Non-response Management During Fieldwork: The 2016 Canadian Census and National Household Survey Strategy

Patrice Mathieu, Chief Methodologist, Social Survey Methods Division, Statistics Canada

Key words: Census of Population, National Household Survey, collection methodology, two-phase sampling, non-response management

1 Introduction

The Census of Population is conducted every five years in Canada. The most recent collection took place between in 2011, with May 10 as reference day, and the next Census will be conducted in 2016. Starting in 2011, a new collection methodology was introduced. Known as the Wave Methodology, it consists of a series of contacts with households. The purpose of those contacts is to remind and encourage households to complete the Census, and in particular, to complete it online.

Also in 2011, the long census questionnaire previously distributed to one household in five was replaced by a voluntary survey, the National Household Survey (NHS), whose questionnaire was distributed to about a third of the population. A wave collection methodology was used for the NHS as well, but was adjusted to fit in to census operations as efficiently as possible. Planning and managing a survey of this size presented some major challenges. There was also a serious risk of bias due to an increase in the non-response rate, and various measures were taken to minimize that risk. This included subsampling the NHS non-respondents to concentrate the follow-up efforts on fewer units in order to reduce and control the workload and the cost of the operations.

Furthermore, a non-response management strategy was implemented in order to have a framework during the collection period for obtaining high and uniform response rates across the country,

This paper outlines in Section 2 the collection methodology of the Canadian Census and NHS, including the NHS sub-sampling strategy. Section 3 discusses the non-response management process.

2 Census and NHS collection methodology

Statistics Canada developed a streamlined approach to communicate with respondents during the collection of census information. This improved approach, called the wave methodology, reminds Canadians to fill out their questionnaire by various methods, at specific times throughout the collection period. It also encourages respondents to use the Internet to complete their questionnaire, while mitigating the risk of a decline in overall response.

2.1 Self-response phase

Census day in 2016 will be May 10. In most of the country, Census and NHS data collection begins with a self-response phase starting on May 2. Respondents are contacted in multiple waves to prompt self-response. The precise content of the waves depends on the geographic area in which a dwelling is located, but for a majority of the dwellings, the wave methodology is as follows:

- *Wave 1:* a letter, to be received on May 2, inviting respondents to complete the Census questionnaire online and offering the possibility of requesting a paper questionnaire;
- *Wave 2:* a reminder letter, received between May 11 and 17, reminding non-respondents to complete the questionnaire (also offering the possibility of requesting a paper questionnaire); and

- *Wave 3:* a questionnaire packaged (combined Census and NHS questionnaire for dwellings selected for the NHS) received between May 19 and 27.
- *NHS Wave:* an NHS-specific reminder letter received between May 26 and 30 and sent only to Census Internet respondents who have not completed the NHS despite being invited to do so at the end of the Census questionnaire.

Respondents choosing the Internet will be invited to continue with the voluntary NHS questions as after they have completed the mandatory Census questions. Similarly, for those who prefer completing a paper questionnaire, the paper form provided to NHS-selected dwellings is a combined questionnaire, containing the mandatory Census questions followed by a section with the voluntary NHS questions.

Late respondents typically have a different socio-demographic profile compared to early respondents, but given that waves 1 and 2 encourage respondents to use the electronic questionnaire and with paper questionnaires being sent out only at wave 3, the profile of later self-respondents tends to differ even more from that of earlier self-respondents. This is something that has to be kept in mind when monitoring the progress and in the next phases of collection.

2.2 Non-response follow in the field

The non-response follow-up (NRFU) operation starts on June 1st for a majority of the areas in the country. Census enumerators telephone and visit households that have not responded to complete the questionnaire. A computerized system accessible over the Internet, the Collection management Portal is used by field staff to facilitate the gathering of collection progress information. It helps managing the task more effectively and making information about progress and results available faster, while facilitating communications between field employees at all levels and with Head Office employees.

NRFU is a heavy drain on human and financial resources. The collection infrastructure consists of several local census offices (25 in 2016) managing collection in their respective geographic areas and about 35,000 field employees. Each local census office also has a telephone support unit that is used strategically to help enumerators by calling non-respondent households in areas where collection is slower. Non-respondent households with a telephone number in the Address Register are often contacted first by that unit and later, if necessary, by enumerators.

2.3 NHS second phase sampling

The Census being mandatory, every household in the country is required to complete a questionnaire and every non-responding dwelling is visited by an enumerator during NRFU. For the 2011 Census, the final response rate was 97.1%. For the NHS, as indicated previously, only a sub-sample of NHS non-respondents is interviewed during the NHS NRFU period in the field. This allows not only reducing and controlling the workload and the cost of the operations but also targeting areas where more resources should be allocated and more efforts deployed in order to reduce the non-response bias.

