

National Data on Immigrants: An Overview

By Ray D. Bollman (RayD.Bollman@sasktel.net)

Highlights

- ◆ *Citizenship and Immigration Canada (CIC) collects data on immigrants when they land in Canada and, for temporary residents, CIC collects data when the individual enters Canada.*
- ◆ *Statistics Canada collects data on censuses and surveys that are reported by individuals.*
- ◆ *In recent years, a number of datasets have been assembled by linking data from more than one source. One database, the Longitudinal Immigrant Database (IMDB) matches the information reported by immigrants when they land with their annual taxfiler data in each year after their admission to Canada.*

Why: An overview of national data

Immigrants represented 21% of Canada's population in 2011. Population projections indicate that immigration will be the key factor of Canadian population growth in the coming years as the number of deaths surpasses the number of births¹. Understanding the role of immigration in Canada is important for both national and community-level policy discussions. Fortunately, a growing number of data-sets are available to assist researchers, policymakers and analysts working in community-based organizations.

The objective of this FactSheet is to provide an overview of national datasets that are helpful for understanding the place of immigrants in Canadian society. This information is based on three presentations² at the November 2014 Pathways to Prosperity conference.

Our discussion is organized in three parts:

- ◆ Data collected at time of arrival in Canada;
- ◆ Data collected from individuals in censuses and surveys; and
- ◆ Data assembled from more than one source.

Data collected at time of arrival in Canada

Citizenship and Immigration Canada (CIC) collects considerable information on immigrants at the time they land in Canada. Summary data are published [here](#).

Some additional data are [here](#).

One list of variables is in the [Glossary of Terms and Concepts for Permanent and Temporary Residents](#).

Another list of variables is in the [Glossary of Terms and Concepts for the CIC administrative process](#).

1. See the 'low-growth' scenario in Statistics Canada. (2014). [Population Projections for Canada \(2013 to 2063\), Provinces and Territories \(2013 to 2038\): Technical Report on Methodology and Assumptions](#) (Ottawa: Statistics Canada. Catalogue number 91-520).

2. Bollman, Ray D. (2014). **National Data on Immigrants: An Overview**. Presentation to the 2014 Pathways to Prosperity Annual Conference, Montreal, November 24; Jantzen, Lorna. (2014). **Moving forward along CIC's data roadmap**. Presentation to the 2014 Pathways to Prosperity Annual Conference, Montreal, November 24; and Barayandema, Athanase. (2014). **Use of the Longitudinal Immigration Database (IMDB) to Inform Immigration Research**, Presentation to the 2014 Pathways to Prosperity Annual Conference, Montreal, November 24. These presentations may be accessed [here](#).

Only a few of these variables are tabulated in the summary data referenced above. The general publication program of CIC data is summarized in Table 1.

Data product	Method of publication	Data of publication
Preliminary estimates	CIC website	Late February*
'Facts and Figures' report	CIC website	Late August*
Digital Library	DVD-ROM	Late September*
Provincial Profiles	Data cubes and PDF	Late September*
Permanent Resident rounded data	Interactive data cubes	Late October*
Quarterly administrative data release	Government of Canada Open Data Portal	3 months after each quarter

* Data refer to previous calendar year

Data collected from the individual

These data are collected on censuses by Statistics Canada and on surveys by Statistics Canada and other organizations, including the P2P.

The largest dataset, which allows data to be tabulated for detailed sub-provincial geographies, is the Form 2B (long-form) Census of Population for 1971 and every five years from 1981 to 2006. This was replaced by the **National Household Survey (NHS)** in 2011. The NHS was a 1/3 sample of all households in the 2011 Census of Population. The questionnaire is [here](#). Immigration status was enumerated in Question 11. Data (including data on immigrants) for each census subdivision³ are published by Statistics Canada. Go to [StatCan website](#) and search (in the upper right-hand box) for a census subdivision by name. Also, 12 tables of variables cross-tabulated with immigrant status are available [here](#). All tables are tabulated for Canada and Provinces and Territories. Some are tabulated for sub-provincial areas. The Census Profile that presents basic demographic data from the 2011 Census of Population for various levels of geography is published [here](#). Special tabulations may be requested from infostats@statcan.gc.ca. Researchers can access the micro-data file at a Statistics Canada Research Data Centre. Also, an NHS Public Use Microdata File (PUMF) is available [here](#).

In January, 2006, Statistics Canada's monthly **Labour Force Survey (LFS)** added four questions in order to identify the immigrant population:

- ◆ country of birth of the respondent;
- ◆ whether or not the respondent was a "landed immigrant";
- ◆ the year the respondent became a landed immigrant (and the month requested for immigrants who arrived within 5 years of the date of the survey enumeration); and
- ◆ the country where the respondent received his/her highest level of education.

