

NEW COLLECTION METHODOLOGY IN THE 2006 CENSUS OF POPULATION

Limei Sun and Laurent Roy ¹

ABSTRACT

Several new collection methods were implemented in the 2006 Census of Population. For the first time, a block canvass operation was performed to update the Address Register, which made it feasible to mail the questionnaires through Canada Post to 70% of households across Canada. Completed questionnaires were mailed back to a central data processing centre where they were scanned and subjected to automated editing for completeness. Failed edit follow-ups were done from three Census Help Line Centres. The 2006 Census also provided the option for respondents to complete their questionnaires on the Internet. These new methods and some results will be discussed.

KEY WORDS: Census, Collection, Internet response

RÉSUMÉ

Plusieurs nouvelles méthodes de collecte ont été mises en place lors du Recensement de la population 2006. Pour la première fois, une opération de prospection des îlots a eu lieu pour mettre à jour le Registre des Adresses, ce qui a rendu possible la livraison par Postes Canada des questionnaires pour environ 70% des logements au Canada. Les questionnaires complétés étaient retournés au Centre de Traitement des Données pour la saisie des données à l'aide du balayage optique et ils étaient sujets à des vérifications automatisées pour la complétude des données. Les suivis pour les questionnaires rejetés au contrôle ont été faits à partir des trois centres d'Assistance Téléphonique du Recensement. Le Recensement 2006 a offert aux répondants l'option de compléter le questionnaire via l'Internet. Ces nouvelles méthodes et certains résultats seront présentés.

MOTS CLÉS: Collecte, recensement, réponse Internet

1. OVERVIEW OF MAJOR CHANGES TO THE CENSUS

The Canadian Census of population is conducted every 5 years. The Census provides a statistical portrait of the social and economic situation of Canada and its people. It is considered to be a photograph of the country at one specific point in time. Short-form questionnaires are sent to 80% of dwellings and long-form questionnaires are sent to 20% of dwellings. Short-form questionnaires contain 7 questions related to topics such as gender, date of birth, marital status, and mother tongue. Long-form questionnaires contain all the questions asked in a short form, plus 45 other questions such as education and income. It also contains a section on dwelling information which has questions such as the number of rooms, year of construction, and utility payment. Both types of questionnaires are self-administered. It is mandatory for all Canadian residents to fill out a Census questionnaire.

The most recent Census was conducted in 2006 and Census day was May 16, 2006. This Census involved quite a few methodological changes in data collection. It was the first time that questionnaires were mailed to 70% of dwellings through Canada Post. The Internet response option was offered for the first time to almost all respondents. A master control system was built to have centralized control during the whole data collection process for the first time.

Prior to the 2006 Census, the majority of questionnaires were delivered using a list/leave method: enumerators went into the field to list dwellings in sequence and leave questionnaires at the door. In the 2001 Census, about 98% of all dwellings in Canada were included in list/leave areas. The remaining 2% of dwellings belonged to canvasser areas, where enumerators conducted interviews during their visit. Canvasser areas were mainly Indian reserves and northern territories. The list/leave method involves hiring a large temporary workforce which was difficult. Fortunately, the extent and quality

¹ Limei Sun and Laurent Roy, 15th floor, R.H. Coats building, Statistics Canada, 100 Tunney's Pasture driveway, Ottawa, ON, Canada K1A 0T6
limei.sun@statcan.ca, laurent.roy@statcan.ca.

