

Profiles of Responses Over the 2006 Canadian Census Collection Period: What are the Differences Between Early and Late Responses?

Laurent Roy¹, Catherine Corriveau¹, Laura McFarlane¹
¹Statistics Canada, 170 Tunney's Pasture Driveway, JT-1, Ottawa, ON,
Canada, K1A 0T6

Abstract

The data collection of the 2006 Canadian Census of Population took place between May and August 2006. An analysis was conducted to examine characteristics of responses by period of collection. The objectives were to compare profiles of late responses (mainly from follow-ups) to those of responses from the first month of collection (self-reported) and to look at the impact of late responses on specific groups of the population. Differences between late and early responses were observed mostly for age, immigration status, visible minority groups and Aboriginal people. Overall, approximately 3% of responses were received in the last month of collection but for some specific groups, over 7% of responses were received in the last month alone. Findings from the analysis will be used to study possible changes to the imputation strategy for the 2011 Census of Population.

Key Words: Census, non-response, collection, imputation

1. Overview of the 2006 Canadian Census of Population

The Canadian Census of Population is conducted every five years. The last one was held in May 2006. Approximately 13 million dwellings and a population of more than 31 million were enumerated in the 2006 Census. Starting in early May of 2006, the self-administered questionnaires were mailed to 70% of the dwellings (mail-out collection methodology) and 29% were dropped off by census enumerators (list-leave collection methodology). Canvasser methodology, where an enumerator completes the questionnaire while interviewing a household member, was used in some parts of the country (about 1% of all dwellings), mostly in northern communities and Indian Reserves.

Two main types of questionnaires were used to collect the majority of the census data. The short form, with eight questions, was distributed to four out of every five households (80%). The long form, with 53 person-level questions (including the eight questions from the short form) and eight dwelling questions, was distributed to the remaining 20% of households. For the canvasser methodology, only the long form was used. The short form contained questions covering demographic information for each member of the household. The long form contained the demographic questions from the short form, as well as questions on Aboriginal identity, education, occupation, language knowledge, income, etc. The following is a link to the Canadian 2006 Census of Population long form: http://www.statcan.gc.ca/imdb-bmdi/instrument/3901_Q2_V3-eng.pdf.

Questionnaires were delivered starting May 1; Census day was May 16, 2006. Respondents in mail-out and list-leave areas returned their questionnaires by mail or had the choice of completing the questionnaire on-line using the Internet or by phone through the Census Help Line (CHL). Non-response follow-up (NRFU) targeted dwellings where no questionnaire had been returned. Interviewers contacted residents at those dwellings to collect their responses over the phone or in person. NRFU started in late May and finished at various dates in August, depending on the region.

Whatever the method of return, the responses were directed to a central data processing center (DPC). The registration of each returned questionnaire was flagged on the Master Control System (MCS) at Statistics Canada. All questionnaires were registered centrally, and information about the date the questionnaires were registered was kept in the MCS. Prior to 2006, census questionnaires were collected by enumerators and held in local offices before being shipped en masse to the DPC. For the 2006 Census, the changes made it possible to assign a return date and examine the characteristics of the data by the date it was received.

2. Study of Characteristics of Responses by Period of Collection

2.1 Introduction

The analysis was conducted to examine characteristics of persons reported on the census questionnaire by period of collection. The objectives were to compare profiles of late responses to those of responses from the first month of collection and to look at the impact of late responses on specific groups of the population. Most questionnaires completed in the first month of collection were by self-enumeration (paper, Internet, CHL). NRFU started in late May 2006. About 65% of questionnaires received from the month of June onwards were received through the NRFU operation; the remaining questionnaires received during those months were mailed back, completed through the Internet or the CHL. This study did not distinguish between responses obtained by NRFU and those reported by respondents themselves.

Prior to the analysis, differences in characteristics of persons who reported early and later in collection were expected, but one objective of the study was to assess this assumption and quantify those differences.

2.2 Methods

Answers to most questions from the short and long forms were examined for this analysis. Distributions were examined by month of response. The exact date that respondents completed the questionnaires is unknown. We approximated this date using the date that the questionnaire was registered at the DPC. For paper questionnaires mailed back to the DPC, the registration date is an approximation of the completion date. The date the respondent actually finished answering the questionnaire could be several days or even a week prior to the registration date. This type of delay is due to the time a respondent might take to mail back the form or the time it takes for the form to be picked up, processed and sent to the DPC by the Canadian postal service.