These questions are comparable to those used in the 2006 long-form Census questionnaire and the 2011 National Household Survey.

Data tables can be downloaded free from Statistics Canada's CANSIM database by searching for

3. A census subdivision is an incorporated town or municipality or equivalent, such as Local Government District or Indian Reserve. For details, see [this link](#).

“immigrant” [here](#). Data are tabulated for a 3-month moving average and thus the first data point is March, 2006. For further information and for requests for special tabulations, contact labour@statcan.gc.ca. Also, researchers can access the micro-data files at a Statistics Canada Research Data Centre and an LFS PUMF is available [here](#).

Statistics Canada’s **Canadian Community Health Survey (CCHS)** aims:

- ◆ to support health surveillance programs by providing health data at the national, provincial and intra-provincial levels;
- ◆ to provide a single data source for health research on small populations and rare characteristics;
- ◆ to provide a timely release of information easily accessible to a diverse community of users; and
- ◆ to create a flexible survey instrument that includes a rapid response option to address emerging issues related to the health of the population.

For details, see [this link](#).

The CCHS is a cross-sectional survey that collects information related to:

- ◆ health status;
- ◆ health care utilization; and
- ◆ health determinants.

Note that the so-called “social determinants of health” can be analyzed specifically for the immigrant population.

The CCHS had a sample of approximately 130,000 respondents during the reference periods of 2001, 2003 and 2005. The sample size was changed to 65,000 respondents each year starting in 2007. For a summary of changes over time, see [here](#). For the details for each time period, see [here](#).

Generally, each questionnaire asks:

- ◆ In what country were you born?
- ◆ In what year did you first come to Canada to live?
- ◆ Plus questions on the size of the household, marital status, self-declared health status, life satisfaction, level of stress, sense of belonging to your local community, utilization of health care services, physical activities, smoking, alcohol use, industry and occupation of job, educational attainment, income level, whether dwelling is owned or rented, etc.

See the complete questionnaire for 2012 [here](#).

CIC is examining the feasibility of a linkage between the CCHS and the Immigration Landing File.

For published CCHS data (including some data by health regions), search for “health” on Statistics Canada’s CANSIM database ([link](#)). Thematic maps showing CCHS results by health region are available [here](#).

For information on obtaining special tabulations or accessing the CCHS, contact cchs-esc@statcan.gc.ca. Also, researchers can access CCHS microdata at Statistics Canada’s Research Data Centres and a CCHS PUMF is available [here](#).

Statistics Canada’s **General Social Survey (GSS)** was established in 1985 as a series of

independent, annual, cross-sectional surveys, each covering one topic in-depth. The survey program aims to gather data on social trends in order to monitor changes in the living conditions and the well-being of Canadians and to provide information on specific social policy issues. For an overview of this annual survey, see [this link](#) where Appendix A lists the focus of each annual survey since 1985.

Regarding immigrants, this survey has asked:

- ◆ In what country were you born?
- ◆ In what year did you first come to Canada to live?
- ◆ Are you now, or have you ever been, a landed immigrant in Canada? and
- ◆ In what year did you first become a landed immigrant in Canada?

The focus of the 2014 GSS was "victimization." See the complete 2014 questionnaire [here](#). The focus of the 2013 GSS was "giving, volunteering and participating." See the complete 2013 questionnaire [here](#).

The focus of the 2012 GSS was "social support and aging." See the complete 2012 questionnaire [here](#). Requests for special tabulations may be forwarded to sasd-dssea@statcan.gc.ca. Researchers can access the micro-data at Statistics Canada's Research Data Centres and GSS PUMF files are available upon request to sasd-dssea@statcan.gc.ca.

Statistics Canada's new **Canadian Income Survey** (CIS) for 2012 is designed to collect cross-sectional data to replace the cancelled Survey of Labour and Income Dynamics (SLID). The questionnaire can be seen [here](#). The CIS is based on the LFS. Data for each respondent to the CIS is combined with the respondent's data on the LFS. Thus, all variables, including immigrant status, are available for each respondent to the CIS. The CIS enumerates some additional context questions but the main income data are obtained from a link to taxfiler data. Special tabulations may be requested from infostats@statcan.gc.ca. Researchers may access microdata at a Statistics Canada Research Data Centre and a PUMF is available (at this point, only for historical SLID data).

Data assembled from more than one source

The **Longitudinal Immigration Database** (IMDB) was created to respond to the need for detailed and reliable data on the performance and impact of Canada's national and provincial immigration programs. The database is managed by Statistics Canada on behalf of a federal-provincial consortium led by CIC (See [here](#).)

