WP. 10
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Ljubljana, Slovenia, 9-11 May 2011)

Topic (ii): Editing of electronic collections

The Evolution of Edits in the Canadian Census of Population Online Questionnaires

Key Invited Paper

Prepared by Danielle Laroche, Statistics Canada

I. Introduction

1. In May 2011, Statistic Canada is encouraging the population of Canada to fill out a Census of Population questionnaire via the Internet. To that end, 60 % of Canadian households will receive a letter inviting them to complete their questionnaire online: they will not receive the traditional paper form. This, paired with a far-reaching public relations campaign to promote Internet response, is expected to encourage 40% of Canadian households to use the online option.

2. This paper provides an overview of the Internet data collection methodology to be used in Canada's 2011 Census of Population. It describes the context of the census, and reviews the main components of the electronic questionnaires. It then presents the various response validation edits. Finally, the paper gives, for each type of edit, a rationale, the edit's effect on data quality and, if applicable, changes made as a result of lessons learned from the 2006 Census of Population, previous tests and Internet evaluation studies.

II. Background for the Canadian Census of Population

A. Overview of the 2006 Census

3. In Canada, a census must be conducted every five years, and all households must complete a census questionnaire. The last census took place in 2006. Approximately 31.6 million people and 12.4 million households were enumerated at their usual place of residence in Canada. Self-enumeration was used to count most of the population. Two main types of questionnaire, offered in both official languages, were used to collect the majority of census data:

i) a short form, containing eight questions, was distributed to 80% of households;
ii) a long form, containing 53 questions about people and 9 questions about dwellings, was distributed to the remaining 20% of households. All questions from the short form were included in the long form.

4. A mail-out methodology was used in areas where addresses were available for each dwelling and when a questionnaire could be mailed to the dwelling (approximately two-third of all dwellings). A list-leave methodology was used in areas where dwellings were listed and for which questionnaires were dropped off by an enumerator (approximately one-third of all dwellings). Finally, a canvasser methodology (personal interview) was used in remote and northern areas for a small portion of the population (approximately 2% of all dwellings).

5. An Internet option was offered for the first time to all private dwellings. The questionnaires contained the web address, or URL, and a secure access code. Except for a few cases, the online questionnaires replicated as closely as possible the paper questionnaires to facilitate the integration of data and to minimize the mode effect. Editing in the online questionnaires was relatively simple, and was performed on one question at a time. No consistency edits were performed between questions. The edit messages were provided mostly for non-response, partial response (for example, missing part of an address such as street, city, province or postal code) and for invalid response in numeric or dollar-amount fields when respondents entered a number outside of the range established for a question.

B. What we learned from the 2006

6. The Internet option was a success—18.3% or approximately 2.2 million of private occupied dwellings completed and returned their census questionnaire online.

(a) Impact of the Internet response on data quality

7. This online option had many benefits and was considered superior to the other modes of response. The online questionnaires were completed more thoroughly and required less follow-up (Table 1). As well, we received them without delay, and they required no data capture. Online questionnaires also reduced manual coding of write-in responses, maintained participation (more than 81% of all responses were from self-enumeration) and responded to public expectations.

Table 1: Comparison of rejection rates for paper and online questionnaires

Type of questionnaire	Rejection rate (%)	
	Paper	Online
Short	5.6	2.5
Long	39.1	5.7

8. The features of the online questionnaires such as edit messages, radio buttons, automated skips and explanations of why we asked each question were the most important reasons for lower online rejection rates.

(b) Internet promotion tools

9. The front cover of the paper questionnaire served poorly as an Internet promotion tool; often, respondents don't read, they just start filling white response spaces. Consequently, many of them (about 17% of households) were not aware of the Internet option. We have learned from follow-up surveys that many households with an Internet connection did not use the option because they had a paper questionnaire and a stamped envelope handy. Thus, sending a paper questionnaire actually discouraged Internet responses.

(c) Impact of sending a letter instead of a questionnaire to promote Internet response

10. A small test was conducted during the 2006 Census to explore another way to promote the online option. Households most likely to respond via Internet were targeted to receive a letter instead of a paper

questionnaire. The letter included a URL, a secure access code and a number to call to request a paper questionnaire. This strategy was very effective: 61.8% of the households who received the letter returned their questionnaire online, compared with 23.8% of households in a control group that received a paper questionnaire. In addition to the stronger use of the Internet, this method also indicated a potential for more non-response follow-up (NRFU), if households receive no questionnaire.

