

Methodology

This part describes the concepts, data sources and methodology used to produce the population estimates. Population estimates are produced to measure the population counts according to various characteristics and geographies between two censuses.

Related methodology notes

The *two-way raking method* is also referred to as the “Deming method”, the “method of iterative proportions”, and calibration (see Shryock, Siegel *et al.*, 1976: 547-549).

Unless otherwise noted, the term *preliminary* include both preliminary and updated estimates

The *T1 family file (TIFF)* is derived from the *Canada Revenue Agency (CRA)* T1 file by Small Area and Administrative Data Division of Statistics Canada.

Differences between census counts and population estimates

This population estimates differ from the 2006 Census results published on March 13, 2007, in two respects. First, the estimates are based on 2001 Census counts, adjusted for net census undercoverage (NCU) and updated between censuses using data on births, deaths and migration from administrative sources.

Second, the 2006 Census provides population figures on May 16, 2006, whereas the population estimates present population counts for a later date. Population estimates based on 2006 Census counts adjusted for NCU will be released in the fall of 2008. Census counts adjusted for NCU will be released in the fall of 2008.

Population Estimates

Estimates of the total population and estimates by age and sex

Types of estimates

Population estimates can be either intercensal or postcensal. Intercensal estimates are produced using counts from two consecutive censuses adjusted for net census undercoverage (NCU) and postcensal estimates. The production of intercensal estimates involves updating the postcensal estimates using the counts from a new census adjusted for NCU.

Postcensal estimates are produced using data from the most recent census adjusted for NCU and the components of population growth. In terms of timeliness, postcensal estimates are more up-to-date than data from the most recent census adjusted for NCU, but as they get farther from the date of that census, they become more variable.

Levels of estimates

Updating population estimates between censuses entails the use of data from administrative files or surveys. The quality of population estimates therefore depends

on the availability of a number of administrative data files that are provided to Statistics Canada by Canadian and foreign government departments. Since some components are not available until several months after the reference date, three kinds of postcensal estimates are produced: preliminary postcensal (PP), updated postcensal (PR) and final postcensal (PD). The time lag between the reference date and the release date is three to four months for preliminary estimates and two to three years for final estimates. Though it requires more vigilance on the part of users, the production of three successive series of postcensal estimates is the strategy that best satisfies the need for both timeliness and accuracy of the estimates.

Calculation of postcensal population estimates

Population estimates – preliminary, updated and final – are produced by the component method. This method consists in taking the population figures from the most recent census, adjusted for net census undercoverage (NCU) (census undercoverage minus census overcoverage), and adding or subtracting the number of births, deaths, and components of international and internal migration.

A. Subprovincial estimates of total population

Population estimates for census metropolitan areas and census divisions

The component method is used to produce estimates for census metropolitan areas (CMAs) and census divisions (CDs).

The component-method formula for estimating the total populations of CMAs and CDs is as follows:

$$P_{(t+i)} = P_{(t)} + B_{(t,t+i)} - D_{(t,t+i)} + I_{(t,t+i)} - [E_{(t,t+i)} + \Delta TE_{(t,t+i)}] + RE_{(t,t+i)} + \Delta NPR_{(t,t+i)} + \Delta N_{(t,t+i)} + \Delta Ninfra_{(t,t+i)}$$

where, for each subprovincial area :

- $(t,t+i)$ = interval between times t and $t+i$;
- $P_{(t+i)}$ = estimate of the population at time $t+i$;
- $P_{(t)}$ = base population at time t (census adjusted for net census undercoverage (NCU), or most recent estimate);
- B = number of births;
- D = number of deaths;
- I = number of immigrants;
- E = number of emigrants;
- ΔTE = net temporary emigration;
- RE = number of returning emigrants;
- ΔNPR = net non-permanent residents;
- ΔN = net interprovincial migration;
- $\Delta Ninfra$ = net intraprovincial migration.

Prorating is used to ensure that the province's total population equals the sum of the populations of the subprovincial areas. The technique adjusts the total population counts and the components of subprovincial areas. It implies that the difference between the province's total population and the sum of the populations of the subprovincial areas is distributed proportionally across those areas.

