

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Ottawa, Canada, 16-18 May 2005)

Topic (iii): Electronic data reporting – editing nearer source and multimode collections

EVALUATION REPORT ON INTERNET OPTION OF 2004 CENSUS TEST
**CHARACTERISTICS OF ELECTRONIC QUESTIONNAIRES, NON-RESPONSE
RATES, FOLLOW-UP RATES AND QUALITATIVE STUDIES**

Supporting Paper

Submitted by Statistics Canada¹

1. Introduction

By law, Statistics Canada (STC) is required to conduct a census every five years, and all households are required to complete a census questionnaire. The last census took place in 2001, while the next is scheduled for May 16, 2006. The *de jure* method is followed, meaning that people are counted at their usual residence in Canada regardless of their location on Census Day. In 2001, Canada had nearly 12 million households and a population of more than 30 million. Approximately 98% of households were counted using the self-enumeration method and the remaining 2% counted via interview.

Major changes are being implemented for the 2006 Census in relation to data collection and processing methods. These changes are primarily as follows: questionnaires will be mailed out to two-thirds of dwellings, with enumerators delivering questionnaires to remaining dwellings as in the past; completed questionnaires will be returned to a single processing centre rather than to enumerators; questionnaires returned via the mail will be scanned and their data captured automatically; telephone follow-up for incomplete questionnaires will be conducted from the Census Help Line sites using a Computer Assisted Telephone Interview (CATI) application; and, last but not least, all households in private dwellings will have the option to complete and submit their questionnaire via the Internet.

Use of the Internet in data collection is not new to Statistics Canada; several business surveys have been conducted using this collection method. However, the Internet option for household surveys is relatively recent. That is, it has only been utilized in the Census, first with a major test during the 2001 Census, then as part of the Census Test in May of 2004 when Statistics Canada carried out the testing of almost all systems and operations to be used during the 2006 Census.

As for the Census, two types of questionnaires were used to collect the majority of the Census Test data. The short-form questionnaire, or form 2A, was distributed to four households out of every five. The long-form questionnaire, referred to as form 2B, was distributed to one of every five households. This questionnaire contains all of the questions appearing on form 2A plus questions on a range of various

¹ Prepared by Danielle Laroche, Social Survey Methods Division (Danielle.laroche@statcan.ca).

topics. The 2004 Census Test was conducted within a limited number of test regions in Nova Scotia, Quebec, Manitoba and Saskatchewan. These sectors were selected based on their socioeconomic characteristics, proportion of francophone and anglophone households and availability of farms with a view to establishing a sample of both mail-out and list/leave areas.

This report describes the characteristics of the electronic questionnaires, presents the main results of the qualitative tests and summarizes the analyses of partial non-response and failed-edit follow-up rates for the 2004 Census test questionnaires. Assessment of the preliminary quality of the data collected consisted of analysis of partial non-response rates and the percentage of questionnaires sent to failed-edit follow-up. These two indicators are key factors in determining the initial data quality that can reveal problems related to specific questions.

2. Characteristics of Electronic Questionnaires

In general, the electronic questionnaire is identical to the paper questionnaire in terms of question wording, instructions and response options. In terms of functionality, respondents have the option to save an Internet 2B questionnaire and complete it over multiple sessions from different computers. Respondents can also switch the questionnaire language, i.e., French or English. Additionally, the Internet application was made to be as consistent as possible with standards and guidelines for presentation of federal government Web sites. All Web pages of the Government of Canada strive to have similar characteristics, and the electronic questionnaires follow the same straightforward, efficient format. The Government of Canada places great emphasis on the uniform presentation of its Web sites, which provides the benefit of ensuring users a consistent experience visit after visit. The use of strict standards also enhances the satisfaction level of respondents using government sites.

The first screen is a **welcome page** giving users the opportunity to verify their computer requirements and settings. If respondents have an adequately recent and properly configured browser, then the next screen is displayed. If not, then respondents can click links to view troubleshooting information to help them modify their browser settings or download a more recent version of the required browser or Java Virtual Machine (JVM).