(d) Mode effect

11. Since the Internet option was first tested during the 2001 Census, significant efforts have been made identifying and measure this 'mode effect'—differences in responses owing to the method used to answer a question—in the Census. These efforts comprised qualitative testing, comparing data reported for each mode, comparing data from the 2006 Census (paper and Internet) to data from the 2001 Census (paper only) for the same person, comparing standardized results and using the propensity score method to detect mode effect. These studies generally showed that the mode effect of the Internet was positive regarding data quality. However, mode effects are not easy to detect. Each method had important limitations, and sometimes distinguishing between mode effect and time effect was difficult. Up to now, negative impacts with the online option were mainly due to the technology: no major effect was observed on the response itself. Some problems were identified but were easy to fix for the next census. Section III provides more details on the mode effect (drop-down menus for provinces, non-response edit for postal codes and invalid response edit for the number of hours worked).

C. The 2009 Census Test

12. The 2009 Census Test comprised two samples. A primary sample (non-probabilistic) of 110,000 dwellings was selected in two provinces (Quebec and Alberta) to test all systems and procedures. A supplementary, probabilistic sample of 25,000 dwellings divided into five panels was selected to test the new collection methodology and the Internet promotion letter. A detailed description of the 2009 Census Test samples can be found in Taylor (2009).

(a) The 'wave methodology'

13. This new methodology was designed to enable maximizing Internet response without increasing total non-response. This improved methodology reminds respondents to complete their questionnaire via various modes at specific times during data collection. The following table outlines the waves that were tested in the 2009 Census Test. In mail-out areas, there were two groups: one received the Internet promotion letter; the other received the paper questionnaire (Table 2).

Table 2: Census Test waves and activities

	Wave	Time to census day	Activity
Mail-out	1	7 days before	Mail-out (letter or questionnaire)
	2	2 days after	Reminder letter to non-respondents
	3	10 days after	Reminder questionnaire to non-respondents or telephone voice broadcast message, if a phone number is available
	4	20 days after	Non-response follow-up

14. In **list-leave** areas, Wave 1 consisted of enumerators dropping off a questionnaire at the dwellings. At Wave 2, a thank you/reminder card was distributed to these dwellings. Non-response follow-up (NRFU) started at Wave 3.

(b) The main findings

15. The main findings of the 2009 Census Test were:

- Internet returns confirmed to be much higher in Panels 1 and 2 receiving a letter (41.1% and 39.6%) rather than receiving a questionnaire (benchmark 14.8%).
- Among all respondents, Internet responses were also higher in panels using the wave methodology. Responses ranged from 56.6% to 69.7% compared to 28.6% for the benchmark.
- A telephone voice broadcast message was an efficient and inexpensive follow-up method.
- The presence of a phone number on the census frame (master control system) was linked to dwellings that responded more; the absence of a phone number was linked to dwellings that responded less.
- Regarding NRFU, more non-response was observed initially with the new methodology compared with the old one. But in the end, both methods produced the same results. (To decrease the NRFU workload, the Wave 3 questionnaire will be sent earlier in the 2011 Census collection.
- Online questionnaires yielded better data quality than did paper ones (Table 3). The Internet rejection rate for the short questionnaire was comparable with 2006 (2.4% and 2.5% respectively). The rejection rate for the long questionnaires was higher than in 2006 (12.4% and 5.7% respectively). There were two reasons for the higher Internet rejection rate. The first was a processing issue: conversion problems occurred when integrating Internet responses to the usual channels. The second reason was related to the 'aging process'— respondents saved their questionnaire but never came back to finish it and their questionnaire was submitted on their behalf (see section III-B (d), "Stop and finish later"). In 2006, most of these questionnaires were submitted towards the end of data collection; thus, many of these households completed another questionnaire during the NRFU. In 2009, almost two-thirds of online questionnaires that were rejected came from this 'aging process'. These saved and incomplete questionnaires were submitted earlier on the respondent's behalf and sent to failed edit telephone follow-up in order to lighten the NRFU workload.
- The average times to complete the census online questionnaires were 7 minutes 19 seconds for the short form and 35 minutes 41seconds for the long form.

Table 3: Comparison of rejection rate between paper and online questionnaires

Type of questionnaire	Rejection rate (%)	
	Paper	**Online**
Short	4.5	2.4
Long	45.9	12.4

16. Therefore, the promotion letter, combined with the wave methodology was successful in promoting Internet responses; both were recommended for the 2011 Census.

D. Modifications made to the 2011 Census

17. Last summer (2010), the Canadian federal government announced the abolition of the mandatory long-form census questionnaire. Privacy concerns were one of the reasons for the decision. The voluntary National Household Survey (NHS) replaced the mandatory long-form census questionnaire. The new survey contains all of the questions that would have been asked in the 2011 census long-form questionnaire. A few modifications were made to questionnaires mainly to reinforce the reference day, May 10, 2011. Two language questions were added to the short census form to enable the government to fulfill its responsibilities under the *Official Languages Act*. All dwellings will be required to complete the mandatory short form. For the NHS, a sample of 4.5 million dwellings (about 30% of all dwellings) will

be selected to participate in the survey (an increase from one in five households in 2006 to one in three in 2011).