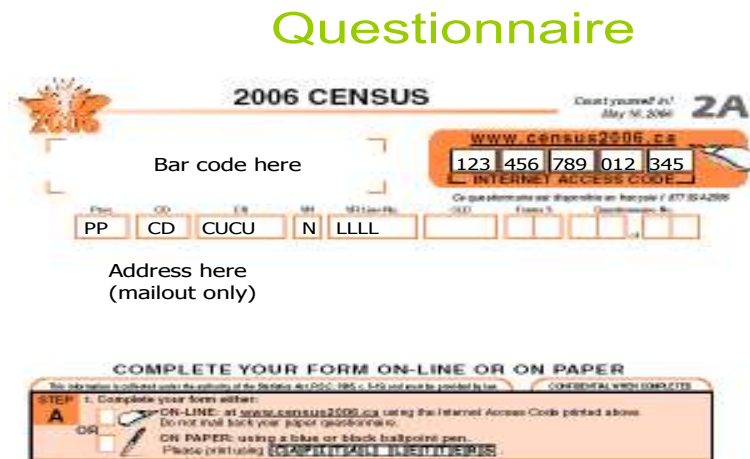
of Statistics Canada's Address Register had improved enough that mailout delivery had become a possible option. So to alleviate the difficulties of list/leave, in the 2006 Census questionnaires were mailed to 70% of all dwellings through Canada Post offices. These dwellings mostly belonged to urban areas where the Address Register had adequate coverage. To make this mailing operation successful, it was required that the mailing list used be up-to-date and accurate. A block canvass operation was performed to update the Address Register, and corresponding quality control was conducted to ensure the quality of the updates. For another 29% of dwellings, the list/leave method was used to deliver questionnaires as in the 2001 Census, while the remaining 1% of dwellings belonged to the canvasser areas.

In Censuses prior to 2006, both short and long questionnaires were in paper format only. For non-canvasser areas in the 2001 Census, completed questionnaires were mailed back to local Census offices and then forwarded back to the local enumerators for review, or respondents could call the Census Help Line (CHL) and the operators would complete paper questionnaires for them. Involving local enumerators in the collection of completed questionnaires increased public concerns about the privacy of their information. To resolve this issue, in the 2006 Census completed questionnaires were mailed back to a central data processing centre (DPC).

Internet use by households has grown rapidly in recent years. Thus in the 2006 Census, electronic questionnaires were developed and, for the first time, an Internet questionnaire was offered as a response option. It was introduced with the goals of reducing cost, improving quality, and to comply with the government policy on offering service electronically. Therefore, for non-canvasser areas, respondents could complete paper questionnaires and mail them back to the DPC, complete questionnaires on-line, or call CHL and the operators would complete electronic questionnaires for them.

Figure 1 illustrates the cover page of a paper questionnaire. The cover page contains the unique geographic identification Province code (prov), Census Division code (CD), Collection Unit code (CU) and the Visitation record line number (Vr line no). It has a unique Internet access code to be used for Internet responses, and it contains a pre-printed address for dwellings in mail-out areas. 'Complete your form on-line or on paper' is clearly stated in bold characters.

Figure 1. Cover page of short form.



The use of multiple response channels makes it necessary to have centralized control of completed questionnaires. For the 2006 Census, all completed questionnaires were returned to a centralized location in Ottawa for processing, and a management information system was developed to produce accurate and timely information on data collection. In this way, it was possible to measure progress, productivity and cost, and to help in the identification of different issues.

Completed paper questionnaires were data captured manually in the 2001 Census while in the 2006 Census, the format of the questionnaires was modified so that they could be scanned and data-captured by an image recognition system. Questionnaires with responses that were difficult for the system to recognize were identified and data-captured by clerks. Completeness and coverage checks of returned questionnaires were applied by an automated system, and questionnaires that failed coverage checks were reviewed by clerks interactively. Follow-up was performed from CHL for questionnaires that failed edits so that the labour force required was smaller than for follow-up done by local enumerators.

A Master Control System (MCS) was built as a dwelling frame. It contains all dwellings in Canada and it assigned a unique identifier and an Internet access code to every dwelling. A completed questionnaire could be linked back to the dwelling frame through the identifier so that its status could be updated; in this way the MCS would control and track questionnaires as they moved through the steps of collection and processing. More details on this process are provided by Dolson (2007).

The following sections present details about some major changes in collection methods for the 2006 Census.