The second screen is the **access code page**. Access codes are printed on the front page of the paper questionnaires. They are unique, randomly generated, 15 digit numbers segmented into five groups of three digits to make them more user-friendly. After entering their unique access code, respondents click the *Start* button at the bottom of the page to validate their code. If the code is valid, then the application automatically selects the appropriate questionnaire type in the appropriate language and displays the first page of the questionnaire.

In the first part of the census questionnaire, respondents are asked to provide their telephone number and household **address**. If the household received its questionnaire via the mail, then the access code is associated with the address in the Master Control System (i.e. master list of dwellings which includes addresses used for mail-out), and the address is displayed for the respondent to confirm. The paper and electronic questionnaires are identical in this regard. If the dwelling is in a list/leave area where the questionnaires are delivered by enumerators, the respondent is required to provide the household address. The Province field has a drop-down list containing the possible response options. Respondents simply click this field to view a list of provinces and territories.

In the electronic questionnaire, all fields for province selection use this **drop-down list** format. The month and day of the date of birth question follow the same format.

A bar representing respondents' **completion status** or progress in the questionnaire appears in the left-hand column on all screens. This feature provides respondents an indication of how much of the survey they have completed and how much remains.

A **help function** appears in the form of a link in the left-hand column on the screen under the completion status bar. Respondents can click the help link for assistance with the current question. The help function contains instructions and examples to help respondents ensure that their responses are as accurate as possible to all questions. In addition to the help link, the left-hand column is configured with explanations for respondents as to **why they are being asked each question** and how the information they provide will be used. This information has been found to enrich the user experience when completing the online census.

Questions and response options appear in a box at the centre of the screen to make them stand out. Each question is displayed on a background colour corresponding to the colours used in the paper questionnaire.

Internet **standards** are followed, including **check boxes** to indicate that multiple responses are possible and circles, or **radio buttons**, to indicate that only one response is possible. When a radio button associated with a write-in response is selected, the cursor moves automatically to the field in which respondents are expected to type their answers. If respondents simply start typing a response in a write-in field, then the corresponding radio button is selected automatically.

Respondents navigate within the questionnaire using **control buttons** located at the bottom of each screen. When respondents click the *Next* button, the data are encrypted and sent to Statistics Canada's secure server, where they are then decrypted and verified. If no problem is detected with the answers, the server sends to the respondent's browser the next appropriate screen. If there is a problem with the respondent's answers, the server encrypts the current screen's data, sends it back to the respondent's browser where it is decrypted and displayed with a validation message. This process is called two-way encryption. That is, data entered by respondents are encrypted by their browser and sent to the server where they are decrypted. Data coming from the server are encrypted and sent to the respondent's browser where they are decrypted. Respondents may also return to the previous screen by clicking the *Go Back* button. The *Stop and Finish Later* button gives respondents the option to save a partially completed questionnaire and fill in the remainder later. After clicking this button, respondents are prompted to choose a password or to let the application assign one. When they return to finish their questionnaire, they are prompted to enter their original access code and then given five attempts to enter their password correctly. For security purposes, if respondents are unsuccessful or if they do not log back in within the prescribed period, then their partially completed questionnaire is submitted on their behalf. Finally, respondents can click the *Cancel* button to simply exit a session. No data are saved if they exit in this manner.