E. Overview of the 2011 Census methodology

(a) Census wave methodology

18. For the 2011 Census, the mail-out methodology was expanded to cover 80% of the population. Dwellings most likely to respond via the Internet will be targeted to receive an Internet promotion letter. The wave methodology is summarized in Table 4.

Table 4: Census waves and activities

	Wave	Time to census day (May 10)	Activity	
			Group 1	**Group 2**
Mail-out 80%	1	1 week before	Mail-out letter (75%)	Mail out questionnaire (25%)
	2	Census day	Reminder letter to non-respondents	
	3	8 days after	Reminder questionnaire to non-respondents	Voice broadcast message (if phone number is available)
	4	20 days after	Non-response follow-up	
List-leave 18%	1	1week before	Questionnaires dropped off (100%)	
	2	Census day	Thank you/Reminder card	
	3	20 days after	Non-response follow-up	
Canvasser 2%	_	_	Personal interview (100%)	

19. In **Wave 1 in mail-out** areas, about 75% of the dwellings (comprising 60% of all dwellings) will receive a letter and 25% (comprising 20% of all dwellings) will receive a paper questionnaire. The Internet promotion letter will contain a 15-digit secure access code, the web site address and a toll free-number to request a paper questionnaire through an automated system. The questionnaires will also contain the 15-digit secure access code and the web site address on the front cover. In **list-leave** areas (approximately 18% of all dwellings), a questionnaire will be delivered by an enumerator. Finally, 2% of the dwellings will be enumerated by personal interviews (**canvasser** methodology for Indian reserves, northern and remote areas). At **Wave 2**, reminder letters will be sent to non-respondents in mail-out areas; dwellings in list-leave areas will receive a thank you/reminder card. At **Wave 3** in mail-out areas, non-respondents who received a letter at Wave 1 will receive a questionnaire; those who received a questionnaire will receive a telephone voice broadcast message if a telephone number is available in the census frame. There will be no Wave 3 for dwellings in list-leave areas. At **Wave 4**, a phone call or a personal visit will be made by an enumerator (NRFU).

(b) National Household Survey

20. About 4.5 million dwellings, or 30% of all dwellings, will be selected to participate in the voluntary National Household Survey (NHS). A systematic sampling design by province and collection method will be used to select the sample in mail-out and list-leave areas. All households in canvasser areas will be asked to participate. The survey will be conducted at the same time as the census, and will use the same infrastructure. The NHS wave methodology will depend on how households respond to the census. There are three survey groups:

- **Survey Group 1: Census Internet respondents**

Respondents who are selected for the NHS, and who answered their census questionnaire online, will be automatically given the NHS online questionnaire. These respondents will skip the questions that were already asked in the 2011 Census and their answers to those questions will be automatically transferred to the NHS online questionnaire. A reminder letter will be sent to non-respondents households on June 7. For reasons of cost- and response-rate efficiency, these households will never receive an NHS questionnaire.

- **Survey Group 2: Other census respondents**
  An NHS questionnaire will be mailed (in mail-out areas) or dropped off (in list-leave areas) during the first week of June. A reminder letter will be sent to non-respondent households one week later in mail-out areas. If they have not filled it out by July, a second questionnaire will be mailed out. Lastly, an enumerator will visit a sample of these non-respondent households.
- **Survey Group 3: Census non-respondents**
  Because these households will not have responded to the mandatory census, sending further material for a voluntary survey would not be efficient. For this reason, NRFU will be conducted by phone or in person at the same time for non-respondents to both the census and the NHS. After a specific date, a sub-sample of NHS non-respondents will be selected for further NRFU.

21. Budget and capacity constraints permit follow-up to be conducted only on a sub-sample of the non-respondents. This strategy aims at reducing non-response bias as much as possible. A detailed description of the Census and NHS wave methodology can be found in Mathieu and Morin (2010).

III. The Internet questionnaires

22. The Internet questionnaires comprise two distinct components. The first, the 'web module' is a communication tool that provides general information about the census and information for completing the questionnaire over the Internet. The other component, the 'online application' comprises the questionnaires that enable respondents to complete and return their information over the Internet.

The Internet offers three types of questionnaire in both official languages, English and French:

i) The short census form (2A) is the main census questionnaire.
ii) The individual census form (3A) is used for people who do not wish to be enumerated on their household questionnaire, such as a roommate, lodger or boarder.
iii) The long National Household Survey form (N1) is used to collect additional data for about 30% of Canadian households.

