C 3

Catalogue no. 91F0003GPE

Introduction to population estimates

Demography Division

Statistics Canada Statistique Canada

Canada

ONE DAY WORKSHOP ON ESTIMATES

AN INTRODUCTION TO POPULATION ESTIMATES

HOW TO: $\left\{ \begin{array}{l} \text{PREPARE THEM;} \\ \text{USE THEM;} \\ \text{INTERPRET THEM.} \end{array} \right.$

Workshop designed and developed by Rosemary Bender and Ravi Verma.

STATISTICS CANADA

1996

ACKNOWLEDGMENTS

The authors wish to thank Yolande Lavoie, Senior Consultant in demography, for her thorough and very competent review of the content and format of this manuscript. We also acknowledge our colleagues of the Population Estimates Section upon whose work Modules 2 and 4 are based. Our thanks also to Danielle St-Germain who prepared this manuscript with remarkable patience.

The authors are solely responsible for any error, imprecision or omission still present in the text. Suggestions for improving this workshop are most welcome. Send your comments to Rosemary Bender, Demography Division (613) 951-2324 - FAX (613) 951-2307.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS - continued

TABLE OF CONTENTS - concluded

TABLE OF CONTENTS - concluded

Page

List of Tables - concluded

Workshop on Estimates

Objective

The purpose of this workshop is to present the basics of population estimation in a way that will enable participants to make good use of existing estimates and, from this starting point, move on to estimating populations not covered by the Statistics Canada Estimation Program.

Structure and Content

For ease of use, this manual is divided into stand-alone modules.

Module 1 discusses the main uses of population estimates and presents an overview of estimation methods.

Modules 2 and 3 describe in detail two methods used at Statistics Canada: the component method and the regression method. The regression method is used particularly for small area estimation where data on the components of population change are either not available or of inferior quality. Module 3 also describes the extrapolation method presently used for custom estimates.

Module 4 discusses the quality of population estimates, focusing on the criteria of accuracy and consistency, and gives the release times for the various types of estimated produced by Statistics Canada.

Common problems encountered by demographers when estimating the population of local areas are discussed in Module 5.

Module 6 is designed to show the users of population estimates how to track demographic events over time and to interpret and analyze estimate results. Module 7 provides information on current research and development in the field of population estimates being done by the Demography Division of Statistics Canada and on the population estimates and products available from the Division.

Module 7, after describing the Statistics Canada Estimation Program, discusses recent significant changes to the Program (e.g. change in population universe, replacement of family allowances as data source on components).

All modules include a number of examples and exercises to enhance the participants' understanding of the theoretical material and prepare them for developing their own population estimates.

MODULE 1

POPULATION ESTIMATES: DEFINITION, USES AND METHODS

Population Estimates: Definition, Uses and Methods

There has always been an interest in population counts. The earliest indication of population enumeration dates back to before 3500 B.C. in Babylonia. Ancient Greece and Rome are known to have conducted censuses. One of the best known censuses is found in the opening chapters of the Gospel according to Luke, which records that the Jews were told to return to their birthplace to be counted. Although some information on the socio-economic characteristics of the people was collected from the censuses, their main purpose was to assess military potential. The first modern census in Canada, which was held in New France (now Quebec) in 1666, enumerated slightly over 3,000 settlers.

Today, demographic information is an invaluable tool for government managers and business persons. Until recently, the data obtained from national censuses provided sufficient information. Now however, data are required on a much more frequent and timely basis. To meet this need, current population estimates are produced at national and sub-national levels and broken down by various demographic and socio-economic characteristics.

1.1 Nature of Population Estimates

Since Canada keeps no permanent population registers, the population data for the dates not covered by the five-year censuses must be estimated.

Estimating the population entails the use of data on marital status, data on migration and other types of data useful in translating population changes. The estimation methods may be demographic or mathematical, and the dates for which the estimates are produced cover the past, the present and the future. From this viewpoint, population projections are estimates. Generally speaking however the term <u>estimate</u> refers to data produced for intercensal periods (i.e. in the past), or postcensal periods up to the present. The former are interpolations, the latter (like projections as such) extrapolations.

1.2 Uses of Estimates

Population estimates have wide-ranging applications to planning, marketing and program evaluation in both the public and private sectors. The following are some specific uses to which estimates can be put:

- as a denominator in calculating social and economic indicators (e.g. birth rate, death rate, unemployment rate);
- in reference publications such as the United Nations Year Book or encyclopedias;
- to obtain a current demographic profile of a region as part of a market study for introducing a new product;
- to determine whether there are sufficient hospitals and nursing homes to meet a community's needs;
- as a base for population projections produced for Canada, provinces and territories, metropolitan areas, urban centres, municipalities, administrative areas such as federal employment centres and

municipal school board districts; population projections are especially useful in estimating future demands on pensions, housing, roads, schools and hospitals;

- to calculate weights for use in surveys such as the Statistics Canada Labour Force Survey, designed on the basis of decennial census data;
- as input in the development of government programs (Employment and Immigration Canada uses population estimates to help establish levels of immigration); and
- in the calculation of transfer payments, grants and cost-sharing agreements between various levels of government.

The list of the users of the population estimates produced by Statistics Canada, is fairly long and includes the names of federal, provincial and municipal agencies as well as major corporations and small companies. The following list of users is not comprehensive.

