

QUALITY MANAGEMENT IN THE 2006 CANADIAN CENSUS OF POPULATION

Paul Poirier, Jean Dumais and Brad Hawkes
Statistics Canada, Ottawa, Ontario, K1A 0T6, Canada

KEY WORDS: Quality Management, Quality Assurance, Census of Population

Introduction

In 1666, Jean Talon arrived in New France with directions to help plan and develop the colony. In order to achieve this goal, Talon conducted what we know to be the very first census in Canada, doing much of the data collection personally by visiting settlers throughout the colony. Not only did Talon's census enumerate the colony's 3,215 inhabitants, but it also recorded their age, sex, marital status, and occupation.

Census-taking in Canada has undergone a few changes during the 350 years since Talon's endeavour. Still, the key information collected every five years by the Census is much the same, and it proves to be essential to the planning and development of the country.

Censuses Past

The most recent significant change to the Census occurred in 1971, over the thirty year period since then Statistics Canada has conducted seven censuses, introducing relatively minor changes to the collection and processing activities of these censuses. Its decentralised collection and data capture methodologies, and linear approach towards handling census questionnaires in batches, passing each batch from one processing activity to the next, was demonstrably very successful given the technology of the time.

The census cycle begins with a series of consultations with the general public, corporations, and government departments at all levels, in order to identify the topics to be covered by the census. Following the consultation process, a series of qualitative and quantitative tests are administered, using mock questionnaires, questions, and response categories. The results of these testing activities are analysed, and recommendations are then made to the Government of Canada on the exact content of the census questionnaires. Once approved, the census questionnaires are sent to be printed.

There are four primary questionnaires used in administering the census. One questionnaire, the short version, is a self-

enumeration questionnaire which is distributed to approximately 80% of the population. The short version of the questionnaire contains questions covering basic demographic characteristics such as age, sex, marital and common-law statuses, the relationship of household members to a reference person, and mother tongue.

There are also three long version questionnaires, which are virtually identical in their content, but distributed to different subpopulations. The long version questionnaires contain all of the questions found on the short version, and also include questions covering citizenship and immigration, knowledge of languages, ethnic origins, education, labour force participation and characteristics, sources of income, as well as questions related to dwelling characteristics. One of the long version questionnaires is administered by personal interview to Canadians living on Indian reserves or in Canada's northernmost communities. The remaining two are self-enumeration questionnaires which are distributed either to Canadian government employees and Armed Forces personnel residing outside of Canada at the time of the census, and to a 20% sample of Canadian households.

In administering the 2001 Census, over 40,000 enumerators were hired, trained, and assigned a geographic area. Enumerators were responsible for all enumeration activities within their area: they visited each dwelling, recorded each address in their control document, and delivered the appropriate questionnaire package. In urban areas, the control document was also compared to an Address Register booklet to help reduce the undercoverage of dwellings. The control document was used to monitor the delivery and return of each questionnaire, for all dwellings within the assigned geographic area. Under this methodology, census field work was controlled at the level of the enumerator assignment.

Respondents completed the questionnaire they received, returned it by mail, and in turn the questionnaire was forwarded to the enumerator who originally dropped it off. Each enumerator checked returned questionnaires for completeness. The enumerator

followed up on incomplete questionnaires, in person or by telephone, so as to obtain missing information and thus complete the questionnaires. When questionnaires delivered to dwellings listed within an enumerator's control document were not received, the enumerator conducted non-response follow-up visits in order to complete the questionnaires by personal interview. As enumerators were following questionnaire collection and completion procedures, their supervisors and quality control technicians were sampling the quality of their work, ensuring that all procedures were strictly observed.

Once the enumerator had satisfied the collection and completion procedures for each of the questionnaires pertaining to their geographic area, the questionnaires were collated into batches. Each batch and the corresponding control document used to monitor the delivery and return of each questionnaire were then forwarded to one of a number of Data Capture Centres.

Once the control documents and their corresponding batches of questionnaires were received and registered at one of the Data Capture Centres, the responses provided within the questionnaires were captured. Within each batch of census questionnaires, individual questionnaires were sampled and re-keyed by a quality control keyer. A third keyer adjudicated discrepancies and entire batches of questionnaires would either pass and move on to the next processing stage, or fail and be subjected to re-keying. In capturing census responses by key entry, a total of slightly more than 3.8 billion key strokes were entered.

Following the capture of census responses, the next stage of census processing is known as coding, whereby written responses to questions asking respondents to describe, for example, their ethnic origins, major field of study or occupation, are converted to numeric codes. Hundreds of carefully trained coding clerks were hired to convert over 40 million written responses to numeric codes. The quality of the work performed by coding clerks was monitored and controlled using an acceptance sampling methodology. The written responses coded by clerks were batched, sampled, and re-coded by a quality control coder. A third coder adjudicated discrepancies and could cause entire batches of written responses either to pass and move on to the next processing stage, or to fail and be subjected to re-coding.