A. Modifications made to the 2006 online questionnaires

23. The initial investment in hardware and software for offering an Internet option is very costly. Nevertheless, Internet data collection is viewed as a way to reduce costs over the longer term, especially through reducing the size of field and processing operations—provided that you get high Internet return rates. After the 2006 Census, Statistics Canada decided to reuse the 2006 system where possible. However, investing in technology upgrades and expanded capacity for the Internet option never ceases. Although we reused the 2006 online application for 2011, the electronic questionnaires needed substantial modifications. The main reason for these changes was the 'Common Look and Feel' (CLF) requirements. The CLF are standards and guidelines for presenting content on Canadian federal government websites. The new version (2.0) came into effect in 2007. The goals of CLF are to enhance recognition of Government of Canada sites and increase general confidence in the information Canadians receive from their government online. These standards are related to usability, interpretability and accessibility. They

aim to enable federal websites to reach the widest possible audience and the broadest possible range of hardware and software platforms, from adaptive technologies to emerging technologies.

24. Two CLF-based modifications had a large impact on the 2006 Internet application. The first was the removal of tables for positional layout due to accessibility: to achieve positional layout, we used cascading style sheets instead. The second was to ensure functionality of the questionnaires without JavaScript which is commonly used to make web pages more interactive. Without JavaScript, the potential for inconsistent responses in the online questionnaires is increased, because JavaScript is used to perform some edits while the respondent is entering data. Overall, CLF affects the general layout and many features and functionality of the Internet web module and questionnaires, including the presentation of pages, edit messages and their presentation, and the placement of the online help and the progress bar.

25. The 2006 online questionnaires were also modified to address problems identified in 2006, content modifications following the 2008 and 2009 Census Tests, as well as recommendations from qualitative testing. While the application required a number of modifications, some edits were developed to test more complex editing strategies, notably for implementing a fourth level of edit rules (the three other levels of edits are non-response, partial response and invalid response). These new edit rules aim to improve the quality of the specific answers while at the same time mitigating potential respondent burden (see section C 4 for a description of these new edits).

B. Questionnaire design and features

26. When the 2006 online questionnaires were designed in early 2002, concerns were raised about the possibility that respondents might give different answers to the same question depending on whether they used a paper or an online questionnaire. The comparability of historical data is an important aspect of a census. We debated internally between two schools of thought: keeping the same questionnaire format for both modes (paper and Internet) or using the full potential of each mode of response. There was no consensus in the literature about the best way to design questionnaires in mixed-mode surveys. During the 2001 Census, a small Internet test was conducted; we looked at its evaluation and recommendations. Early in 2002, we developed a few options and conducted, with the help of usability experts, several tests with parts of the questionnaire. We also did formal and informal qualitative testing; consulted with questionnaire design experts inside and outside of Statistics Canada; and conducted a quantitative test in 2004 and a few follow-up surveys with Internet and non-Internet respondents. At the end, we kept what seemed to work best for us. Our overall approach was to adhere to the paper form regarding the wording of the questions, instructions and presentation of response choices while incorporating many web questionnaire standards and functionalities.

27. Questions and response options appear in a box at the centre of the page to make them stand out. Each question is displayed on a background colour. The online questionnaires use an interactive multi-page design. With this design, the questionnaire is presented screen by screen, each screen displaying a group of questions. A new screen is presented when there is a skip or when a new group of questions needs to be displayed (for example, activities of daily living, sociocultural information, mobility, education, labour market activities, income and housing). In the end, the short questionnaire had 10 pages; the long questionnaire had 45.

(a) Formats (matrix and sequential)

28. As with the paper questionnaire, a matrix format is mainly used in the online questionnaires. With this format, the question is asked once and the response choices are repeated under the name of each person in the household. A sequential format is used in the long form for labour market activities and income questions (Questions 40 to 46 and 55). Qualitative testing revealed that respondents found it easier

to concentrate on one person at a time. With the sequential format, the name of the person is included in the question and a group of questions is asked one person at a time.

(b) Radio buttons

29. Radio buttons are used in the online questionnaires for two main reasons. First, they are well-recognized by the general public since they are similar to what is used on paper questionnaires. Respondents generally have no issue with this type of response choice. Second, radio buttons force respondents to provide only one response. This increases data quality by avoiding inconsistencies. Despite these advantages, radio buttons pose accessibility issues. Because of the small target, users with mobility impairments might have difficulties selecting radio buttons. To address this, we made the text associated with the radio buttons in the online questionnaires 'clickable'. Second, radio buttons are not scalable under some browser versions, which might hamper vision-impaired users. The short and long questionnaires have many long response choices associated with radio buttons that prevent the use of select lists, which are recommended in the CLF guidelines.

30. Another negative issue concerning usability: once respondents have selected an answer, they cannot leave the question blank. With radio buttons, when respondents change their answer by selecting another response, the radio button for the initial answer automatically clears. With select lists, the entire row of the select option is highlighted: users with low vision can easily determine which option is selected. Similarly, users with mobility impairments can easily select an option because the target is scalable and the option can be navigated and selected using a keyboard.

