

## Article

Symposium 2008:  
Data Collection: Challenges, Achievements and New Directions

### Testing for the 2011 Census of Canada

by Marie-Noëlle Parent and Heather Lathe

2009



Statistics  
Canada

Statistique  
Canada

Canada

## Testing for the 2011 Census of Canada

Marie-Noëlle Parent and Heather Lathe<sup>1</sup>

### Abstract

The next census will be conducted in May 2011. Being a major survey, it presents a formidable challenge for Statistics Canada and requires a great deal of time and resources. Careful planning has been done to ensure that all deadlines are met. A number of steps have been planned in the questionnaire testing process. These tests apply to both census content and the proposed communications strategy. This paper presents an overview of the strategy, with a focus on combining qualitative studies with the 2008 quantitative study so that the results can be analyzed and the proposals properly evaluated.

Key Words: Questionnaire, Census, Qualitative studies, Quantitative study, Experimental plan, Analysis.

### 1. Introduction

#### 1.1 Overview of the Census

The Census applies to every man, woman and child living in Canada on Census Day as well as to Canadian citizens who are temporarily out of the country. It is conducted every five years and is one of the few compulsory social surveys administered by Statistics Canada. Two main questionnaires are used in the Census: a short questionnaire (the 2A, sent to 80% of households) and a long questionnaire (the 2B, sent to 20% of households).

In Canada, the Census is the only reliable source of detailed data on small groups, and it also provides access to data on areas as small as neighbourhoods and as large as the entire country. The next Census of Population will take place on May 10, 2011.

As in other surveys that are repeated over time, the trade-off between comparability and relevance arises when determining the content for each census. Indeed, the need for chronological data suggests that the questions asked in a national census should always be the same. However, revisions sometimes have to be made because of new statistical needs or changes in society. It is necessary to take such changes into account in order to develop, sustain and monitor laws and programs for the effective management of the country's affairs. For this reason, the Census has undergone various modifications over the years, particularly in its terminology and definitions.

#### 1.2 Steps in questionnaire development

Few changes were made to the questionnaires for the last census in 2006. Indeed, the operational side was deliberately given priority because a number of changes were expected in that area. As a result, the resources available for qualitative studies and other studies on census content were limited in 2006. Consequently, for 2011, the focus has been on the decision-making process regarding questions that will be added, removed or revised.

The main goals of the content determination team were to consult key users of census data, conduct studies, carry out a detailed analysis of potential content changes and ensure that the various client divisions involved in the process use harmonized concepts.

---

<sup>1</sup> Marie-Noëlle Parent, Statistics Canada, Ottawa, Canada, K1A 0T6 (marie-noelle.parent@statcan.gc.ca); Heather Lathe, Statistics Canada, Ottawa, Canada, K1A 0T6 (heather.lathe@statcan.gc.ca).

Various steps toward content determination were planned. First, consultations on the 2011 Census started in Spring 2007. Next, three major phases of qualitative studies on questionnaire content were carried out in Fall 2007 and Spring 2008, followed by a quantitative study in May 2008. At the same time, three phases of qualitative studies on the proposed communications strategy were carried out. A number of recommendations came out of those studies, some of which were implemented between two successive study phases. Other proposed changes will be tested in the May 2009 Census Test. Lastly, the content of the census questionnaires will be submitted for final approval in Fall 2009 in preparation for the 2011 Census. The schedule is not as tight for the communications strategy, and there may be an additional qualitative study phase in Winter 2009.

### **1.3 Consultation process**

Before each census, Statistics Canada asks data users and other interested parties across Canada for their views on the nature and quantity of information they would like to have in the census database. The purpose of these consultations is to ensure that Statistics Canada considers users' emerging needs and new social or economic questions.

Three major strategies were used in the consultation process to obtain comments from the various parties concerning the 2011 Census questionnaires. First, face-to-face meetings were held, mostly with data users such as not-for-profit organizations, federal government departments and university researchers. Second, e-mails were sent out to announce the publication of the consultation guide and invite the 900 known census data users to submit their comments by mail or via the Internet. In fact, all Canadians had the opportunity to take part in this process and send in suggestions.

The consultation process began in April 2007, and most comments had been received by November 2007. Since consultation is an on-going process, a number of other suggestions were sent in later on. Some of these were considered for the 2011 Census. Due to the tight schedule, others will have to wait until future censuses.