Once the written responses were successfully converted to numeric codes within the coding operation, all responses entered the edit and imputation process, whereby numerous logical edits were employed to identify and resolve problems of validity, consistency, and missing data. Missing data were automatically imputed in cases where an enumerator was unsuccessful in obtaining information or resolving data inconsistencies during follow-up with respondents. The subject-matter specialists monitored and analysed response data as they were processed and transformed where necessary by the edit and imputation process.

Upon completion of the edit and imputation process, the subject-matter specialists would embark on the certification process, whereby they would assess the validity of aggregated data by theme, and at various geographic levels. Data were not only examined within the context of the current census, but were also compared with data from previous censuses as well as with data from external sources.

Finally, once census responses were collected, captured, coded, edited, imputed, and certified, they were ready for public use. Disseminated data were available for a variety of geographic levels, and they were presented in various standard formats and through custom tabulations. Prior to being disseminated, the data contained within aggregated tabulations were verified by dissemination personnel against control counts produced by subject-matter analysts.

Three studies address coverage errors. In the first study, a sample of dwellings listed by census enumerators as being vacant or as dwellings where they encountered non-response is revisited, in order to establish how many dwellings were in fact occupied on Census Day and the number of persons who were living in these dwellings. Estimates of the total number of households and persons missed are obtained in this way, and the census results are adjusted accordingly.

The remaining two studies provide estimates of gross undercoverage and overcoverage, but are not used as a basis for correcting census results. The Reverse Record Check study estimates gross undercoverage by using administrative records to identify a sample of people, collecting all addresses where these people may have been enumerated, and then checking census questionnaires received from these addresses to see if these people were actually enumerated.

The census also includes a study to measure gross overcoverage. The Overcoverage Study contains three components. The first and main component is the Automated Match Study which attempts to match all households in the census database against each other; the detected matches are classified to strata and a sample of matches within each stratum is verified against census questionnaire information to confirm overcoverage. The second component is the Reverse Record Check study, which collects all addresses where a selected person may have been enumerated; each address is verified in order to detect multiple enumeration. The third component is the Collective Dwelling Study, which verifies if a person enumerated in a non-institutional collective dwelling, like a religious community, may have been enumerated at a private dwelling where this person lived previously.

The overcoverage and undercoverage studies are used to produce estimates of net coverage error. These estimates are produced for the main demographic groups and geographic areas. Net coverage error is added to the census counts to determine the population figures that will enter into the calculation of federal-provincial and -territorial transfer payments.

Changes to the Census Methodologies

There are three principal reasons why changes in procedures are being considered for the 2006 Census. Firstly, it is becoming more difficult to collect the information by way of the traditional drop-off and mail-back methods. Though hiring Census enumerators from the neighbourhood makes practical and economical sense, this practice seems to heighten citizens' concerns about privacy. Secondly, at the time of planning, it was unsure whether the government agency which was traditionally hired to capture census data would move towards electronic data receipt for its own data; this possibility meant that our usual data capture operators would no longer be available. Thirdly, the federal government's commitment to make it possible for individuals and businesses use information and communication technology to access government services on-line dictates that changes to the census collection procedures are needed.

Proposed changes to the 2006 Census can be summarized along three dimensions: collection, processing, and testing. With respect to collection, the traditional drop-off methodology will be replaced with a mail-out strategy for approximately two thirds of

households, that is, for those households located in areas for which the quality of the Address Register is deemed sufficient. In the interest of mitigating possible negative effects on coverage caused by the absence of enumerator contact in these areas around Census Day, a comprehensive public communications program will be used. For the remaining one third of dwellings for which the Address Register is not considered reliable enough or where dwellings lack mailable civic number addresses, the traditional drop-off methodology will be used. Mail-back questionnaires will be routed to a centralised Data Processing Centre instead of one of the 3,000 local census offices as was previously the case. In addition to the usual mail-back option, respondents will also be able to complete their census questionnaire via the Internet or over the telephone. To register and control all the possible incoming sources of completed questionnaires, a Master Control System using unique dwelling address identifiers will be used. Automated completeness checks and a centralized computer-assisted telephone follow-up operation to handle incomplete questionnaires are being developed. Thus, questionnaires will no longer be received and processed by enumeration area. These changes are intended to significantly alleviate the current privacy concerns and reduce reliance on a large, decentralised workforce. A high-quality Address Register as well as a secure Internet application are essential to the success of these initiatives.