2. BLOCK CANVASS

In the 2006 Census, questionnaires were mailed to 70% of all dwellings and the Statistics Canada Address Register was used as the mailing list. Therefore, the accuracy and coverage of the Address Register were critical to the success of the Census. In mail-out areas, a block canvass operation was performed to update the Address Register. Geographic areas were split into small Collection Units (CU) and each CU contained about 300 dwellings. The information from the Address Register for each CU was pre-printed in a booklet. This booklet was known as the Block Canvass Register (BCR). All in-scope CUs for the block canvass operation were assigned to enumerators. Enumerators walked the streets of a particular CU sequentially matching and comparing dwellings listed in the register to dwellings in the field. They were trained to perform the following tasks: verify the addresses, make corrections, add new dwellings, delete dwellings that were on the register but did not exist anymore; verify the dwelling type and classification; verify that dwellings were coded to the correct block; record key contact information for collective dwellings; and update maps.

Corresponding quality control procedures were developed to ensure the quality of the block canvass operation. It targeted all in-scope CUs for the block canvass and looked at different aspects of the enumerator's work: whether the Block Canvass Register was completed correctly, whether all dwellings received codes for dwelling status and dwelling type (valid dwelling, invalid dwelling, row-house, etc), whether all maps were updated correctly when needed, whether enumerators were able to find new dwellings (dwellings not pre-printed) when needed and whether enumerators were correctly deleting invalid dwellings pre-printed in the register. To evaluate the last two aspects, a sample of up to 12 addresses that were believed to exist were suppressed from the pre-printed list (suppressed sample), and a list of up to 4 non-existing/demolished dwellings were added to the pre-printed list for each in-scope CU (overcoverage sample). If the enumerator listed these suppressed dwellings and deleted the overcoverage addresses, it is likely that he/she was effective in finding other unlisted dwellings and deleting other pre-printed addresses which were invalid.

The quality control procedure was divided into two reviews: in the first review, quality control technicians first performed desk checks to verify the work of the block canvass enumerator using a variety of lists and forms. If any discrepancies were found, field visits would be conducted to determine which one was correct. Each CU was accepted or rejected according to the number and type of errors identified. If a CU was accepted, the errors would be corrected on the register. If a CU was rejected, the CU would be repaired when only parts of the assignment area required further work to correct identified errors, or re-canvassed if the quality problems were severe. A second quality control review was performed on items that failed the first review. Quality control results were documented and returned to a crew leader who would pass feedback to enumerators when needed.

Block canvass was conducted from September to October 2005. There were 9.5 million dwellings pre-printed in all BCR. As a result of block canvass, about 570,000 new dwellings were added and about 520,000 dwellings were deleted (including dwellings that were pre-printed in the wrong CUs). The number of updates performed indicates that it was necessary to have the block canvass conducted to ensure the quality of the Address Register.