For questionnaires completed through the Internet, the registration date is the date the respondent submitted the questionnaire and, in most cases, this date corresponds to when the respondent completed the questionnaire on-line, i.e. in most cases, the respondent completed and submitted the Internet questionnaire within one session. For

questionnaires completed through CHL, the registration date is the date the respondent called the CHL. For questionnaires completed through the NRFU, the completion date is the date the enumerator entered the status of the dwelling (questionnaire completed, for example) in the MCS, which would be very close to the actual date the enumerator completed the questionnaire at the door or over the phone.

To investigate whether early responses were similar to late responses, this study eliminated all questionnaires completed through the canvasser methodology. Those questionnaires were mostly from dwellings in northern or remote communities, Indian Reserve dwellings, or collective dwellings (for example, hotels, correctional institutions, nursing homes). The registration date refers only to the date questionnaires were received from the enumerators and in most cases the enumerator would have shipped all completed questionnaires at once. Comparing characteristics of persons reported by month of collection would not be relevant, as the date is more related to the logistics of the collection methodology than to persons reported. With those exclusions, we excluded about 7% of the total Canadian population. The population in scope for this study, then, is the population from private dwellings enumerated through the mail-out or list-leave collection methodology.

For the purpose of this study, we focussed on Canada-level data. We also examined the data at a lower level of geography for large agglomerations such as Montreal, Toronto and Vancouver. The analysis found that the trends in most cases were similar to those at the Canada level.

2.3 Results

The analysis examined the monthly distributions of most variables on the short and long forms. This section presents the variables with the most significant differences between early and late responses.

2.3.1 Monthly Distributions for Selected Variables

Table 1 shows the distribution of persons in scope for this study by the month of collection. Data for approximately 76% of the total was received in the month of May, while data for roughly three percent was received in August. For this study, it is important to note that we are referring to ‘persons reported’ as opposed to ‘respondents’. For the census questionnaire, usually one person with knowledge will answer for all members in the household, especially when all persons are related. When analysing the census data, it is not possible to identify the respondent among all members of the household, i.e. the person who completed the questionnaire.

Table 1: Distribution of Responses by Month of Collection

	May	June	July	August	Total
Number of persons (in ‘000s)	22,385	4,031	2,271	748	29,435
% of total	76.1%	13.7%	7.7%	2.5%	100%

When comparing the variable distributions, the study looked at the distribution of characteristics for each month, using the number of responses received each month as the total. All tables and figures in this section show distributions of monthly responses expressed as percentages, calculated using the total number of persons reported in a given month as the denominator (see Table 1). For any month, the percentages represent the number of persons with the characteristic of interest (responses) divided by the total number of persons reported in the same month. The monthly total is always 100%.

Figure 1 shows the distribution of monthly responses by sex. The data shows that more females were enumerated in May than males, but for each of the following three months of collection, males increased as a percentage of monthly responses while females decreased. Figure 1 demonstrates a reversal of male/female distribution between May and August. Over the full collection period, males made up 49 % of the total and females made up 51%.

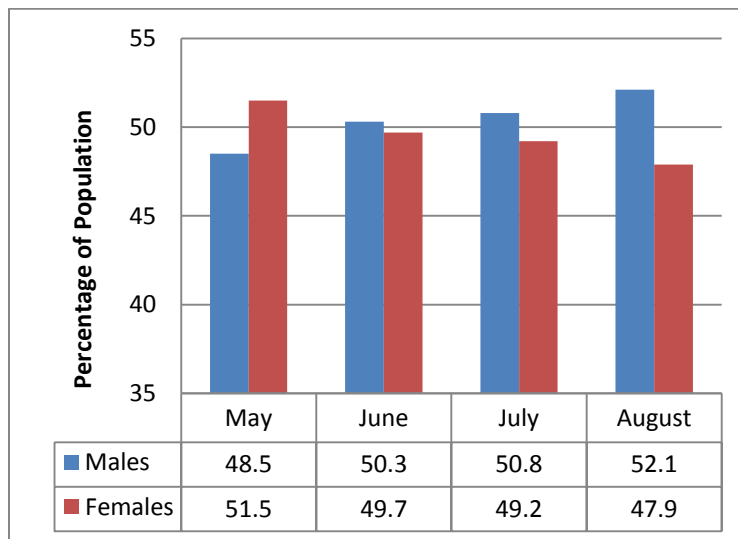


Figure 1: Distribution of Monthly Responses (Population) by Sex

Figure 2 illustrates the distribution of monthly responses by marital status. In the month of May, single and married persons made up 42.5% and 43.9% of the total respectively. In the latter months of collection, single persons predominated, making up 61.1% of August responses. The distribution over the four-month collection period was 46.1% single and 40.3% married persons.