(c) Check boxes and write-in boxes

31. Checkboxes are used to indicate that multiple responses are possible. For questions that have an "*other*" radio button associated with a "*specify*" write-in box, the cursor automatically appeared in the text field when the user selected the radio button. Inversely, when the respondent just started to type a response in a write-in box, the corresponding radio button was automatically selected. This functionality was also performed by JavaScript.

(d) Drop-down menus

32. Drop-down menus contain the possible response options for a question. They were used in the 2006 online questionnaire for selecting one of the13 Canadian provinces or territories as well as respondent's day and month of birth. Respondents like drop-down menus because they perceive the menus as easy to use and they need not enter text. Data quality from questions with drop-down menus was generally good. Nevertheless, these menus could be problematic when using the wheel on the mouse to scroll down. When respondents needed to scroll down to the next question after selecting an answer from a drop-down menu, they had to first click outside the menu to deactivate it. Otherwise, the selected response could change without the respondent noticing. Failure to properly move from the menu to the next field resulted in a response being changed to another selection in the list, often the last one on the list. We detected evidence of this mode effect for a small portion of respondents for two of the six questions for which drop-down menus were used in the 2006 Census online questionnaires. Edit and imputation strategies were put in place to detect and correct this situation. (For the other questions where no scrolling was needed, there was no evidence of the problem.) This problem was corrected by confining drop-down menus only to selected questions, and by automatically deactivating the menu after response.

(e) Automated skip

33. The automated skip is used to reduce response burden by ensuring that respondents are not presented with irrelevant questions. This feature is used only in the long questionnaire. There are two types of automated skip. The first applies to the entire households when all persons in the household must

skip one or more questions. A message is presented at the top of the screen. For example: "*Based on your response to Question 31, Question 32 is not applicable. Therefore, proceed with Question 33.*" The other type of automated skip applies when a question is applicable to one or more persons but not the entire household. In this case, a message appears in the answer space below each person's name for which the question is irrelevant. For example: "*Not applicable based on a previous answer.*"

(f) Online help

34. There are two possible ways to obtain help in the online questionnaires. Respondents can click the Help button of the top menu bar to obtain a national toll-free number that provides assistance for any question or problem with the Internet application. A help icon also appears at the top-right corner of the page for each question. The corresponding help is in two parts. The first part provides instructions and examples to help respondents answer the question. The second part explains why the question is being asked and how the information will be used. Although respondents thought the online help is a nice feature to have, we know that they used it considerably less often than needed. In 2006, the online help was presented as a link in the left column beside the question. Less than 10% of respondents used it.

(g) Progress bar

35. A bar representing the respondent's progress (or their completion status) in the online questionnaire appears on the right of the screen beside the title of each page. This feature was improved for the 2011 Census. The change of location from 2006 makes it less obvious for respondents, as it is now buried in the blue background. However, respondents like it and consider it a nice feature to have.

(h) Stop and finish later

36. This feature is available in the long questionnaire only. It enables respondents to save their questionnaire and to complete it at a later time, over multiple sessions, and even from different locations. To use this functionality, respondents can create their own password, or let the application assign them one. When respondents return to finish their questionnaire, they are prompted to enter their secure access code and their password. They have up to five attempts to enter their password correctly. For security reasons, if they are unsuccessful or if they do not log back in within a prescribed period of time, their partly-completed questionnaire is submitted on their behalf (this is called 'aging process'). In 2006, approximately 17% of all Internet users who logged in to the long form used the 'Stop and finish later' functionality; 90% of them submitted their questionnaire. In 2009, 13% saved their questionnaire and 84%of them submitted it.

C. Edit messages

37. We tried to keep a balance between the number of edits and respondent burden. We wanted the edits to help respondents give complete and accurate data without annoying or frustrating them. Qualitative testing and follow-up surveys enabled us to better understand respondent's behaviour toward this new response channel. Our experience with online edits indicates that respondents are generally receptive to messages originating from edits. In fact, Internet users expect the online questionnaire to be 'intelligent', easy and quick to complete—they want an interactive experience.

38. All the online messages follow the same approach. When respondents click the **Continue** button, the information on the current page is validated: if the application detects any problem, its refreshes the page and the question that needs to be corrected is noted in red at the top of the page. If applicable, this text also indicates the person for whom the answer should be corrected. To assist the respondent, the question and field requiring attention appear in red, and a red arrow highlights the missing response. The respondent can then either fill in the missing information or ignore the message and continue to the next

page. If the respondent chooses the latter, the next page appears. If the respondent adds or changes any information, then the responses are validated again. Each message is shown only once.

