



Economic and Social Council

Distr.: General
23 April 2010
English
Original: French

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Thirteenth Meeting

Geneva, 7–9 July 2010

Provisional agenda item 3

Plans for census quality evaluation

Quality Assurance Programme of the 2011 Canadian Census of Population

Note by Statistics Canada

Summary

This article outlines the quality assurance programme of the 2011 Canadian Census of Population and gives an overview of the main changes introduced for the 2011 Census.

Conducted every five years, the Canadian Census of Population is a major undertaking whose planning and implementation spans a period of over eight years. The census is unique because it is the only source of detailed socio-economic and demographic data on small geographical areas including neighbourhoods and communities.

In order to ensure that the population and dwelling counts and the data on the characteristics of the population and dwellings are of sufficiently high quality, a diversified quality assurance programme has been developed over the last few censuses. During these censuses, changes and additions were made to the quality assurance programme in the light of operational, technological and methodological changes as well as a changing reality.

1. At its meeting in Washington, D.C. (United States) on 19 and 20 October 2006, the Bureau of the Conference of European Statisticians (CES) adopted the renewed terms of reference of the Steering Group on Population and Housing Censuses and the plan of future CES activities in the field of population and housing censuses. The Bureau of the CES also confirmed that the Steering Group would coordinate the work associated with the various meetings.
2. The present document has been drawn up at the request of the Steering Group on Population and Housing Censuses for presentation and discussion at the ECE/Eurostat Meeting on Population and Housing Censuses to be held in Geneva from 7 to 9 July 2010.

I. Introduction

3. Conducted every five years, the Canadian Census of Population is a major undertaking whose planning and implementation spans a period of over eight years. The census is unique because it is the only source of detailed socio-economic and demographic data on small geographical areas including neighbourhoods and communities. Most households in Canada (80 per cent) receive a short census questionnaire, which contains seven questions on basic topics such as household and family relationships, age, sex, marital status, and mother tongue, in addition to questions useful for determining the permanent residents of the household. One in five households (20 per cent) receives a long questionnaire, which contains the 7 questions from the short form plus over 50 additional questions on topics such as education, ethnicity, mobility, income and employment. The questionnaires of the 2006 Canadian Census of Population can be viewed by clicking on the following link: <http://www12.statcan.gc.ca/census-recensement/2006/ref/question-guide-eng.cfm>.
4. In order to ensure that the population and dwelling counts and the data on the characteristics of the population and dwellings are of sufficiently high quality, a diversified quality assurance programme has been developed over the last few censuses. During these censuses, changes and additions were made to the quality assurance programme in the light of operational, technological and methodological changes as well as a changing reality. This article outlines the quality assurance programme of the 2011 Canadian Census of Population and gives an overview of the main changes introduced for the 2011 Census.

II. Quality assurance programme of the Census of Population

5. The data quality assurance programme is intended to assess the quality of all census data. The results of this check are used to inform users of the reliability of the data, to improve the current census or future censuses and, in the case of two data quality studies, to correct the official demographic estimates. Quality assurance activities take place throughout the census process, starting before data collection and ending after dissemination. The main activities of the programme are described in the following sections.

A. Collection activities

6. There are two main enumeration modes, namely self-enumeration (around 98 per cent of the population) and canvassing (2 per cent). For self-enumeration there are two methods, namely delivery of a questionnaire by Canada Post or delivery of a questionnaire by a census enumerator. With both methods, householders are asked to complete the questionnaire for all members of the household. They can complete the questionnaire

securely over the Internet or fill in the paper questionnaire and return it in the reply-paid envelope. For the canvasser method, a census enumerator visits a household and completes a questionnaire for the household by a personal interview. This method is normally used in remote and northern areas of the country and on most Indian reserves.

7. In areas where census questionnaires are mailed out, Statistics Canada periodically validates and updates lists of existing addresses. This exercise results in a comprehensive and reliable list of addresses for mailing out census questionnaires. Checks and mechanisms are in place to ensure the lists are as comprehensive as possible and that there are no duplicate addresses. Address lists are updated using various sources of administrative data and on the basis of previous censuses. In areas where enumerators deliver the census questionnaires, the address list is drawn up at the time of delivery, in which case there are several procedures for ensuring that the operation also produces a comprehensive address list. Dwelling counts with a very high geographical resolution are compared with those of the previous census; detailed maps are produced in order to provide full coverage of every geographical area, and local knowledge is also used to identify possible problems.

B. Follow-up of incomplete questionnaires

8. All questionnaires are edited for completeness when they are received at the Data Processing Centre. Those that fail edit because they are missing too much information are sent to be followed up. In order to ensure the highest possible data quality, households with incomplete forms are then contacted by census staff to resolve missing and incomplete responses. Studies have shown that the absence of such follow-up would introduce significant bias into the data (Boudreau, 2009).