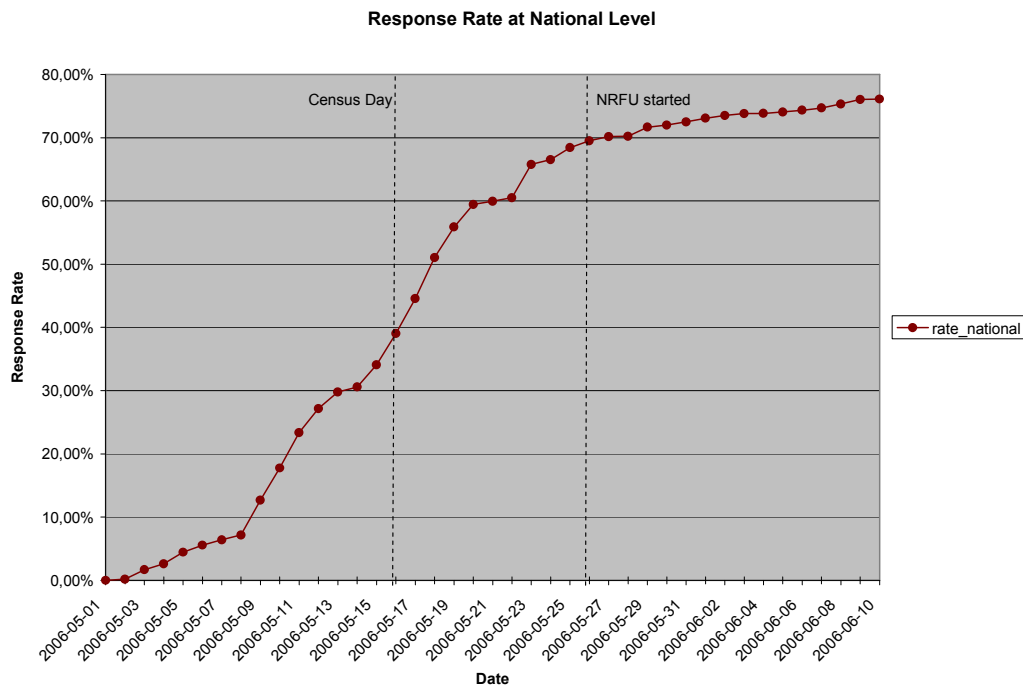
Because the block canvass was performed half a year before Census day (May 16, 2006), some dwellings' status might have changed during this period. Existing dwellings at the time of block canvass that were demolished at Census time would be identified during the non response follow-up (NRFU) period. During the block canvass operation, dwellings that were under construction (or not built yet) at block canvass time but were ready at Census day would be missed from the Address Register and no questionnaire would be mailed to these dwellings if the mailing address was unknown during block canvass. Therefore, it was necessary to follow-up block canvass. The follow-up was conducted in February 2006 and was called late block canvass.

The late block canvass operation revisited CUs with potential for new dwellings. The identification of these CUs took place during the block canvass operation: a total of about 9,000 construction sites were identified for a potential of about 100,000 new dwellings for Census day. The late block canvass operation covered about 4,000 CUs (10% of all mail-out CUs), i.e. all CUs that had at least one construction site identified during block canvass. For late block canvass, enumerators visited all construction sites and a total of 43,000 dwellings with a known mailing address were added to the Address Register in February 2006. Dwellings under construction that had the potential to be available for occupancy by Census day with no mailing address available at the time of late block canvass were listed for further follow-up during non-response follow-up period. Among the 43,000 new dwellings added during the late block canvass operation, the Census results show that 61% of them were occupied, 30% were unoccupied, and 9% of them were still under construction on Census day. Among the 4,000 CUs revisited during late block canvass, 93 CUs contained a total of 18,000 added dwellings. Thus the new dwellings were mostly clustered, which made their inclusion especially important for regional statistics. Overall, these facts indicate that it was important to conduct late block canvass.

3. COLLECTION AND NON-RESPONSE FOLLOW-UP

Census day was May 16th in 2006. In mail-out areas, respondents received their paper questionnaires on May 2nd, 3rd or 4th. In list/leave areas, questionnaires were delivered to the door from May 2nd to May 15th. Each questionnaire had a unique identifier so that its status could be linked back and updated on the MCS promptly through the steps of collection and processing. Figure 2 shows the cumulative response rate at the national level from May 2nd to June 10th to illustrate the changing pattern. The rate is calculated as the number of returned questionnaires (including those completed through the CHL) divided by the final number of occupied dwellings, i.e. dwellings where a questionnaire is expected. The objective of this graph is to show respondent behaviour in May and early June. Unoccupied dwellings are excluded from the numerator and denominator of this rate. The rate increased slowly after mid-June (not shown) and reached about 80% in August (questionnaires completed through non-response follow-up are not included here). The first dotted line is Census day and the second dotted line is the start date for non-response follow-up. The figure shows that the response rate increased fast in the week before and the week after Census day. After that, the rate of increase slowed down. On May 26, the response rate was about 70% at the Canada level. That means we still had not received completed questionnaires from 30% of all occupied dwellings at that time. The non-response follow-up (NRFU) operation was conducted for these dwellings.

Figure2. Cumulative response rate at national level (NRFU response excluded).



