

Distr.: General
04 October 2018

English

United Nations Economic Commission for Europe

Conference of European Statisticians

Work Session on Migration Statistics

Geneva, Switzerland

24-26 October 2018

Item 3 of the provisional agenda

Integration of data from censuses, administrative sources and surveys for measuring migration

Replacing immigration questions with administrative data for the Canadian Census of Population

Note by Statistics Canada*

Abstract

Household surveys conducted by Statistics Canada, including the Census of Population, generally include questions on immigrant status and year of immigration. The responses to these questions permit analysis comparing outcomes for immigrants, by period of immigration, with those for Canadian citizens by birth. In an effort to reduce respondent burden, improve data quality, and open possibilities for the integration of other immigration variables, Statistics Canada is undertaking a feasibility study to examine replacing these questions with linked administrative records for the 2021 Census. This study will examine data quality differences between survey responses and linked administrative values, impacts on processing (edit and imputation) of other variables, how this change could affect comparability with previous results, and how additional immigration content could be integrated. This paper will outline preliminary findings from this study based on results from the 2016 Census.

*Prepared by Scott McLeish, Section Chief, Social and Aboriginal Statistics Division and Athanase Barayandema, Analyst, Social and Aboriginal Statistics Division

I. Background

1. The Census of Population and other household surveys are key data sources on the socio-economic outcomes of immigrants in Canada. Census questions related to immigration include immigrant status, year of immigration, citizenship (Canadian citizenship and other countries of 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 linked administrative data (McLeish 2017). Finally, other immigration variables are derived from responses to the questions listed above, including age at immigration and generation status. Other household surveys (e.g. General Social Survey) typically include a subset of these questions. The immigration questions on the Census of Population and other household surveys allow for comparisons to be made between the outcomes and characteristics of immigrants and non-immigrants.
2. In an effort to reduce respondent burden, Statistics Canada has been investigating alternatives to asking questions on surveys. For example, the 2016 Census replaced income-related questions with linked administrative tax data (Statistics Canada 2017a). As immigration is a process administered by Immigration Refugees and Citizenship Canada (IRCC), administrative data is collected for new temporary (non-permanent) and permanent residents in Canada. IRCC administrative data was already used for the addition of admission category variables to the 2016 Census (Statistics Canada 2017b; McLeish 2017). In preparation for the 2021 Census, Statistics Canada is investigating building on the 2016 Census project by replacing the questions on immigrant status and year of immigration with linked IRCC administrative values.
3. This study will examine methods to use the linked administrative data to replace the related questions, assess the quality of the linked administrative data, and examine how replacing the questions could affect historical comparability. This study will also assess the quality of the existing questions in order to understand the differences in data quality between asking questions and using linked administrative values.
4. With survey questions, errors specific to the questions include item non-response and measurement errors (including response errors). With linked administrative data, assuming the administrative values themselves are correct, the relevant errors are false negatives (missing links), which result in the equivalent of item non-response, and false positives (incorrect links), which result in the equivalent of measurement errors (or incorrect values).
5. The census questions relevant to this study include (as phrased for the 2016 Census):
 - i. Citizenship: Of what country is this person a citizen? (Canada, by birth / Canada, by naturalization / Other country – specify)
 - ii. Immigrant status: Is this person now, or has this person ever been, a landed immigrant? (Yes / No)
 - iii. Year of immigration (only asked of those who responded “Yes” to Immigrant status): In what year did this person first become a landed immigrant?
6. While this study investigates replacing the immigrant status and year of immigration questions, the citizenship question is intrinsically linked to both concepts. Indeed the population residing in Canada can be delineated as:
 - i. Canadian citizens by birth
 - ii. Immigrants

