

Distr.: General
8 April 2016
English

Economic Commission for Europe

Conference of European Statisticians

Work Session on Migration Statistics

Geneva, Switzerland

18-20 May 2016

Item 8 of the provisional agenda

Emerging issues, including new methods to estimate and project migration

Adding immigrant admission category to the Canadian census of population

Note by Statistics Canada*

Abstract

The Canadian Census of Population provides detailed information on the demographic, social and economic characteristics of immigrants living in Canada. Immigration variables have included immigrant status, year of immigration, place of birth, and citizenship. However, information regarding the conditions under which they immigrated is not collected. For the 2016 Census of Population, Statistics Canada is undertaking a project to add immigrant admission category (e.g., economic immigrants, refugees) through a record linkage to administrative immigration data. The results from this record linkage will be used to better understand the socio-economic outcomes of immigrants and also to support and evaluate immigration policies and programs in Canada.

This paper will provide a brief overview of this project and outline the methods being considered to address resulting inconsistencies between the linked administrative variables and self-reported variables traditionally collected on the Census.

I. Introduction

1. The 2016 Canadian Census of Population will collect information regarding the residents of Canada on a range of socio-economic topics, including demography, culture,

*Prepared by Scott McLeish, Senior Research Analyst, Social and Aboriginal Statistics Division

education, labour, income, and housing. Related to immigration, questions will be asked on immigrant status, year of immigration, citizenship and place of birth. Due to the breadth of topics and the sample size, the Canadian Census (and the 2011 National Household Survey) has long been an important data source for researchers studying the outcomes of immigrants to Canada. However, the Canadian Census has never included a question regarding immigrant admission category, or the conditions under which they landed in Canada. In preparation for the 2016 Census of Population, Immigration, Refugees, and Citizenship Canada (IRCC) has requested that Statistics Canada explore the feasibility of including variables related to immigrant admission category, processing them along with other census variables, and disseminating the results to the general public.

2. Immigrant admission category refers to the program (or group of programs) under which immigrants were granted permanent residency in Canada. They mostly fall into 3 main streams: Economic immigrants, Family reunification, and Refugees. Immigrant admission category provides important context that would affect an immigrant's short and long term outcomes after landing. Economic immigrants are selected to contribute to Canada's economy (e.g., as skilled workers, investors, live-in caregivers, etc.) based on criteria such as education, or knowledge of official language. On the other hand, other immigrants land to reunite with family or are refugees seeking sanctuary and may not speak English or French, or have a post-secondary education at arrival (Government of Canada 2001).
3. Including questions regarding admission category can be difficult to implement in a social survey, as many respondents may not know or remember the specific program under which they were granted permanent residency (especially in cases where immigrants landed many years prior to survey collection or where they landed as children under their parents' admission category).
4. IRCC, who is the department responsible for immigration in Canada, collects data on immigrants landing in Canada from their permanent residency visa application¹. This includes, but is not limited to, the admission category under which they were granted permanent residency. IRCC publishes statistics from this administrative data source on a regular basis (for example: Citizenship and Immigration Canada 2015). This administrative immigration data is also used by Statistics Canada, including for the production of demographic population estimates (Statistics Canada 2016) and the Longitudinal Immigration Database (IMDB).
5. The IMDB links an administrative census of immigrants who have landed since 1980 to annual tax records since 1982, allowing researchers to study the short and long-term economic outcomes of immigrants in Canada by their characteristics at landing (including their admission category). For example, the 2013 IMDB release showed that the median employment income in 2013 of immigrant taxfilers who landed in Canada since 1980 as principal applicant skilled workers was approximately \$48,000 compared with \$29,000 for refugees (Statistics Canada 2015).

¹ While IRCC administrative data provides detailed information on immigrants when they land in Canada, it cannot be used to estimate the population of immigrants living in Canada at a specific point in time as there is no follow-up data collection.

6. Adding immigrant admission category to the 2016 Census would provide new analytical opportunities not already available with the IMDB. This project would supplement the IMDB outcomes which are restricted to tax variables such as income, with information on language, labour force status, current occupation, level of education, housing, and detailed living arrangements.
7. This paper will provide an overview of the 2016 Census project to study the feasibility of including immigrant admission category variables. This will include an examination of similar record linkages that have taken place in the past, an outline of the activities required to implement the project for the 2016 Census, and some of the methods being explored for processing the variables.