This approach, initially proposed by Hansen and Hurwitz, is an effective method to use when the second phase per-unit collection cost is high or when more intensive, focused efforts are required for a smaller number of non-respondents in NRFU. This method has been used successfully on a number of occasions, such as for the American Community Survey. The aim for the NHS was to put maximum effort into the cases retained for follow-up rather than dilute the effort over all non-response cases. Since collection is halted for a large portion of the NHS sample, the resulting response rate is obviously lower, but weighted response rates, which take into account the weighting factors of the subsampled units, can be calculated to provide a more accurate picture of collection progress with regard to the risk of non-response bias.

The details of the subsampling design, such as the size and allocation of the sample, cannot be fully determined in advance. They have to be developed within the limits of the operational constraints, mainly the unpredictable available budget and the limited capacity of the local census offices, whose top priority

is still census collection. Thus, the subsampling methodology has to be simple and flexible so that it can be adjusted for any unexpected scenarios that might arise during collection.

The second phase sampling is designed not only to reduce the risks of non-response bias by providing a better response rate for a representative sample of all non-respondents but also to increase the number of respondents for certain target populations. Thus, areas composed largely of populations of interest and whose response rates are typically lower are identified. The objective is to boost the representativeness of those populations by increasing the sampling fraction in the identified areas. The target populations are:

- Aboriginal persons living off reserve
- Recent immigrants and non-permanent residents
- Members of visible minorities
- Persons in the labour force who have no more than a high school diploma
- Households with an income of less than $20,000

Oversampling of the target areas helps increasing the number of respondents in the target populations, which lowers the sampling variance and facilitates donor imputation for those populations. In addition, the target areas are handled differently in the response rate management process described in the next section, which helps reducing the risk of non-response bias for these target populations.

However, it should be noted that the conditions under the second phase sub-sample is selected differ from the ideal Hansen and Hurwitz framework in a few ways. For example, the condensed collection calendar means that self-response generated by the letters and questionnaires delivered to respondents will still be coming in at a good rate at the beginning of NRFU. As a result, there will be many dwellings not in the subsample whose self-response will be received after sub-sampling. Sometimes referred to as "surprise" respondents, these dwellings are of interest because in the usual double-expansion estimator proposed by Hansen and Hurwitz they receive a weight of 0. In 2011, rather than giving surprise respondents a weight of 0, they were given weight of 1 to be self-representative and alternatives are considered for 2016.

3 Response rate management strategy

The main goal of the Census and NHS response rate management strategy is to stop collection in areas where quality targets have been met in order to redirect resources to areas where they are most needed. This strategy can be seen as an adaptive approach where the goal is not necessarily to obtain the lowest non-response rate as possible, but to ensure that the response rates are consistent across Canada at small geographic level, and that the respondents are representative of the population.

Response rate management is done at the level of the collection unit (CU), the enumerators' work unit. As soon as a CU's response rate is considered sufficient (the definition of sufficient is adjusted during collection), collection is halted in that CU, and the crew leader in charge is notified. When a CU is deactivated for collection, efforts are redirected to other CUs.

Each CU's response rate is assessed daily against a criterion specified in the strategy for that CU. The strategy differs for the Census and NHS and both strategies are explained in this section.

3.1 Census response rate management strategy

For the 2011 Census, A four-tier response rate management strategy took into account the fact that in the census, unlike voluntary surveys, there are no criteria for discontinuing follow-up, such as a maximum number of contacts with a household. The four tiers identified situations requiring changes in quality expectations so that resources could be used as efficiently as possible to obtain the most uniform response rates possible. The table below summarizes the four tiers used in 2011.

Tier 1	Scheduled for the month of June, the target is a response rate of 98% in each CU, or 96% if a response rate of 98% is achieved at a higher geography level. During that period, the progress at each local census office and potential quality problems is monitored carefully.
Tier 2	Reduce the volume of NRFU cases if an area is falling behind schedule or resources are becoming inadequate. Collection is halted in CUs where the impact on response rates and quality is smallest. (i.e CUs with the fewest unresolved cases until a reasonable overall impact level is reached are deactivated.)
Tier 3	Lower the targets on the basis of 2006 Census results, but with Winsorization to ensure that the target rates are not too low. This is done to take account of local difficulties encountered in the previous census.
Tier 4	Case-by-case review of the remaining CUs, with a view to making decisions that would yield high enough response rates to meet the minimum publication criteria.

