



Economic and Social Council

Distr.: General
13 July 2022

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Twenty-fourth Meeting

Geneva, 21–23 September 2022

Item 5 of the provisional agenda

Transitions in census methodology; plans, experiences and innovations

Use of Administrative Data in the Canadian Census

Note by Statistics Canada*

Summary

Many countries around the world have recognized the challenges faced by traditional Census collection, notably the efforts needed to maintain response rates and ongoing difficulties in enumerating individuals who respond later in the Census collection period. In Canada, the census response rate has not been an issue so far. The 2021 Census reached a successful 98 per cent collection response rate, similar to the 2016 Census. But Statistics Canada has to prepare now for the moment when the current traditional census collection approach will no longer be sustainable. Research toward a future model for the census with a goal to reduce response burden, increase the efficiency of collection and reduce overall cost of the Census Program is ongoing. Many countries, including Canada, have an extensive nonresponse follow-up (NRFU) approach as part of their collection activities. In recent years, National Statistical Offices have utilized administrative data to supplement NRFU activities and in some cases have developed a Census methodology based solely or partly on administrative data. Within this context, Statistics Canada is also investigating increased usage of administrative data for the 2026 Census following the successful implementation in 2021 of a statistical contingency plan as a response to the public health measures deriving from the global Covid-19 pandemic, and potentially developing future censuses based on a combined census methodology.

* Prepared by Geoff Bowlby, Director General, Census Program, Statistics Canada and Josée Morel, Assistant Director, Statistical Integration Methods Division, Statistics Canada.

Note: The designations employed in this document do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.



I. Introduction

1. The Canadian Census of Population aims to enumerate and to collect information on the entire population of Canada every five years. This is achieved by ensuring a questionnaire is made available to each and every dwelling and that data are obtained for each household. In this regard, the current methodology of the Canadian Census of Population fits within the category of traditional censuses, given full field enumeration is conducted (Department of Economic and Social Affairs of the United Nations, 2017).

2. Enumeration and collection of basic socio-demographic information are performed using short- and long-form questionnaires. Dwelling characteristics and additional demographic and socio-economic characteristics on individuals are mainly collected using the long-form questionnaire, which was distributed to a 25 per cent sample of dwellings in the 2016 and 2021 censuses. Three modes of delivery are used in the Census, namely mail-out, list/leave, and enumeration by canvassers. The dwelling list frame in mail-out areas is built from the Address Register. In mail-out areas, an invitation letter is sent by the postal service with a Secure Access Code (SAC), which allows a household to complete their Census form online. In list/leave areas (so named because census enumerators list new dwellings in the field, and leave a census invitation in the mailbox), census enumerators perform door-to-door delivery of similar invitations to fill out the questionnaire online. In the canvassing areas, census enumerators also prepare dwelling lists and usually complete the questionnaires at the dwelling with the respondents. Various field operations are conducted to ensure appropriate dwelling classification and response level. These operations are important as they help to produce a more accurate occupancy status for a number of dwellings and reduce the workload of the census Non-response Follow-up (NRFU) activity. NRFU is conducted as the final collection activity in mail-out and list/leave areas, where enumerators try to obtain a completed questionnaire from all occupied households that did not return a questionnaire.

3. In Canada, census coverage errors are measured using three statistical activities. The first, the Dwelling Classification Survey (DCS), is conducted in the later stages of the census collection to study misclassification of dwellings on Census Day. In particular, a sample of dwellings for which no census questionnaire was returned is selected. Some of these dwellings will have been coded as unoccupied and others as nonresponse. The DCS collected information on the occupancy status and, if occupied, on the number of usual residents. This information is used to adjust the census counts for dwellings, households and persons. This is done by creating control totals to correct classification errors and adjust household size distributions through imputation for all dwellings that did not complete a questionnaire in the DCS regions (Whole Household Imputation). It is done in time for the initial population count release. The last two activities make up the Census Coverage Measurement Program and are completed the year after the census. The Census Undercoverage Survey (CUS) which was formerly known as the Reverse Record Check, relies on sampling from independent frames, including administrative frames. It estimates the population missed by the Census, referred to as census or population undercoverage, while the third, the Census Overcoverage Study (COS), estimates the population enumerated more than once, referred to as census or population overcoverage. Census data themselves are not adjusted for the coverage errors measured by the CUS and the COS. Rather, Statistics Canada uses net undercoverage estimates along with other adjustments to produce demographic estimates.

4. The Canadian Census of Population has been using administrative data for many cycles. Since 2016, income variables are collected from administrative data for 100 per cent of the population, a technique also applied in 2021 for some immigration variables. Disruption to data collection following wildfires in a northern region of Canada led to an extensive use of administrative data in the 2016 Census as many residents living in this area were evacuated and were not able to fill their census questionnaires.