Federal government	Provincial and municipal governments	Other
Indian and Northern Affairs	Provincial statistics bureaus	3M Canada
Agriculture Canada	City of Calgary Electric System	Canadian Potato Chip and Snack Food Association
National Defence	Ottawa-Carleton Health Department	Association of Canadian Medical Colleges
Elections Canada	Ottawa Civic Hospital (geriatrics)	Ontario Medical Association
Employment and Immigration Canada		Association of Universities and Colleges of Canada
Environment Canada		Bank of Canada
The Conference Board of Canada		Bank of Montreal
Department of Finance		Bell Canada
Fisheries and Oceans		Canadian Dairy Commission
Emergency Preparedness Canada		Embassies of various countries
Health and Welfare Canada		Globe and Mail
Canada Mortgage and Housing Corporation		Financial Times
CBC		Kidney foundation
Solicitor General		INRS Urbanisation
Statistics Canada		London Times
Transport Canada		Canadian Chicken Marketing Board
		Reader's Digest
		Sears Canada
		World Book Encyclopedia

1.3 Estimation Methods - Overview

Various methods are used for population estimation. They range in complexity from crude techniques (e.g. confirmation of most recent census figures) to sophisticated multiple regression models.

The methods presented in this section belong to two main types of estimation. The first, the component method, comprises a single method. The second set of methods are mathematical and include extrapolation techniques and ratio methods. The latter are based on the ratio between population growth in the area considered and the growth of symptomatic indicators or related populations.

The module concludes with a discussion of intercensal estimates, which render the population estimates compatible with censuses preceding and following the reference date.

1.3.1. Component Method

The component method, essentially demographic, is based on the fact that population size and structure change as the result of demographic events (births, deaths, migration). The method entails adding to the population figures for a given point in time (usually a census) the natural and migratory increase observed (or estimated) between that point in time and the reference date.

Basic equation:

$$\hat{P}_{t+i} = P_t + B_{t,t+i} - D_{t,t+i} \pm M_{t,t+i} \qquad (1.1)$$

where
- $\hat{P}_{t+i}$ = population estimate at time t+i;
- P_t = base population at time t;
- $B_{t,t+i}$ = births between times t and t+i;
- $D_{t,t+i}$ = deaths between times t and t+i;
- $M_{t,t+i}$ = net migration between times t and t+i.

Advantages and Limitations

Estimating quality control for this method is simple as it makes a clear distinction between base population and the factors of change. If the initial figures are accurate, the components are the sole source of error. Figures on births and deaths are extremely reliable as registration of these events is mandatory. The only possible source of error is the measurement of migration, where only legal international immigration is consistently recorded. Other international and internal migration are estimated.

The delays in obtaining data on the components, especially at the sub-provincial level, leads to a considerable delay in estimating the population, making development of faster methods essential. The component method is used to produce population estimates for Canada and the provinces within a timeframe of several months. Because these estimates are based on preliminary data on the components, they are only provisional and are adjusted when the final data become available.

Example 1.1

Estimate the population of Hilltop as of June 1, 1993 using the component method, based on the following information:

1991 Census count: 10,000
Between June 1, 1991 and May 31, 1993:
- births (B): 345
- deaths (D): 165
- net migration (M): 150

Solution:

$$\hat{P}_{93} = P_{91} + B_{91,93} - D_{91,93} + M_{91,93}$$
$$= 10{,}000 + 345 - 165 + 150$$
$$= 10{,}330$$

1.3.2 Mathematical Methods

When components of demographic change are not available, or not available on a timely basis, other methods must be sought. Some of these methods model demographic growth observed in the past, others are based on current symptomatic indicators of population change.

These methods are called mathematical because they are based on equations that express growth rates as a function of time rather than on the factors that influence the trend over a given time period.

a) Extrapolation

The principle of extrapolation is a simple one, where the population of an area is estimated for a date subsequent to the preceding census. The population growth is calculated from the two previous censuses. In most cases, it is assumed that growth has followed either a linear or geometric progression.

The method should only be used to obtain gross population estimates for reference dates close to the date of the previous census. The method takes into account neither data on nor factors of population change; furthermore, it assumes that growth is constant, although this is in fact rarely the case. Population change, especially for small areas and for detailed demographic characteristics, can be very volatile.

Linear extrapolation: It is assumed that the population of a given area records a constant annual increase in the number of individuals.

Basic equation: $$\hat{P}_{t+i} = P_t + \frac{i}{5} (P_t - P_{t-5}) \qquad (1.2)$$

where $\hat{P}_{t+i}$ = population estimate at time t+i;

P_t = census count at time t;

P_{t-5} = census count at time t-5;

i = interval of time between preceding census and reference date.

Geometric extrapolation: Generally speaking, population figures grow according to geometric rather than linear patterns, as the people who contribute to the population growth in an area over a given time period are themselves a source of population growth. In this case, it is assumed the population grows as a function of time, based on a constant growth rate rather than a constant number of individuals.

Basic equation: $$\hat{P}_{t+i} = P_t (1 + r)^i \qquad (1.3)$$

where $\hat{P}_{t+i}$ = population estimate at time t+i;

P_t = census count at time t;

i = interval of time between preceding census and reference date;

r = annual geometric growth rate between preceding two censuses

$$\sqrt[5]{\frac{P_t}{P_{t-5}}} - 1$$

P_{t-5} = census count at time t-5.