C. Questionnaires not returned

9. All households that do not return a census questionnaire are contacted by census staff to obtain a completed questionnaire. This process is managed in such a way that the non-response rate is as low as possible. Targets are thus set at various levels of geographical resolution to ensure optimum quality, and resources are used with the aim of reaching these targets. In the case of questionnaires completed by enumerators as part of this process, checks are carried out by supervisors to ensure an acceptable quality level and provide feedback on the work.

D. Dwelling Classification Survey

10. Despite best efforts during census data collection, there always remains a number of households that have not returned a questionnaire and for which no questionnaire has been completed by an enumerator. On the basis of a sample survey of addresses from which no census questionnaire was received, the Dwelling Classification Survey provides adjustments for the initial population and dwelling data to account for persons in these households. Estimates are derived using data from the sample on whether these addresses are occupied or not and, if so, how many persons usually live there. The results of the Dwelling Classification Survey are disseminated with the technical report on coverage. The 2006 Census Technical Report on coverage can be viewed by clicking on the following link: http://www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/rp/coverage-couverture/cov-couv_index-eng.cfm.

E. Census data processing

11. This stage in the census process involves processing all of the completed questionnaires, from the capture of data through to the creation of an accurate and complete census database: questionnaire registration, questionnaire imaging, data capture, error correction and coding. For all of these activities, quality control plans are in place to ensure the highest possible quality. The plans include standard edits such as manual sample edits, automated edits, and performance and evaluation indicators at each stage. Multiple automated edits are carried out on respondent data. The checks include coverage edits and a series of response edits.

F. Edit, imputation and weighting

12. The data collected in any survey or census contain some omissions or inconsistencies. For example, a respondent might be unwilling to answer a question, fail to remember the right answer, or misunderstand the question. Also, census staff may code responses incorrectly or make other mistakes during processing. The final clean-up of the data was done in edit and imputation and was, for the most part, fully automated.

13. An extensive programme is also in place for the evaluation of sampling and weighting. This programme is designed to determine the effect of sampling and weighting on the quality of census sample data. Four studies were carried out to measure the quality of census sample data and estimates, and to provide information for the planning of future censuses. These studies involved: an examination of sampling bias; an evaluation of weighting procedures; an evaluation of sample estimate to population count consistency; and determining the standard errors for various 20 per cent sample characteristics. The 2006 Census Technical Report on sampling and weighting can be viewed by clicking on the following link: http://www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/rp/sw-ep/sw-ep_index-eng.cfm.

III. Major changes planned for the 2011 Census

14. A number of major changes are planned for the 2011 Census, which is expected to rely once again on mail-outs for the majority of dwellings. In 2006, nearly 70 per cent of all dwellings received a paper questionnaire by mail. In 2011, mail-outs will cover nearly 80 per cent of all dwellings. Based on the strength of the Internet response in 2006, the target is to double the Internet response rate in 2011 to 35–40 per cent.

15. A strategy is being developed for recruiting and retaining staff in order to ensure that the required standards are met throughout the collection process. A new computer system has been developed to facilitate, enhance and expedite information communication between enumerators and the Data Processing Centre. Changes are planned with a view to enhancing staff training and making instruction manuals more user-friendly.

16. Among the activities of the census quality assurance programme, some major changes will be in place for the 2011 Census. The following sections give a brief overview of some of these changes.

IV. New measures in the quality assurance programme of 2011

A. Address register

17. The increased proportion of mail-outs (up from 70 to 80 per cent) required improvements to the coverage of the Address Register (AR). Fresh sources of administrative data have been used and will be used by 2011 with a view to enhancing AR coverage. Talks with certain partners have made it possible to reach new agreements on enhancing AR quality. The talks were held at the national level at first, and on a smaller scale thereafter. A series of outreach programmes are currently under way with various Canadian municipalities that are important for the use of their administrative data to check and possibly improve AR coverage. Also for 2011, a new programme has been launched for the purpose of checking and updating the AR on an ongoing basis. The programme is jointly implemented by the Census of Population and other Statistics Canada programmes. Since the programme is being implemented on an ongoing basis, new dwellings in fast-growing areas can be added to the AR in the weeks preceding the census. The programme also uses experienced interviewers who are familiar with the various concepts, aims and principles of such activities.

B. Managing the list of dwellings

18. An important addition to the quality assurance programme has been the creation of a Committee for the management of the list of dwellings. The Committee comprises a number of experts on various census activities and areas. It is tasked with supervising the implementation of a quality assurance programme for the maintenance of the list of dwellings in the mail-out portion and the listing and delivery portion. The Committee ensures proper and efficient integration of all the information and changes in the various collection and processing operations, and will achieve the best possible coverage.