The IMDB is assembled for each landed immigrant from the Immigrant Landing File (collected from the landing form) and from personal income tax data (plus additional family data added from Statistics Canada's T1FF: T1 Family File). It is designed to be a 100% census of landed immigrants with longitudinal annual income data from taxfiler records. However, present linkage methodologies are not able to attain a complete linkage rate. (See also [CIC link](#).)

The IMDB covers all immigrants who have landed since 1980 and who filed an income tax form at least once since 1982. The IMDB is updated annually and, perhaps obviously, is limited to tax-filing immigrants who were linked by the current linkage methodologies. The file contains detailed characteristics at landing such as admission category, education qualifications, language, and country of origin plus detailed tax information such as sources of income, type of deductions and family structure. There is a detailed geographic coding of the intended

destination for settlement and each tax record carries the postal code of residence at the time the income tax form was submitted. Thus, analysts are able to follow immigrant economic outcomes over time and inter- and intra-provincial mobility. The IMDB uses hierarchical deterministic matching which generates very few false positive matches and provides relatively high linkage rates — covering about 87% of all immigrants. For working age adults, the matching rate is in the order of 90% (which is approaching the expected maximum linkage rate).

Analysts can submit requests for multivariate analysis of the microdata by first testing the program on a synthetic file to test the SAS code and submitting the final request to the IMDB project team at Statistics Canada (sasd-dssea@statcan.gc.ca) where the SAS code will be executed on the real database, the outputs will be vetted for confidentiality, and the results will be returned to the researcher. Alternatively, analysts may choose to access the microdata at the Ottawa office of Statistics Canada. This requires that the individual become a deemed employee of Statistics Canada, requiring the appropriate documents to be signed.

The **Canadian Employer-Employee Dynamic Database** (CEEDD) is another recent initiative within Statistics Canada to assemble data from various sources into one database. Information at the level of the business-enterprise will be drawn from the National Accounts Longitudinal Micro-data File (NALMF) while individual- and job-level data will be drawn from T1 files, the T4 Statement of Remuneration Paid file, the Record of Employment (ROE) file, the Longitudinal Immigration Data Base (IMDB), and the Temporary Foreign Work file, for the years 1999 onward. This database will support analytic projects on business start-ups and job creation, with particular emphasis on the role of immigrant entrepreneurs; the distribution of immigrants across business enterprises and how this differs from the distribution of Canadian-born workers; how workforce aging is playing out within business enterprises, including its effect on labour productivity; local labour market information, including hiring rates, separation rates, layoff rates, and aggregate turnover rates within sub-provincial regions and the impacts of organizational changes, such as mergers and acquisitions, on individual-level outcomes. To access this database, contact Statistics Canada's Canadian Centre for Data Development and Economic Research (CDER) (see [this link](#)).

Another initiative to assemble data from various sources into one database is the 2011 **National Household Survey-Immigration Landing File** (NHS-ILF). This database is a record linkage between the 2011 NHS and the Citizenship and Immigration Canada (CIC) Landing File and permits detailed analysis of socio-economic outcomes of immigrants to Canada by entrance characteristics such as admission category (e.g. refugee, family class, etc.). Requests for special tabulations may be directed to the Social and Aboriginal Statistics Division at Statistics Canada at sasd-dssea@statcan.gc.ca.

Another linkage of data to create a new dataset is the **Linkage of the 2008 General Social Survey (GSS) Cycle 22, Social Networks, and the Longitudinal Immigrant Database**. This database permits the analysis of social outcomes (such as civic participation) of immigrants to Canada by entrance characteristics such as admission category (e.g. refugee, family class, etc.). For information on accessing these data, contact the Social and Aboriginal Statistics Division at Statistics Canada at sasd-dssea@statcan.gc.ca.

Comparing data from more than one source

CIC's RDM (Research Data Mart) provides special tabulations for many variables, including the number of landed immigrants by intended census division (CD) of settlement. Statistics