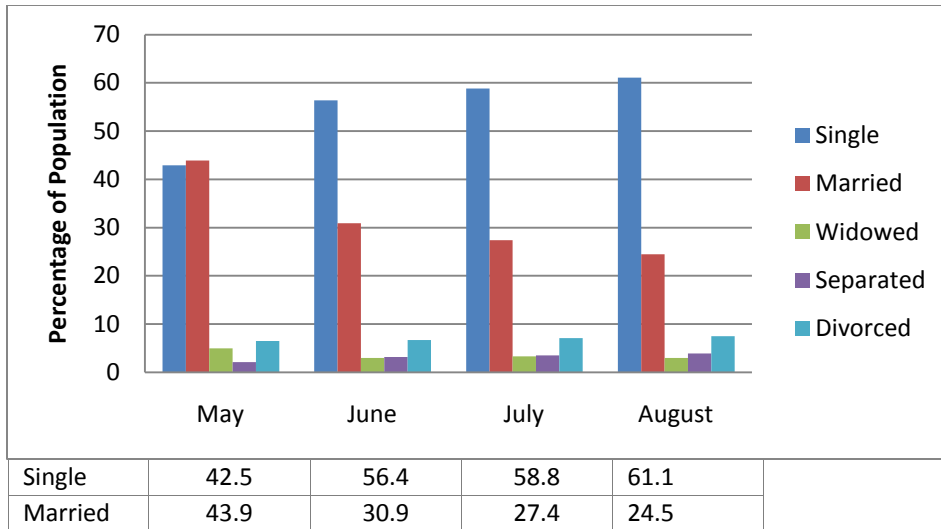


Figure 2: Distribution of Monthly Responses (Population) by Marital Status

To examine the age variable, we grouped responses into 10-year age ranges. In Figure 3, showing the distribution of population by selected age groups, the biggest changes in distribution over the collection period occurred for those aged 70 and over and those aged 20-29. Both of those groups made up close to 11% of the May responses but the younger population increased to 21.1% of the August persons reported, while the older group fell to only 4.0%. The total distribution over the four-month period was 12.9% persons aged 20-29 and 9.0% persons aged 70 and over.

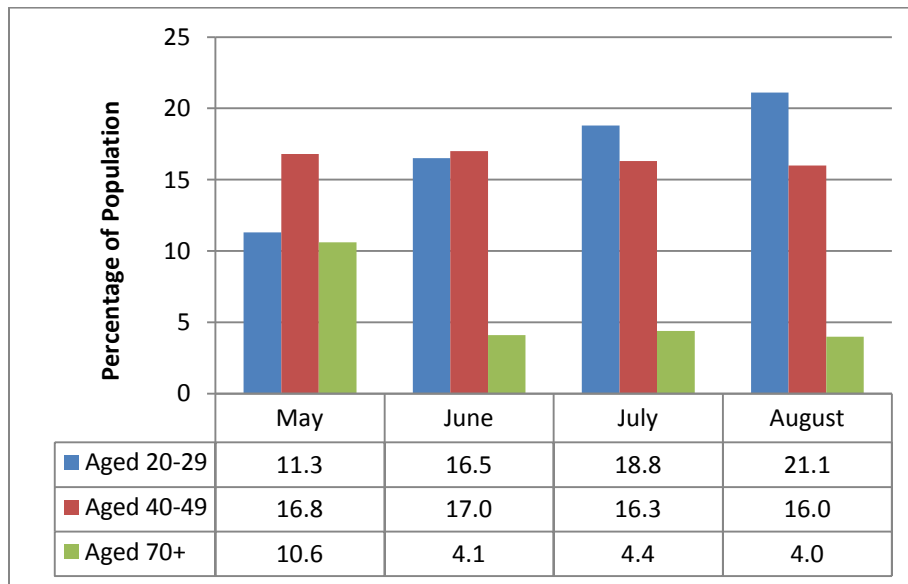


Figure 3: Distribution of Monthly Responses (Population) by Age Group (Selected)

Sex and marital status were asked on both the short and long census questionnaires, and data collected from both those form types was used to calculate the percentages of monthly responses. Percentages presented in the rest of this section are for variables asked on the long form only, weighted up to the total population.