II. 2011 National Household Survey linked with administrative immigration records

A. Background

8. The 2011 National Household Survey (NHS) was a large sample household survey which collected similar content as the 2006 Census (long form). The NHS took place concurrently with the 2011 Census of Population, and used the same frame and reference date (May 10, 2011). While the 2011 Census continued to include questions on demographic characteristics such as age, sex, mother tongue, and living arrangements, variables related to immigration were only collected on the NHS.
9. In order to assess the quality of NHS variables, including the effects of non-response, there was a need for auxiliary variables (McLeish 2014). Since the 2011 NHS and the 2011 Census had a common frame, the two surveys were linked at the person-level, permitting analysis, using census variables, of the differences between NHS respondents and non-respondents. This linkage also facilitated record linkages between administrative data and the NHS (respondents and non-respondents) via the 2011 Census.
10. Building on a feasibility study with the 2006 Census (Brennan 2011, McLeish 2012), a record linkage was undertaken between the 2011 Census and IRCC administrative immigration data (ILF) covering those who landed in Canada since 1980. This made it possible to study certain immigration variables for NHS respondents and non-respondents (e.g. year of immigration and place of birth) which were available on the administrative data. Approximately 86% of NHS respondents who indicated they were in-scope (immigrated to Canada since 1980) were linked to an administrative record.

B. Record linkage quality

11. Building off of the similar exercise done for feasibility study purposes using the 2006 Census, the 2011 Census/NHS – ILF linkage underwent an evaluation which looked at the risk of both false negative (missed) and false positive (incorrect) links. The methods used for the record linkage were conservative, and a qualitative review of the links found little evidence of false positives (Brennan 2013, McLeish 2012).
12. One possible approach to assessing the quality of the record linkage was to use the NHS results. The NHS identified the population in-scope for the linkage with the question on year of immigration. As noted above, using responses to this question for the NHS, roughly 86% of these cases were linked (implying approximately 14% were missed).

13. However, this approach combines different sources of error. As demonstrated (McLeish 2014), there is evidence that respondents do not always provide the precise year of immigration, and could be responding as in scope (or out of scope) incorrectly. Further, the year of immigration variable had a 12.5% imputation rate for the 2011 NHS (Statistics Canada 2013). Imputed values could compound the issue presented by response error and inflate any estimates of false negative or positive links.
14. On the other hand, if interest lies in studying the admission category for immigrants using NHS variables, the total inconsistency between the record linkage results and the NHS values (after imputation) should be considered. Overall, after imputation and using final NHS weights, about 79% of immigrants who landed since 1980 were linked to the ILF, where about 3% of the links made were out of scope according to the NHS.
15. Table 1 demonstrates how these ‘record linkage rates’ can vary for different NHS variables, such as place of birth. Recall that some of these discrepancies are caused by errors due to NHS response or imputation, while others are caused by the record linkage itself.

Table 1 Percentage linked to IRCC immigration data by place of birth for immigrants who landed 1980-2011, Canada

Place of birth of respondent	% Linked to IRCC immigration data
Total - Place of birth of respondent	79.2%
Americas	68.6%
North America	53.8%
Central America	65.5%
Caribbean and Bermuda	70.6%
South America	78.2%
Europe	80.1%
Western Europe	76.3%
Eastern Europe	86.2%
Northern Europe	73.6%
Southern Europe	76.0%
Africa	80.6%
Western Africa	77.9%
Eastern Africa	74.7%
Northern Africa	86.2%
Central Africa	78.9%
Southern Africa	83.6%
Asia	81.8%
West Central Asia and the Middle East	82.1%
Eastern Asia	81.8%
Southeast Asia	81.2%
Southern Asia	82.1%
Oceania	64.8%

Source: Statistics Canada, 2011 National Household Survey linked with IRCC immigration data.

16. Table 2 demonstrates the differences between labour force rates for immigrants who were linked and the in-scope population according to the NHS values for year of immigration. Some differences are observed such as overall stronger rates for the linked population.