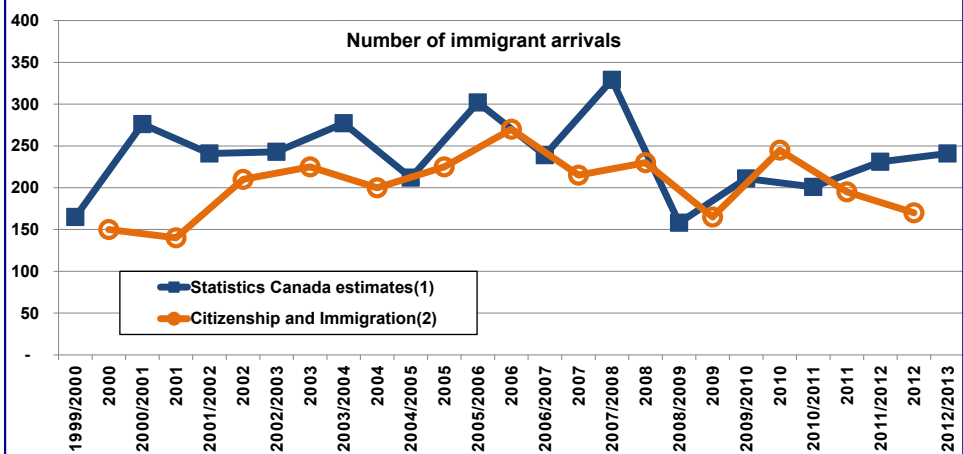
Canada's Annual Demography Program provides estimates of each component of population change for each CD as part of its program to estimate population by CD. See CANSIM Tables 051-0062, 051-0063 and 051-0064 [here](#). Statistics Canada estimates the number of immigrant arrivals in a given CD in a given year by:

- ◆ taking CIC arrival data at the provincial level; and
- ◆ prorating these numbers across CDs by calculating the share of immigrant T1 tax filers for each CD as a percent of all immigrant T1 tax filers in the province; and then
- ◆ multiplying this share by the CIC province-level data for immigrant arrivals in the given year in order to estimate immigrant arrivals by CD.

The CIC data are based on the stated intentions of the immigrant and the Statistics Canada data are based on the address of immigrant taxfilers. The data do not completely align, as shown for the Ontario CD of Brant (see Figure 1 for the annual data and see Figure 2 for a calculated 3-year moving average). In this case, the number of immigrants intending to reside in the Brant CD was larger (up to about 2010) than the estimated number of immigrant arrivals generated from taxfiler data.

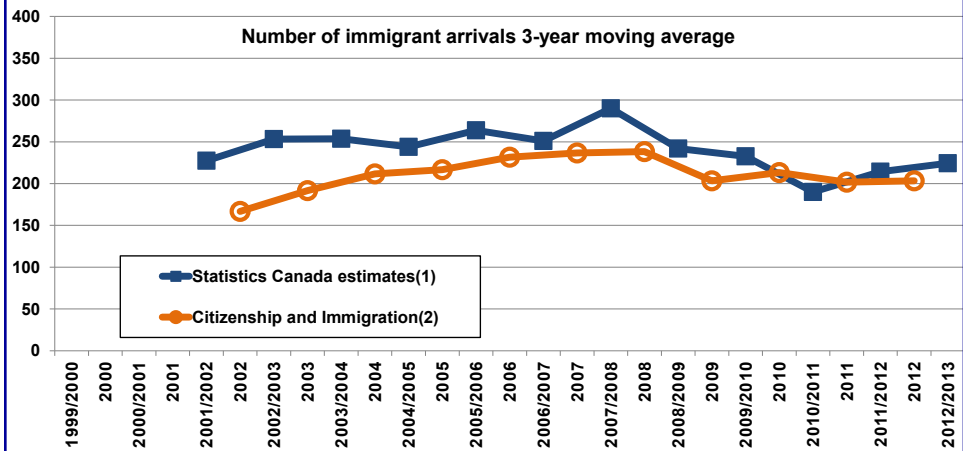
When confronted with the question regarding which estimate is 'correct', the appropriate answer is "They are both correct." One data series shows the stated intention of the immigrant at the time of arrival and the other data series shows the residence of the immigrant at the time that the income tax form was submitted.

Figure 1: Alternative data on the number of immigrant arrivals: Brant Census Division, Ontario



Sources:
 (1) Statistics Canada. Annual Demographic Estimates, CANSIM Table 051-0063. The CIC data on immigrant arrivals at the province level are prorated to each census division using the share of T1 income tax immigrants in each census division as a percent of all T1 immigrants in the province. Estimates are published for the 12 months from July 1 to June 30.
 (2) Citizenship and Immigration Canada. Special Tabulation of Permanent Residents, RDM as of November 2013. Data request tracking number: RE-14-0062. Data are provided for the calendar year.

Figure 2: Alternative data on the number of immigrant arrivals: Brant Census Division, Ontario



Sources:
 (1) Statistics Canada. Annual Demographic Estimates, CANSIM Table 051-0063. The CIC data on immigrant arrivals at the province level are prorated to each census division using the share of T1 income tax immigrants in each census division as a percent of all T1 immigrants in the province. Estimates are published for the 12 months from July 1 to June 30.
 (2) Citizenship and Immigration Canada. Special Tabulation of Permanent Residents, RDM as of November 2013. Data request tracking number: RE-14-0062. Data are provided for the calendar year.