39. Follow-up surveys conducted after the 2006 Census showed that 18% of Internet respondents (13% for the short form and 38% for the long form) received at least one message as a result of editing; 94% of them reported that the messages were useful but 88% did not know that they could have ignored these messages. Qualitative testing showed that respondents to the long form received two or three edit messages while completing their online questionnaire. There are no 'standards' for an appropriate number of edit messages to be presented to a respondent. Too many messages can increase their burden and the time required to complete the questionnaire. However, we want to take advantage of the possibilities that Internet data collection offers. The benefits of editing for data quality must be weighed against greater response burden to find an appropriate balance.

40. Testing was also useful in indicating which edits seemed more appropriate than others. The edits performed on the 2006 online questionnaires were relatively simple. All of them were performed one question at a time. For 2011, we introduced more complex edits to test their feasibility.

(a) Mandatory edits (or hard edits)

41. The electronic questionnaires have two **mandatory questions**: Step B1, the number of persons living at the address on Census Day and Step B2, the names of these persons. These two questions are mandatory because they are used to build and customize the online questionnaires. The first question is used to build the next page: for example, if a respondent indicates "4" in the number of persons page, then on the next page, the application generates four lines to enter the names of each person. These names are subsequently used in all other questions to identify the answer of each person. These two questions are the only mandatory questions on the Internet questionnaires. Besides those two questions, we were opposed to mandatory questions because of the risk of respondents dropping out and because, for a survey the size of a census, any data is better than nothing.

(b) Confirmation pages

42. Two pages were designed to help respondents provide quality data: one for date of birth and the other one for income. The date of birth is particularly important for the census and the NHS. In both questionnaires, persons younger than 15 years of age are subject to different edits for marital and common-law status questions. The non-response message is not presented when these questions are not answered for such people. In the long questionnaire (N1), a person younger than 15 is also not required to respond to portions of the questionnaire, including questions on education, activities in the labour market and income.

43. Upon entering the date of birth, the system calculates the age of each respondent as of Census Day (May 10, 2011). A page called **age confirmation** gives respondents the opportunity to verify the ages of each household member. With the online questionnaire, if a date of birth is not indicated, then the message "*Date of birth not indicated*" appears under the appropriate person's name; if it is invalid, then the message "*Unable to calculate the age*" appears. Respondents can go back to the previous page to modify their response to the date-of-birth question or continue to the next page. If they opt to move without providing a valid date of birth, then the person is presumed to be 15 or older and all applicable questions on the long form are asked.

44. Respondents who do not give Statistics Canada permission to use their income tax file must respond to a 14-part question. A **summary of income** page gives respondents the opportunity to review the amount reported, to go back to the previous page if they deem it necessary, and rectify erroneous responses. The page shows each income source, total income and income tax paid for each person.

(c) Editing and messages used in the online questionnaires

45. The online questionnaires have four levels of edit message. When more than one problem is detected for a question, the message that will be presented used this priority: the non-response, the partial response, the invalid response and the new editing messages.

1. Non-response edits

46. The edit for non-response is the most frequent one; they are performed on almost each question. This edit is the most useful one. There are two types of non-response messages in the online questionnaires. One applies to the household questions, such as "*Please answer Step B3 (a).*" The other applies to the person questions, such as "*Please complete Question 2 for John Doe.*"

47. Past experience shows that the non-response messages are effective in obtaining answers to questions that respondents might have overlooked and directing respondents to correct inadvertent errors. As already mentioned, however, these messages also lead respondents to believe that they must provide an answer to every single question in order to be able to continue. This false impression might sometimes result in undesirable behaviours that impact data quality. The 2006 Census evaluation identified an issue of this nature in some answers that required a postal code for the mobility questions. Remembering a postal code from a previous address where you lived one or five years ago may be difficult. Some respondents reported that they lived in a different city but entered their current postal code instead of leaving the answer blank, as they would probably do with a paper questionnaire. This behaviour was confirmed during subsequent qualitative testing, and the problem was corrected during the coding operation. For 2011, the non-response messages associated with postal code have been disabled to discourage this undesirable behaviour and to avoid getting data that might be valid but incorrect. The verification is now done only if six digits are entered in the two parts of the postal code for answers that are not in the format "ANA NAN" where A is a letter and N is a single-digit number.

2. Partial response edits

48. In the 2006 Census, partial response edits were used in two types of questions. The first type was questions that asked for address which contain multi-part answers (street, city, province or territory and postal code), for example, in Question 46, place of work: *"Please enter Number and street address for Jane Doe."*

49. The second type are questions that contain radio buttons associated with a write-in response such as "other – specify." In 2006, JavaScript had to be enabled in order to use the Internet application. When the radio button was selected, the cursor moved automatically to the write-in field. Inversely, when the respondent started entering a response in a write-in field, then the corresponding radio button was selected automatically. This resulted in having only one possible way for the answer to be partly completed (the radio button "other" selected without a write-in answer).