Four types of **validation messages** are possible. **Non-response** messages appear when respondents have not answered a question. **Partial response** messages appear when respondents provide only a partial response to a question, for example, if they omit the city name from their address. **Invalid response** messages appear for numerical responses when respondents enter a number outside of the range established for a question. Finally, **soft edit** messages appear only for questions relating to money amounts whenever the amount in a response appears unusual. This type of message asks respondents to verify that they have entered the correct amount, for example, "*Please verify the amount you entered for part (f), if correct leave as is*". All of these messages follow the same approach. When respondents click the *Next* button, the information on the current page is validated, and, if necessary, the application displays the same screen again noting any problems at the top of the page in red text, for example, "*Please answer question 5 about John Doe.*" The question and field requiring attention appear in red, and a red arrow highlights the missing response to assist the respondent, who can then either fill in the missing information or continue to the next screen. If the respondent chooses to move on without making any changes, then the next screen is presented. If the respondent adds or changes any information, then the responses are validated again. This approach is consistent with the Common Look and Feel guidelines

prescribed for Canadian government Web sites in that pop-up windows should not be used within pages to convey information to respondents.

The electronic questionnaire follows primarily the **matrix format** but also, in places, the sequential format. With the matrix format, each question appears only once, and response options are repeated under the name of each person in the household. Usability tests have demonstrated that this format reduces the response burden, since respondents have to read each question only once and can then respond for all members of the household. Another advantage of the matrix approach is that it reduces the number of screens and, as a result, requirements with regard to system infrastructure. With the **sequential format**, questions are asked about one person at a time. As a result, questions are repeated as many times as there are persons in the household. The sequential format supports increased customization of questionnaires. For one, it allows a respondent's name to be directly incorporated into each question. The sequential format is used in two places on the electronic version of form 2B: questions 40 to 46 concerning labour market activities and question 52 on income. Usability tests have indicated that it is easier for respondents to focus on one person at a time in responding to these particular questions.

The electronic version of form 2B has two types of automated **skip patterns**. The first relates to all members of the household. With this skip type, any questions deemed non-applicable are not displayed. However, since the questions are numbered, a message appears at the top of the screen for the next applicable question to indicate to the respondent that one or more non-applicable questions were skipped. In this situation, a skip message like this would be presented: *"Based on your response to Question 11, Question 12 is not applicable. Therefore, proceed with Question 13"*. The second type of automated skip relates to questions applicable to one or more persons but not to all persons in the household. With this skip type, the question must be displayed since it applies to some members of the household. In this event, a message appears under the names of persons as appropriate to indicate that the question does not apply to them. By skipping any questions deemed non-applicable, we hope to reduce the response burden, thereby making the user experience more pleasant and less frustrating in comparison to the paper questionnaire. For example, with paper questionnaires respondents sometimes do not follow skip instructions which results in a considerable increase in response burden.

The electronic questionnaire has two **mandatory questions**, one on the number of persons staying at the address on Census Day (Step B1), the other on the names of the members of the household (Step B2). The second screen depends on the first; for example, if a respondent indicates "3" for the number of persons in the household, then the application generates three lines to type the names of these persons. The respondent must provide a first or a last name for each person. These names are subsequently used to customize responses and selected questions. These two questions are the only mandatory questions in the Internet questionnaire.

One of the main **differences** between paper questionnaires and their electronic versions relates to the population coverage Steps questions. In the electronic questionnaire, respondents are required to enter the names of all persons staying at their address on Census Day and three additional questions are then used to eliminate any temporary residents (persons whose usual residence is at another address in Canada) or foreign residents (visitors or government representatives from another country). On the paper questionnaire, respondents list only the names of usual residents at their address. The three questions of the electronic questionnaire are presented in the form of instructions on the paper questionnaire. In either response mode, if all residents at an address are temporary or foreign, then respondents (i.e. the persons filling out the questionnaire) list their name and usual telephone number, and there is no need to respond to any further questions. The application can accommodate up to 36 persons, although it is possible to indicate more persons in the corresponding question.

Age is particularly important in the long-form questionnaire (2B), as persons less than 15 years of age are subject to different validation for the questions on marital or common-law status and are not required to respond to a portion of the questionnaire, including questions on mobility status, education, activities in the labour market and income. **Age confirmation** provides respondents the opportunity to verify the ages calculated based on the responses given to the question on date of birth. This confirmation occurs only

for the electronic questionnaire. If a date of birth is not indicated, then the message "*Date of birth not indicated*" appears next to the appropriate person's name; if it is invalid, then the message "*Impossible to calculate age*" appears. Respondents can go back to the previous screen to modify their response to the date of birth question or go on to the next screen. If they opt to move on or not to provide a date of birth, then the person in question is presumed to be more than 15 years of age, and all applicable questions on the form 2B are asked.