Table 2 Participation rate, employment rate, and unemployment rate for immigrants aged 15 and over who landed between 1980 and 2011 or who landed between 2006 and 2011 by linkage status, Canada

Period of immigration and linkage status	Participation rate	Employment rate	Unemployment rate
Immigrants landed between 1980 and 2011 - Total	70.6%	64.1%	9.2%
Immigrants landed between 1980 and 2011 - Linked	71.7%	65.1%	9.1%
Immigrants landed between 2006 and 2011 - Total	66.4%	57.6%	13.3%
Immigrants landed between 2006 and 2011 - Linked	67.4%	58.5%	13.2%

Source: Statistics Canada, 2011 National Household Survey linked with IRCC immigration data.

C. Use for analysis

17. IRCC approached Statistics Canada to explore the feasibility of building an analytical database using the record linkage results between the immigration administrative data (ILF) and the 2011 NHS. The goal of this linked database was to study NHS outcomes for linked immigrants by their characteristics at landing, including admission category. This database was designed to mirror the NHS, therefore all linkage results were conditional upon being in-scope according to final NHS values for year of immigration. In addition, NHS weights and dissemination procedures were applied.
18. The differences between the linkage results and the NHS were not addressed with any supplemental processing or re-weighting, meaning that no adjustment was made for those immigrants who landed since 1980 and were not linked to the ILF. Instead, a 'Not linked' category was included in all tabulations. Links made to NHS respondents deemed out of scope were excluded from the database.
19. Following the creation of the database, Statistics Canada prepared over 150 custom tabulations for IRCC. In addition, analysis resulting from this linkage has covered topics such as labour force status and occupation of economic immigrants (Jantzen 2015) and outcomes of childhood immigrants by admission category (Hou and Bonikowska 2016).
20. Table 3 provides an example of a tabulation that would be produced for IRCC. The variable coming from the ILF is immigrant admission category. It is embedded in NHS variables which define the in-scope population as immigrants landed since 1980, and includes a category for those who are not linked. The actual immigrant admission category variables are only available for those who were linked.

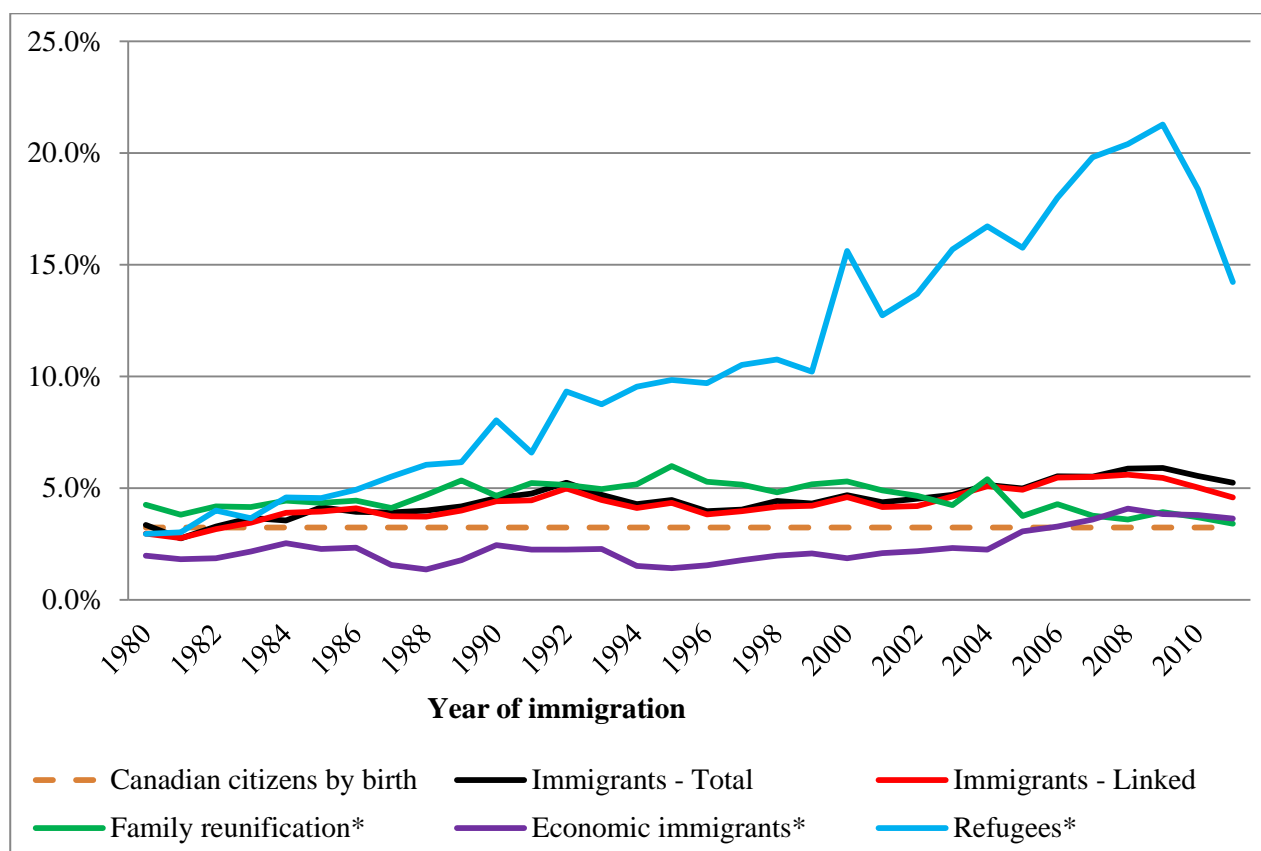
Table 3 Immigrant status, period of immigration and immigrant admission category, Canada

