

Distr.: General
29 April 2024

English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Migration Statistics

Geneva, Switzerland, 7–8 May 2024

Item 6 of the provisional agenda

Improvements in use of administrative data for migration statistics

Replacing immigration question with administrative data in the Canadian Census of Population

Note by Statistics Canada

Abstract

The Canadian Census of Population is a key data source on the socio-economic outcomes of immigrants in Canada. Census questions related to immigration include immigrant status, year of immigration, citizenship, place of birth, and place of birth of parents. As a strategy to reduce burden and improve data quality, Statistics Canada replaced the immigrant status and year of immigration variables on the 2021 Census questionnaire with administrative records. Immigration is a process administered by Immigration, Refugees and Citizenship Canada (IRCC), administrative data is collected for temporary (non-permanent) and permanent residents in Canada. Having successfully replaced income questions with administrative data and integrating new immigration variables in 2016 there was a strong precedent for this approach. This paper will outline the results from using administrative data to replace immigration question on the 2021 Census of Population providing an overview of the benefits and challenges with administrative data integration.

*Prepared by Kathryn Spence and Athanase Barayandema

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 Census of Population is a key data source on the socio-economic outcomes of immigrants in Canada. Census questions related to immigration include immigrant status, year of immigration, citizenship, place of birth, and place of birth of parents. The 2016 Census also included variables related to admission category (e.g., refugees, economic immigrants, etc.) using administrative data (McLeish 2017).
2. As a strategy to reduce respondent burden and improve data quality, Statistics Canada integrated administrative data from Immigration Refugees and Citizenship Canada (IRCC) to replace the questions on immigrant status and year of immigration building on the success of integrating administrative data in the 2016 Census (Statistics Canada 2017b; McLeish 2017).
3. This paper will present the results of integrating the administrative data to replace the immigration questions, assessing the quality of the administrative data, and examining how replacing the questions could affect historical comparability. This paper will also assess the quality of the existing questions to understand the differences in data quality between asking questions and using administrative values outlining the benefits and challenges.