Population estimates for economic regions

A different method is used to produce population estimates for economic regions (ERs): the census division (CD) aggregation method. First, the ERs are defined in terms of CDs using Standard Geographical Classification (SGC) specifications. When the geographic delineation of the CDs matches that of the ER, no adjustment is required; the population estimates for the CDs that make up the ER are simply added together.

However, when the geographic delineation of the CD does not match that of the ER – i.e., when a CD is in more than one ER – allocation of the CD's demographic components is prorated on the basis of its proportion of each ER's population. The proportions are referred to as *conversion factors*. They are calculated using the most recent census counts.

Thus, demographic components (births, deaths and migration) initially measured at the CD level can be allocated to each ER. In other words, the population and demographic components of ERs can be estimated by aggregating the CD data based on the ERs' geographic delineation.

However, using CD aggregation to estimate the components of intraprovincial migration for ERs does not produce the right numbers of in-migrants and out-migrants. It overestimates those figures. In-migrants to a given CD from another CD in the same ER should not be counted, since the migration occurred within the ERs' boundaries. They are false in-migrants. The same is true for out-migrants from one CD to another CD in the same ER. They are false out-migrants. However, combining the in-migration and out-migration figures produced by the CD aggregation method produces a consistent result since the false in-migrants and false out-migrants cancel out. Hence, only the net intraprovincial migration of ER's can be estimated accurately by the CD aggregation method. That is why the numbers of intraprovincial in-migrants and out-migrants are not available for ERs.

Special treatment for preliminary postcensal estimates for Quebec and British Columbia

A different method is used to calculate preliminary postcensal population estimates for census divisions (CDs) and census metropolitan areas (CMAs) in Quebec and British Columbia. For Quebec, the total population estimates produced by the "*Institut de la statistique du Québec (ISQ)*" are used. Those estimates are based on data from the insured persons "*Fichier d'inscription des personnes assurées (FIPA) de la Régie de l'assurance-maladie du Québec (RAMQ)*". They are controlled to Demography Division's estimate of Quebec's total population. The same approach is

followed for the ISQs' economic region (ER)-level estimates, which are derived from its CD-level estimates.

For British Columbia, preliminary postcensal estimates are computed using CD and CMA growth rates provided by *British Columbia's Statistical Agency* (BC STATS), for total population only. To produce new population estimates for each CD and CMA, the rates are applied to Demography Division's estimates of total population for the previous year. Prorating is used to ensure that the province's total population equals the sum of the populations of the subprovincial areas.

The British Columbia population estimates used to calculate the rates are produced using the difference-correlation method, a regression model based on residential electrical (Hydro) connections and *Ministry of Health Client Registry* data as symptomatic indicators.

B. Subprovincial estimates by age and sex

Postcensal estimates by age and sex for census divisions (CDs), census metropolitan areas (CMAs) and economic regions (ERs) are produced by applying the component method to each age-sex cohort in the base population.

Two-way raking is used to ensure consistency between subprovincial estimates and provincial / territorial estimates by age and sex. It maintains consistency between subprovincial and provincial / territorial estimates for each age-sex combination.

Special treatment for preliminary postcensal estimates for Quebec and British Columbia

A different method is used to calculate preliminary postcensal population estimates for census divisions (CDs), census metropolitan areas (CMAs) and economic regions (ERs) in Quebec. The population estimates by age and sex produced by the "*Institut de la statistique du Québec* (ISQ)" are used. Those estimates are based on data from the insured persons registration file "*Fichier d'inscription des personnes assurées* (FIPA) de la *Régie de l'assurance-maladie du Québec* (RAMQ)". They correspond to Statistics Canada Demography Division's total estimate by age and sex for Quebec.

For British Columbia, the estimates by age and sex are produced by applying the age-sex distribution derived by the component method to the total population estimates for CMAs and CDs obtained by applying the growth rates supplied by *British Columbia's statistical agency* (BC STATS).

Two-way raking is used to ensure that the estimates are consistent with the provincial totals by age and sex. The same approach is followed for the ER-level estimates, which are derived from the CD-level estimates.