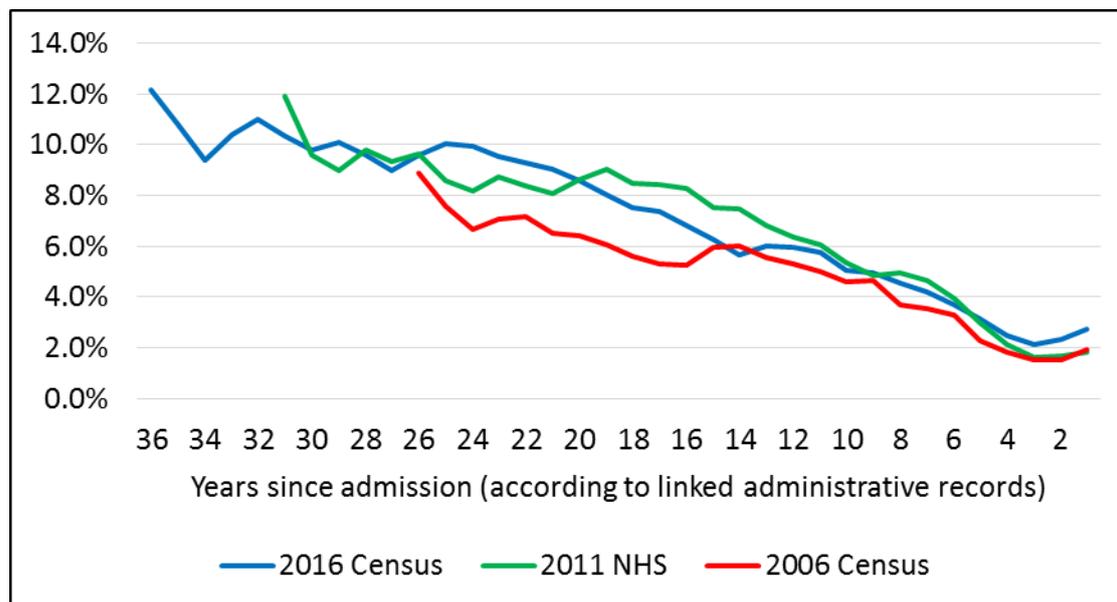
1. Canadian citizens by naturalization
 2. Permanent residents (not Canadian citizens)
 - iii. Non-permanent residents.
7. Anyone who is a Canadian citizen by birth cannot be an immigrant. Alternatively, anyone who is a Canadian citizen by naturalization must be an immigrant. Those who do not hold Canadian citizenship are either permanent or non-permanent residents. In the case of administrative data replacement, responses to the citizenship question identify those respondents theoretically in scope to be linked (either to an immigrant or non-permanent resident record).

II. Limitations of questions and administrative data

A. Limitations of questions on immigration

8. As has been documented in previous papers (McLeish 2017; McLeish 2014), questions are not always answered by respondents, and answers provided are not always accurate. For immigration and citizenship questions specifically, there are a number of common issues that have been observed during the certification or data quality evaluation of census results:
- i. The longer immigrants have been in Canada, the less likely they are to respond affirmatively to the immigrant status question (despite it being worded in such a way as to include anyone who has ever been a landed immigrant in Canada).
 - ii. Respondents do not always provide the correct response to the year of immigration question, and accuracy appears to decrease the longer the respondent has been in Canada.
 - iii. Certain respondents appear to provide their year of arrival instead of their year of immigration (this appears to be more of an issue for asylum claimants who become permanent residents).
 - iv. 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.
9. Edits are applied to ensure consistency between responses and donor imputation methods are used to address item non-response (Crowe 2017; Guertin 2014).
10. Using linkages between past censuses and IRCC administrative data, some of the issues mentioned above have been analyzed as part of the certification process.
11. Figure 1 illustrates issue 7.i, where 7.6% of 2016 Census respondents who were linked to an administrative immigrant record (for immigrants admitted since 1980) responded “No” to the immigrant status question. This proportion increases the longer the linked respondents have been in Canada. Most of these cases are resolved during edit and imputation, as they usually respond that they are Canadian citizens by naturalization (i.e. immigrants). However, since the immigrant status questions serves as a filter for the year of immigration question, this issue results in a high imputation rate for the latter, as shown in Table 1.

Figure 1 Percentage of census respondents linked to IRCC immigrant records who responded that they are not immigrants by years since admission