II. Limitations on questions and administrative data

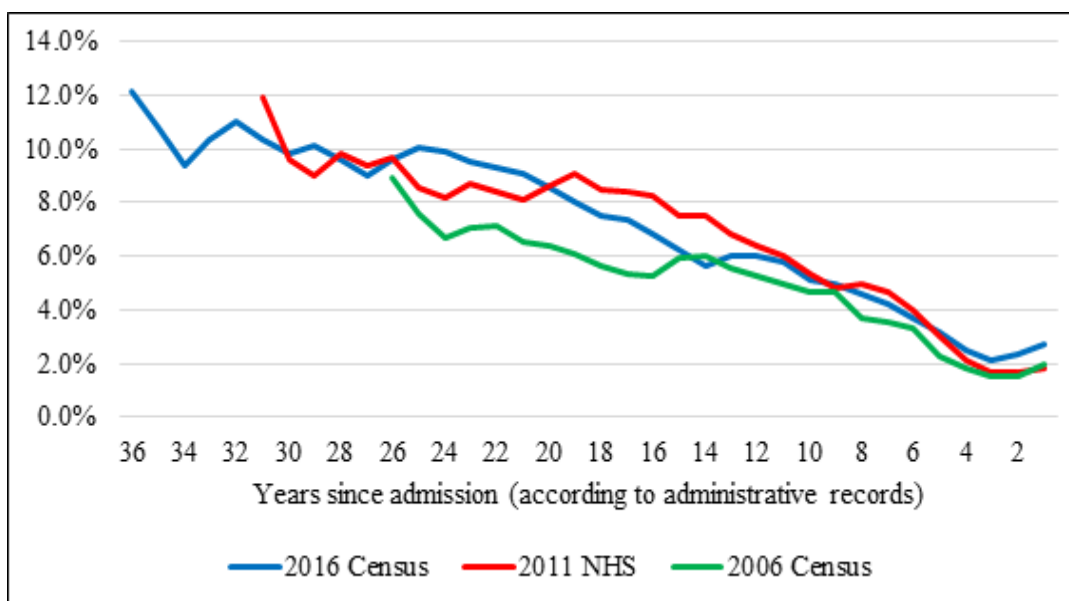
A. Limitations on questions

4. Data on immigrant status and year of immigration have been collected on the census questionnaire since 1901. As has been documented in previous papers (McLeish 2017; McLeish 2014), questions are not always answered by respondents, and answers provided are not always precise. For immigration and citizenship questions specifically, there are some common issues that have been observed during the certification of the data quality of census results:
 - a. the longer immigrants have been in Canada, the less likely they are to respond affirmatively to the immigrant status question,
 - b. respondents do not always provide the precise response to the year of immigration question, and accuracy appears to decrease the longer the respondent has been in Canada,
 - c. certain respondents appear to provide their year of arrival instead of their year of immigration, and
 - d. Canadian citizens by birth born abroad, that is those who are entitled to Canadian citizenship because of their parentage, appear to sometimes respond that they are a Canadian citizen by naturalization.
5. During data processing, edits are applied to ensure consistency between responses and donor imputation methods are used to address item non-response (Crowe 2017; Guertin 2014).
6. Figure 1 illustrates issue 4a) where 7.6% of 2016 Census respondents who were linked to administrative immigration records (for immigrants admitted since 1980) responded “No” to the immigrant status question. This proportion increased the longer the linked respondents have been in

Canada. Most of these cases were resolved during edit and imputation. However, since the immigrant status questions serves as a filter for the year of immigration question, this issue results in a higher imputation rate for the year of immigration¹.

Figure 1

Percentage of census respondents linked to IRCC immigrant records who responded that they are not immigrants by years since admission.



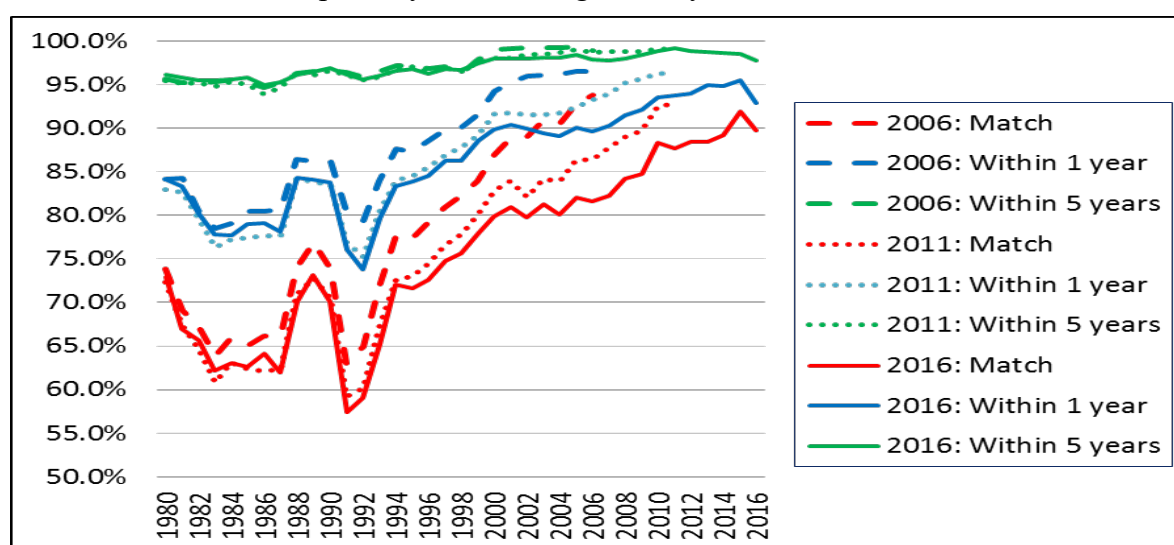
Source: Statistics Canada, 2016 and 2006 Censuses linked to IRCC administrative immigrant data since 1980 and 2011 National Household Survey (NHS) linked to IRCC administrative immigrant data since 1980