C. Levels of estimates

For Quebec and British Columbia, the methods described in sections **A** and **B** of **Calculation of postcensal population estimates**, use only for preliminary postcensal estimates. For updated and final postcensal estimates, the component method is used.

The difference between preliminary and final postcensal population estimates lies in the timeliness of the components. When all the components are preliminary, the population estimate is described as preliminary postcensal (PP). When they are all final, the estimate is referred to as final postcensal (PD). Any other combination of levels is referred to as updated postcensal (PR).

Population estimates are final intercensal up to 2000, final postcensal from 2001 to 2004, updated postcensal for 2005 and 2006, and preliminary postcensal for 2007.

Base population and components of population growth

D. Base population

The base populations are derived from the quinquennial censuses between 1971 and 2001. The population universe of the 2001 Census includes the following groups:

- Canadian citizens (by birth or by naturalization) and landed immigrants with a usual place of residence in Canada;
- Canadian citizens (by birth or by naturalization) and landed immigrants who are abroad, either on a military base or attached to a diplomatic mission;
- Canadian citizens (by birth or by naturalization) and landed immigrants at sea or in port aboard merchant vessels under Canadian registry;
- persons with a usual place of residence in Canada who are claiming refugee status and members of their families living with them;
- persons with a usual place of residence in Canada who hold student authorizations (student visas or student permits) and members of their families living with them;
- persons with a usual place of residence in Canada who hold employment authorizations (or work permits) and members of their families living with them;
- persons with a usual place of residence in Canada who hold Minister's permits (including extensions) and members of their families living with them.

For census purposes, the last four groups in this list are referred to as non-permanent residents (NPR).

Foreign residents have not been enumerated since 1991. Foreign residents are persons who belong to the following groups:

- government representatives of another country attached to the embassy, high commission or other diplomatic body of that country in Canada, and members of their families living with them;
- members of the Armed Forces of another country who are stationed in Canada, and members of their families living with them;
- residents of another country visiting Canada temporarily (for example, a foreign visitor on vacation or on business, with or without a visitor's permit).

These populations are adjusted as follows (unless otherwise noted, adjustments to the base population apply to both provincial/territorial and subprovincial levels):

- adjustment of the population for net undercoverage (NUC);
- addition of independent estimates for incompletely enumerated Indian reserves in 1991, 1996 and 2001;
- adjustment for early enumeration in 1991 and 1996 in parts of northern Quebec, Newfoundland and Labrador, the Yukon Territory and the Northwest Territories;
- addition of estimates of NPRs in 1971, 1976, 1981 and 1986. Since 1991, NPRs included in the census universe;
- estimation of the July 1 base population by addition or subtraction of the components of growth between Census Day and June 30. At the subprovincial level, the estimate of the July 1 base population is obtained by applying the subprovincial age-sex distribution of the adjusted census to the provincial / territorial population estimate.

Adjustment for net census undercoverage (NCU)

The adjustment for NCU is important. NCU is the difference between the number of persons who should have been enumerated but were missed (undercoverage) and the number of persons who were enumerated but should not have been or who were counted more than once (overcoverage).

Coverage studies provide undercoverage estimates for the 1991, 1996 and 2001 Censuses at the provincial and territorial levels, and for the 1971, 1976, 1981 and 1986 Censuses at the provincial level only. Estimates of overcoverage at the provincial and territorial levels are available only for the last three censuses (1991, 1996 and 2001). Overcoverage for previous censuses was estimated by assuming that the overcoverage-to-undercoverage ratio for each census between 1971 and 1986 was the same as in 1991. The NCU for the Yukon Territory and the Northwest Territories prior to 1991 was estimated by assuming that the ratio between the NCU for each territory and the 10 provinces for each census between 1971 and 1986 was the same as in 1991.

For consistency, 1991 Census undercoverage and overcoverage were revised in 1998 to take into account the methodological improvements made in the 1996 Census coverage studies. This revision altered the NCU in all censuses between 1971 and 1986. Similarly, 1996 Census undercoverage and overcoverage were revised in 2003.