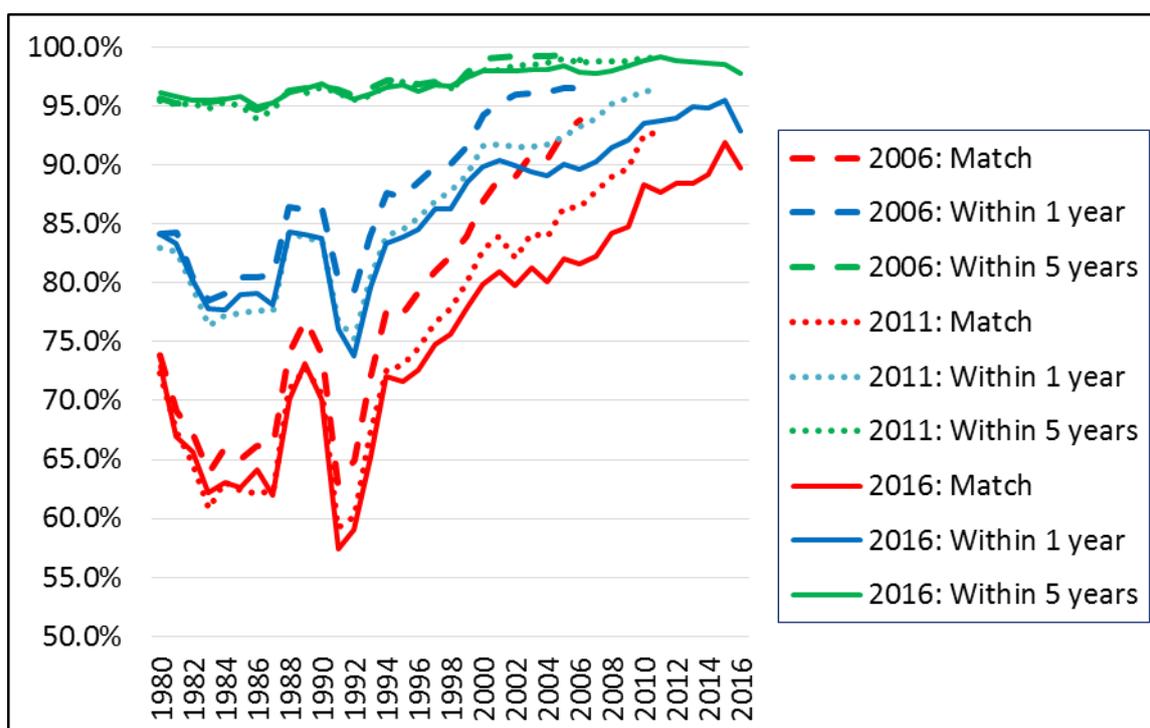
Source: Statistics Canada, 2016 and 2006 Censuses linked to IRCC administrative immigrant data since 1980 and 2011 National Household Survey (NHS) linked to IRCC administrative immigrant data since 1980

Table 1: Imputation rates for selected variables, 2016 Census of Population

Variable	Imputation rate (%)
Citizenship	1.3
Place of birth	1.0
Place of birth of mother	1.6
Place of birth of father	1.7
Immigrant status	0.7
Year of immigration	9.4

Source: Statistics Canada, Place of Birth, Generation Status, Citizenship and Immigration Reference Guide, 2016 Census of Population

12. Figure 2 illustrates issue 7ii, where over 21% of 2016 Census respondents linked to an immigrant record since 1980 provided a year of immigration response which did not match their linked administrative value. This proportion increases the longer the linked respondents have been in Canada. Most of these differences in value are less than 5 years.

Figure 2: Match rate for census reported year of immigration by linked 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

13. Issue 7iii 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 persons 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.

14. Issue 7iv cannot be shown directly with links to administrative immigration values as it pertains to Canadian citizens by birth. However, among those who respond as Canadian citizens by naturalization and respond “No” to the immigrant status question, there is a subset which does not link to immigration records and has Canada as the place of birth of parents.

B. Limitations of administrative immigration data

15. As the department responsible for immigration in Canada, Immigration, Refugees and Citizenship Canada (IRCC) collects a range of data on immigrants and non-permanent residents in Canada. IRCC administrative data currently available to Statistics Canada includes:

- i. Detailed immigration records from 1980 to present,
- ii. Limited immigration records from 1952 to 1979, and
- iii. Non-permanent resident permit records from 1980 to present.