7. Figure 2 illustrates issue 4b), where over 21% of 2016 Census respondents linked to immigration records since 1980 provided a year of immigration response which did not match their administrative value. This proportion increases the longer the respondents have been in Canada. However, most of these differences in value are less than 5 years.
8. Issue 4c) is also shown in Figure 2, where there is a notable drop in the congruence between administrative and census year of immigration for the years 1991 to 1993. This is explained by a large number of people who claimed asylum in Canada in 1989 and 1990. While their official years of immigration are between 1991 and 1993, their responses reflect their year of arrival or asylum claim.

¹ Imputation rates for the 2016 Census citizenship and immigration variables: <https://www12.statcan.gc.ca/census-recensement/2016/ref/guides/007/98-500-x2016007-eng.cfm#tbl1>

Figure 2

Match rate for census reported year of immigration by administrative value.



Source: Statistics Canada, 2016 and 2006 Censuses linked to IRCC administrative immigrant data since 1980 and 2011 National Household Survey (NHS) linked to IRCC administrative immigrant data since 1980

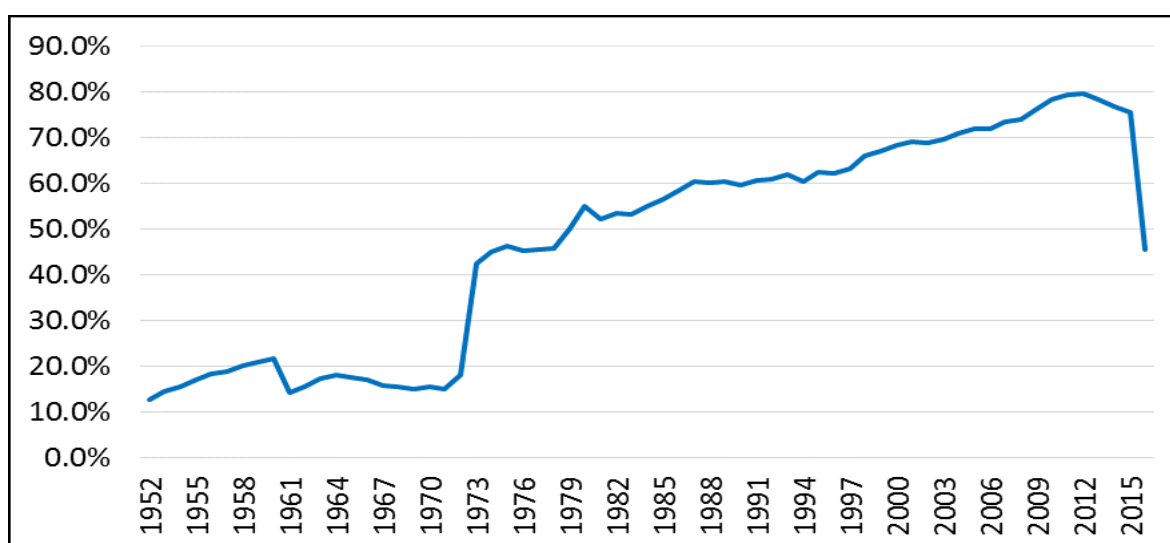
B. Limitations of administrative data

9. IRCC administrative data currently available to Statistics Canada includes;
 - a. Detailed immigration records from 1980 to present,
 - b. Limited immigration records from 1952 to 1979,
 - c. Non-permanent resident permit records from 1980 to present,
 - d. Citizenship records from 2004 to present, and
 - e. Visitor records from 2004 to present.
10. While IRCC data reflect an administrative census of all incoming immigrants and non-permanent residents to Canada, they are not updated to capture deaths or outmigration. Therefore, they cannot be used in isolation to estimate the current population of immigrants living in Canada.
11. For the purposes of replacing the two immigration questions on the Census of Population, the coverage of IRCC data introduces three limitations.
 - a. Firstly, there are no immigration records available prior to 1952. While this is a decreasing population, it still represented 111,000 immigrants (or 1.5% of all immigrants) living in Canada according to the 2016 Census.
 - b. Secondly, non-permanent resident records contain information on permit holders only; any accompanying family members who do not hold permits are not covered.
 - c. Thirdly, another limitation rests with the quality of the immigration records prior to 1980. Overall, there is less information included on these records that can be used for record linkage purposes. Records from 1961 to 1972 contain an incomplete date of birth which leads to lower linkage rates.

12. Figure 3 demonstrates the percentage of the administrative immigration records linked to 2016 Census respondents by administrative year of immigration. In general, the linkage rates are lower for older cohorts of immigrants, as these individuals are more likely to be out-of-scope for the 2016 Census (e.g., dead or no longer residing in Canada). However, there is a notable drop in the linkage rate between 1961 and 1972.

Figure 3

Percentage of administrative immigration records linked to the 2016 Census by administrative year of immigration.



Source: Statistics Canada, 2016 Census linked to IRCC administrative immigration data since 1952

III. Methods used for replacement, solutions, and challenges

C. Data linkage and methods used for replacement