### **1.4 QDRC involvement and methods used**

Once the consultation process was well under way, qualitative studies were planned to test some of the recommendations. The Questionnaire Design Resource Centre (QDRC) was asked to assist with these studies. Indeed, one of the QDRC's mandates is to play an active role in reviewing the questionnaires used in the various surveys administered by Statistics Canada. Accordingly, the QDRC was heavily involved in all the qualitative studies carried out for the Census, regarding both the communications strategy and questionnaire testing. For the communications strategy, the aim was to test various introductory letters to be mailed out and the way in which questionnaire completion options would be presented to respondents. Next, a number of tests of the 2A and 2B questionnaires were conducted. The studies were conducted in several urban centres across Canada, using two principal methods: cognitive interviews and focus groups. While most of the meetings were held in locations where a number of external clients could observe, some of the interviews took place in people's homes.

In the tests, QDRC staff attempted to comprehend and analyze the participants' cognitive response process (Tourangeau, 1984). To that end, their understanding of the instructions, concepts and questions was probed in detail to determine the causes of response errors.

## **2. Communications strategy**

### **2.1 Importance**

Communication is a vital component of the census process. For example, it is critical that the very first letter sent to the majority of Canadian homes be well received and, above all, well understood. The communications strategy therefore focuses on the way in which the Census will be presented to the public and not on the questionnaires as such. The first step is to evaluate the envelope and letter to be sent out to announce the Census and explain how to

participate. Next, the overall strategy has to be analyzed: the reminder letters to be mailed out, the telephone messages or other means that might be used if required, and the proposed schedule. As mentioned, one of the major challenges is to clearly demonstrate the importance of the Census to Canadians. It is also necessary to ensure that a good rapport is established with respondents to encourage them to cooperate, without overemphasizing the fact that the Census is compulsory as this could discourage some participants. Obviously it is important to ensure that the message is clear and to devise a suitable follow-up procedure. A number of contact strategies have been evaluated against these criteria.

## **2.2 Issues explored**

There was a major innovation in the collection method used in the 2006 Census. For the first time, Canadians had the option of completing their census questionnaire on-line. Even so, most households were sent a paper questionnaire in the mail, which was standard procedure in previous censuses. The new option was a great success in 2006 (more than 18% of responses were submitted via the Internet). This result, supported by a convincing test of about 20,000 households that received only a letter in 2006 (Laroche, 2007), led to a reconsideration of the strategy used. Indeed, for 2011, the plan is to promote the Internet response method and mail only a letter (not the paper questionnaire) to about 60% of households. In most cases, these households will be in large urban centres for which the Internet response rate was high in 2006. The Internet collection method is preferred due to the speed and quality of the responses and because data capture is part of the response process. It is also preferable from an environmental perspective because far fewer questionnaires will have to be printed.

The initial letter would tell people to complete the census questionnaire on-line (with a unique secure access code) or to call (an automated system) for a paper questionnaire if preferred. During the qualitative studies, the reaction to this suggestion was mixed. Many computer-literate people viewed it as a logical step, as well as a way to save paper. For others, it was perceived as an additional burden. Indeed, many said they would not or could not use the Internet for various reasons and considered it inconvenient to have to telephone and then wait a few days to receive and complete the paper questionnaire. Not surprisingly, reactions varied markedly according to the age of the people interviewed. Older people who were accustomed to receiving the paper questionnaire and were less familiar with computers in general were particularly resistant to the suggested approach. It may be difficult to grasp the considerable, large-scale benefits of this new approach. It is therefore expected that more extensive follow-up will be required for non-respondents, which makes it important to mail out reminder letters fairly promptly when required.

A number of other issues were raised in the studies. For example, there were a number of comments on the value of specifying a deadline for completing the census questionnaire. While some people like to have a date for reference, others admit that it might diminish the urgency of the Census and that they might take longer to complete it. There were also a number of discussions on the fact that the Census is compulsory. It is difficult to say to what extent the legal aspect should be addressed in the various letters. Specifically, the challenge is to encourage people to participate willingly in the Census without putting too much emphasis on the possibility of legal consequences if they refuse. The results of the studies suggest that a gradual approach would be best. The initial letter intended for all respondents should strike a balance between describing the benefits of the Census and informing people of the obligation to respond, and subsequent letters and messages should convey an increasingly explicit message about the compulsory nature of the Census. In addition, there were questions about the number of reminder letters that should be sent, the interval between the letters, and even the colour of the envelope to be used.