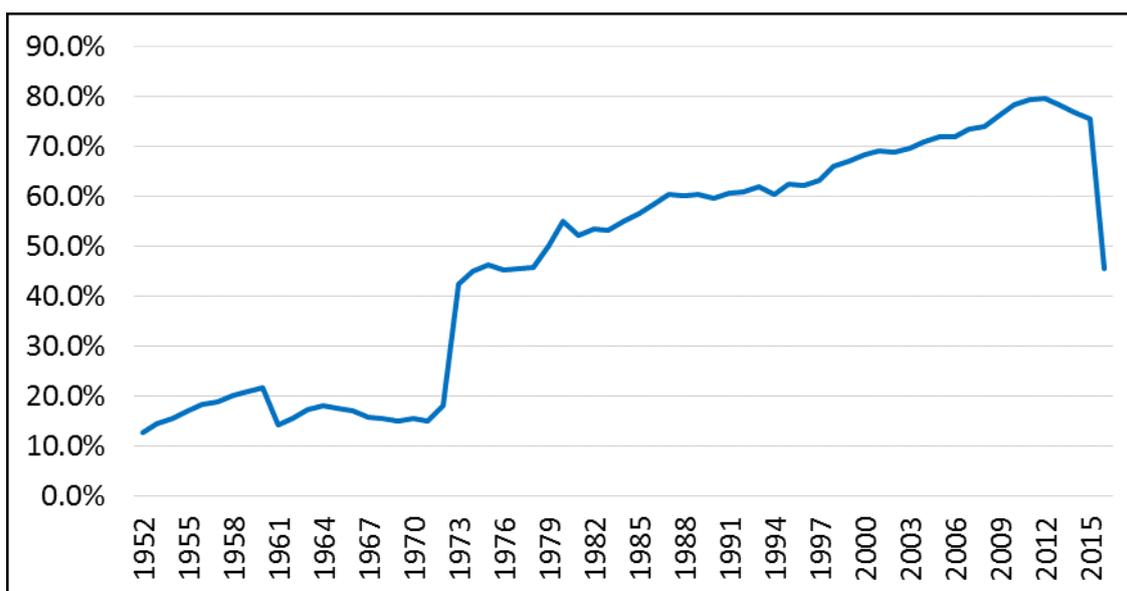
16. These files provide the actual administrative immigrant status and year of immigration, where applicable, for the immigrants and non-permanent residents they cover.

17. 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.

18. For the purposes of replacing the two immigration questions on the Census of Population, the coverage of IRCC data introduces two limitations. 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. Secondly, non-permanent resident records contain information on permit holders only; any accompanying family members who do not hold permits in their own right are not covered.

19. Beyond coverage, 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. In particular, records from 1961 to 1972 contain an incomplete date of birth which leads to lower linkage rates. Figure 3 demonstrates the percentage of the administrative immigration records linked to a 2016 Census respondent 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.

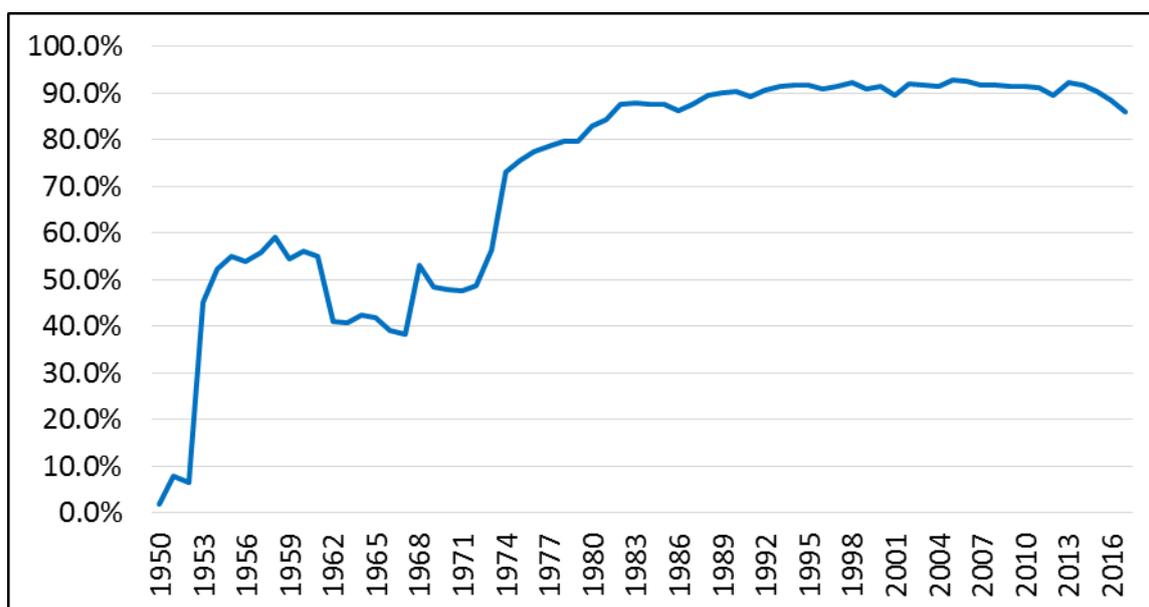
Figure 3: Percentage of administrative immigration records linked to the 2016 Census by administrative year of immigration