The processing changes include greater use of automation in the areas of questionnaire receipt and registration, intelligent character recognition, editing, and non-response caseload generation and follow-up. This automation will significantly reduce labour-intensive keying and editing costs, provide for consistency and improve the timeliness of data releases. However, investment in questionnaire design, hardware, software, staff development and accommodations are required.

These methodological changes are accompanied by an environmental change: outsourcing. Statistics Canada has contracted out the development, testing, maintenance and operation of many technical tasks, primarily those requiring highly specialized skills or equipment: questionnaire printing, development of an Internet application, and automated data capture are among those.

Since these are the most significant collection and processing changes since self-enumeration was introduced in 1971, extensive

testing will be required. As a partial response to this need for testing, a full-scale census test is planned for May 2004. In the longer term, it is hoped that increased Address Register coverage and increased Internet response will further reduce the cost of processing and interviewing.

As planning is underway and as most of the issues are being resolved, an appropriate balance of risk and change is being sought. To help the Census Management Team in obtaining a viable compromise, a decision was made to develop and implement a global quality management plan (QMP).

2006 Census Quality Management Plan

The goal of the QMP is to ensure that each step along the way to the final product meets the objectives and adheres to the policies of Statistics Canada.

The *Quality Assurance Framework*, originally drafted in 1997 and last adopted by Statistics Canada in 2002, sets the stage for quality management. All statistical output produced by Statistics Canada must conform to the Framework. As it has been defined in earlier policies, quality is to be assessed along six dimensions: relevance, timeliness, accuracy, interpretability, coherence and accessibility. Brackstone (2000) stated that “without relevance the other five dimensions are unimportant (...) Only when relevance, timeliness and accessibility are satisfied, do accuracy, interpretability and coherence become important.” Thus, within the census, each project, each task, from early consultations with users to the dissemination of the final data, should be within the scope of the QMP.

The significant changes being implemented for 2006 Census collection and processing, combined with the number of operations being outsourced, call for a sustained effort of quality assurance and control. Whereas in previous censuses, projects and tasks had to meet their own objectives of quality and timeliness and to stay within their own dedicated budgets, the new management philosophy tends to integrate operations, tasks, resources and budgets, steering emphasis away from the “parts” and towards the “whole”. Savings in one area can be redirected more expediently towards areas in need, that is, those where the needs are greater or where resources appear insufficient to attain mutually shared goals.

The purpose of the QMP is to support the Census Management Team in ensuring that the quality of the 2006 Census is at least that of

the 2001 Census. The Census Management Team is accomplishing this by demonstrating leadership, by monitoring the quality of the various processes and by communicating in a timely and relevant manner with the Census Project Management Group.

The 2006 Census comprises a great number of projects and tasks. The Quality Management Team continually meets with each project and task manager in order to agree upon the quality levels to be attained in the 2006 Census, and to discuss the means by which the quality will be monitored and appraised (batch quality control, process control, etc.). In many instances, entire tasks can be seen as quality control of an earlier task. For example, the certification process and content studies, which happen in the two years following Census Day, can be seen as an assessment of integrated consultations, which happen three to four years prior to Census Day. A more specific example is found in the set of cognitive and volume testing that is needed to appraise the questionnaire design. Both examples address the aspects of relevance, interpretability and coherence of the census.

Discussions on quality are ongoing, and the results are being integrated within the 2006 Census QMP. In essence, the QMP is a work-in-progress, and will evolve with the advancement of the census program in order to remain both relevant and useful. At this time, the QMP lists the tasks and projects comprising the 2006 Census, the various activities undertaken under the umbrellas of quality assurance and quality control within each task and project, and the pre-defined levels of outgoing error expected from each task or project, where appropriate. A broad array of carefully designed production reports will help managers monitor and assess the progress of their projects and tasks. Ultimately, the information gathered on an ongoing basis by the QMP should help census management adjust resources allocated across projects and determine the ensuing impact on quality.

Lessons Learned

Although it is still very early in the process of implementing a QMP for the 2006 Census, we have already come across many challenges.

One of our most significant realizations has been that the task of gathering the knowledge required to properly implement such a plan is far from trivial. In the past, we have compartmentalized the various steps required to conduct the census into somewhat independent

components. In general terms, we collected the data first; then, we processed the data; and finally, we disseminated the data. Because of how the census was structured in the past, the vast majority of census employees have very detailed knowledge of a few of the components and only some varying degree of knowledge of other components (in total, the complete census cycle had approximately 15 to 20 major components). Since it was not as critical in the past, very few employees were required to have a detailed understanding of most components. In the vast majority of cases, quality was controlled the same way, that is, within individual components and not very often across components. In the 2006 Census, most of the collection and processing steps will be more integrated than ever before, and these processes will therefore be more interdependent.