Following all these studies, which took place between July 2007 and December 2008, internal reports were prepared to help the 2011 Census communications team make informed decisions. Two somewhat different approaches will be used in the May 2009 Census Test, which may be followed by another qualitative study phase, to provide the communications team with quantitative results and clarify the requirements for finalizing the strategy for 2011.

### **3. Qualitative studies for the census questionnaires**

#### **3.1 Background**

With regard to typical questionnaire testing, more than 400 people from the general population took part in the QDRC's qualitative studies over the past year. To test specific questions and take changes in Canadian society into consideration, some groups were asked for their opinions more often than others. Specifically, the consultations focused in particular on Aboriginal people, members of large blended families, recent immigrants, people living in same-sex relationships, people of various faiths, and people with various levels of education.

#### **3.2 Instructions**

The census questionnaire is normally completed by one member of the household without assistance from an interviewer. Therefore, the readability, clarity and comprehensibility of the instructions in the questionnaire are crucial. In particular, the major challenge is to ensure that respondents understand in what order the members of their household are to be listed (there is a specific order to follow) and, most importantly, who is to be included in the list. Due to the nature of the census, it is vital to provide very precise instructions to ensure that every person who should be counted is counted and that no one is enumerated in two different households (for example, a child in joint custody).

The instructions concerning the order in which household members should be listed were reviewed in the most recent qualitative studies in an attempt to improve the quality of census data. The wording was altered slightly and additional notes were provided, but the changes do not seem to have had the desired effect and were not conclusive. The additional notes were simply overlooked in many cases.

The detailed instructions on making the list, and a space for the very few households that are not required to complete the questionnaire for various reasons, appear on a separate page, the location of which changed over the course of the qualitative studies (it was initially at the very beginning and was later placed after the list). The impact of placing it after the list was notable in that a number of respondents ignored it completely. It is worth noting that for the majority of households, determining who should be included in the list is very easy. In those cases, reading the instructions fully is less important, though that may seem counterintuitive. On the other hand, it is necessary to ensure that people with ambiguous or difficult cases consult the instructions as required. In the most recent qualitative studies, it seemed to work even though a number of people whose situation was clear overlooked the instructions page. In addition, the proportion of people who erroneously filled in the space reserved for households that did not have to complete the questionnaire seems to have declined substantially.

Finally, it must be noted that the detailed instructions page contains a great deal of text, which may be a burden for people. The visual layout of the page was altered somewhat during the studies, and it is always a challenge to make information as concise and simple as possible.

#### **3.3 Marital status**

In the 2006 Census, one of the suggested response categories for marital status was "Never legally married (single)". In the qualitative studies, it was noted that that question seemed slightly offensive to people. Increasing numbers of Canadians are living in common-law relationships. Most of them have been living common-law for years and have children; consequently, they do not see themselves as single. While the problem did not generate any response errors (people still selected the appropriate response), it seemed important to try to correct it. The suggested solution was simply to remove the word "single" from the response category. The new wording appeared to win greater acceptance in subsequent qualitative studies.

### **3.4 Relationship to person 1**

It is difficult to collect information about relationships between household members in a self-administered questionnaire. The format of the question is slightly different, and there is a long list of response categories from which respondents can choose. In 2006, there were categories for spouses (husband or wife of Person 1) and children (son or daughter of Person 1) living in the household. There was also an “Other” category, of which the frequency of use was analyzed. It was found that 30% of the people who used the “Other” category wrote in that the child was their spouse’s son or daughter (and not their own). That fact is not surprising given there are growing numbers of blended families in Canada. It was pointed out that the members of such families were always a little hesitant about this question concerning the children in the household because there was no category that reflected their situation. To remedy this problem, three new response categories were created to replace the original: “Son or daughter of Persons 1 and 2”, “Son or daughter of Person 1 only” and “Son or daughter of Person 2 only”. These new categories seemed to be very widely accepted and properly used by participants. A number of people liked being able to clearly indicate each child’s relationship to the adults in the household, without the burden of looking for the relationship or writing it in as they had to do in the past.