5. The next sections will describe the use of administrative data in the 2021 Census, the plan to increase its usage in the 2026 Census, and research on combined census models for the future censuses. The focus in the following section is on administrative data use in field operations or post-collection stages. At the time of writing, changes to administrative data

use for question replacement on the census form are not known for 2026 and as a result are not the focus of this paper.

II. Statistical Contingency Plan for the 2021 Census

6. In early 2020 as a response to the global Covid-19 pandemic emergency and associated public health measures, plans to implement administrative data in the Canadian Census context were greatly accelerated. Statistics Canada developed and implemented a “statistical contingency plan” based on the use of administrative data for post-collection imputation of 2021 Census non-responding households. The plan was developed in the event that Covid-19 or other disasters prevented the traditional census from a satisfactory enumeration. Only administrative files from government sources were used. Security measures for linkage keys and administrative files respected the policies, directives and guidelines for information technology security at Statistics Canada. When record linkage was required, it was done using anonymized statistical identifiers (“linkage keys”) and, as a result, no linked file contained personal identifiers such as name, phone number and address (excluding postal code).

7. Statistical models were used extensively to form households from individual government-provided administrative information (e.g., tax records) and quality indicators were developed to determine household eligibility for inclusion in imputation processes. Such administrative households were then incorporated into existing Census short-form imputation steps to compensate for nonresponse.

8. While administrative information is primarily person-based (e.g., individual tax records, immigration files), the Census 2021 statistical contingency used various statistical models to group individual records into “administrative households” and determine their level of quality for inclusion within the Census 2021 processes. Adapting the approaches of other countries (Keller and al., 2018; Bycroft and Matheson-Dunning, 2020), quality indicators at the household (dwelling) level were created to evaluate and rank eligible administrative households. Generally, this approach is referred to as the Household Model, consisting of three stages: i) the person-place model assessing individual coherence between traditional census responses and administrative data, ii) the household composition model which verifies dwelling-level coherence and iii) a distance metric which combines predictions from both models to determine a measure of overall quality. Both logistic and multinomial regression models were used to obtain predictions for the models mentioned in i) and ii).

9. Since the impact of the pandemic on census response rates was unknown in the planning stage, the use of administrative data was earmarked for imputation of 2021 Census non-responding households after other collection activities had ceased. In particular, administrative households of high quality were given priority in the Whole Household Imputation (WHI) process, which imputed from administrative data occupancy status, household information and short-form variables (e.g., household size, age, sex at birth, language, etc.) for non-responding dwellings, and long-form variables using traditional imputation (donor) methods. The existing WHI methodology is largely based on control totals provided by estimates of the DCS. Just prior to the field operations of Census 2021, simulations were conducted using the administrative-assisted WHI methodology which showed reduced bias in estimates compared to the existing donor imputation methods when nonresponse increased. Further, the most gain was found in geographical areas where the response rate was lower than 90 per cent. An analysis of age and sex distributions also noted closer values than existing donor imputation methods, with the exception of young adults and adults over 80 years of age.

10. In the end, the 2021 Census had a successful enumeration, with 98 per cent of Canadians responding to the census. As a result, widespread use of the contingency plan was not necessary. However, because some localized areas of the country showed response rates well below the national rate, administrative data were used to support the imputation of non-responding households in these areas. About 1,045 collection units (out of about 49,000 in Canada), which are small geographical areas, showed a response rate below 90 per cent.

Dwellings in these areas that had good quality administrative data were in-scope for this imputation plan. Approximately 12,000 non-responding households were imputed using administrative data, representing less than 0.1 per cent of occupied private dwellings in Canada. Note that the contingency plan used data already provided to Statistics Canada, and meets the highest standards of privacy, confidentiality and data security.

III. Increased Usage of Administrative Data for the 2026 Census

11. The 2026 Canadian Census of Population will still be a traditional census although research is ongoing towards a combined census model in the future (see section IV.). The objective for 2026 is to leverage the 2021 success and continue to provide very high-quality data, maintain cost effectiveness on a per-dwelling basis, pay more attention to its environmental (carbon footprint and paper consumption) impact, and earn and maintain citizen trust by providing very high levels of service and reducing burden.

12. This section will look at how the Census Program could increase its use of administrative data and identify more efficiencies in 2026 while maintaining the highest quality standard expected from the census. Social acceptability of an increased usage of administrative data in the census and the impact on the current administrative data-based coverage studies will need to be assessed, as well as operational considerations. A testing strategy (significant census tests will be conducted in 2024, with other testing throughout the intercensal period) is currently under development and will cover the general census operations, the questionnaire content, as well as the increased usage of administrative data in 2026.