19. A Collection Management System has also been developed to manage the list of dwellings. The System will be used mainly by enumerators and their supervisors, as well as for adding addresses to the listing and delivery portion, for adding new dwellings to the mail-out portion, and for updating the status of dwellings or certain other relevant data. It includes interactive checks to ensure that address and dwelling status information is complete, accurate and as up-to-date as possible. The System will enable direct management of the progress of the collection and accordingly increase the efficiency of non-response follow-ups, thereby ultimately increasing the response rate.

C. Internet response option

20. In 2006, a little over 18 per cent of all questionnaires were completed over the Internet. Studies have shown that data collected using this response option were of higher quality than with the traditional paper response option. Although these studies showed that respondents who chose the Internet were not the same as those who used paper, they also showed that the Internet response mode by its very nature made it possible to achieve a higher data quality (Grondin & Sun, 2008). The online help functions, validation messages and automated logical leaps are all features that lead to higher data quality with the Internet response option. For 2011, a new collection methodology will be brought on-stream with the aim of achieving a significant increase in the number of questionnaires completed online. This new methodology is based on the results of research carried out in 2006. Around 60 per cent of all households will receive a letter inviting them to complete the census questionnaire online. The letter will include a unique secure access code for each

dwelling. Simple instructions for those wishing to receive a paper questionnaire will also be provided in the letter. This collection method, known as the wave method, will also involve major changes affecting the collection so that a high response rate is stimulated and achieved. It is therefore expected that reminder letters will be sent and targeted telephone messages will be sent out even before the non-response follow-up operation begins. The aim of all these changes is essentially twofold: to stimulate response and to increase the number of questionnaires completed online. The latter objective has many advantages and a major impact on data quality. Data will be more complete and consistent, and there will be no input errors.

D. Content determination

21. Content determination in the 2011 Census has two main aims: to alter the content in response to new requirements (new question, new choice of answers); and to address data quality issues in certain areas identified in the course of the 2006 Census. Indeed, analyses of the data of the 2006 Census led to the identification of a number of points that needed changing in order to enhance quality. For example, an abnormal increase in the number of centenarians was observed in the 2006 preliminary data. Corrective measures were therefore introduced at the data checking stage. Content development for 2011 has also involved adding a question to check the age of each person. Other changes have been made to the content in order to improve data in areas where corrective measures had been needed or warnings had been issued to users regarding the quality of some data.

E. Dwelling occupancy status

22. A major aspect of quality assurance for population and dwelling counts is the determination of the occupancy status of a dwelling. A dwelling can have one of three possible statuses: a dwelling occupied by usual residents, a dwelling occupied only by temporary or foreign residents, and an unoccupied dwelling. The status is determined on the basis of the reference date. During the non-response follow-up process in the 2006 Census, enumerators had to assign an occupancy status to every dwelling that had not returned a questionnaire. Obviously, for occupied dwellings, enumerators had to complete a questionnaire. In the 2006 Census, the non-response follow-up period lasted for 10 to 14 weeks, depending on the region. Analyses showed that the error rate in dwelling occupancy statuses was higher when the status of a dwelling had been assigned towards the end of the follow-up period, and thus further away from the reference date for the 2006 Census. The dwelling occupancy status error rate should be minimized and estimated accurately before any necessary adjustments are made. A number of changes and improvements will be introduced in the collection activities of the 2011 Census in order to improve this aspect. In 2011, a new operation will be built into the mail-out portion prior to the start of the non-response follow-up period, in order to improve the determination of dwelling occupancy status. This operation will last approximately two weeks and enable efforts to be focused on determining the occupancy status of private dwellings as near as possible to the Census reference date. It will not be possible to identify all unoccupied dwellings during the operation, but the aim will be to identify as many as possible. For this purpose, predetermined clearly defined geographical areas will be selected according to various criteria based on the possible concentration of unoccupied dwellings (as determined using data from the previous census or administrative data). Enumerators will visit these areas and determine the occupancy status of non-responding dwellings. The primary aim is therefore to determine the dwelling occupancy status even before the start of non-response follow-up operations. This operation is therefore expected to lead to an improvement in the quality of the population and dwelling counts in 2011.

23. The Census counts persons at their usual place of residence. Some dwellings are occupied only by temporary or foreign residents. Such dwellings are identified by respondents who must follow the instructions given in the questionnaire. Formerly, all dwellings indicated by respondents as being occupied only by temporary or foreign residents were the subject of a telephone follow-up. Unfortunately, it was not always possible to contact all of these dwellings. If a respondent gave a wrong answer and the follow-up failed, the result could be population undercoverage. Additional improvements were made to the 2011 questionnaire in order to reduce the potential response error rate and, in particular, to improve the targeting of follow-ups in such cases. The change will lead to better concentration and targeting of follow-up efforts on cases where there may be wrong answers in the relevant portion of the questionnaire.