13. A record linkage between the census or survey respondents and IRCC administrative data took place immediately after collection, and before processing. The probabilistic method was used to integrate IRCC's immigration data to the 2021 Census data. The global linkage rate for immigration records was 97.3%. The linkage rate for permanent residents was 99.4% and 92.4% for non-permanent residents. Respondents to the census who are linked to an administrative record are either an immigrant if linked to the landing file or a non-permanent resident if linked to a temporary resident file (work or study permit) or asylum claim.
14. The administrative data and the census question on Canadian citizenship were used to derive immigrant status variable to determine whether the person is a non-immigrant, an immigrant or a non-permanent resident as shown in Table 1. The responses to the citizenship question were used to determine which individuals required imputation to the immigrant status variable due to missing links. A person who is Canadian citizen by birth cannot be an immigrant, a person who is Canadian

citizen by naturalization are immigrants and a person who is not a Canadian citizen is either an immigrant (permanent resident) or non-permanent resident.

Table 1

Immigrant status and citizenship, Canada

Citizenship status	Immigrant	Non-immigrant
Canadian by birth		Canadian by birth
Canadian by naturalization	Canadian by naturalization	
Not a Canadian citizen	Permanent resident	Non-permanent resident

18. In general, the same edit and imputation methods used in the past were still applied using administrative values in lieu of responses to questions. In processing, those who were not linked and who gave a citizenship response of Canadian citizen by birth were considered to be non-immigrants and required no donor imputation for immigration status. Canadian citizens by naturalization and non-Canadian citizens were considered in-scope and required imputation if no administrative record was available.

D. Solutions to administrative data limitations

19. To resolve the absence of administrative records prior to 1952, the administrative files were supplemented to include records for past census respondents who responded with a year of immigration prior to 1952. These values would include the same response errors described above and would only be available for a sample of the total population who reported having immigrated prior to 1952. However, they are the only records available for this subpopulation.
20. Concerning the low linkage quality for years 1961 to 1972, an approach of using the multipliers directly to assign a year of immigration in this period for a specified number of unlinked records was used. The multipliers are the ratios of predicted probabilities of being linked if the date of birth was not missing. Different counts of unlinked persons, who did not identify themselves as Canadian citizen by birth on the Census, were assigned to be converted to each specific year for different age / place of birth groupings (cells). This was done before any Census donor imputation was performed. This approach has the great advantage of implementing the spirit of the multipliers to alleviate the linked year of immigration distribution. Another advantage of this approach is that it would be immediately and clearly generalizable to future Censuses.
21. Additional administrative data from IRCC such as visitor records, were used to supplement the temporary resident records for family members of work and study permit holders since some family members such as children, parents and grandparents may not have their own permit.