A. Operational efficiencies during the Non-Response Follow Up (NRFU)

1. Identifying Unoccupied Dwellings

13. As in past censuses, the 2021 Census spent time and resources attempting to obtain responses from dwellings/structures where it was eventually determined there were no usual residents. Initial research has shown that the Household Model (HM) described in Section II has good potential to identify many unoccupied dwellings with reasonable accuracy. In addition, the Dwelling Occupancy Model (DOM) used before the start of NRFU to identify a sample of dwellings with high probability of being unoccupied and to be verified with field work has similar potential. For the 2026 Census, these models will be used to reduce field efforts away from the in-person classification of an unoccupied, cancelled or duplicate status, not only prior to NRFU but also during NRFU. Comparisons between the DOM and HM are ongoing and show promising preliminary results on the best use of both models.

2. Optimizing Priority Cases

14. NRFU is a very costly part of the overall Canadian Census collection activities, especially considering the diminishing returns on expenditures for the small number of individuals and households yielded by NRFU. However, NRFU is essential to capture the last 10 per cent or more of the most difficult-to-reach census respondents, in light of Statistics Canada's mandate to enumerate the whole population and produce the highest possible quality population counts. There is potential for optimizing NRFU activities by using administrative data for non-responding households when available, allowing the reallocation of NRFU personnel and resources towards areas and households where administrative data cannot be used. The quality indicators from the HM would be used to identify the dwellings that have high quality administrative data and therefore could be flagged as candidates for imputation using those administrative data, when a Census form is not completed at the end of the collection period.

B. Identifying cases for the Failed Edit follow-up (FEFU)

15. The Canadian Census of Population conducts a FEFU operation that involves re-contacting census respondents by telephone to enumerate potentially missed persons. FEFU

is conducted during collection and the targeted households are those that had ambiguities on the inclusion of usual residents in the dwelling. This operation successfully adds in tens of thousands of usual residents to census responding dwellings that otherwise would have been missed, reducing the census undercoverage. In the 2021 Census, over 40,000 persons were added from that operation.

16. Up until 2021, the dwellings identified for FEFU have been identified through the respondent-provided information or paradata from collection. For example, census respondents indicating that the dwelling contains residents “staying at the address temporarily” could be flagged for FEFU. A new research initiative for 2026 will examine using administrative data to identify additional dwellings for FEFU with potentially missed residents. These dwellings would be identified by comparing good quality administrative data to the dwelling’s usual residents on the completed census questionnaire. In short, the concept is to use the administrative data to address undercoverage through interview follow-up with the responding census dwellings during census collection. Dwellings with administrative data which indicate they are likely to have a missed person would be called to verify the usual residents and to enumerate missed persons.

C. Post-collection: Imputation of non-responding households

17. As described in Section II, in 2021 administrative data households of high quality were used in the initial imputation steps in the Whole Household Imputation (WHI) process, before donor imputation. The objective for 2026 is to extend this operation. In 2021, it was used mainly in the mail out areas. The timeline of some record linkage activities could be improved to allow imputation for non-mailout areas.

18. As dwellings would be flagged for imputation with administrative data during NRFU, this will reduce the pool of potential donors for the WHI performed post-collection. Other options are being explored to make up for that issue. The current methodology necessitates that each administrative record be linked to an address on the census database. The administrative records not linked could still be considered as donors if in the same collection area, for example. This could lead to an improvement of the pool of donors.

19. Another potential option is for the DCS, described in the introduction, to benefit from using administrative data. Population counts from administrative data at a certain level of geography could be used as control totals for census population counts and help to determine where to strategically impute. This could ensure high-quality population and dwelling counts in areas where census collection was affected by low response rates.

D. Social and stakeholder acceptability

20. The Census of Population is vital for the application of many laws, regulations, and programs at all levels of government, and is a trusted source of population statistics for many public, academic, business, and civil-society data users. The Census is highly visible in the public conversation, well publicised during the year of collection and it is well-known by all Canadians, many of whom take pride in completing their questionnaires. For these reasons, one risk of an increased use of administrative data for the census-taking methodology lies in its social acceptability from the public, the stakeholders and the data users.

21. Maintaining and increasing the trust of Canadians is essential. Even if socially-acceptable to the public, there could be a risk that the use of administrative data causes higher levels of non-response. Upon learning about Statistics Canada’s plan to use their administrative data, dwelling occupants might avoid completing their Census questionnaire on the assumption that administrative data will be substituted for their questionnaires anyway. An understanding of the social acceptability for an increased use of administrative data is a prerequisite for any change to census collection methods, and therefore a comprehensive consultation plan is being developed to engage and communicate with Canadians and various stakeholders on these and other issues.

