months instead of one. Monthly samples were still downloaded at the beginning of each month, but

the interviewers now had two months to complete the monthly workload. While this resulted in some overlap of monthly samples, the interviewers had more flexibility (time) to carry out the follow-up procedures for non-response cases. The decision was even made to stop downloading the non-response cases the next month; they were set aside and kept for the 13th month of collection.

Because such a large workload had accumulated in the first few months of collection, the size of the January 2001 sample was reduced substantially. To allow the interviewers to catch up and the regional office managers to hire and train more staff, the field team's workload was cut by 80% at the Vancouver Regional Office, 66% in Toronto and 50% in Montréal and Halifax. Except for Ontario's cases, surveying of the remainder of the sample was simply postponed to the fall of 2001, and a 14th month of collection (October 2001) was added at the very end. In Ontario, where the situation was more serious, all of the unsurveyed cases (about 2,000), which were from the area frame, were dropped from the sample and replaced with units from the telephone frame.

3.2 Call centres team

As in the case of field collection, the response rate targets for the telephone frame units were not met. Three factors played a major role: the inexperience of the new interviewers, the uneven distribution of call centre workloads and the low hit rates in some parts of the country.

As noted earlier, the start of CCHS collection operations coincided with the establishment of five call centres in the Agency's five regional offices. As a result of this major change, a large number of new interviewers had to be hired to conduct all the telephone interviewing required for household surveys, including the CCHS. With a new survey and new interviewers, response rates were bound to be affected.

In addition, because of the complexity of the CCHS's sample design, the number of units required from the telephone frame varied widely from region to region. Proportionately, this created a major imbalance in total workload among the various regional offices. For example, compared with Halifax, the Toronto/Sturgeon Falls call centres, which serve Ontario, received a monthly workload that was much too large for the number of available interviewers. The Vancouver and Quebec centres were working flat out. At the Edmonton call centre, which served the three Prairie provinces, the problem was different. The hit rates – the percentage of cases in which a valid household was reached – were far lower than expected. The average hit rate for the northern parts of the three provinces was less than 25%, with some regional rates as low as 16%. The low population density in these large areas and the method used to generate the sample of telephone numbers led to the selection of a great many invalid numbers, which wasted the interviewers' time and energy. In practice, many banks of telephone numbers (area code plus the first five digits) that made up the RDD strata contained few valid numbers. After three months of collection, the national response rate for telephone frame units was a mere 74%.

A number of steps were taken to remedy the situation at the call centres. Naturally, the regional offices hired and trained additional personnel. Other measures were implemented from Head Office. In particular, the telephone numbers selected from the RDD frame were matched with the corresponding civic addresses using various computer files available at the Agency. An introductory letter describing the importance of taking part in the survey was then mailed out two weeks before the initial call. Although this process was more time-consuming and less effective than for area frame units, it was necessary. On the basis of tests carried out as part of the General Social Survey collection operations, in which telephone interviewing was used, response rates improved by 2% to 3%. (It should be noted, however, that the impact was not accurately measured in the case of the CCHS.)

Also using various computer files, additional information was provided to the interviewers. A flag (possibly residential, possibly commercial or possibly invalid) was assigned to each selected telephone number downloaded to the call centres. This information was very useful to the interviewers because their workload could be prioritized by telephone number type. For example, possibly commercial numbers were assigned to the day shift, and possibly residential ones to the evening shift. Possibly invalid numbers were tried only

four or five times before being sent to the telephone company for verification. Monthly workloads were redistributed; half of the Toronto/Sturgeon Falls cases were sent to Halifax. The effect on the response rates for cases assigned to the Toronto/Sturgeon Falls call centres was immediate; the rates were 5% higher in February than in January. On the other hand, there was a slight decline in Halifax, but the rate bounced back up the next month.

In addition to the measures described above, a priority list at the health region level was prepared on a monthly basis. The health regions were grouped by category based on their cumulative response rates (1, response rate \leq 50%; 2, 50-60%; 3, 60-70%; 4, 70-85%; and 5, \geq 85%). Within a call centre, the cases from high-priority health regions were then assigned to the best interviewers in an effort to improve the lowest response rates for those regions.

Of all the measures taken to rectify the situation, the most significant was probably the change in sampling strategy for the RDD frame. The switch was made from a strategy based on selecting telephone numbers by the elimination of non-working banks method to simple random sampling of numbers from a list frame of possibly valid numbers. This major change was introduced gradually over a period of three months in health regions where the area frame was also being used. The reason for this was very simple: although the telephone number list frame was not perfect, the use of the area frame in the background provided adequate coverage of the target population. (It is important to note that an adjustment for telephone number list frame undercoverage, based on various computer files available in-house, was incorporated into the weighting strategy.) This change had a direct impact on hit rates, as it substantially reduced the number of attempts needed to contact the required number of respondents. In some regions, the hit rate jumped from 20% to 60%.