On the last page of the electronic questionnaire, the completion status bar indicates that the questionnaire is complete. Respondents then have the option to document any suggestions or **comments** in the designated space. The **Submit** button is located at the bottom of the page. When respondents click this button, the application submits the questionnaire and displays an acknowledgment page containing a **confirmation number**. Respondents can click a button at the bottom of this page to print the number to retain as evidence that they have submitted their questionnaire in case an enumerator telephones or knocks at their door.

3. Results of Qualitative Studies (Internet Respondents)

Qualitative studies of the electronic version of form 2B were carried out in respondents' homes in May and June of 2004, in parallel with the 2004 Census test. Approximately 50 interviews were conducted in both official languages.

Most respondents stated that they enjoyed using the application and that they found their experience to be positive and pleasant. They found the electronic questionnaire easy to complete, efficient and user-friendly. Some respondents stated that they appreciated not having to fill out the form manually, while others said it was easier to correct mistakes on the electronic copy. The majority of respondents liked the interactive nature of the application and its "smart" features (personalized questions and answers with the respondents' names, age confirmation, automated skips, etc.). Moreover, the collection method saved respondents the trouble of going to the post office or finding a mailbox.

One of the problems found was not related to the questionnaire itself but rather to the process for gaining access to it. This is because respondents who had to modify their computer settings were generally unable to do so without assistance from an interviewer. The majority of respondents did not know what type or version of browser they had or how to find this information. This could mean lower response rates at census time as a result of unsuccessful attempts to gain access to the electronic questionnaire.

During qualitative tests, it seemed to us that the 2B Internet questionnaire did not offer respondents significant time savings compared to paper, and may in fact have taken longer in some cases. However, data collected by the Internet application, that is, from the time each respondent logged-in to the time the respondent submitted their form, did not confirm this. In fact, the electronic questionnaires take less time to complete than the paper version, except in cases where Internet respondents do not use a high-speed connection. In addition, most Internet respondents, even those with "dial up" connections, reported that they thought the electronic version was faster to use and much more efficient than the paper version. This perception was likely due in part to the interactive aspects of the application and to the fact that after respondents complete the electronic questionnaire, they can submit it instantly as opposed to the paper version, which must be mailed.

Validation messages, particularly those for non-response or partial response, were effective in that they helped to ensure that respondents provided responses for questions they had inadvertently skipped. In spite of this, respondents found the messages to have a negative connotation in that they only appeared when respondents did something "wrong." In addition, generally held conceptions of Internet respondents with regard to how forms on the Internet should be completed can lead them to believe that they must provide a response to every single question, without exception, before moving on. For the Census, this false impression could result in undesirable behaviours that might impact data quality. For example, some

respondents might invent responses or pick them at random in order to move on. An issue of this nature may be difficult to detect.