E. Key Challenges

22. The principal challenges associated with replacing the census immigration questions with administrative values pertain to the limitations of the administrative data itself. The absence of records prior to 1952 require the supplement of past census responses. However, past census responses will not cover this entire subpopulation (the immigration questions have only been asked of a sample of census respondents), and it could be underestimated, as a result.
23. The limited linked information on the immigration records from 1952-1979 (especially from 1961-1972) could lead to an underestimation of this subpopulation as well. Since the final results (post edit and imputation) are anchored by responses to the citizenship question, any underestimation of certain periods of immigration may lead to an overestimation of others (especially recent immigrants, whose records are of higher quality for the purposes of linkage).
24. Replacing the questions with administrative data would affect historical comparability. In particular, shifts in the distribution of year of immigration may occur as perceived year of immigration (e.g., year of arrival) is replaced with actual year of immigration.

IV. 2021 Census Results

F. Data Quality

25. A number of quality indicators such as the non-response rate and the imputation rate per question were produced and analyzed during the 2021 Census of Population data quality assessment.
26. The non-response rate for the immigration variables largely measures the proportion of immigrants and non-permanent residents for whom an administrative value was not available because the record was not linked to an administrative record. It also includes some inconsistencies between the census responses to the other questions and the linked administrative values. For example, immigrants may have reported a birth year on the census questionnaire that was before their year of immigration.
27. At the national level, the immigrant status variable had an imputation rate of 2.2% and a non-response rate of 12.1%. The difference between these two rates is because non-responses were resolved early during data processing because a single resolution was possible based on the answers provided to other questions, such as the citizenship question, making imputation unnecessary. For more details please refer to [Place of Birth, Generation Status, Citizenship and Immigration Reference Guide, Census of Population, 2021](#).

G. Comparability Over Time

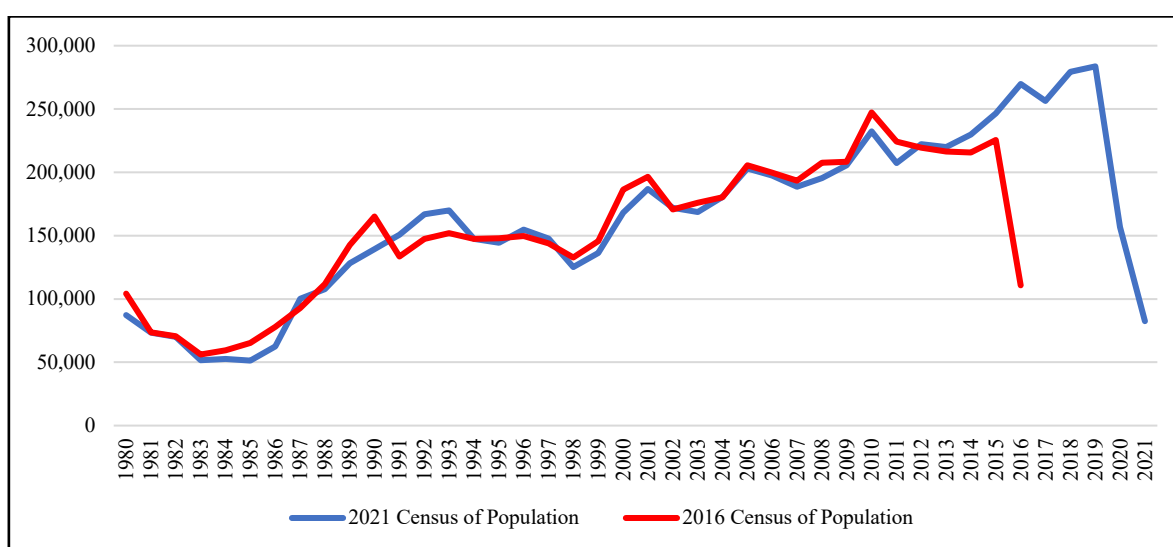
28. The methods used to collect data for immigrant status and year of immigration are different (questionnaire response vs. administrative value), therefore historical comparability with previous censuses will be affected. In particular, there are shifts in the distribution of year of immigration, as perceived year of immigration (e.g., year of arrival) is replaced with the administrative year of

immigration. A comparison of the number of immigrants by year of immigration between the 2021 Census and the 2016 Census (Figure 4) shows these shifts.