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

20. While Figure 3 demonstrates the linkage rates from the perspective of the administrative data, Figure 4 shows the perspective from the 2016 Census. Overall, 82.5% of immigrants according to the 2016 Census were linked to an IRCC administrative record. Figure 4 demonstrates how these linkage rates vary by census year of immigration, with higher linkage rates for immigrants since 1980.

Figure 4: Percentage of 2016 Census immigrants linked to administrative immigration records by 2016 Census year of immigration



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

21. The false positive rate for the linkage between the IRCC administrative data and the 2016 Census was estimated at 0.29% (Biernot 2017). While false positive links imply that two individuals were incorrectly linked, these cases may not necessarily have incorrect values for immigrant status and year of immigration.

III. Proposed methods for replacement and challenges

A. Record linkage and supplemented administrative data

22. Using similar methods that have been employed for past censuses (Biernot 2017; Brennan 2013; Brennan 2011), a record linkage between census or survey respondents and IRCC administrative data would take place immediately after collection, and before any edit and imputation processes. This record linkage would use all common linkage fields available, such as name, date of birth, address, etc.

23. All census respondents would be considered in scope for the purpose of the record linkage itself. That is to say that no records would be excluded from the record linkage based on how they responded to the census questions.

24. In order to resolve the absence of administrative records prior to 1952, the administrative files would be 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.

B. Edit and imputation methods

25. In general, the same edit and imputation methods used in the past can still be applied using linked administrative values in lieu of responses to questions. The important difference is that, in the context of a survey, item non-response is clear – a question is either answered or not. However, in the context of using linked data, the absence of a linked value could be the result of a missing link (or the equivalent of item non-response), or could be reflective of respondents who are out-of-scope for the linkage. In the specific case of immigrant status, a missing link could be a missing value or could be a valid value of “No”. This distinction is important in order to determine which records require imputation, and which do not.

26. While an effort will be made to link all respondents to the administrative data, responses to the citizenship question will be used to determine which individuals require imputation due to missing links. For those who are not linked, a citizenship response of Canadian citizen by birth will be considered to be non-immigrants and require no edit or imputation. Canadian citizens by naturalization and non-Canadian citizens will be considered in-scope and will require imputation if not linked.

27. In certain cases, there will be inconsistencies between linked administrative values and responses to census questions. This could include those who, for example, are linked but respond as Canadian citizens by birth, or those whose linked administrative year of immigration is prior to their response for year of birth. Consistency edits and imputation will be required to resolve these issues.

28. In the case of those who incorrectly respond that they are Canadian citizens by naturalization (when they are, in fact, Canadian citizens by birth), the combination of the absence of a link and place of birth of parents can be used to help resolve this issue.

29. In the case of non-permanent residents, there are no IRCC administrative records for their non-permit holding accompanying family members. However, where a link is made to only a subset of a family (i.e. permit holders only), imputation can be used to assign the same immigrant status value to all family members.

C. Key challenges

30. The principal challenges associated with replacing the census immigration questions with linked 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. The limited linkage 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).

31. Additionally, moving away from asking the questions and using the linked administrative values directly means that edit and imputation processes can no longer benefit from having highly correlated auxiliary variables, as was the case in the 2016 Census (Crowe 2017; McLeish 2017), and the linked data can no longer be used to certify the results. Finally, replacing the questions with linked 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. Preliminary testing

A. Testing environment

32. In order to study the possible impact of replacing the immigration questions with linked administrative data, a testing environment was created based on the 2016 Census. This parallel database included the original 2016 Census responses for every question except immigrant status and year of immigration, which instead came directly from the linked administrative data. This test environment is linked at the record-level to the original 2016 Census database in order to analyse how the results may differ.