Example 1.2

Estimate the population of Hilltop as of June 1, 1993 using
i) linear extrapolation
ii) geometric extrapolation

and the following information:

1986 Census count: 9,000
1991 Census count: 10,000

Solution: i) linear extrapolation

$$\hat{P}_{93} = P_{91} + \frac{2}{5}(P_{91} - P_{86})$$

$$= 10{,}000 + \frac{2}{5}(10{,}000 - 9{,}000)$$

$$= 10{,}400$$

ii) geometric extrapolation

$$\hat{P}_{93} = P_{91}(1 + r)^2$$

$$r = \sqrt[5]{\frac{P_{91}}{P_{86}}} - 1$$

$$= 0.0213$$

$$\hat{P}_{93} = 10{,}000\ (1.0213)^2$$

$$= 10{,}431$$

b) Ratio Method

This method allocates to the population to be estimated a growth rate calculated for another population or derived from the change in a symptomatic indicator. One of its advantages is that, rather than extrapolating past growth, it relates growth of the population in question to growth of another related population, or to symptomatic indicators for which data are available for both the date of the previous census and the reference date. The procedure is the same whether the basic assumption is similarity of growth rates for the target population and the other population, or similarity of the target population and an indicator.

- *Method based on ratio between growth in a given area and growth in a related area*

In this method, the growth rate of an area is assumed to be equal to that of a larger or overlapping area for which current demographic estimates are available for the reference date.

Although the method does take into account population growth in the larger and overlapping area, between the previous census and the reference date, it is also to be used with much caution. Population growth can vary considerably across geographic areas, and the growth in one large area is not necessarily representative of the growth occurring in all of its smaller areas.

Basic equation: $$\hat{P}_{t+i} = (P'_{t+i} / P'_t) \bullet P_t \tag{1.4}$$

where $\hat{P}_{t+i}$ = population estimate at time t+i;

P_t = census count at time t;

P'_{t+i} = available population estimate for larger or overlapping area at time t+i;

P'_t = census count for this area at time t.

- *Method based on the ratio of population growth to selected symptomatic indicators*

This method uses current indicators of population change available for the area of estimation in question. These indicators may include data on school enrolment, hydro connections, births, deaths or number of Child Tax Benefit recipients. The method calculates the ratio of these indicators to the population counts observed for a census year. Using this ratio and the values of the indicators for the reference date, the current population of the region may be estimated.

Basic equation: $$\hat{P}_{t+i} = (I_{t+i} / I_t) \bullet P_t \tag{1.5}$$

where $\hat{P}_{t+i}$ = population estimate at time t+i;

P_t = census count at time t;

I_{t+i} = value of indicator at time t+i;

I_t = value of indicator at time t.

Example 1.3 Estimate the population of Hilltop as of June 1, 993 using the rate of growth method, based on the following information:

Solution: $$\hat{P}_{93} = \frac{P'_{93}}{P'_{91}} \bullet P_{91} = \frac{52{,}500}{50{,}000} \bullet 10{,}000 = 10{,}500$$

Region	Census Count	Population Estimate
Hilltop	10,000	-
North County	50,000	52,500

Example 1.4

Estimate the population of Hilltop aged 5-14 as of June 1, 1993 using the proportional allocation method, based on the following information:

Persons aged from 5 to 14 years:

Year	Census Count Ages 5-14	School Enrolment (SE) Ages 5-14
1991	1000	985
1992	-	1,004

Solution: $$\hat{P}_{93,5-14} = \frac{SE_{93,5-14}}{SE_{91,5-14}} \bullet P_{91,5-14}$$

$$= \frac{1,004}{985} \bullet 1,000$$

$$= 1,019$$

- *Housing units method*

A variant of the ratio method uses number of housing units as an indicator of the number of households. When the number of housing units is multiplied by average household size, the result is the population size.

Basic equation: $$\hat{P}_{t+i} = H_{t+i} \bullet AHS \quad (1.6)$$

where $\hat{P}_{t+i}$ = population estimate at time t+i;

H_{t+i} = estimated number of households at time t+i;

AHS = average household size observed in the census preceding the reference date.

Example 1.5

Estimate the population of Hilltop as of June 1, 1993 using the housing unit method, based on the following information:

1991 Census count: 10,000
1991 Census number of households: 4,080
Number of households in 1993: 4,250

Solution: $\hat{P}_{93} = H_{93}\,(\,AHS_{91}\,)$

where $AHS_{91} = \dfrac{P_{91}}{H_{91}} = \dfrac{10,000}{4,080} = 2.45$

$\hat{P}_{93} = 4,250 \bullet 2.45$

$= 10,431$

c) The regression method

The idea behind the regression technique is similar to that for proportional allocation: information on population counts and symptomatic indicators for a census year is applied to current symptomatic data to obtain current population estimates. With regression, however, the relationship between population counts and symptomatic indicators is modelled simultaneously for all the small areas making up a larger area. The resulting equation is then used for current population estimates for each small area.

At Statistics Canada, two types of regression techniques (ratio-correlation and difference-correlation) are used to estimate the population of census divisions.

Module 3 expands on the regression methodology, which is based on the following equation.

Basic equation:

$$\hat{y} = \alpha + \beta x + \varepsilon \qquad (1.7)$$

where $\hat{y}$ = dependent variable (population);

x = independent variable (symptomatic indicator);

α,β = regression coefficients; and

ε = model error.

d) Birth and death rate (vital rates) method

The birth and death rate method, also ratio based, requires data on births and deaths to be available for the area in question for the reference period. It is assumed that the birth rates for the area follow the same pattern as those for a related area. The method estimates the population based on births and deaths separately, and then averages the results. Any inconsistencies between the sum of these estimates and those available for larger areas are prorated among the smaller areas.

The first step in the method consists in calculating, for the census year, the crude birth rate (CBR) and crude death rate (CDR) in the small area(s), for which the population is to be estimated, and in the larger area, for which the population count is known. These rates must also be available for the reference year for the larger area. Using the rule of three, the current estimate of the CBR and the CDR for the area(s) can then be derived for the reference year. The most recent birth and death figures are then divided by the corresponding rates to produce the population estimates sought.