Immigrant status, period of immigration and immigrant admission category	Frequency
Total - Immigration admission category	32,852,325
Canadian citizens by birth	25,720,175
Immigrants landed before 1980	2,021,335
Immigrants landed 1980-2011	4,754,435
Immigrants landed 1980-2011 - Not linked	988,935
Immigrants landed 1980-2011 - Linked	3,765,495
Economic immigrants (linked immigrants)	1,970,350
Family class (linked immigrants)	1,095,965
Refugees (linked immigrants)	532,960
Other immigrants (linked immigrants)	166,115

Source: Statistics Canada, 2011 National Household Survey linked with IRCC immigration data.

21. Figure 1 provides another example of a tabulation possible from this database. Similar to table 3, immigrant admission category is taken from the ILF. The figure shows the percentage of the population living in subsidized housing by year of immigration and immigrant admission category. The outcomes for Canadian citizens by birth are shown as a comparison. Again, the characteristics of the linked and the in-scope population can be compared. It can be observed that for this particular outcome, by year of immigration, the linked and in-scope population have similar results.

Figure 1 Percentage living in subsidized housing by year of immigration and immigrant admission category, Canada.



* Linked to IRCC immigration data

Source: 2011 National Household Survey linked with IRCC immigration data.

III. 2016 Census of Population project overview

A. Project outline

22. In late 2014, IRCC and Statistics Canada began studying the feasibility of adding variables on immigrant admission category to the 2016 Census of Population (long form). These would include the admission category and the applicant type (to identify the principal applicants) of each immigrant from their immigration (permanent resident) visa application. This would follow a record linkage between the 2016 Census and the IRCC immigration data, as well as build on the 2011 NHS linkage project in several ways:
 - i. Address inconsistencies between the linked values and the results collected from census respondents.
 - ii. Impute missing values for those in-scope immigrants who were not linked.
 - iii. Prepare and publish reference material on the record linkage and the results.
 - iv. Make the variables available to the general public within the 2016 Census database.

23. The admission category variables would be the only ones integrated into the 2016 Census in such a manner, but it would be possible to explore including other variables, such as intended occupation, from IRCC administrative records in the future.