4. Comparison of Partial Non-Response Rates for Internet and Paper Questionnaires (Unweighted)

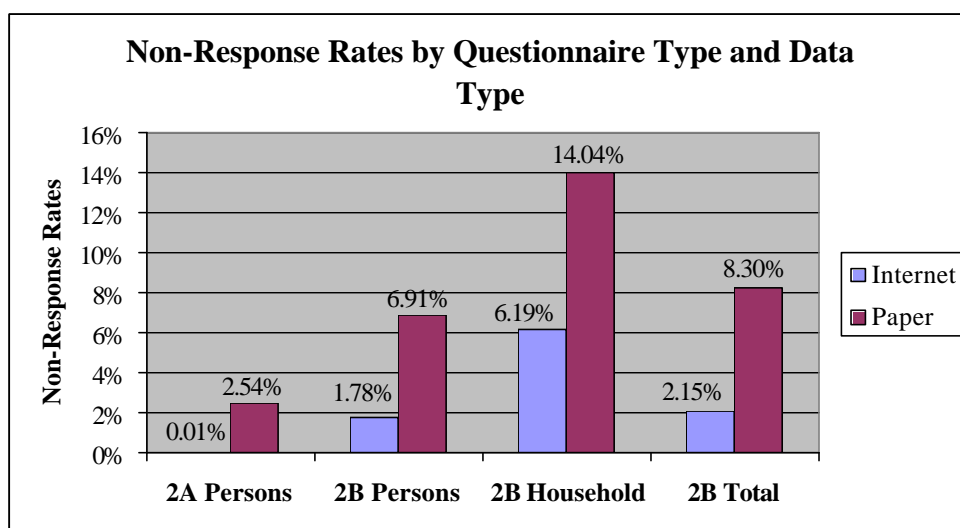
The rate of partial non-response to questions can provide an indication of the difficulties encountered by respondents. It is useful in assessing the overall level of understanding of questions and provides a preliminary indication of data quality. A question is generally deemed to be missing a response when a response is required but none has been provided. However, the situation may not always be as obvious. For example, if question 11 (landed immigrant) has no response, it is impossible to know whether a respondent was or was not supposed to respond to question 12 (year of immigration). In this event, it is presumed that the respondent was not supposed to respond to the question. As such, the rate of partial non-response studied in this section corresponds to actual rates in that only responses for which we were certain there should have been a response were taken into consideration. For some questions, the rates also include incomplete and invalid responses. These rates are unweighted and were calculated based on all non-blank questionnaires returned.

Wherever possible, the Internet questionnaire mirrored the paper form in terms of question wording, instructions provided and response options. However, in comparing non-response rates, certain characteristics of the electronic questionnaires need to be taken into account. Firstly, the use of radio buttons in the response options for certain questions makes it impossible to select more than one response for these questions. Secondly, when respondents click "Continue" at the bottom of each page, any appropriate non-response, partial response, invalid response (for numerical responses) messages or soft-edit messages (for amounts) are displayed automatically. Respondents can then choose to enter or correct a response or to simply move on. In addition, all skips are automated.

Form 2A is divided into two parts, coverage steps and persons data. **Form 2B** has three parts: coverage steps, person data and household data. In terms of coverage steps, we did not evaluate non-response but rather only situations requiring confirmation from a household. These results appear in the section on rejection rates.

The following chart illustrates the non-response rates calculated for both collection methods by questionnaire type (i.e. 2A or 2B) and data type (i.e. person data or household data).

Chart 1: Partial Non-Response Rates by Questionnaire Type and Data Type



As the chart indicates, the non-response rate was almost nil for the Internet version of form 2A (0.01%), while the same rate was 2.54% for the paper questionnaire. For form 2B, the non-response rate for person data was four times as high for the paper questionnaire (6.91%) as for the Internet version (1.78%). With regard to household data, the non-response rate for the paper version of form 2B was twice as high (14.04%) as for the Internet version (6.19%). When all questions are considered together, the rate of partial non-response of 2B was nearly four times as high for the paper questionnaire (8.30%) as for the Internet version (2.15%).

It appears, therefore, that the validation messages used for the Internet version are effective in reducing non-response. Qualitative studies appear to confirm this hypothesis. However, these studies have also revealed that many Internet respondents believe that they must provide a response to every single question before moving on. For some respondents, non-response messages are interpreted as a confirmation of this. Nevertheless, it would appear that the non-response messages assist inattentive respondents in correcting the majority of unintentional errors or omissions.