The 2006 Census long form had one question about immigration status (landed immigrant or not a landed immigrant) and one question about the year of immigration for landed immigrants. Figure 4 demonstrates that the proportion of immigrants rose steadily through the collection period from 19.2% of May responses to almost 25.7% of August responses. The remainder of the population are non-immigrants. Immigrants made up 20.1% of the total responses over the four months of collection.

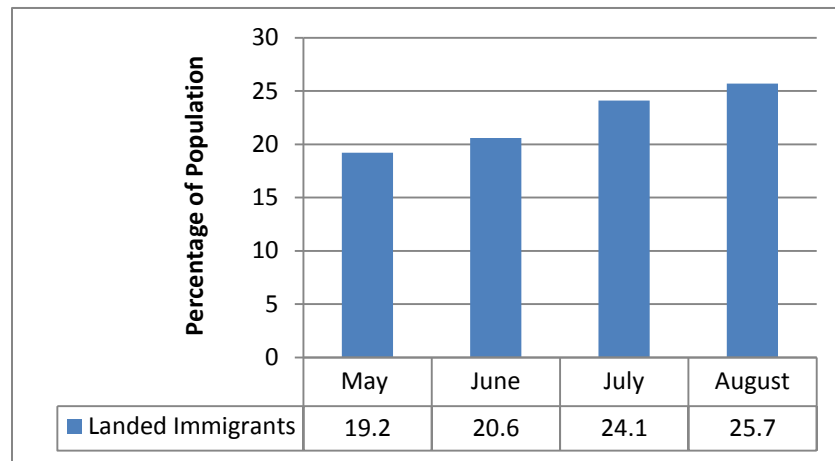


Figure 4: Distribution of Monthly Responses (Population) by Immigrant Status

When looking at year of immigration distributions in Figure 5, we found that recent immigrants (in Canada for 10 years or less) made up more of the late responses (10.3% in August) than those who had been here longer. Those who had been in Canada for 30 years or more declined as a percentage of the monthly totals, while more recent immigrants increased in later months of collection. Over the total collection period, the distribution of immigrants who immigrated before 1976 is 6.5%, between 1986 and 1995 is 4.8% and between 1996 and 2006 is 6.3%. As a reminder, the denominators used in the calculation of percentages are the total in-scope population enumerated in a given month.

The 2006 Census long form had three questions on Aboriginal identity. As specified in section 1, the analysis excluded data collected by the canvasser methodology, which was used for northern communities and Indian Reserves. Therefore, the analysis of data from those questions included only persons enumerated outside Indian reserves and northern communities.

Table 2 illustrates the distribution of monthly responses (percentages and counts) by Aboriginal status. The table shows a significant increase of persons who identified as an Aboriginal person or Registered Indian in the months of July and August. For the population in scope for this study, a total of 716,000 were reported as Aboriginal people living off-reserve. Of this group, 23% (113,000 + 51,000) were reported in the last two

months of collection. The distribution over the total four-month period of persons (living off-reserve and not in northern communities) reporting Aboriginal identity is 2.4% and reporting Registered Indian status is 1.0%. The increase from May to August in the percentage of persons reporting as Aboriginal (more than doubling from 1.8% to 5.2%) is one of the largest and therefore most significant of all variables analysed.