50. With the new CLF standards, the online questionnaire must also still function if JavaScript is turned off or disabled. This introduces the possibility of multiple or inconsistent responses, which was not an issue in 2006. An inconsistent partial response could result when respondents provide a write-in without selecting the corresponding radio button. Consequently, messages for these types of situations had to be developed. Instead of implementing another type of partial non-response, we modified the existing one to cover the new possibility. These messages were changed to a more generic one such as: "*The answer to Question 16 for Jane Doe is incomplete. Please ensure that "Other" is selected and a response is entered in the "Specify" field*". We also had to customize these new messages because different labels are associated with radio button or check boxes. For example, in addition to the other – specify types of

answer, we also have Yes–specify, Yes, other–specify, Born outside Canada–specify, Outside Canada–specify, etc. Even if we expanded the partial non-response messages, we expect the impact to be minimal: in 2006, we estimated that only 5% of the Internet users have JavaScript disabled on their browsers. This solution might not have been the optimal one—we always try to develop precise messages—but schedules were too tight to develop and test other new messages for inconsistent and partial responses.

3. Invalid response edits

51. In 2006, invalid response edits were used only for postal code and numeric fields, including dollar-related questions. An invalid response message appeared for numeric fields when respondents entered a number outside the range established for a question. For example, if a respondent entered "60" as the number of weeks worked in the previous year, the following message was presented: "*You should enter a number between 0 and 52. Please re-enter the number for John Doe.*" For income question, the message was more specific, for example: *"You should enter an amount between 1 and 99,999,999 in the Total Income" field. Please re-enter the amount. If you entered "0", please select "No."*

52. In 2006, when comparing the distribution of responses between the Internet and paper, a mode effect was found with the number of hours worked in the week preceding the Census. Although the numbers were quite small, we observed that online respondents were more likely to report "1" hour worked in comparison to paper respondents. Respondents who did not work had to check a checkbox labelled "None" located below the numerical field for the number of hours. Our experience with the paper questionnaire has shown that sometimes people did not see the check box and entered "0" in the numeric field instead of checking the "None" box. Internet respondents received an invalid response message, because the valid range was between 1 and 168, the maximum number of hours in a week. Consequently, because they felt they had to correct their answer, they changed their answer to "1" hour, as the message requested. To remedy this, after the 2006 Census all numeric fields were modified to accept a "0" value in the text field. The corresponding messages were modified and the questionnaire flow (automated skip), when applicable, was also corrected to present the next appropriate question.

53. Again, because the online questionnaire must work with JavaScript turned off, we had to develop another type of invalid message for a situation that did not exist before: invalid multiple responses (questions with checkboxes used with a write-in field, for which only some combinations of multiple responses are acceptable). For example, Question 18–Aboriginal Identity, the answers Yes, First Nations, Yes, Métis and Yes, Inuit are possible but not in combination with No, not an aboriginal. In this case, the editing message would be: "*Please correct the answer to Question 18 for John Doe, "No" and "Yes" responses should not be selected together.*"

4. New edits

54. Three types of new edits were developed for the 2011 Census to access the feasibility and the acceptability of more complex edits.

i) Consistency between two questions

55. This edit was developed to address inconsistency between two questions: the year of birth and the year of immigration. In 2006, some immigrants reported a year of immigration occurring before the year of birth. Analysis showed the response error occurred for all immigrants, not only for recent ones. The message is as follows: "*The year reported for John Doe in Question 12 is earlier than his or her year of birth. Please verify your answer.*"

ii) Looking for a more accurate response

56. These edits were developed to seek more precision to the industry and occupation questions. Answers such as "construction", "engineer" and "technician" are not precise enough: more specific information is needed to be able to code to the corresponding 4-digit North American Industry Classification System code. The answers to these two questions were edited using a look-up table containing different spellings of the same word as well as the English and French version of each word. This edit is intended to enhance data specificity and quality. For example, this message for Question 41, Industry when "construction" is given as the answer is: "*Please specify the kind of construction business (if applicable) in Question 41 for John Doe. For example, road and highway construction, pipeline construction, condominium construction, drywall installation contractor, plumbing contractor, excavation contractor, etc*." The message for Q42, Occupation when engineer given as an answer is: "*Please be more specific (if applicable) in Question 42 for Jane Doe. For example, chemical engineer, mechanical engineer, civil engineering technologist, electronics engineering technician, stationary engineer, 2nd class power engineer, etc."*

iii) Avoiding double reporting of amounts

57. This edit is intended to address potential double-counting between parents for the question about child care expenses. The question is new for 2011, but was tested in the 2009 Census Test. The results showed that 82% of households who reported multiple amounts (for example, an amount for each of the two parents), reported duplicate amounts. This edit is expected to reduce data inconsistency and enhance data accuracy. The message is as follows: "*The same amount has been entered in Question 52 for more than one person. If each amount represents an equal share of the total for your household, simply leave as is. If not, please modify the answers to avoid double-counting.*"

IV. Conclusion

58. In summary, data collection via the Internet is becoming more and more popular. Many census developers as well as business and social researchers are looking at ways to incorporate the Internet in their data-collection modes. In the near future, the Internet will probably become the primary collection mode for the Census of Population and perhaps, for other Statistics Canada surveys.