33. The same edit and imputation processes as were run for the 2016 Census were applied to this test environment, with the few key exceptions described above. It is important to note that since imputation includes a random component, it is possible for imputed values to change each time the process is run even when keeping the inputs the same.

B. Preliminary results

34. Table 2 illustrates high-level results from preliminary testing for immigrant status and citizenship. Canadian citizens by birth or by naturalization (a subset of immigrants) are derived from the citizenship question. Permanent residents (not Canadian citizens but immigrants) and non-permanent residents (not Canadian citizens and not immigrants) are derived from both variables.

Table 2: Immigrant status and citizenship, 2016 Census and preliminary test results

Immigration status and citizenship	2016 Census	Admin. data replacement	Difference	
			Number	%
Canadian citizens by birth	26,412,610	26,415,100	2,490	0.0%
Immigrants, total	7,540,830	7,563,150	22,320	0.3%
Permanent residents	1,918,855	1,945,945	27,090	1.4%
Canadian citizens by naturalization	5,621,975	5,617,205	-4,770	-0.1%
Non-permanent residents	506,625	481,815	-24,810	-4.9%

Source: Statistics Canada, 2016 Census and preliminary test results using linked IRCC administrative immigration data

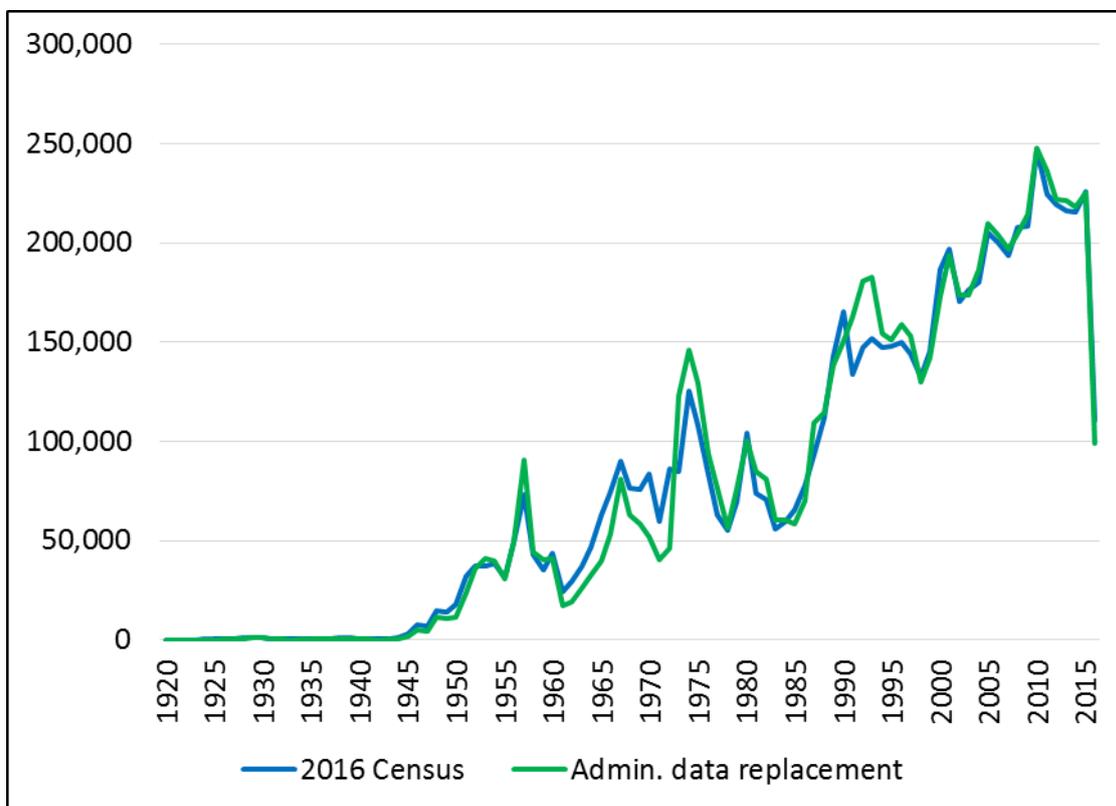
35. Overall, there is little difference between the 2016 Census and the test results for the number of Canadian citizens by birth. This is not unexpected, as this is still driven by the responses to the questionnaire, and is only affected in the context of this project when there is a conflict in the test (e.g. response of Canadian citizen by birth but a link to an immigration record) or when there is a conflict with the 2016 Census responses (e.g. response of Canadian citizen by birth but a response of 'Yes' to the immigrant status question).