The objectives of the NRFU operation were to determine the occupancy status of the non-response dwellings, to obtain missing questionnaires for occupied dwellings, and to reach an acceptable level of response in each CU.

The NRFU operation was performed in the 2001 Census also. Back then, completed questionnaires were returned back to the enumerator who listed dwellings in that area. The enumerator updated the status of the questionnaires on the dwellings listing book (visitation record). Then the enumerator followed up the dwellings that were listed but had not returned a questionnaire. This method only allows the NRFU to be performed and managed locally.

In the 2006 Census, centralized control was established and all completed questionnaires were returned to a central location for processing. The status of completed questionnaires was updated on the MCS once they were received. Therefore, it was feasible in 2006 to have a central control on the progress of completed questionnaires returned, to create the list of dwellings for follow-up from the MCS, and to send a notification list to enumerators if dwellings on the NRFU list had questionnaires returned. One enumerator could be responsible for multiple CUs. An automated system was also developed for reporting NRFU progress: it calculated the daily counts of complete questionnaires, incomplete questionnaires, refusals, unoccupied dwellings, etc, and calculated daily non-response rates. These reports made it possible to monitor the progress in different regions and identify quality issues.

The NRFU operation started 10 days after Census day – May 26, 2006. It was expected to be finished in late July but was actually finished between Aug 8th to 31st, 2006 depending on the region. This extension was needed to make sure the 2006 Census non-response reached acceptable levels of data quality and coverage. Two main reasons why NRFU lasted longer than expected are that: the response rate at the beginning of the NRFU was lower than expected which led to a larger workload; and there were difficulties in hiring and maintaining sufficient staff during the operation, mostly in areas with competitive labour markets. Figure 2 shows that the response rate increased slower around and after May 26th, which indicated that the NRFU operation started at the right time. Starting the operation when a large number of questionnaires were still being received would not have been efficient.

4. INTERNET RESPONSE MODE

It was the first time in Canadian Census history that the Internet was offered as a response option to almost all dwellings. In the 2006 Census, respondents from all private households except canvasser areas could choose to fill out paper questionnaires and mail them back to Statistics Canada, to call the Census help line operator and fill out questionnaires on the phone, or to fill out questionnaires on-line.

The Internet application was designed with high security in mind. It was even more secure than the on-line banking system for most banks. Respondents needed to type in their Internet access code to log in. The Internet access code was printed on the cover page of the paper questionnaire. The Internet questionnaire was available in both official languages. It was identical to the paper questionnaire in terms of question wording, instructions, and response options. To help respondents answer the questions, an on-line help function was built to provide instructions and examples. Automatic skip patterns were designed. A validation message would appear to warn respondents about non-response, partial non-response, or invalid response for a question.

In the 2006 Census, 63% of respondents completed a paper questionnaire, 18% filled a questionnaire on-line, 2% via CHL, and 17% via NRFU. Although it was the first time that the Internet was offered as a response mode in the Canadian Census, the 18% Internet rate was actually the highest ever achieved for a Census among all countries up to now (Laroche, 2006).

Tables 1 to 3 give some preliminary results on the distribution of characteristics of people who reported by paper or by Internet. Table 1 has the distribution of households by size (number of persons) and by response mode. Table 2 has the distribution of persons by age group and by response mode. Table 3 has the distribution of persons by mother tongue and by response mode. Tables 1 and 2 show that larger households and households with younger persons were more reported on the Internet questionnaires compared to paper questionnaires. For households reported on paper, less than 40% have at least 3 persons, while for households reported on the Internet, 50% have at least 3 persons. For persons reported on paper, 32% are 55 years or older; for persons reported on the Internet, only 14% are 55 years or older; in addition, 15% of all persons reported on paper are less than 15 years old as compared to 21% for those reported on the Internet. Table 3 shows

that for persons reported on paper, 24% have French as their mother tongue while it is only 18% for persons reported on the Internet. Furthermore, 13% of persons reported on paper have a non-official language as mother tongue, while 19% have the same for the Internet. These results are in-line with other Statistics Canada surveys that show a lower Internet usage rate in Québec compared to other large provinces such as Ontario or Alberta. Most people with French as their mother tongue live in the province of Québec.