Various methods were used to produce the estimates of NCU by age and sex for 1991, 1996 and 2001. First, the national estimates of NCU based on the coverage studies by age and sex were smoothed. Then an Empirical Bayes regression model was used to generate provincial and territorial estimates of NCU by broad age groups, and a synthetic model produced estimates by single year of age. Lastly, two-way raking¹ was used to ensure that the NCU estimates were consistent with the provincial and territorial NCU totals and the national estimates by age and sex.

For the 1971-1986 period, the NCU estimates by age and sex were simply prorated to the revised NCU estimates for the total population.

To estimate NCU at the subprovincial level, provincial and territorial NCU rates by age and sex were applied to all geographic regions (census metropolitan areas (CMAs) and census divisions (CDs) in the province).

E. Births and deaths

As it is the case for provincial and territorial estimates, the numbers of births and deaths at the census division (CD) levels are taken directly from the database of Statistics Canada's Health Statistics Division.

A different method is used to produce estimates of births and deaths for census metropolitan areas (CMAs): the census division conversion method. According to this method, each CMA is first defined in terms of CDs (complete or in part) using the Standard Geographical Classification (SGC) specifications.

In cases when the geographic delineation of a group of CDs matches that of a CMA, no adjustment is required; the births and deaths of each CD that make up the CMA are simply added together.

However, when the geographic delineation of the CD does not match that of the CMA –*i.e.* when a CD is not totally included in a CMA – a proportion of the CD's births and deaths is allocated to the CMA. This proportion is equal to the fraction of the CD's population living on the CMA territory.

These proportions are referred to as *conversion factors*. They are calculated using the most recent census counts. With these factors and the census division conversion method, it is possible to allocate to each CMA, the births and deaths initially measured at the CD level.

Where appropriate, the estimates of births and deaths are categorized as final. To ensure their consistency, the estimates are subsequently controlled to the provincial totals using two-way raking.

When no data are available for births and deaths, subprovincial estimates are produced by disaggregating the preliminary provincial or territorial estimates on the basis of the most recent subprovincial distribution derived from Health Statistics Division's vital statistics. In such case, the estimates of births and deaths are

categorized as preliminary. To ensure their consistency, the estimates are subsequently controlled to the provincial totals using two-way raking.

Levels of estimates

Estimates of births are final up to 2004-2005, updated for 2005-2006 and preliminary for 2006-2007. Estimates of deaths are final up to 2003-2004, updated for 2004-2005 and 2005-2006, and preliminary for 2006-2007.

F. Immigration

Since *Citizenship and Immigration Canada's* (CIC's) subprovincial immigration data are not used, subprovincial estimates are produced by disaggregating the preliminary and final provincial or territorial estimates on the basis of the most recent subprovincial distribution derived from T1FF. The data are available only by broad age groups (0-17, 18-24, 25-44, 45-64, 65+) and must be disaggregated by sex and single year of age based on the provincial distribution. To ensure their consistency, they are subsequently controlled to the provincial totals using two-way raking.

Levels of estimates

The difference between preliminary and final estimates lies in the timeliness of the sources used to estimate this component. Since the estimates of the number of immigrants are derived from provincial / territorial estimates, the level of the subprovincial estimates will be the same. Immigration estimates are preliminary the first year and updated the following year. They become final two years after the reference year.

The numbers of immigrants are final up to 2004-2005, updated for 2005-2006, and preliminary for 2006-2007.

G. Net non-permanent residents

At the subprovincial level, there are no reliable administrative data available to estimate NPRs. To compensate for the lack of data, the provincial / territorial estimates by age and sex are disaggregated by subprovincial area on the basis of the subprovincial distribution in the most recent census. To ensure their consistency, they are subsequently controlled to the provincial totals using two-way raking.

Levels of estimates

Since the estimates of the net number of NPR are derived from provincial / territorial estimates, the level of the subprovincial estimates will be the same.

Estimates of the net number of NPR are final up to 2003-2004, updated for 2004-2005 and 2005-2006, and preliminary for 2006-2007.

H. Emigration

As in the case of immigrants, the number of emigrants at the subprovincial level is derived from the T1FF. The estimates are available only by broad age groups (0-17, 18-24, 25-44, 45-64, 65+) and must be disaggregated by sex and single year of age based on the provincial distribution. To ensure their consistency, they are subsequently controlled to the provincial totals using two-way raking.