36. The most notable preliminary finding for immigrant status and citizenship is the apparent shift from non-permanent residents to permanent residents in the test. It is not clear whether this could be the result of overestimation in the 2016 Census or underestimation due to the use of linked administrative data. Further investigation is required to better understand this shift.

37. Figure 5 demonstrates the comparison between the 2016 Census and the test results by single year of immigration. Overall, the resulting trends are similar. The test yielded lower

estimates for the years prior to 1952 and between 1961 and 1972. This consequently leads to higher estimates for the more recent years of immigration in the test, particularly in the 1970s.

Figure 5 Single year of immigration, 2016 Census and preliminary test results



Source: Statistics Canada, 2016 Census and preliminary test results using linked IRCC administrative immigration data

38. Between 1989 and 1993, responses to the question were more heavily distributed to 1989 and 1990, while the administrative values were more heavily distributed from 1991 to 1993. This is in line with expectations as well, due to respondents providing their year of arrival or asylum claim as opposed to their year of immigration.

V. Conclusion and next steps

39. This project seeks to replace census questions on immigrant status and year of immigration with linked administrative values. This is effectively taking the next step following the successful use of administrative immigration data to improve the edit and imputation processes for immigration variables and integrating new variables related to admission category into the 2016 Census. If successful, this will result in asking two fewer questions on the 2021 Census questionnaire, and it could yield better quality results and potentially lead to the integration of additional administrative variables (beyond admission category) such as pre-admission experience in Canada.

40. Before proposing the direction for the 2021 Census, a complete evaluation of the costs and benefits of replacing the immigration questions needs to be undertaken. This investigation must examine the differences in data quality for the two approaches, and seek to understand the possible

effects on historical comparability. To that end, a testing environment has been created and preliminary testing results have been obtained.

41. The success of using the administrative data will hinge on the methods used to mitigate the limitations of the data, and ensuring that responses to the citizenship question are as accurate as possible. The absence of records before 1952 has been somewhat alleviated by the use of prior census responses. The lower linkage quality for the period 1961-1972, however, remains an outstanding issue. Statistics Canada will continue to investigate methods to address these particular issues using the test environment to better understand the quality of the results.

42. The results of this project could have consequences beyond the Census of Population. If successful, other household surveys at Statistics Canada may 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).

VI. Acknowledgments

43. This project would not be possible without support and collaboration from many individuals representing different areas of expertise including subject-matter analysts, methodologists and database managers. In particular, Statistics Canada would like to acknowledge the ongoing contributions from IRCC including but limited to their ongoing support in interpreting their administrative data. Within Statistics Canada, the authors would like to specifically recognize the ongoing efforts of Kathryn Spence, Tristan Cayn and Eric Mongrain from Social and Aboriginal Statistics Division, Sean Crowe, Caroline Cauchon, Andrew Stelmack and Lyne Guertin from Social Survey Methods Division, and many others who continue to contribute to the development of immigration content for the 2021 Census of Population.

VII. References

44. Biernot, P. 2017. "External linkage between the Immigration File (1952-2016) and the 2016 Census Response Database." Unpublished report. Ottawa: Statistics Canada.
45. Brennan, J. 2011. "CIC Landing File to Census 2006 Linkage." Unpublished report. Ottawa: Statistics Canada.
46. Brennan, J. 2013. "CIC Landing File to Census 2011/NHS Linkage." Unpublished report. Ottawa: Statistics Canada.
47. Crowe, S. and Janes, D. 2017. "Edit and Imputation Report: Ethnocultural Process." Unpublished report. Ottawa: Statistics Canada.
48. 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.
49. 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.
50. 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.

51. Statistics Canada. 2016. "2016 Census 2A-L Questionnaire" 2016 Census of Population. Ottawa: Statistics Canada.
52. Statistics Canada. 2017. "Income Reference Guide." 2016 Census of Population. Catalogue no. 98-500-X2016004. Ottawa: Statistics Canada.
53. Statistics Canada. 2017. "Place of Birth, Generation Status, Citizenship and Immigration Reference Guide." 2016 Census of Population. Catalogue no. 98-500-X2016007. Ottawa: Statistics Canada.

—