Moving towards this more integrated environment, employees with an overall understanding of the full range of census processes are needed to assume crosscutting roles such as those required in the building of a QMP. Assembling a group of employees with overall census-taking knowledge has proven to be a challenge. This challenge requires timely communication, to keep track of all quality-related decisions that are being made as the redesign of the census takes place and to share the specialized knowledge involved in each of these components.

In the same manner, convincing specialized employees that the overall quality of the census is more important than the quality of the individual component they are working on is also challenging. Understandably, many find it hard to let go of the old way of doing things, and furthermore, they do not necessarily have the opportunity to gain enough overall knowledge to see that the old methodology used for their assigned component may no longer be appropriate in the new redesign. Without an exhaustive, holistic QMP in this newly redesigned environment, there would be a real risk of missing an important aspect of the new methodology that may have a significantly adverse effect on the overall quality.

Finally, other changes in the way we manage our operations were introduced when we entered into a partnership with a contractor for a major portion of our work, for the first time. The collaborative planning has been coordinated through Integrated Project Teams (IPTs). Although a quality management IPT has been created, the detailed work of establishing quality

targets and developing appropriate quality assurance procedures has largely been completed by those IPTs responsible for specific components of the overall system, such as data capture and coding. However, strong and proactive quality management leadership is beneficial in ensuring that coherent trade-offs are made.

Challenges that Lie Ahead

In the former environment, most managers have used the "let's-do-the-best-we-can-with-the-time-and-money-we-have" approach to managing quality. Since previous censuses had more independent components, this approach was not unreasonable. However, in the more integrated approach to census collection and processing that is envisaged for 2006, most managers will need to agree on targets, since the quality of the component they are responsible for may affect the quality of many other components. These targets will be monitored and sometimes measured by the Quality Management Team, who will be asked to oversee the progress of the overall quality of the census program.

In the same manner, agreeing on a clear definition of an error is also challenging. Some managers will prefer definitions that are simple to measure operationally (e.g., two coders did not code the information the same way) while others will prefer more theoretically pure and sometimes subjective definitions (e.g., we did not capture the data that the respondent intended to communicate).

Another challenge is to identify the key processes that have the greatest potential to affect overall quality. Once determined, these key processes should be controlled and monitored more closely than the other processes.

Finally, establishing a team that will be able to advise Census management on quality issues in a timely manner will be very important. To do so, this team will need to fully understand all of the census components in sufficient detail, in order to have a good understanding of the real cost and timing of the quality improvement, quality assurance, quality control and quality measurement processes. A solid knowledge of these processes and their respective costs is required in order to provide effective recommendations on the following decisions:

- 1) In a financially stable context, which areas should be modified so as to improve overall quality;

- 2) In a context of budget restrictions, which areas could take a reduction in budget with minimal impact on overall quality;
- 3) In a context of time constraints, which areas could sustain a decrease in quality targets with minimal impact on overall quality and while cutting significantly on processing time.

Conclusion

The collection methodology for the 2006 Census relies on the use of proven technology for repetitive tasks. It also allows for greater integration with processing functions, provides respondents with a choice of response media, resolves current privacy and confidentiality concerns, and offers the prospect of long-term cost control. The new methodology is not without risks, but they are manageable. In order to minimise risks, a large scale census test is planned for May 2004 in which an overall QMP will be implemented. The objective is to attain or surpass the level of quality achieved in previous censuses.

The QMP will first be implemented in the May 2004 census test, therefore providing an opportunity to evaluate the effectiveness of the QMP and to make improvements in advance of the 2006 Census.

References

- Brackstone, G.J. (2000). "Managing Data Quality at Statistics Canada." Statistical Quality Seminar, Cheju, Korea, December 2000.
- Poirier, P. (2003). "The Future of Canada's Census." *Liaison* (Statistical Society of Canada), 17,1 (February): 34-37.
- Statistics Canada (1998). *Statistics Canada Quality Guidelines* (Statistics Canada Catalogue No. 12-539-XIE). Ottawa.
- Statistics Canada (2001). Advisory Committee on Statistical Methods, Meeting No. 32, April 23-24, 2001.
- Statistics Canada (2002). *2001 Census Handbook* (Statistics Canada Catalogue No. 92-379-XPB). Ottawa.
- Statistics Canada (2002). Advisory Committee on Statistical Methods, Meeting No. 34, April 29-30, 2002.
- Statistics Canada (2002). *Quality Assurance Framework* (Statistics Canada Catalogue No. 12-586-XIE). Ottawa.