29. For the period from 1989 to 1993, Figure 4 shows that 2016 Census data (responses) were more heavily distributed in 1989 and 1990, while the 2021 Census data (from administrative data) were more heavily distributed from 1991 to 1993. This is caused by respondents providing their year of arrival or asylum claim (in the 2016 Census), as opposed to their year of immigration (in the 2021 Census).

Figure 4

Year of immigration for the immigrant population who were admitted between 1980 and 2021, Canada, Census of Population, 2016 and 2021



Source: Statistics Canada, Census of Population, 2016 and 2021

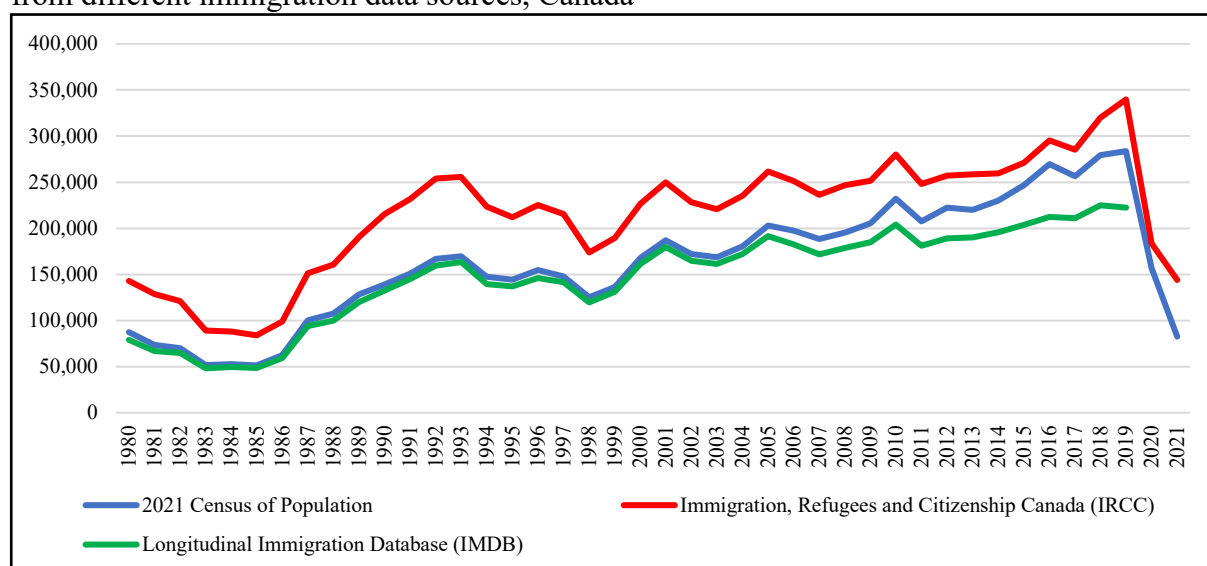
H. Comparability with other data sources

30. Many statistical sources provide information on immigration to Canada, covering different reference periods and different subpopulations, thereby meeting different informational needs which needs to be considered when comparing data sources. The 2021 Census of Population estimates the immigrant population living in private households in Canada on Census Day, May 11, 2021.
31. In comparison, the administrative data from IRCC provide information on the total number of immigrants admitted to Canada each year as permanent residents. The IRCC data cannot be used to estimate the population of immigrants living in Canada as they do not account for any outflows, such as deaths or emigration. Since they provide the total number of all those who have ever been permanent residents in Canada, the IRCC administrative data counts are higher than the census estimates of immigrants living in Canada at a given point in time.