Step 1: Calculate crude birth and crude death rates

The crude birth rate is expressed as:

$$CBR = (B / P) \cdot 1{,}000 \qquad (1.8a)$$

where CBR = crude birth rates;

B = annual number of births;

P = mid-year population.

To calculate the death rate, replace the annual number of births by the annual number of deaths. The crude rates for the area in question (reference year) are calculated as follows:

$$CBR_{t+i} = (CBR'_{t+i} / CBR'_t) \cdot CBR_t \qquad (1.8b)$$

where CBR = birth rate for area in question;

CBR' = birth rate for larger area;

$t+i$ = reference year

t = census year

The crude death rate is similarly calculated.

Step 2: Estimate the population using crude birth and crude death rates

$$\hat{P}_B = (B_{t+i} / CBR_{t+i}) \cdot 1{,}000 \tag{1.8c}$$

where $\hat{P}_B$ = population estimate at time t+i based on crude birth rate;

B_{t+i} = number of births, year t+i;

CBR = crude birth rate, year t+i.

The population estimate based on the crude death rate is similarly calculated, producing two population estimates.

Step 3: Average the two estimates

$$\hat{P}_{t+i} = \frac{\hat{P}_B + \hat{P}_D}{2} \tag{1.8d}$$

Example 1.6
Estimate the population of Hilltop as of June 1, 1993 using the vital rates method, based on the following information:

	Hilltop		North County	
	1991	1993	1991	1993
Births	160	175	825	875
Deaths	80	85	405	425
Population	10,000	-	50,000	52,500

Step 1 - Calculate crude birth and crude death rates
Equation 1.8a can be used for North County and Hilltop (1991 only) to obtain the following crude birth and death rates (per thousand)

	Hilltop	North County	
	1991	1991	1993
CBR	16.0	16.5	16.7
CDR	8.0	8.1	8.1

Example 1.6 - concluded

For Hilltop:

$$CBR_{93} = \frac{CBR'_{93}}{CBR'_{91}} \cdot CBR_{91} \qquad CDR_{93} = \frac{CDR'_{93}}{CDR'_{91}} \cdot CDR_{91}$$

$$= \frac{16.7}{16.5} \cdot 16.0 \qquad = \frac{8.1}{8.1} \cdot 8.0$$

$$= 16.2 \qquad = 8.0$$

Step 2 - Estimate population using each rate:

$$\hat{P}_B = \frac{B_{93}}{CBR_{93}} \cdot 1{,}000 \qquad \hat{P}_D = \frac{D_{93}}{CDR_{93}} \cdot 1{,}000$$

$$= \frac{175}{16.2} \cdot 1{,}000 \qquad = \frac{85}{8.0} \cdot 1{,}000$$

$$= 10{,}802 \qquad = 10{,}625$$

Step 3 - Average the two estimates

$$\hat{P}_{93} = \frac{10{,}802 + 10{,}625}{2} = 10{,}714$$

1.3.3 Intercensal Estimates

All of the estimates described above are know as *postcensal* estimates as they are based on data from the census preceding the reference date. These estimates are not necessarily consistent with those based on the following census, due to errors and inconsistencies in the methodologies and/or data sources. Although census results cannot be 100% accurate (response errors, overcoverage, undercoverage), they are always assumed to be of better quality than estimate results.

Consequently, when data from a following census are available, the postcensal estimates are retroactively adjusted to render them consistent with the censuses preceding and following the estimate reference date. These new estimates are known as *intercensal* estimates. Intercensal estimates can also be derived directly using interpolation techniques. The two basic methods are presented below.

a) Retroactive adjustment

The information derived on the population change for the intercensal period is maintained. The difference between the postcensal estimate for the census reference date at the end of the period and the census count is called the *error of closure*.

$$E = P_{t+5} - \hat{P}_{t+5} \tag{1.9}$$

where E = error of closure;

$\hat{P}_{t+5}$ = postcensal estimate at time t+5;

P_{t+5} = census count, year t+5.

This error is distributed linearly over the population estimates for the intercensal period, as follows:

$$\hat{I}_{t+i} = \hat{P}_{t+i} + (i / 5) \bullet E \tag{1.10}$$

where $\hat{I}_{t+i}$ = intercensal estimate at time t+i;

$\hat{P}_{t+i}$ = postcensal estimate at time t+i;

i = time interval (in years) between preceding census and reference date.

The estimates of population between two censuses are made consistent with the census counts using the above equation. If the components of demographic change are also available, they could in theory also be adjusted. However it is not usually possible to determine the amount of error attributable to each component. Some components are known to be of high quality and are not likely to be the source of great error (births, deaths and immigration in larger areas).

Internal migration, emigration, non-permanent residents and returning Canadians must be indirectly estimated from administrative files and are most subject to error, particularly for small areas or detailed demographic characteristics.

Another factor contributing to errors of closure relates to the quality of the census base population. Differential undercoverage, discussed more fully in Module 2, can create fictitious population growth or loss, affecting the size of the error. Because of the difficulty in determining the contribution of each source of error, total error of closure is generally treated as a residual.

b) Interpolation

In contrast to retroactive adjustment, which makes use of the postcensal estimates, interpolation techniques are based solely on census data. Linear interpolation assumes a constant growth in population between the two censuses.

$$\hat{I}_{t+i} = P_t + i / 5 (P_{t+5} - P_t) \tag{1.11}$$

where $\hat{I}_{t+i}$ = intercensal estimate at time t+i;

P_t = census count at time t;

P_{t+5} = census count at time t+5;

i = time interval between preceding census and reference date.