Analysis of rates of partial non-response for each question indicates that some question types are more likely to have non-response than others for either collection method. These are generally questions that respondents deem non-applicable (for example, question 5 on common-law status for married respondents or young children, question 8B on activity limitations at work or at school for respondents not working or attending school, question 30 on major field of study for respondents without certificate or for those who only have a secondary certificate, question 33 on the number of hours spent caring for children for respondents without children or question 39 on the last date worked for respondents who are not working or have never worked). Other questions, meanwhile, are more difficult to respond to as a proxy respondent, i.e., on behalf of another person (for example, question 17 on ethnic origin, question 43 on the main activities performed at a person's work or question 46 on place of work). Finally, still other questions are simply more difficult to respond to in that they depend on memory recall (for example, question 24 on mobility status over the previous five years or question H5 on period of construction of dwelling) or require retrieval of documentation (for example, question 52 on income or the questions in section H6 on amounts paid for electricity, oil, gas, wood, water or other municipal services) or in that respondents do not know or refuse to provide the answer (for example, question 52 on income or on income tax paid). In most cases, the types of questions listed above tend to have higher non-response rates for both collection methods. However, these rates are consistently twice as high for paper as for electronic questionnaires.

According to this analysis, non-response rates also increase as we progress on the questionnaire regardless of collection method. Internet questionnaires include both questionnaires submitted by respondents and those submitted automatically by the system (that is when a respondent saves a questionnaire and then does not return within the prescribed time, the system submits the questionnaire on the respondent's behalf). In addition, respondents who did not respond to the question on date of birth were required to respond to all questions on form 2B (although this proportion was negligible at 15 cases out of 7,526, or 0.02%). This explains in part why non-response rates tend to increase as we progress in an Internet questionnaire. Additional factors to explain this increase might include increasing question difficulty, increasing effort required or respondent fatigue.

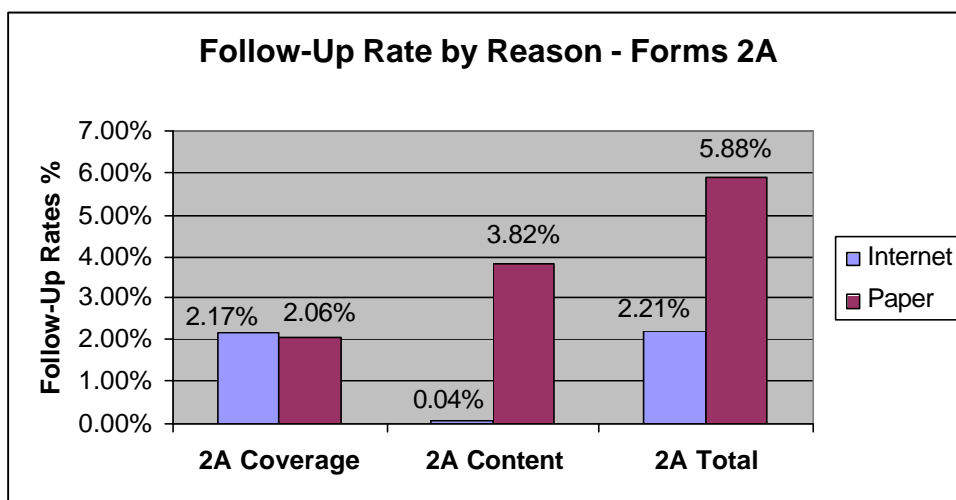
5. Comparison of Follow-Up Rates for Internet and Paper Questionnaires (Unweighted)

After data entry, questionnaires are transmitted to data processing where automated edits are performed. These edits identify any questions requiring follow-up and compile a score for each household. Follow-up requirements are calculated based on these scores and weights assigned to each question. The data for households with a score exceeding a predefined value are forwarded to follow-up. These households are grouped to classify follow-ups by priority; those with the highest scores are given higher priority in terms of number of contact attempts. In other words, the higher a household's score, the higher its priority for follow-up.