B. Challenges and required activities

24. Adding immigrant admission category variables to the 2016 Census of Population does come with a suite of challenges and required activities.
25. All record linkage projects at Statistics Canada must follow a prescribed review and approval process. The record linkage proposal to add immigrant admission category variables to the 2016 Census of Population from IRCC immigration data was approved by the Chief Statistician of Canada in December 2015. More information on record linkages at Statistics Canada can be found on the official website:
<http://www.statcan.gc.ca/eng/record/gen>
26. Adding a new variable to the 2016 Census of Population cannot be accomplished in a vacuum, and possible impacts on the traditionally collected census variables need to be assessed as well. This project has intended, from the beginning, to have as little impact as possible on the production and dissemination of the other census variables.
27. As a result, it was determined that the immigrant admission category variables would be processed conditionally on the final values from the other variables. This implied that some of the discrepancies associated with imputing year of immigration would need to be addressed when editing or imputing immigrant admission category. Therefore, the edit and imputation methods for the immigrant admission category variables will be addressing census response and imputation error as well as linkage error.
28. As with census variables, it is important that the categories are well understood by a general audience of users, and not just policy analysts. Statistics Canada and IRCC have been collaborating on developing final categories and definitions for these variables including plain language labels in English and French. In addition, reference material is being developed, which will provide brief descriptions of each category, background information on the record linkage, and the IRCC administrative data itself. More detailed information about each category will be available from IRCC directly.
29. Considering IRCC publishes figures on the number of immigrants landing in Canada every year, by admission category, it is important to develop clear guidelines explaining the differences between the two sources of data (e.g. survey estimates of current population versus administrative counts of inflow).
30. In addition, a comprehensive certification strategy is being developed to assess the quality and confront and validate the results. While the record linkage itself will undergo its own evaluation, the edit and imputation processes will also need to be examined. Confrontation sources will include the IRCC immigration data (in isolation), past linkage results, and the IMDB.
31. In the end, the certification results will determine the dissemination strategy for these variables, and the breadth to which the results will be published.

IV. Addressing inconsistencies between census responses and administrative values

A. Incoherencies with linked administrative values

32. Edit rules are being developed to address inconsistencies between linked values for immigrant admission category and the census values for other variables. For example, if a program did not come into force until 2008, then any year of immigration prior to 2008 would be an inconsistency. As mentioned in ‘Challenges and required activities’, these processes are conditional on the final census values, which means that any inconsistencies are addressed by changing the linked value.
33. Edit rules are largely being developed based on the IRCC immigration data, which includes several variables in common with the 2016 Census such as age at landing, sex, place of birth, and year of immigration. Proposed rules are being assessed against IRCC’s own business rules to ensure consistency.
34. All proposed edits are being tested using the 2011 NHS linkage results.

B. Missing administrative values

35. Perhaps the most important exercise for this project is addressing the missing immigrant admission category values for the in-scope population. As described above, these cases could be the result of linkage error (respondent was in scope but failed to link), response error (respondent incorrectly identifies as in-scope), or imputation error (respondent is incorrectly imputed as in-scope). If these errors remain similar to the 2011 NHS, an imputation rate of roughly 21% can be expected for these variables.
36. Different strategies are being considered to mitigate the relatively large imputation rate, including less conservative linkage methods. Alternatively, the 2016 Census will employ linked IRCC variables (including year of immigration) when imputing year of immigration. This should improve the quality of the imputation (at least for linked records) and decrease the number of cases falsely imputed as in-scope.
37. For those that will be imputed, CANCEIS (Canadian Census Edit and Imputation System) will be used. CANCEIS employs nearest-neighbour donor imputation methods (Guertin 2014), and it is the official system used to process all census variables.
38. For immigrant admission category variables, donor imputation is being stratified by living arrangements: persons not in a census family, lone parents, couples (both in scope), couples (one in scope), children of lone parent in scope, children of a lone parent out of scope, children of couples (at least one in scope), children of couples (neither in scope). This stratification is used to ensure that familial associations are maintained (especially since many families land together), and to borrow strength from already linked family members.
39. However, it is important to note that living arrangements evolve and many immigrants may not be living in the same family in 2016 as they did when they landed. Aside from living arrangements, other variables being considered for donor selection include:
 - i. Year of immigration;
 - ii. Place of birth;

- iii. Age at immigration;
 - iv. Sex;
 - v. Province of residence;
 - vi. Location of study (inside or outside Canada);
 - vii. Marital status;
 - viii. Level of education;
 - ix. Current occupation skill level; and
 - x. Sources of income (e.g., employment, investment, self-employment, government, etc.).
40. Determining the use and relative importance of auxiliary variables for donor selections is a significant component of this project. While subject-matter input plays an important role, results from different measures of association, such as adjusted odds ratios, are being considered as well.
41. The imputation processes are being tested using the 2011 NHS linkage results. This includes validating against the observed values for a test data set. A report on the edit and imputation methods and the testing results will be prepared and used to inform the final decision on dissemination strategy for the immigrant admission category variables.