F. Coverage of dwellings

24. Although every effort has been made to ensure that every dwelling in the mail-out portion is on the address list or that every dwelling in the listing and delivery portion is listed by an enumerator, some dwellings may still be missed. This could give rise to problems if there are any clusters of such dwellings. For 2011, efforts will be made to ensure that population and dwelling counts are reliable wherever such problems arise. Multiple information sources have been identified as potential means of pinpointing the problems. Data on telephone calls received by the Census Helpline will be analysed to determine the number of requests for questionnaires by respondents claiming not to have received a questionnaire or letter. Other data such as administrative data on recent growth in housing (new sectors being developed) will also be used, as will data from previous censuses. It will be possible to analyse all of these data very quickly on a daily basis, and to determine what actions are necessary. Relevant and accurate information will then be transmitted to operational staff to ensure that if any dwellings have been missed, measures are taken to add them to the respective address lists. Similar information was used in the past, but it was more difficult to use it for centralized management of the analysis and corresponding actions to be taken. The development of the collection management system will enable better monitoring of the entire operation.

G. Minimizing data deletion due to non-response

25. A quality indicator is computed in association with dissemination in every geographical area. The indicator is based on an overall non-response rate matching the percentage of required responses that were not supplied by respondents. When the overall rate is higher than 25 per cent, characteristics are not disseminated but population and dwelling counts are. The overall rate takes account of complete non-response (an occupied dwelling that has not returned a questionnaire) as well as partial non-response (an occupied dwelling that has returned a questionnaire with some questions left unanswered). In 2011, with the aim of minimizing data deletion, rapid access to data from returned questionnaires will enable the overall non-response rate to be estimated fairly accurately without waiting for the end of the collection. It will therefore be possible to have a more exact real-time indication of data quality during collection activities in every geographical area, and thus to target follow-up efforts in order to achieve an acceptable overall non-response rate and avoid data deletion.

H. Imputation

26. In 2006, the Census non-response rate was 3.5 per cent, up from 1.6 per cent in 2001. As defined herein, the non-response rate is the percentage of occupied dwellings that have not returned a questionnaire. Imputation is performed for all such dwellings. As the non-response rate rises, the non-response bias, if any, may be presumed to increase. Changes are currently being considered to evaluate possible imputation methods that might allow this non-response bias to be corrected. Possible changes include the option of selecting a donor for the imputation of the total non-response, taking into account the return date of the questionnaire associated with each potential donor. For donor selection, greater importance could therefore be given to questionnaires received at the end of the collection period.

27. Furthermore, the approach to data edit and imputation was altered by emphasizing the integration and harmonization of strategies across the various data subject areas. A set of best practices has been developed and will be applied in all specialized areas of the Census. Standard imputation flags will, *inter alia*, be generated in order to enhance activity monitoring.

I. Data evaluation

28. In 2006, automated data input systems were used with paper questionnaires. This change not only ensured higher data quality but also enabled data to be accessed far more quickly than with a conventional manual capture of data from paper questionnaires. The introduction of an Internet response option has also contributed to this rapid data access during collection. In 2011, given the anticipated increase in the number of Internet questionnaires, a very large corpus of data will be rapidly accessible at a very early stage in the collection period. Lead times between the receipt of (paper and Internet) questionnaires and the moment when analysts will be able to access the data via user-friendly databases will be considerably reduced. Various tools are being developed to enable analysts in 2011 to access the data quickly and easily within days of the start of collection. These tools will make it easier to compare 2011 Census data distributions with data from other sources and from previous censuses. This advantage will therefore enable potential data quality problems to be identified more rapidly and provide more opportunities for taking the necessary corrective measures. Such corrective measures could be in the form of a manual correction in one subgroup of the population, or of adjustments in the data checking and imputation programmes.

29. Other activities are also being planned for the 2011 Census. A technical report on imputation is currently at the planning stage. This report could help to keep users better informed of data quality as far as imputation rates by variable and by geographical area are concerned.

V. Conclusion

30. Most of the activities of the Census of Population affect the quality of the resulting information. For the 2011 Census, Statistics Canada will continue to improve its quality assurance programme and all related activities. Further improvements not covered in the present document are also being planned or examined. Quality management will continue to be a key component of the 2011 Census.

VI. References

Bender R. (2009), “2011 Canadian Census of Population”. 2010 World Population and Housing Census Programme, United Nations Statistics Division, Newsletter March 2009.

Boudreau J.R. (2009), “Failed edit follow-up and item non-response bias”. In-house document, Statistics Canada.

Grondin C., and Sun L. (2008), “2006 Census Internet Mode Effect Study”. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 1255–1262.