However, as mentioned above, growth in population tends to follow geometric (rather than linear) patterns. The geometric interpolation is given by:

$$\hat{I}_{t+i} = P_t \, (1 + r)^i \tag{1.12}$$

where $\hat{I}_{t+i}$ = intercensal estimate at time t+i;

P_t = census count at time t;

i = time interval between preceding census and reference date;

r = annual geometric growth rate between two censuses

$$= \sqrt[5]{\frac{P_{t+5}}{P_t}} - 1 \; ;$$

P_{t+5} = census count at time t+5.

Example 1.7

The postcensal estimate of the population of Hilltop as of June 1, 1989 is 9,400. Estimate the intercensal population for the same date using the following intercensal estimation techniques and the data in example 1.2.

i) Retroactive adjustment (1991 postcensal estimate) is 10,200.
ii) Linear interpolation
iii) Geometric interpolation

Solution: i) Retroactive adjustment

$$\hat{I}_{89} = \hat{P}_{89} + \frac{3}{5}\left(P_{91} - \hat{P}_{91} \right)$$

$$= 9{,}400 + \frac{3}{5}\left(10{,}000 - 10{,}200 \right)$$

$$= 9{,}280$$

ii) Linear interpolation

$$\hat{I}_{89} = P_{86} + \frac{3}{5}\left(P_{91} - P_{86} \right)$$

$$= 9{,}000 + \frac{3}{5}\left(10{,}000 - 9{,}000 \right)$$

$$= 9{,}600$$

iii) Geometric interpolation

$$\hat{I}_{89} = P_{86}\,(1 + r)^3$$

where $= \quad r - \sqrt[5]{\frac{P_{91}}{P_{86}}} - 1$

$$= 0.0213$$

$$\hat{I}_{89} = 9{,}000 \bullet (1.0213)^3$$

$$= 9{,}587$$

1.4 Method Selection

Given unlimited time and resources and quality input data, the component method is preferred over all others. The resulting population estimates are generally most accurate as they are based on *events that change population figures* rather than those *assumptions* inherent in all mathematical models. The information on the components of population increase are also invaluable by-products for demographic and socio-economic analysis. Unfortunately unlimited time and resources are rare, and input data quality is often deficient, especially for small areas. Thus compromises are necessary, and the choice of methods will depend on one or more of the following factors:

- use to be made of the estimates;
- boundaries;
- timeliness;
- resources and data availability;
- type of population estimates;
- consistency with other estimates.

Use to be made of the Estimates

The use of the population estimates determines to a large extent the amount of time and resources allocated to their estimation. This in turn dictates the type of estimation technique used. Students, for example, may only be interested in the approximate size of an area's population, or the ranking of areas according to population size. In this case, simple interpolation from previous census data will suffice. On the other hand, those doing market studies are interested in getting as good a demographic profile as possible of potential clients in an area, and may be willing to invest more time and money in techniques that provide a more accurate picture. At the other end of the scale, federal, provincial and municipal governments use population data to calculate transfer payments: in this case it is very important to obtain population estimates that are as timely and accurate as possible.

Boundaries

Depending on the geographic boundaries of the area, population estimates may be available for other areas with similar boundaries. For example, population estimates produced by Statistics Canada for census divisions and census subdivisions are used by many users for their own needs, either by combining them to form larger areas or using population ratios observed in the previous census.

Timeliness

The time allocated for population estimation also determines to a great extent the type of method and data sources used. If one is asked to provide a current estimate of the population of a small area within a few hours, extrapolating census data is probably the best course. If, on the other hand, one is involved in an in-depth demographic study in preparation for setting government immigration policy, time is taken to determine the best methods and data sources.

Resources and Data Availability

The resources that may be required to estimate population include human resources (i.e. person-years), computer resources (e.g. hardware, software, processing costs) and other costs (e.g. input data costs). The component method requires data on births and deaths, and on international, interprovincial and/or intraprovincial migration. Some of these data are not always available and must be estimated usually by an indirect method. The ratio methods are based on symptomatic indicators of population change, which at times must be adjusted due to boundary problems or inconsistencies in coverage or concepts. Techniques such as extrapolating from census data, on the other hand, require very little time or resources.

Type of Demographic Estimates

Estimates of the total population of a given area can be done using any of the methods described above. For more detailed demographic characteristics (e.g. age, sex, marital status) not all methods are applicable. In some instances, a combination of methods may be in order. For example, population estimates by age and sex can be determined using the component method, while the breakdown by mother tongue can be based on characteristics observed in the previous census. Some sources of data provide information for a specific subset of the population and can be used directly or in proportional allocation methods. Thus, estimates of the population under the age of fifteen can be based on Child Tax Benefit data from Revenue Canada.

Consistency with other Estimates

Another factor that might determine the type of method used is one of consistency with other estimates. It may be important for evaluation or analysis purposes to have, for subprovincial areas, the same indicators of demographic change that exist for provinces and territories. Also, it might be desirable to estimate the population for two different but overlapping sets of subprovincial areas using, wherever possible, the same methodology and data sources.

MODULE 2

THE COMPONENT METHOD

The Component Method

2.1 Methodology

The component method produces the most accurate picture of population dynamics and is a popular estimation technique used by national statistical agencies in countries where good demographic data are readily available.

The base population is taken from census counts, to which is added the change in population size between the census date and the reference date.

General equation:

$$POP(t+1) = POP(t) + BIRTHS - DEATHS \pm NET\ MIGRATION$$

This change is produced by the components that contribute to the natural increase of a population (i.e. births minus deaths) and those that involve a permanent change of residence (movement into or out of the area). Net migration is defined as the number of people entering an area less those moving out of the area. This may include international, interregional and/or intraregional migration.