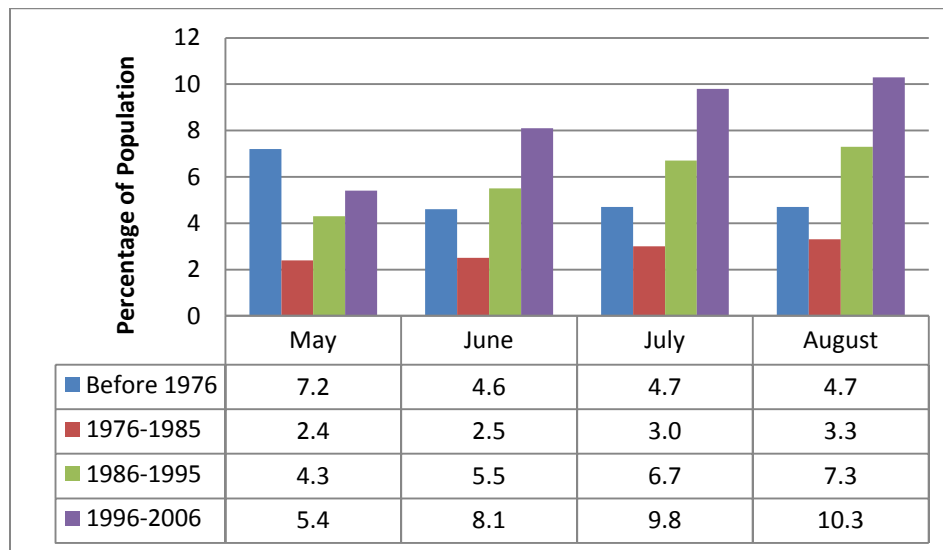


Figure 5: Distribution of Monthly Responses (Population) by Year of Immigration

Table 2: Distribution of Monthly Responses (Population) by Aboriginal Status (Off-Reserve and excluding northern communities)

Status		May	June	July	August
Aboriginal (identify with at least one aboriginal group)	% of total pop.	1.8%	3.6%	4.2%	5.2%
	Count (in 000's)	397	155	113	51
Registered or Treaty Indians (registered under the Indian Act of Canada)	% of total pop.	0.6%	1.6%	2.1%	2.5%
	Count (in 000's)	132	26	16	6

Visible minorities are defined in the *Employment Equity Act* as 'persons, other than Aboriginal peoples, who are non-Caucasian in race or non-white in colour'. The long form has one question about the self-identification of persons in one of the visible minority groups. In Figure 6, showing the distribution of the total visible minority population, members of all visible minority groups combined increased in proportion over the collection period from 14.5% in May to 26.3% in August. Over the four-month period, visible minorities made up 16.5% of the total responses.

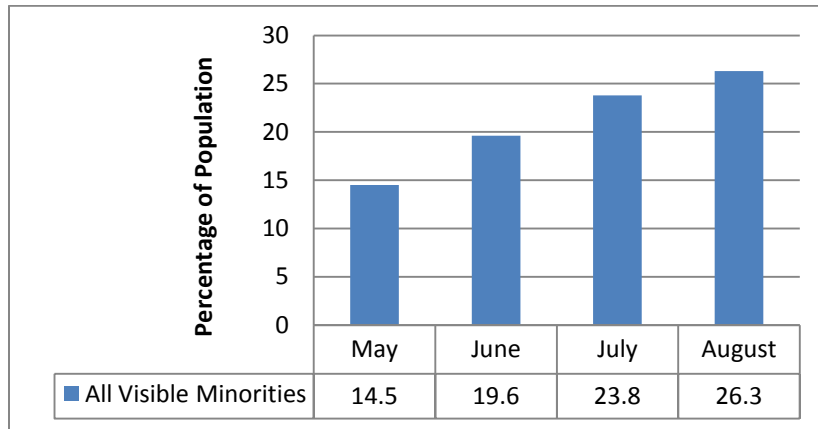


Figure 6: Distribution of Monthly Responses (Population) by Visible Minority Group (All)

The analysis also looked at the monthly responses for the three largest visible minority groups in Canada. As shown in Figure 7, the Black and South Asian populations increased as a percentage of the monthly totals from May to August, but the percentage of Chinese decreased. In fact, Chinese were the only single visible minority group to decline as a percentage of the monthly totals. The overall distribution for persons reporting was: Black 2.3%, South Asian 4.0% and Chinese 3.8%.