A new category was also added for spouses in order to reflect the current situation in Canadian society. Marriage between persons of the same sex was legalized in Canada in 2005, but the change was not echoed in the 2006 Census. The two response choices are now “Opposite-sex spouse of Person 1” and “Same-sex spouse of Person 1”. The change was greatly appreciated by the people concerned. In addition, the somewhat sensitive nature of the new category did not appear to present any problems for other Canadians who were consulted.

### **3.5 Aboriginal identity**

There have been a number of consultations about the Aboriginal identity question in recent years and in previous censuses. It is very difficult to find a wording that will be properly understood by and relevant to the various Aboriginal groups across Canada. It is especially important to avoid false positives (people who might incorrectly identify themselves as Aboriginal if the question is not clear enough) and false negatives (people who really are Aboriginal but do not see themselves in the definition provided).

In response to suggestions from users of the data on Aboriginals, as well as to reflect many recent qualitative studies, the wording of the question and the terms used in it have been changed. For example, one of the proposed changes is to use “Inuk (Inuit)” instead of “Inuit (Eskimo)”. Thus, current terminology will be used as it seems to be better understood by the three main groups concerned: North American Indians (First Nations), Inuit, and Métis. Recent analyses suggest that all persons will be in a better position to identify themselves with the new wording.

### **3.6 Education**

While it may seem straightforward at first glance, it is not easy to formulate the questions for the education section. Specifically, this is because education systems vary across Canada. CEGEPs, for example, are unique to the province of Quebec. In addition, education systems have changed over the years, which may make comparisons difficult depending on the person’s age. The education section is also very difficult for immigrants who have not been in Canada for long and received most of their education in another country. For them, answering questions that are geared to the Canadian education system is very complicated. The section continues to be the subject of qualitative studies and analyses aimed at determining what the needs are and how to meet them effectively.

## **4. Quantitative study: 2008 content test**

### **4.1 Experimental plan**

As described in this paper, qualitative studies help to identify potential problem questions and formulate hypotheses while suggesting possible solutions. The hypotheses can then be tested by quantitative studies based on an experimental plan. An experimental plan makes it possible to compare and quantify changes from one questionnaire to another. The results are analyzed, and a decision can then be made about how to formulate the final questions.

For the 2011 Census, the study used was the May 2008 content test. Two different versions of the questionnaires were used to assess the impact on data quality. One set (the control version) consisted of the 2006 questionnaires (2A and 2B), while the other (the alternative version) was composed of revised questionnaires (2A and 2B) that incorporated the changes proposed following the qualitative studies.

### **4.2 Target population and sample**

For this study, six independent, representative panels were selected in the 10 provinces. There were four panels of 4,750 dwellings each in the general population, of which two received the control version (one the 2A and the other the 2B) and two received the alternative version. The other two panels were each made up of 3,500 dwellings, in which at least one person had been identified as an Aboriginal in 2006. Their main purpose was to provide data for a more effective analysis of the changes in the Aboriginal identity questions. For these last two panels, only off-reserve dwellings were considered. One received the control 2B questionnaire while the other received the alternative 2B questionnaire.

Unlike the Census, the 2008 content test was not compulsory. Its response rate was 46% for the 2B questionnaires and 52% for the 2A questionnaires. Consequently, the analysis was based on data from about 12,000 respondent households (28,000 individuals) out of the 26,000 households sampled.

### **4.3 Analyses and results**

This section deals with some topics that have already been covered in this paper, such as marital status and relationship to Person 1. It also deals with some new variables: sex and date of birth.

First, the response categories for the sex variable (male or female) were listed vertically in 2006 (one below the other). No change was planned for 2011, but because of a space shortage due to changes in other questions, the categories were arranged horizontally (one beside the other) in the studies. No problems were noted in the qualitative studies; people saw the choices and answered the question. There was, however, a slight increase in non-response in the quantitative study. The increase is probably due to the fact that some people simply did not see the question because it now takes up less space in the questionnaire.

Date of birth is one of the key pieces of information collected by the Census. Because it is broken into three fields (day, month and year), it is subject to reporting errors by the respondent and capture errors due to illegible handwriting. Moreover, in the 2006 Census, there were some problems with automated data capture. The distribution of errors is less uniform for automated data capture than for manual data capture. The difference appears to have had a negative impact on certain small population groups based on age. To mitigate the problems and improve data quality for date of birth and age in general, it was suggested that a field be added for the person's age as well as his or her date of birth. The additional information about age would be used mostly to validate the date of birth and could be considered for imputation purposes. The approach is already in use in other countries. On the other hand, the addition of the age variable is one more burden for the respondent. In the qualitative studies, most people responded to both variables (date of birth and age) without comment. Some people were later asked explicitly about the value of including both questions, and very few commented that it was indeed redundant and that there was no need to ask for both. In the quantitative study, there was a very slight increase in non-response to the date-of-birth question in the alternative version of the questionnaire (the one with the new question on age), but

96% of respondents answered both questions. This indicates that data quality could probably be improved, without increasing the response burden excessively, by adding the age question.