This method offers many advantages.

Government administrative files are good sources of information on many of the components of population change, particularly when the recording of demographic events is required by law, as it is in Canada for births, deaths and immigration.

When the data are sufficiently detailed, as national and provincial data usually are, the demographic characteristics of the population (births, deaths, marital status) can be estimated fairly accurately.

The breakdown of population change into its separate demographic components provides greater scope for in-depth analysis of the various factors affecting population change. For each component, estimates can be derived from current information, past trends, the age and sex structure of the base population, and an understanding of the socio-economic and demographic forces influencing the behaviour of this component.

Finally, the component method is easily understood by the lay person. The basic equation is a simple summation of terms representing demographic events with which he or she can easily identify.

Countries where the component method is used include[1]:

Albania	El Salvador	Luxembourg
Australia	Finland	Malta
Austria	France	Mauritius
Belgium	Germany	New Zealand
Bulgaria	Greece	Poland
Canada	Guatemala	Portugal
Caribbean Islands	Hong Kong	Romania
Chile	Hungary	Switzerland
Commonwealth of Indpendent States	Ireland	United Kingdom
	Israel	United States
Cyprus	Italy	Yugoslavia
Czechoslovakia	Japan	

2.1.1 Estimating Total Population

The equation for estimating the total population is straightforward.[2]

$$\hat{P} = P_t + B_{t,t+i} - D_{t,t+i} \pm M_{t,t+i} \tag{2.1}$$

where $\hat{P}_{t+i}$ = population estimate at time t+i;

P_t = base population (census) at time t;

between time t and t+i

$B_{t,t+i}$ = number of births;

$D_{t,t+i}$ = number of deaths;

$M_{t,t+i}$ = net migration.

Depending on the type of area, the migration component can be further broken down into five categories of migration, three of which can be again broken down into two types of movement (in and out):

international migration (immigrants / emigrants),
interprovincial migration (in / out) and
intraprovincial migration (in / out).

---

[1]Source: Demographic Yearbook, 1989. United Nations, New York, 1991.

[2]Census counts are adjusted for net census undercoverage (see Section 2.2.1).

Net flows of non permanent residents were introduced as a fourth migration category in order to take into account the expansion of the census universe in 1991. The last category concerns Canadians who left the country permanently and have subsequently returned.

To estimate the net migration for a given province (i.e. between the census and the reference date) the migration variable is expressed as:

$$M_{t,t+i} = I_{t,t+i} + E_{t,t+i} + NI_{t,t+i} - NO_{t,t+i} + NPR_{t,t+i} + RC_{t,t+i} \quad (2.2)$$

where, between times t and t+i

I = number of immigrants;

E = number of emigrants;

NI = number of interprovincial migrants entering the province from another province or territory;

NO = number of interprovincial migrants leaving the province for another province or territory;

NPR = net flow of non-permanent residents;

RC = number of returning Canadians.

Example 2.1

Estimate the population of Prince Edward Island as of June 1, 1987 adjusted for net census undercoverage using the component method, based on the following information:

1. Base population as of July 1, 1986 (1986 adjusted Census count): 128,832

2. Components of growth between July 1, 1986 and June 30, 1987

Births:	1,916	Emigration:	60
Deaths:	1,129	Interprovincial-in:	2,685
Immigration:	165	Interprovincial-out:	2,932
NPR:	66	Returning Canadians:	40

Solution: Using equation 2.2 and the above information, we can calculate the net migration for PEI for the censal year 1986-1987.

$$M_{86\text{-}87} = 165 - 60 + 2{,}685 - 2{,}932 + 66 + 40$$
$$= -36$$

The net migration can be used in equation 2.1 with the other components to obtain an estimate of the total population. In practice, equations 2.1 and 2.2 are combined into a single step.

$$\hat{P}_{87} = 128{,}832 + 1{,}916 - 1{,}129 - 36$$
$$= 129{,}583$$

2.1.2 Estimating Population by Age and Sex

Population breakdown by age and sex forms the backbone of most demographic studies. The structure by age and sex varies with time and place; at the same time, demographic behaviour is a function of age and sex. Estimates of fertility are based on the number of women in the reproductive ages. Mortality rates are heavily skewed toward the older generations. Higher migration rates are associated with young adults and their children, as they move for personal and economic reasons. Government and private sector planning and policies are largely driven by the age and sex profiles of the target populations.

a) Estimates by Year of Age

The population by age and sex can be estimated using the cohort-component method. This variant of the component method used for total population also factors in the aging of the cohorts through time. For example, individuals aged 20 one year will be age 21 the following year. Therefore the base population for estimates of age 21 is the population aged 20.

$$P_{t+1,21} = P_{t,20} \pm components$$

Estimation, at t+1, of the population counts for the 0 age group (base population = newborns born between t and t+1) and the last age group (usually 90+) uses equations slightly different from the one used for the intermediate ages. The following equations are used for each sex.

For age 0:
$$\hat{P}_{t+1,0} = B_{(t,t+1)} - D_{(t,t+1),-1} \pm M_{(t,t_1),-1} \tag{2.3}$$

where age (-1) designates persons born between t and t+1.

For ages 1 to 89:
$$\hat{P}_{t+1,a+1} = \hat{P}_{t,a} - D_{(t,t+1),a} \pm M_{(t,t+1),a} \tag{2.4}$$

where (a) is age at time t.

For the age group 90+:
$$\hat{P}_{t+1,90+} = \hat{P}_{t,89+} - D_{(t,t+1),89+} \pm M_{(t,t+1),89+} \tag{2.5}$$

where 89+ is the age group at time t.