V. Conclusion and next steps

42. This project builds on two prior record linkages: 2006 Census for feasibility study purposes and the 2011 Census and NHS. This project would go beyond what was done for the 2011 NHS linkage, by processing admission category variables, addressing inconsistencies and missing values, disseminating the data to a broad range of users, and developing reference material to support such general usage.
43. Perhaps the largest challenge of the project is the imputation strategy, which will use CANCEIS, the official system for edit and imputation processes for the Canadian Census of Population. Missing values, in this case, are caused by census response and imputation error for year of immigration combined with record linkage error. With a possible imputation rate of roughly 21%, the accuracy of the donor imputation will have a large impact on the final processed results.
44. The edit and imputation methods are continuing to be developed and tested, and will be finalized in 2016.
45. Other ongoing activities include the development of standard variable categories and labels, preparation of reference materials, and finalizing the certification strategy for the final results. The record linkage itself is scheduled to take place in the fall of 2016 once collection activities have been completed. The results from this linkage will be used to assist processing of other census variables including year of immigration and place of birth.
46. Adding immigrant admission category variables to the 2016 Canadian Census of Population would introduce new analytical possibilities about immigrants living in Canada. This project would allow data users to better understand the associations between an immigrant's conditions at landing and their short- and long-term socio-economic outcomes.

VI. Acknowledgments

47. 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 of Lorna Jantzen and her colleagues at IRCC who pursued this project and continue to provide important guidance on the use and interpretation of their administrative data. Within Statistics Canada, staff from Social and Aboriginal Statistics Division, Social Survey Methods Division, Household Survey Methods Division, Census Subject-Matter Secretariat, and Census Operations Division have all provided input to this endeavour. The author would like to recognize the ongoing efforts of Eric Mongrain, Andrew Stelmack, H el ene Maheux, Laetitia Martin, Mireille V ezina, and many others who continue to move this project forward.

VII. References

48. Brennan, J. 2011. "CIC Landing File to Census 2006 Linkage." Unpublished report. Ottawa: Statistics Canada.
49. Brennan, J. 2013. "CIC Landing File to Census 2011/NHS Linkage." Unpublished report. Ottawa: Statistics Canada.
50. Citizenship and Immigration Canada. 2015. "Facts and Figures 2014: Immigrant Overview - Permanent Residents 2014." Available at: www.cic.gc.ca/english/resources/statistics/facts2014/index.asp.
51. 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.
52. Hou, F. and Bonikowska, A. 2016. "Educational and Labour Market Outcomes of Childhood Immigrants by Immigration Class." *Analytical Studies Branch Research Paper Series 377*. Catalogue no. 11F0019M. Ottawa: Statistics Canada.
53. Jantzen, L. 2015. "Do Economic Principal Applicants Work in Their Intended Regulated Occupation? Introducing the National Household Survey and Immigration Landing File Linkage Database." *Canadian Public Policy* 41 (Supplement 1): S48-S63.
54. Government of Canada. 2001. *Immigration and Refugee Protection Act (IRPA)*. Available at: laws-lois.justice.gc.ca/eng/acts/I-2.5/.
55. McLeish, S. 2012. "Feasibility Study: IMDB for Statistical Purposes." Unpublished report. Ottawa: Statistics Canada.
56. 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.
57. Statistics Canada. 2013. "Place of Birth, Generation Status, Citizenship and Immigration Reference Guide." 2011 National Household Survey. Catalogue no. 99-010-XIE2011008. Ottawa: Statistics Canada.

58. Statistics Canada. 2015. "Income and mobility of immigrants, 2013." *The Daily*, Catalogue no. 11-001-X. Ottawa: Statistics Canada.
 59. Statistics Canada. 2016. *Quarterly Demographic Estimates*. Catalogue no. 91-002-X. Ottawa: Statistics Canada.
-