32. To some extent, the counts from the IMDB consider deaths and emigration. The 2019 IMDB combines linked administrative data from IRCC with tax data files. The counts presented in Figure 5 are for those immigrants who filed tax returns in 2019. Since not all immigrants living in Canada would have filed tax returns, because of age or other factors, counts from the IMDB are expected to be lower than the estimates from the 2021 Census.
33. The number of immigrants by year of immigration in the 2021 Census data, the IRCC data and the IMDB data (Figure 5) shows similar trends in all three sources. In the earlier years of immigration, the census estimates are closer to the IMDB counts, as immigrants who landed between 1980 and 2005 would most likely be taxfilers still living in Canada at the time of the census. As the year of immigration moves towards 2021, the census estimates start to move closer to the counts from IRCC, as the number of non-taxfilers, such as children, increases, while the number of deaths and emigrants would be lower.

Figure 5

Year of immigration for the immigrant population who were admitted between 1980 and 2021, from different immigration data sources, Canada



Source: Statistics Canada, Census of Population, 2021, and IMDB, tax year 2019; and IRCC

V. Conclusion

34. The replacement of census questions on immigrant status and year of immigration with administrative values in the 2021 Census of Population was a success. Considering the good quality of administrative immigration data (particularly from 1980 to present) and the limitations of Census questions on immigration, the benefits of the replacement are data quality improvement, reduction in response burden, and the integration of additional variables from the administrative data such as non-

- permanent resident type (asylum claimant, work or study permit holder), pre-admission experience in Canada, province or territory of intended destination and year of arrival.
35. Another advantage of the replacement of census questions on immigrant status and year of immigration with administrative values is the impact beyond the Census of Population. Other household surveys may also follow in the same direction. Similarly, it could lead to a reduction in response burden on those surveys, and open possibilities of integrating the additional immigration content available on administrative files (such as admission category).
 36. Replacing immigration questions by administrative data has some disadvantages. First, because the method used to collect the data for immigrant status and year of immigration is different, the historical comparability will be affected for these questions. Specifically, the year of immigration will be more reflective of administrative value versus perceived value which reflects the true value of the concept being measured. Secondly, the replacement will imply a loss of the linked data as certification data source. However, Statistics Canada has the benefit of a wealth of available sources of data, such as Immigration Landing File, the Longitudinal Immigration Database (IMDB), and other social surveys such as Labour Force Survey and the General Social Survey Program that are used for certification.

References

- Biernot, P. 2017. "External linkage between the Immigration File (1952-2016) and the 2016 Census Response Database." Unpublished report. Ottawa: Statistics Canada.
- Brennan, J. 2011. "CIC Landing File to Census 2006 Linkage." Unpublished report. Ottawa: Statistics Canada.
- Brennan, J. 2013. "CIC Landing File to Census 2011/NHS Linkage." Unpublished report. Ottawa: Statistics Canada.
- Crowe, S. and Janes, D. 2017. "Edit and Imputation Report: Ethnocultural Process." Unpublished report. Ottawa: Statistics Canada.
- Guertin, L., Bureau, M., and Morel, J. 2014. "Editing the 2011 Census data with CANCEIS and options considered for 2016." *2014 Work Session on Statistical Data Editing*. Conference of European Statisticians. United Nations Economic Commission for Europe.
- McLeish, S. 2014. "Using administrative data to evaluate data quality: Immigrants in the 2011 National Household Survey", *Proceedings of Statistics Canada Symposium 2013*, Catalogue no. 11-522-X. Ottawa: Statistics Canada.
- McLeish, S. 2017 "2016 Census of Population of Canada: Integration of immigration administrative data" *2017 Work Session on Migration Statistics*. Conference of European Statisticians. United Nations Economic Commission for Europe.
- Statistics Canada. 2017. "Place of Birth, Generation Status, Citizenship and Immigration Reference Guide." 2021 Census of Population. Catalogue no. 98-500-X2016007. Ottawa: Statistics Canada.