Example 2.2 **Method of Computing Population Figures as of t + 1 (using as an example, the estimate of women in British Columbia as of June 1, 1991)**

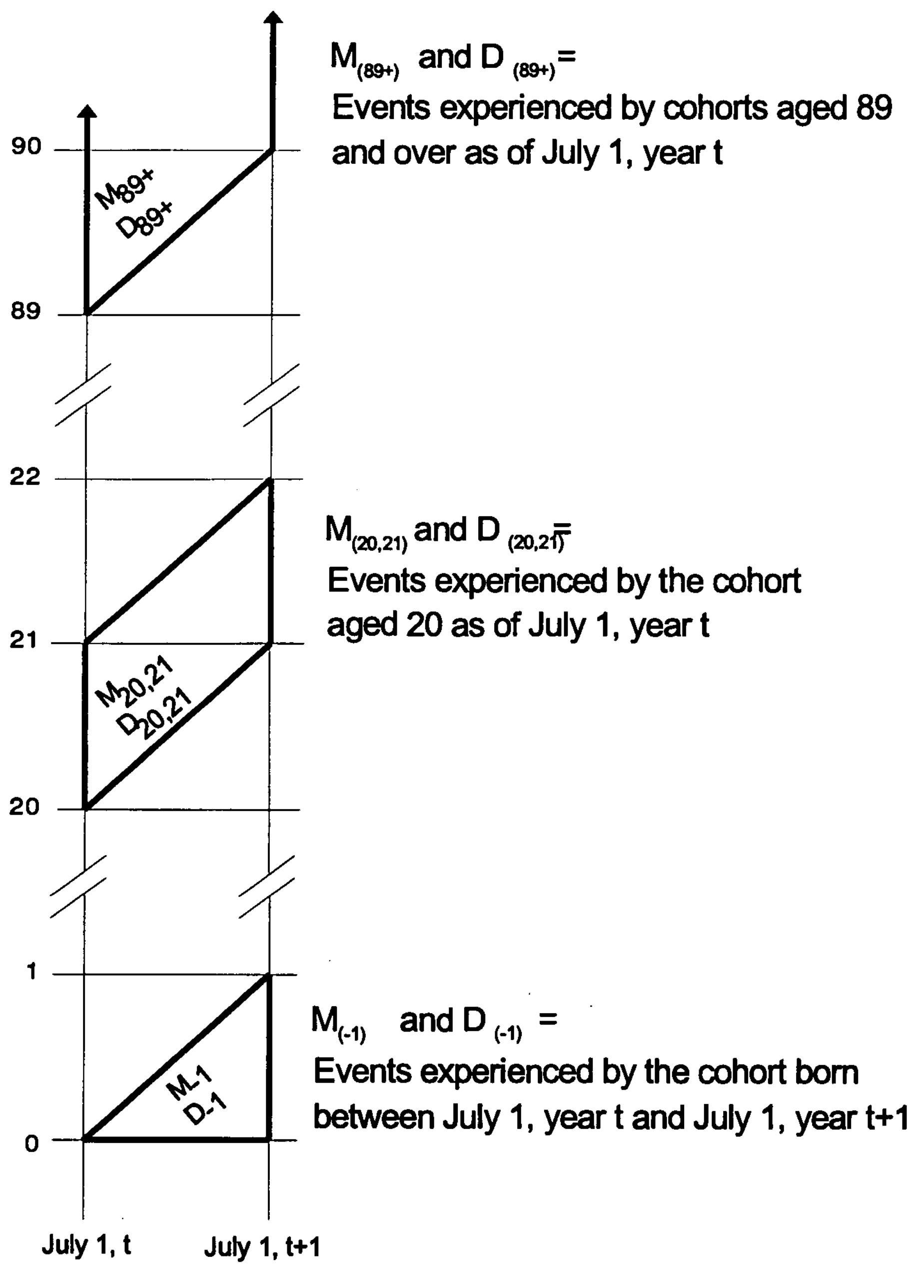
Age	Population as of t* (01/06/1990)*	Deaths**	Immigrants**	Emigrants**	In-migrants**	Out-migrants**		Population as of t+1 (01/06/1991)
-1***	21,932	145	63	6	156	102	↘	
0	21,918	31	127	13	365	240	↘	21,898
1	21,273	7	130	23	607	370	↘	22,126
2	21,033	6	114	28	658	387		21,610
.								21,384
.								
19	22,185	8	275	36	845	555	↘	
20	22,162	12	307	40	970	602	↘	22,706
21	21,916	12	311	45	1,095	649	↘	22,785
.								22,616
.								
87	2,985	276	5	1	15	11	↘	
88	2,593	298	3	0	11	6	↘	2,717
89	2,223	283	1	0	9	4	↘	2,303
90+	8,119	1,622	3	2	13	8	→	8,449
Total	**1,583,199**	**10,825**	**15,171**	**2,572**	**43,177**	**23,721**		**1626361**

* Final Postcensal Estimates

** Updated Postcensal Estimates

*** The Population -1 years of age as of t is equal to the number of births between t and t+1 (updated data). When using equation 2.1, this figure corresponds to B (births) and must be taken into account to produce the population figure of 1,626,361 as of t+1.