For marital status, the two versions differed only in whether they used the term “single”, as discussed previously. The distribution of responses was very similar between the two versions. The analysis suggests that it would be appropriate to remove the term “single”, since it would have no impact on data quality. It would also be more appropriate in view of the comments received from respondents in the qualitative studies.

For the relationship to Person 1, the new categories proposed in Section 3.4 were added to the alternative version. For example, the number of sons and daughters was very similar between the two versions, and the distribution of the new response categories seemed reasonable. It appears that the new categories would be relevant and well understood and that they would not generate an additional response burden – far from it. The new categories would also provide census data on blended families, in contrast to the 2006 question. Lastly, the change would significantly decrease the number of people who have to use the “Other” category and take the time to write in the relationship. The analysis suggests that it would be positive and appropriate to go ahead with this proposal.

A number of other questions were revised, and analyses are continuing so that informed decisions can be made about the content of the next census questionnaires.

## **5. Conclusion**

### **5.1 Next steps for the census**

As mentioned in Section 2.2, two different communications strategies will be tested and compared in the May 2009 Census Test. For the moment, the plan is still to send only a letter to 60% of households, particularly those in areas with a high Internet response rate in 2006. This letter would be followed by a first reminder letter (about 10 days later). Next, there would probably be a recorded telephone message reminding people to complete their questionnaire. Finally, a second reminder letter and a paper questionnaire would be mailed to everyone who had not yet responded. The legal message would be present but discreet on the envelopes and in the letters, with some emphasis in the reminder letters. There will be no deadline date as such, but people will be encouraged to complete their questionnaire within 10 days of receiving the initial letter. In fact, one of the differences between the two approaches in the next test will be the use of different deadlines so that the impact on the time taken to complete the questionnaire can be analyzed. Depending on the results of the test, the overall strategy might be adjusted slightly once again.

At the same time, analyses of the quantitative study will be carried out for each question in which a change has been proposed, with a view to determining the final content of the questionnaires used in the May 2009 Census Test. Further qualitative studies on specific questions are also likely to be conducted over the next few months. In addition, although the purpose of the May 2009 test is mainly operational, the data collected will be analyzed, and changes could be made if there are obvious problems. All that will culminate in the final submission of the questions planned for the fall of 2009.

All the proposed changes that have not been tested sufficiently (for lack of time or resources, or because they were of lower priority) to warrant a change in the 2011 Census will nevertheless be noted and reassessed for future censuses.

### **5.2 Combining the two types of studies**

As explained at the beginning of this paper, there are constraints on the Census because of the need to maintain comparability over time. There are intensive consultations on any changes made between censuses. The public, governments, the private sector, the academic community and a number of other parties are consulted to obtain a clearer picture of census data needs. The purpose of any changes is to provide higher-quality data that are relevant to Canada’s current economic and social situations.



Proposed revisions to census content and the decisions made on the proposals are also based on a number of qualitative studies, which in turn are supported by quantitative studies wherever possible. It seems much more practical and efficient to use both types of studies to accurately gauge the effects on data quality. No study would be able to identify every potential problem, but combining the two in this context adds a dimension that is essential to the decision-making process.

Unfortunately, the resources and time needed to carry out both types of studies are not available for most other Statistics Canada surveys. In many cases, they are much smaller surveys and have cycles measured only in months. However, it would certainly be useful to at least consider the possibility for surveys with samples of a significant size and cycles covering several years.

## References

- Laroche, D. (2007), 2006 Census Internet Data Collection. Technical report presented at Statistics Canada's Advisory Committee on Statistical Methods, April 30-May 1, 2007.
- Tourangeau, R. (1984). Cognitive Sciences and Survey Methods. *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*, (Eds. T. Jabine, E. Loftus, M. Straf, J. Tanur, and R. Tourangeau), Washington, DC: National Academy of Science.