59. The Internet option provides respondents more flexibility on when and how to complete their questionnaire. This flexibility may encourage some of them to participate in a survey. Our experience indicates that Internet users expect a positive and interactive experience when using the Web to complete a census questionnaire. They also expect the online questionnaire to be easy and quick to complete. They are generally receptive to messages. As shown in this report, through the evolution of the census edit messages, we attempted to keep a balance between the number of edit messages presented to respondents and the necessity of having respondents correct an error. We also saw that, in some circumstances, it's better to have no-response to a question rather than a valid but incorrect answer, while in other circumstances, a respondent's best estimate is more desirable. In the 2006 Census, we were able to detect cases where respondents provided a valid but wrong answer for specific questions. However, we do not want to encourage this behaviour: response errors are very hard to detect. Nevertheless, the edit messages are effective and they, along with the automated skips and the radio buttons, collectively increase the quality of census data.

60. Typically, respondents don't see each edit message. When respondents see too many messages, the edit rules are probably too strict, or some questions might have problems. The long questionnaire asks questions on a vast array of subjects. Each subject has a limited number of questions. In addition, related questions might be far from each other—for example, Question 3, date of birth and Question 30, university degree. Therefore, this particular context limits the actual level of editing in the census and

especially the National Household Survey, given its voluntary status. Within the next year, we will explore further the quality and the reasons for differences among the data collection modes, and we will have more opportunities to assess data validity. Mode effects in the census are quite sensitive. It is crucial to choose wording that transmits to respondents the same meaning across modes.

A. Analysis for 2011

61. To access the efficiency and effectiveness of the 2011 Census and NHS edit messages, we created a tool that will record, for each question, which edit messages were presented to respondents. Up to four messages will be recorded. For example, '130' will indicate that the respondent originally did not respond to the question ('1'); following the non-response message, the respondent supplied an invalid entry ('3'); and following the invalid message, the respondent provided a valid response ('0'). This tool will enable us to assess the impact of each edit message for all questions.

B. Future plans

62. Statistics Canada approved a new 'Corporate Business Architecture' project in April 2010. This project focuses on two electronic services, the e-file transfer service and the e-questionnaire service. The goal of the latter is to develop and implement e-questionnaires as the primary collection mode for the majority of Statistics Canada surveys. Over 130 surveys have been identified, including the 2016 Census. Thus, e-questionnaires will become the first step in a sequential multi-mode collection approach. The key features of the new e-questionnaire will be ease of use for respondents, an absence of footprint (nothing left on the respondent's computer), a secure solution (industry-standard encryption algorithms), flexibility, scalability, accessibility as well as compliance with Common Look and Feel.

63. By 2015-2016, the online census questionnaires will be part of this project. A committee is now looking at the editing strategy.

V. References

Grondin, Chantal and Limei Sun. 2008. "2006 Census Internet Mode Effect Study". *Proceedings of the Joint Statistical Meeting 2008*.

Karaganis, Milana. 2011. "Development and Implementation of E-questionnaires as a Primary Collection Mode at Statistics Canada". Internal document. Statistics Canada.

Laroche, Danielle. 2007. "2006 Census Internet Data Collection". Technical report presented at Statistics Canada's Advisory Committee on Statistical Methods, April 30th - May 1st, 2007.

Laroche, Danielle and Chantal Grondin. 2008. "Impact of Online Edits and Internet Features in the 2006 Canadian Census". United Nations Statistical Commission and Economic Commission for Europe. Work Session on Statistical Data Editing. April 2008.

Mathieu, Patrice and Jean-Pierre Morin. 2010. "National Household Survey, Non-response Follow-up". Technical Report presented at Statistics Canada's Advisory Committee on Statistical Methods. October 2010.

Moore, Tracy, Lorna Bailie and Graeme Gilmour. 2009. "Building a Business Case for Census Internet Data Collection". Proceedings of Statistics Canada Symposium 2008.

Rodrigre, Jean-François and Jennifer Taylor. 2010. “2009 Census Test: Overview of the Results and Findings from the Primary and Supplementary Samples”. Internal presentation at Statistics Canada’s Data Quality Review Board, April 14, 2010.

Taylor, Jennifer. 2009. “2009 Census Test Supplementary Study of the New Census Collection Methodology”. Proceedings of Statistics Canada Symposium 2010