Levels of estimates

The difference between preliminary and final estimates lies in the timeliness of the sources used to estimate this component. Since the estimates of the number of emigrants are derived from provincial / territorial estimates, the level of the subprovincial estimates will be the same.

The numbers of emigrants are final up to 2003-2004, updated for 2004-2005 and 2005-2006, and preliminary for 2006-2007.

I. Net temporary emigration

At the subprovincial level, provincial / territorial net temporary emigration is disaggregated on the basis of the age-sex distribution of subprovincial emigrants. To ensure their consistency, the estimates are subsequently controlled to the provincial totals using two-way raking.

Levels of estimates

Since the census is the only data source, there is no difference between the preliminary and final estimates for this component.

The estimates of net temporary emigration are final up to 2003-2004, updated for 2004-2005 and 2005-2006, and preliminary for 2006-2007.

J. Returning emigrants

To estimate the numbers of returning emigrants at the subprovincial level, provincial numbers of returning emigrants are disaggregated on the basis of the age-sex distribution of subprovincial immigrants. To ensure their consistency, they are subsequently controlled to the provincial totals using two-way raking.

Levels of estimates

The difference between preliminary and final estimates lies in the timeliness of the sources used to estimate this component. Since the estimates of the number of returning emigrants are derived from provincial / territorial estimates, the level of the subprovincial estimates will be the same.

The numbers of returning emigrants are final up to 2003-2004, updated for 2004-2005 and 2005-2006, and preliminary for 2006-2007.

K. Interprovincial migration

Interprovincial migration by broad age group and sex for subprovincial areas is derived from the T1FF. The estimates by broad age group and sex are disaggregated into single years of age using distributions from the 2001 Census one-year mobility question. To ensure their consistency, they are subsequently controlled to the provincial totals using two-way raking.

Data from the T1FF are used to produce the final estimates.

Levels of estimates

The difference between preliminary and final estimates lies in the timeliness of the sources used to estimate this component. Since the estimates of the number of migrants are derived from provincial / territorial estimates, the level of the subprovincial estimates will be the same.

The estimate of interprovincial migration is final up to 2005-2006 and preliminary for 2006-2007.

L. Intraprovincial migration

As in the case of interprovincial migration, the components of intraprovincial migration by broad age group and sex are derived from the T1FF. The estimates by broad age group and sex are disaggregated into single years of age using distributions from the 2001 Census one-year mobility question.

These sources are used for both preliminary and final estimates.

Levels of estimates

The difference between preliminary and final estimates lies in the timeliness of the T1FF data used to estimate this component.

The estimates of intraprovincial migration are final up to 2005-2006 and preliminary for 2006-2007. Given the fact there are no reliable data for preliminary intraprovincial migration estimates, the data of the most recent year, for which final estimates are available, is used. Hence, for preliminary data, the following assumption is made: intraprovincial migratory behaviors of the current year are similar to those of the last year for which final estimates are available.

M. Intercensal population estimates

Intercensal estimates – population estimates for reference dates between two censuses – are produced following each census. They reconcile previous postcensal estimates with the new census counts.

There are three main steps in the production of intercensal estimates:

- the correspondence of the boundaries between both censuses;
- calculation of the error of closure;
- linear distribution of the error of closure.

To ensure the correspondence of the boundaries for the 1986-2001 period, the base populations and components of population growth from 1986 to 2001 had to be adjusted for the 2001 Census boundaries. For areas whose boundaries changed between the two censuses (1996 and 2001 Standard Geographical Classification (SGC), *conversion factors* based on 2001 Census subdivisions were used. In general, the corrections to census divisions (CDs), census metropolitan areas (CMAs) and economic regions (ERs) were minor.

Error of closure is defined as the difference between the postcensal population estimates on Census Day and the population enumerated in that census (after adjustment for net census undercoverage (NCU)).

The error of closure is spread uniformly over the intercensal period.

Intercensal estimates by age and sex are adjusted in the same way, i.e., by distributing the error of closure uniformly across the age-sex cohorts.

Like the postcensal estimates, the subprovincial intercensal estimates by age and sex are adjusted to ensure consistency with the provincial estimates using two-way raking.