4. SOME QUALITY INDICATORS AND CONCLUDING REMARKS

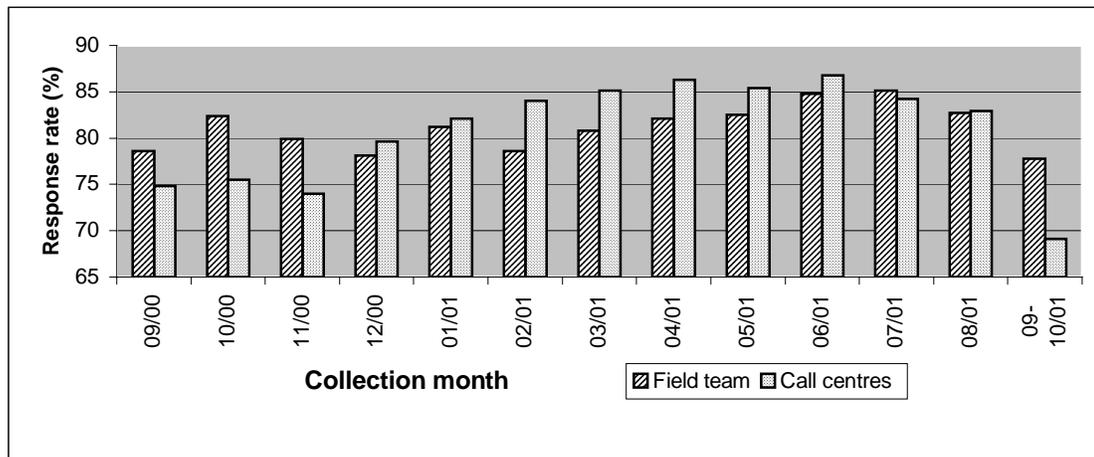
Although it is difficult or even impossible to assess the impact that each of the corrective measures described above had individually on response rates, since many of them overlapped, the overall effect was quite evident. The average national response rate was nearly 85% after 14 months of collection. Table 1 shows the final response rate by collection method (field team and call centres team) for each province and territory. It also shows the proxy response rates for each province and territory. The response rates for health regions, not shown here, range from 70% to 95%.

Table 1. Final response rate by collection method, and proxy response rate, CCHS

Province/Territory	Response rate (%)			Proxy response rate
	Field team	Call centres	Total	
Newfoundland	86.6	89.3	86.8	4.9
Prince Edward Island	87.7	82.6	84.7	4.2
Nova Scotia	88.8	89.3	88.8	3.9
New Brunswick	88.4	92.4	88.5	9.1
Quebec	85.7	84.8	85.6	5.2
Ontario	82.8	79.5	82.0	5.4
Manitoba	90.0	85.0	89.5	8.3
Saskatchewan	87.0	85.4	86.8	5.9
Alberta	85.2	84.9	85.1	5.9
British Columbia	83.9	86.7	84.7	9.5
Yukon	79.3	95.6	82.7	8.4
Northwest Territories	89.6	85.4	89.2	12.7
Nunavut	66.3	34.6	62.5	6.9
Canada	85.1	83.1	84.7	6.3

Chart 1 shows the monthly response rates for Canada by collection method. Clearly visible are the slow start of CCHS collection activities and the gradual improvement in response rates as the corrective measure were introduced. It is important to note that the 12,636 non-response cases accumulated in the first 11 months of collection and downloaded again in mid-August 2001 for a final conversion attempt were not included in the calculation of the monthly response rates in Chart 1. Of these cases, 42.9% were successfully converted by field team interviewers, and 29.6% by call centre interviewers. Non-response cases are much more difficult to convert by telephone than in person.

Chart 1. Monthly response rates by collection method, Canada



Data collection for the Canadian Community Health Survey presented a formidable challenge for Statistics Canada. The workload was so heavy that collection resources had to be restructured to accommodate it. A number of corrective measures were taken during collection to control the level of non-response. All these changes were made worthwhile by the success of the CCHS. The lessons learned in this initial version of the CCHS will help prevent non-response in future cycles of this survey. Detailed health data are now available for over 131,000 Canadians aged 12 and over in 136 health regions. With those data, decision-makers and health professionals will be able to establish benchmarks and monitor the progress of health promotion programs in their regions. Further details concerning the CCHS are available on Statistics Canada's Web site: http://www.statcan.ca/health_surveys.

REFERENCES

- Béland, Y., Dufour, J., Lessard, S., Morano, M., Saint-Denis, O. and Thivierge, S. (2000), "Plan de sondage de la composante régionale de l'Enquête sur la santé dans les collectivités canadiennes", Proceeding of the survey methods section, Statistical Society of Canada, p. 155-60.
- Canadian Institute for Health Information (1999), "Health Information Roadmap : Responding to needs". (1-895581-30-3)
- Morano, M., Lessard, S. and Béland, Y. (2000), "Creation of a dual frame design for the Canadian Community Health Survey", Proceeding of the survey methods section, Statistical Society of Canada, p. 249-54.
- Norris, D.A., Paton, D.G. (1991), Canada's General Social Survey : five Years of Experience. *Survey Methodology* (Statistics Canada, Catalogue 12-001); 17, pp. 227-240.

Statistique Canada (1998), *Methodology of the Canadian Labour Force Survey*, Statistics Canada. Cat. No. 71-526-XPB.

Statistique Canada (1999), "Task Force - Survey Operations Capacity and Infrastructure capabilities".

Thivierge, S. (1999). "Selection of two persons in some households of the Canadian Community Health Survey", Household Survey Methods Division, internal document.