Table 1. Distribution of households by household size, by form type and by response mode.

Household size	Paper mail back		Internet Response	
	Short form	Long form	Short form	Long form
1 - 2 persons	63.0%	64.1%	49.8%	50.0%
3 - 4 persons	29.5%	29.4%	38.3%	37.1%
5 or more	7.5%	6.7%	11.8%	12.8%
total	100%	100%	100%	100%

Table 2. Distribution of persons by age group and by response mode.

Response mode	Age group						total	
	0 - 14	15 -24	25 -34	35 -44	45 -54	55 -64		65+
Paper mailed back	15.1%	12.1%	10.0%	13.9%	17.0%	14.3%	17.7%	100%
Internet	21.2%	14.7%	16.7%	18.2%	15.4%	8.7%	5.2%	100%

Table 3. Distribution of persons by mother tongue and by response mode.

Response mode	Mother tongue					total
	English only	French only	2 Languages or more*	Non official language	Non response	
Paper mailed back	57.8%	23.8%	3.1%	13.3%	2.1%	100%
Internet	59.8%	17.8%	3.6%	18.8%	0.1%	100%

*Here '2 languages or more' means 2 languages or more, with at least one official language.

5. FAILED EDIT FOLLOW-UP

Returned questionnaires were edited for completeness and coverage checks. A score function (Bornais and Boudreau, 2003) was developed to evaluate the severity of edit failure. If a returned questionnaire had lots of unanswered questions or lots of invalid answers, or if some specific questions were not answered, then the score would be higher than the pre-determined threshold and the questionnaire would fail the edit and be rejected. It would be sent for follow-up with the respondent. This operation was named Failed Edit Follow-Up (FEFU). In 2006, failed cases were first flagged on the MCS, and the list was then sent to Census help line regional offices. Census help line operators performed FEFU through Computer Assisted Telephone Interview (CATI). All completed cases were sent back for processing so that their status could be updated and tracked via the MCS.

Obviously the lower the failed-edit rate, the lower the cost of the FEFU operation. Table 4 illustrates the failed edit rate for short form and long form by response mode. For short forms, the failed edit rate was 5.6% for paper mailed-back questionnaires and 2.5% for Internet questionnaires. For long forms, the failed edit rate was 39.1% for paper but 5.8% for Internet. Thus, the Internet questionnaires had a much lower failed-edit rate than paper questionnaires. This might be due to the use of some features on the Internet such as validation messages, automatic skips, on-line help and explanations of why each question was asked. The fact that the internet system prevented the respondent from making certain types of coverage errors that could be made in paper responses likely contributed, and different behaviour by different respondents may have also had an impact. More studies are needed to know how much of the decrease in the failed edit rate is directly related to the collection mode.

Table 4. Failed edit rate by form type and response mode.

Response mode	Short form	Long form
Paper mailed back	5.6%	39.1%
Internet	2.5%	5.8%

6. SUMMARY

The 2006 Census involved quite a few changes in collection methodology. It increased automation and centralization in many steps and implemented a mail-out strategy and an Internet response mode. It reached a non-response rate similar to the 2001 Census. The Internet response option was a success, and the Internet return rate was the highest for a Census among all countries.

ACKNOWLEDGEMENTS

For assistance in the preparation of this paper, thanks to J. Taylor, D. Laroche, A.M. Côte, and L. Mackenzie.

REFERENCES

- Bornais K., and Boudreau, J.R. (2003), "Conclusions of the Edit and Follow-up Study", internal report. Statistics Canada.
- Dolson D. (2007), "Efficient Multi Mode Data Collection in a Census Context". *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 2965-2972.
- Laroche, D. (2006), "2006 Census Internet Data Collection", internal report. Statistics Canada