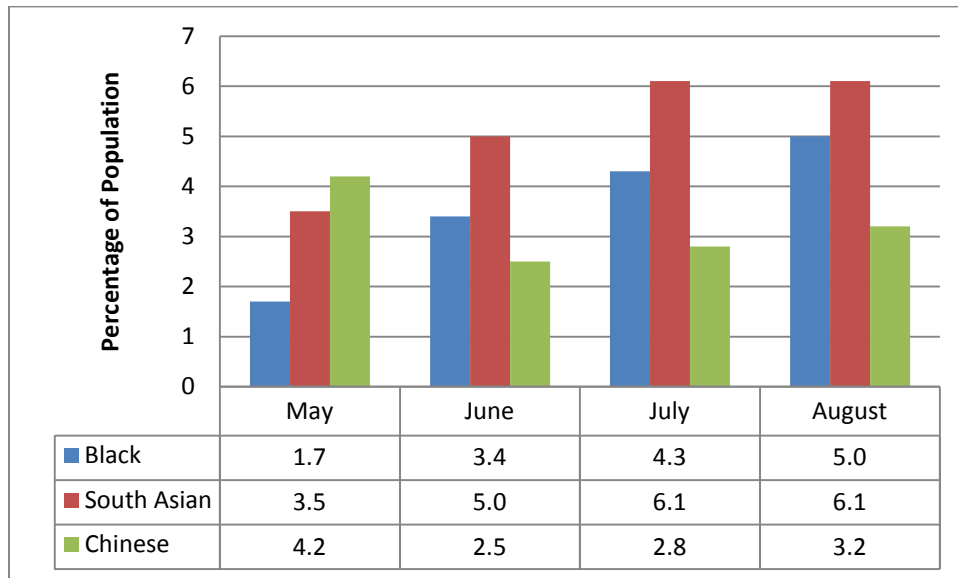


Figure 7: Distribution of Monthly Responses (Population) by Visible Minority Group (Selected)

Another variable where differences between early and late responses were apparent was mobility. The Census long form had two questions about mobility: one question about the

place of residence one year prior to the 2006 Census day and another about the place of residence five years prior to the 2006 Census day. Table 3 illustrates the distribution of monthly responses to those questions. Persons who were living at the same address one year ago declined in proportion through the collection period, while persons who had moved increased, especially those who had moved within the same census sub-division. The same trend occurs for place of residence five years ago.

Table 3: Distribution of Monthly Responses (Population) by Place of Residence

Place of Residence 1 year ago

	May	June	July	August
Same Address	87.7%	80.4%	76.7%	73.5%
Moved, Same Census Sub-Division	6.5%	10.7%	13.7%	16.0%
Moved, Different Census Sub-Division	4.1%	6.3%	6.7%	7.5%
Moved, Outside Canada	0.8%	1.2%	1.5%	1.7%
Not applicable (persons < 1 year old)	0.9%	1.4%	1.4%	1.3%

Place of Residence 5 years ago

	May	June	July	August
Same Address	60.8%	46.5%	40.9%	36.7%
Moved, Same Census Sub-Division	18.5%	24.6%	28.9%	32.1%
Moved, Different Census Sub-Division	12.9%	16.9%	17.4%	18.2%
Moved, Outside Canada	3.1%	5.0%	5.7%	6.4%
Not applicable (persons < 5 years old)	4.7%	7.0%	7.1%	6.6%

Response distribution by month was analysed for other variables on both the short and long census forms. Differences between early and late responses were also evident for variables such as education (highest degree attained), income, and labour activity. The results presented in this section have highlighted variables where some of the most significant differences were observed over the four-month collection period.

2.3.2 Logistic Analysis

Results from the previous sections showed different distributions of responses for the four-month collection period. Obviously, responses to some questions are correlated and those results don't show how the responses influenced the return date or what characteristics best explained the return date. For example, is age more important than marital status to explain a late return? Is immigration status more important than visible minority status to explain a late return? Logistic regression would enable us to investigate the relationship between a categorical outcome (month of return) and a set of explanatory variables (answers on the short or long form).

An analysis by logistic regression was conducted to find the characteristics (questions) that best explain the return date and to try to answer the above questions. Because the return date is an ordinal response, we fitted a proportional odds model. The dependant variable was the return date (month) and the model had about 20 characteristics as independent variables. All those 20 characteristics were categorical variables. Some characteristics (questions) were removed from the logistic model. As an example, there are four education questions about the degree attained: one question for the high school diploma attained one question for the trade school diploma attained, one question for the

college diploma attained and one question for the university degree attained. Based on the answers from these four questions, the highest degree attained is derived for each person and it is this derived variable that we used in the model instead of four separate variables (one for each education question).