In terms of coverage steps, we do not evaluate non-response but rather the situations requiring verification with households. The first situation in this regard concerns households identifying themselves as temporary or foreign resident households (Step B). It is necessary to confirm that these households actually do include only temporary or foreign residents. The second situation relates to doubts concerning the exclusion of a person from the questionnaire (Step C). In the event of doubt concerning whether to exclude a person from a questionnaire, respondents are supposed to indicate the person's name, the reason for exclusion and the person's relationship with Person 1. It might be necessary to contact the household to confirm that the person is, in fact, not a usual resident at the address. Should one of these situations arise that cannot be resolved through analysis of the questionnaire, then verification must take place with the household.

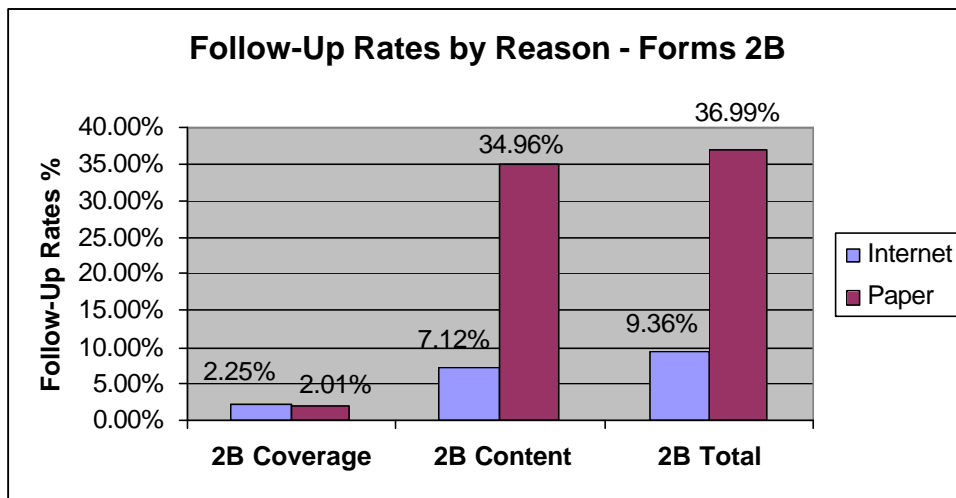
The following chart illustrates the follow-up rates by reason for forms 2A.

Chart 2: Follow-Up Rates by Reason – Forms 2A



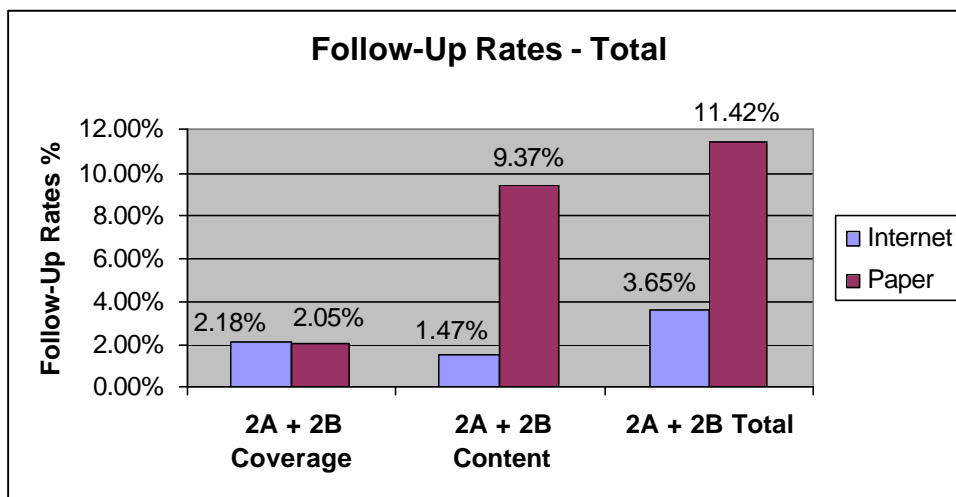
In terms of coverage steps, the Internet and paper versions of form 2A rank more or less equivalent (2.17% and 2.06%). For either collection method, nearly three-quarters of all coverage issues relate to Step C. In terms of content, on the other hand, follow-up rates are negligible for the electronic version of form 2A (0.04%) in comparison to the paper version (3.82%). Overall, the paper versions of form 2A require follow-up three times as often as the Internet versions (5.88% versus 2.21% respectively).