3.2 NHS response rate management strategy

The 2011 NHS response rate management was similar to the Census strategy, but because of the subsampling done on non-response cases in the NHS, weighted response rates, mainly at the CU level, were used to manage the response progress. Since NHS collection progress varied substantially from CU to CU, resources could be redeployed to minimize the risk of non-response bias. However, that was subject to the Census's requirements, which always took priority.

The first step in the response rate management process was to obtain a weighted response rate of 90% in each CU, but without requiring a response for more than 90% of the outstanding cases in the CU. When a CU did not have enough unresolved cases to meet the target, the CU target was lowered, but the targets in other CUs in the same province were increased so that a weighted response rate of 90% could be achieved at the provincial level. As a result, the target set for some CUs was a weighted response rate of more than 90%. Though overly ambitious in many cases, the targets were helpful in organizing the field work.

After a few weeks, CUs were assigned one of three levels of importance based on the presence of populations at risk (target CUs described in the previous section) and on the heterogeneity of the population's characteristics. The measure of CU heterogeneity was made using a method that measured the complexity of the combinations of a wide variety of population characteristics derived from the 2006 long questionnaire. The CUs were assigned to categories ranging from the most heterogeneous to the most homogeneous, and the target weighted response rates were revised downward on the basis of the CU's level of importance. The target CUs were treated as having the same level as the most heterogeneous CUs. The use of levels of importance helped speed up the deactivation of CUs while minimizing the impact on the non-response bias.

Close to the end of collection, with a significant decrease in human and financial resources looming, it was decided to begin terminating collection in all CUs of a number of local census offices across the country. The offices were selected on the basis of the progress made to that point with regard to the weighted response rate and of the distribution of their CUs across the levels of importance. This ensured that the remaining resources were used as efficiently as possible.

A few days later, collection was halted in all of the most homogeneous CUs in some of Canada's major cities. This made it possible to concentrate collection efforts in the most heterogeneous neighbourhoods in those cities to reduce the risk of non-response bias.

3.3 Non-response management strategy for the 2016 Census and NHS

The 2011 response rate management strategy worked very well in general. It was beneficial to have rates set at the CU level. The national response rate achieved was in line with the goal and response rates were

generally uniform throughout the country. Having pre-defined tiers meant that programs were developed in advance and at any time the impact of applying a tier could be evaluated; however other scenarios were still considered. For the 2016 Census, it may be beneficial to have a more flexible plan whereby a variety of options are considered and one or a combination of options is applied.

3.4 Proposal for 2016 Census and NHS

In 2011, the response rate criteria were applied at the CU level because collection was managed at the CU level. However, the dissemination area (DA) is the geographic area of interest for determining whether or not results will be disseminated. It would therefore be ideal to also manage non-response at the DA level.

The proposal for 2016 consists of five tiers. It assumes that collection will proceed well. If there are serious issues (i.e., very low self-response rate) then some tiers may be shortened or skipped altogether. The five tiers are described below.

Tier 1	Monitor progress: The first few weeks of regular NRFU, monitor the progress of NRFU collection for potential quality issues such as very high rates of unoccupied dwellings or refusals.
Tier 2	Aim high: After a few weeks of NRFU so that enough time has passed since the start of collection to have a grasp of how NRFU is progressing and can expect to progress. For the Census, start with a target response rate 0.5 percentage points higher than the goal for collection; thus target 98.5%. To be closed, the CU must also meet criteria related to the dwelling count, unoccupied dwellings, incomplete questionnaires, refusals and absent household rates.
Tier 3	Aim for the goal: All outstanding cases in the area have a minimum number of attempts, determined in conjunction with the NRFU strategy. Drop the target response rate to the collection goal (98% for the Census). To be closed, the CU must also meet criteria related to the dwelling count, unoccupied dwellings, incomplete questionnaires, refusals and absent household rates (as in Tier 2).
Tier 4	Concentrate efforts: A (large) area has a lot of work left to do, low response rate and it is not expected that they will meet the collection deadline within the available budget. Efforts are redirected in the most problematic portions of the area.
Tier 5	Final efforts: At the end of collection when a last push is needed in areas that will not be able to publish data or with most to gain and greatest impact. Identify key areas (lowest response rate, most outstanding cases, highest heterogeneity score, lowest attempts per case, areas with populations of interest). Must focus efforts in order to be able to publish data and to reduce response bias.

The Census and NHS collection will be better integrated in 2016 and the response rate management strategies will be more similar. The homogeneity score and the presence of target populations should be used more extensively, and not only for the NHS, but also for the Census. A research project is also on-going in order to use administrative or auxiliary data about respondents in the management of non-response follow-up.

Also, from the perspective of controlling the risk of non-response bias for the NHS, a second measure becomes important, the response rate within the sub-sample.