The analysis of the maximum likelihood estimates produced for each parameter in the model was used to determine the most important characteristics to explain the return date. The model used reference groups for each characteristic. The determination of each reference group was based on the distribution of May responses and in most cases the reference group is the group with the highest proportion of persons. Based on the maximum likelihood estimates and the selection of reference groups, the characteristics that best explain the return date were, in order of importance:

1. Age: persons 60 years old and over (increase in May); reference group: persons aged 40-49.
2. Aboriginal status: persons self-identified as Aboriginal (decrease in May); reference group: not Aboriginal.
3. Citizenship status: persons with no Canadian citizenship (decrease in May); reference group: persons with Canadian citizenship.
4. Visible Minority status: persons from visible minority groups (decrease in May); reference group: persons not in any visible minority group.
5. Immigration status: immigrants (decrease in May); reference group: non immigrants.
6. Highest degree: person with the highest degree being high school diploma or less (decrease in May); reference group: persons with any university degree.

Note that the order doesn't take into account the variance of each maximum likelihood estimate; that is, the difference between the first and the second variable may not be significant.

2.3.3 Imputation for complete non-response

Although the 2006 Census collection tried to get a questionnaire from all households, non-response did occur. In 2006, 3.5% of all private occupied dwellings did not return a questionnaire, up from 1.6% for the 2001 Census of Population. In 2006, responses for all private occupied dwellings with complete non-response (no questionnaire) were imputed. Donor imputation was used as the imputation method. The imputation selected a donor household by geography and by household size (when possible) and 'cloned' the records into the household with a complete non-response. The level of geography used for this process is a Collection Unit (CU). Typically, a CU has on average 250-450 dwellings. CUs are often created to form a homogenous groups of dwellings and to help the logistics of collection.

The analysis compared the characteristics of persons within households imputed for complete non-response to those of persons reported during collection. We found that the variable distributions of the imputed population looked very similar to the distributions over the full four-month collection period. If we assume that households who did not return a questionnaire are similar to households who completed a questionnaire in the last months of collection, one could suggest using the return date as a criteria for selecting a donor, i.e. to select a donor household, in addition to the geography and household size criteria, we could use the return date. Households who returned a questionnaire late in collection would have a higher probability of selection as a donor.

Although the assumption seems to suggest this methodology would make sense, analysis must be done to see if the results of such a change would improve the data at lower levels of geography. The analysis presented in this paper focused mostly on Canada-level data and Table 1 shows a significant number of persons enumerated in July and August. The donor imputation methodology in 2006 used a low level of geography, i.e. the CU level (250-450 dwellings). The collection of data in 2006 was completed for many CUs by mid-June 2006, and therefore no person was enumerated in July or August for a large number of CUs. Applying this type of change to the donor imputation at this very low level of geography may have unexpected results, as the distribution of responses by month of collection at a lower level of geography may not be in line with the Canada-level patterns. Further study will need to be conducted to measure the impact of this potential change and its feasibility.

3. Conclusion

In examining the variable distributions by month of response, we found that the population who are young, single, recent immigrants, visible minorities, Aboriginal people, and mobile make up a larger percentage of July/August responses (late responses) than May/June responses (early responses).

There are many assumptions to explain why persons with those characteristics might tend to be reported later in collection. Qualitative testing of the census forms supports some of those assumptions. Some recent immigrants are not able to read the short or long forms distributed in English and French (although other languages are available upon request or on the Internet) or they don't understand some of the concepts on the forms. In both cases, they require the help of a census enumerator to complete the form (NRFU). Some persons without Canadian citizenship think the census of population is only for Canadian citizens (although instructions on whom to include specify including not only people with Canadian citizenship). Other people with literacy difficulties have problems reading and understanding the form and wait until they can get the help of a census enumerator. In addition, young adults are more mobile and difficult to reach and are thus sometimes reported later.

More analysis will be needed to measure the impact of using additional information such as the questionnaire return date to select a donor to impute for complete non-response. Information from this study may also be used for the planning of the Canadian 2011 Census of Population.

Acknowledgements

The authors would like to thank Jane Badets, Dany Faucher, Linda Howatson-Leo, Yves Lafortune and Laurie Reedman for their helpful comments.