The following chart illustrates the follow-up rates by reason for forms 2B.

Chart 3: Follow-Up Rates by Reason – Forms 2B

With regard to coverage steps, follow-up rates for form 2B are similar for either collection method (2.25% for electronic versions, 2.01% for paper). These rates are also similar to those for form 2A, with the majority of coverage issues relating to Step C for either collection method. With regard to content, the follow-up rate was nearly five times as high for the paper questionnaire (34.96%) as for the Internet version (7.12%). Overall, the follow-up rate was nearly four times as high for the paper questionnaire (36.99%) as for the electronic version (9.36%).

The following chart illustrates the follow-up rates by reason for all questionnaires combined.

Chart 4: Follow-Up Rates by Reason – Forms 2A and 2B Combined

In terms of overall coverage steps, follow-up rates were similar for the electronic and paper versions (2.18% and 2.05% respectively). Follow-up rates relating to content were six times higher for paper questionnaires than for their electronic versions, at 9.37% and 1.47% respectively. Overall, the follow-up rate was three times as high for the paper questionnaire (11.42%) as for the electronic version (3.65%).

The analysis of non-response and follow-up rates leads to the conclusion that the electronic questionnaires are more complete than the paper questionnaires and consequently less expensive in terms of follow-up than the paper questionnaires.

6. Conclusion

Data collection via the Internet offers a range of both new possibilities and new challenges. As demonstrated in this report, partial non-response rates and follow-up rates (i.e. failed edit questionnaires) are much lower for electronic versions of questionnaires than for their paper versions. With regard to the validation messages, it is possible to conclude that in general, they are effective in obtaining answers to questions respondents might otherwise have overlooked and in having them correct errors they inadvertently committed. Along with the automated skips of non-applicable questions, these messages result in a general perception among respondents that the electronic questionnaire is “smart”. This responds in part to the high expectations the general public have with regard to Internet questionnaires.

The results provided in this report are thus highly encouraging. They demonstrate that data collected via the Internet are more complete and consequently less expensive from a failed edit and follow-up perspective than data collected via paper questionnaires. Moreover, electronic questionnaires are ready for processing upon submission since they are already in electronic form. A number of countries have invested significant financial and other resources into this new collection method, as the new technology offers hope in terms of reducing the cost of future censuses if we obtain a high take-up rate using this methodology. The infrastructure for collecting data via paper will always be required, at least for a Census. With an on-line option, organizations must invest additional funds for a second infrastructure. The saving can only be realized if the Internet take-up rate is high enough to reduce the paper processing infrastructure and eliminate operations costs related to processing paper questionnaires.

However, it is also appropriate to explore more fully the reasons for the differences in non-response and follow-up rates between these two data collection methods. Do the differences in format between electronic and paper questionnaires influence the quality of the responses collected? Are data collected from households via the Internet and via paper questionnaires strictly comparable? Is the quality of responses equivalent from one collection method to the next and do we need to implement additional processing steps to control for the differences? More detailed studies are necessary and are currently underway with a view to finding the answers to these questions.

References

Laroche, Danielle (2004) *Census of Population Data Collection Via Internet*. Statistics Canada, SSC Montréal, May 31, 2004.

Boudreau, J.-R. and Bornais K. (2003) *Les conclusions de l'étude du contrôle et du suivi*. Statistics Canada, 20 février 2003.

Statistics Canada, 2006 Census Processing *Response Integration and Verification Task – Coverage and Collective Edit Systems Requirements*. Version 9 – February, 2005.

Statistic Canada, Bornais, K. (2004) *Rates by response Channel*. Internet questionnaires derived from Booklet Id. E-mail October 21, 2004.