22. The objectives of this consultation and testing plan are to both measure and promote the social acceptability for an increased use of administrative data, through repeated research and public engagement. The results of these tests will inform the design of Statistics Canada's administrative data strategy and public communications about the Census. The research and testing methods will collect both qualitative and quantitative information using public opinion surveys, interviews, focus groups, and deliberative public engagement research events. The latter will be an adaptation of the usual focus group method, with the difference being that the event moderators will deliver educational presentations to allow the research participants to reach an informed consensus opinion.

23. Updated census methods that make increased use of administrative data must also be acceptable to stakeholders, researchers, and expert users of official statistics. These stakeholders have different priorities, compared to the general public, and therefore the methods of testing and promoting stakeholder acceptability will be different. The objectives of stakeholder acceptability testing will be to generate the evidence, arguments, and justifications supporting an increased use of administrative data, and to assure stakeholder organizations that new census methods will still produce the high-quality data that will allow them to fulfil their mandates. Testing stakeholder acceptability will include extensive consultations with each stakeholder organization, such as with Canada's Office of the Privacy Commissioner (OPC), Provincial and Territorial representatives, national Indigenous organizations, academic and research communities, and various civil society and privacy advocacy groups.

24. In the case of FEFU, using an external source such as administrative data to revisit (or question) the responses Canadians have provided could pose risks to the trust relationship established with Canadians. Any risks associated with activities that might negatively impact the trust from Canadians or their willingness to complete their Census questionnaires must be seriously considered. The results of these public and stakeholder consultations will inform the design of Census communications that will resonate with Canadians and effectively communicate the benefits of an increased use of administrative data.

E. Impact on the current undercoverage estimation quality

25. Depending on the extent of the increased use of administrative data, there could be an impact on the design-independence of the undercoverage program, meaning that the same information would be used for the purpose of the census, as well as for the measurement of undercoverage. For example, the administrative data sources that would be used to identify the potential missed persons in FEFU or for the imputation may also be used in the Census Undercoverage Survey (CUS). As described in the Section I, the CUS relies on sampling from independent frames, including administrative frames. Design-Independence of the CUS methods and the census collection has been a necessary condition for the quality of the CUS. While the FEFU operation would rely on respondent confirmation and enumeration, considering all aspects to ensure that the CUS can still accurately estimate the Census undercoverage is critical to producing reliable and accurate demographic estimates.

IV. Conclusion

26. Statistics Canada has to prepare now for the moment when the current traditional census collection approach will no longer be sustainable. A few years ago, the Census Program brought together a team – the Census Program Transformation Project – dedicated to develop a future model for the census with a goal to reduce response burden, increase the efficiency of collection and reduce overall cost of the Census Program, while maintaining the quality and relevance of the Census data. In fact, the statistical contingency plan that was developed for the 2021 Census and described in Section II benefited from this research.

27. A combined census, consisting in integrating administrative data with traditional data collection to produce population and dwelling counts is being considered post-2026. The administrative data would be leveraged to reduce the number of short-form questionnaires Canadians would need to fill out. A long-form questionnaire, as the one in the current census,

would still be used to collect detailed characteristics as they are not available from the administrative data. Research is ongoing and various combined census models are being studied, as well as evaluation of appropriate coverage studies. The Household Model described in Section II would be at the heart of the new methodology.

28. It is proposed that components of a combined census be tested during the 2026 Census, to the extent possible. It will help determine the degree to which the administrative data on dwellings and people could replace (to a greater extent than what is proposed in 3.1.2) some of the traditional short-form census collection in the future and the social and stakeholder acceptability of such an approach. The operationalization of some procedures and evaluation of new methods to measure coverage can also be tested. The results of the test would be used a year later to help determine if the 2031 Census could be the first in Canada to be a combined census. In the lead up to this test, challenges and prerequisites related to a combined census such as the current content of the short-form questionnaire, revisiting current coverage studies and consultations related to social and stakeholder acceptability will continue to be addressed.

References

- Guide to the Census of Population, 2021, <https://www.census.gc.ca/census-recensement/2021/ref/98-304/2021001/app-ann1-7-eng.cfm>.
- Coverage Technical Report, Census of Population, 2016, <https://www12.statcan.gc.ca/census-recensement/2016/ref/98-303/intro-eng.cfm>.
- Bycroft, C. and Matheson-Dunning, N., 2020, Use of administrative records for non-response in the New Zealand 2018 Census. *Statistical Journal of the IAOS* 36(1), pp 107–116.
- Keller, A., Mule, V.T., Morris, D.S. and Konicki, S., 2018, A distance metric for modelling the quality of administrative records for use in the 2020 US Census. *Journal of Official Statistics* 34(3), pp 599–624.
-