Figure 2.1 Estimate of Population as of t+1, selected ages



b) Estimates by Age Group

More often than not, population estimates are required by 5-year, 10-year or other broad age grouping rather than by single year of age. Input data are often only available by broad age group. When they are available, data by single year of age or by 5-year age group are often used as building blocks to generate larger groupings suited to specific user needs. For example, funding to local school boards is based on the number of children of school age. Real estate agents studying the market for first-time home buyers are mainly interested in the population between the ages of say 25 and 45, perhaps broken down by five-year age group. When estimates are not detailed enough to form the required grouping, data by single year of age can be generated from broad age groups using some form of interpolation technique.

When estimating population by broad age group, it is preferable to use periods of estimation equal in length to the age interval. For example, estimates by 5-year age group should be calculated every 5 years. The base population for a given age category in this case is the preceding age category.

Because the method tracks cohorts, component data must be classified by cohort (or cohort group). Otherwise the data must be adjusted to meet method requirements.

The following equations: 2.6 for the first group, 2.8 for the last and 2.7 for the intermediate age groups, correspond to equations 2.3, 2.5 et 2.4.

For the (0,4) age group:

$$\hat{P}_{t+5,(0.4)} = B_{t,t+5} - D_{t,t+5,(0-4)} \pm M_{t,t+5,(0-4)} \tag{2.6}$$

For the (5-9) to (85-89) age groups:

$$\hat{P}_{t,t+5,(a+5,a+9)} = \hat{P}_{t,(a,a+4)} - D_{t,t+5,(a,a+4)} \pm M_{t,t+5,(a,a+4)} \tag{2.7}$$

For the (90+) age group:

$$\hat{P}_{t+5,(90+)} = \hat{P}_{t,(85+)} - D_{t,t+5,(85+)} \pm M_{t,t+5,(85+)} \tag{2.8}$$

where the (0,4) group refers to those born between t and t+5, and (a,a+4) refers to cohort ages at time t.

2.1.3 Estimating Population by Other Demographic Characteristics

In addition to age and sex, users often require estimates of population according to other characteristics (e.g. marital status, ethnic origin, mother tongue). In the case of marital status, vital statistics provide annual data at the national and provincial levels on the number of deaths by age, sex and marital status; the number of marriages by age, sex and previous marital status; and the number of divorces by age and sex. Immigration statistics are also available for these areas by age, sex and marital status.

Thus, most of the components are available for estimating the population of Canada, the provinces and territories by age, sex and marital status using the component method. Details on the methods and data sources used by Statistics Canada to estimate these populations can be found in Population Estimation Methods, Canada (Statistics Canada, 1987, Catalogue 91-528).

For many other characteristics there is unfortunately little current information available on the components of demographic change. One solution is to apply the age and sex distributions of the population for these characteristics from the preceding census to the current population estimates by age and sex. If it is felt that these distributions are changing over time, they can be extrapolated from the two previous censuses.

Example 2.3
Estimate the number of married men in Canada as of July 1, 1987 using the following two approaches:
i) Distribution from the most recent census, adjusted for net census undercoverage
ii) Component method

Solution:
i) Distribution from the most recent census

$$\hat{P}^m_{87} = \hat{P}_{87} \bullet \frac{P^m_{86}}{P_{86}}$$

$$= 13{,}179{,}630 \bullet \frac{6{,}411{,}444}{13{,}012{,}940}$$

$$= 6{,}493{,}571 \text{ married men}$$

where $\hat{P}^m_{87}$ = estimate of the number of married men, as of July 1, 1987;

$\hat{P}_{87}$ = estimate of the total number of males, as of July 1, 1987;

P^m_{86} = 1986 adjusted Census count of married men;

P_{86} = 1986 adjusted Census count of males.

ii) Component method

$$\hat{P}^m_{87} = P^m_{86} + M_{86,87} - V_{86,87} - W_{86,87} - D^m_{86,87} + I^m_{86,87} - E_{86,87}m + NPR^m_{86,87} + RC^m_{86,87}$$

= 6,411,444 + 179,657 - 91,321 - 26,340 - 61,988 + 29,951 - 11,873 + 13,744 + 6,341
= 6,449,615 married men

where, between July 1st, 1986 and June 30, 1987,

M = number of men who married;

V = number of men who divorced;

W = number of new widowers;

D^m = number of deaths of married men;

I^m = number of immigrant married men;

E^m = number of emigrant married men.

NPR^m = net flow of married non-permanent residents

RC^m = number of married returning Canadians

With only a single year between the census and reference date, the methods differ by 0.68%. Any changes over time to the distribution of the population by marital status are only captured by the component method.

2.2 Data Sources

This section describes the principle sources of data that are used to estimate population by the component method. Some, like those for base population, births and deaths, are the same for most types of geographic areas for which estimates are required. Data on migration are not as readily available, and their number and source varies with the type of geographic area (national, provincial or subprovincial).

Variable	Sources
1. Base Population	Census
2. Vital Events	
- Births	
- Deaths	
- Marriages }	Canadian Centre for Health Information,
- Divorces }	Statistics Canada
3. Migration	Census
	Administrative Files
	- Employment and Immigration Canada
	- Revenue Canada
	- Health and Welfare Canada
	- Provincial Governments
	- Municipal Governments
4. Socio-Demographic Characteristics	Census

2.2.1 Base Population

The base population is generally taken from the quinquennial Census of Canada, which covers a wide variety of demographic and social characteristics. The universe of the *de jure* censuses[3] extends to the following individuals[4]:

- all persons whose usual place of residence is somewhere in Canada;
- Canadian government employees and their families stationed abroad;
- members of the Canadian Armed Forces and their families stationed abroad; and
- crews of Canadian merchant vessels;
- persons residing in Canada and holding a ministerial permit, and their families;
- persons residing in Canada who have applied for refugee status, and their families;

---

[3]A de jure census is designed to count persons according to their usual place of residence.

[4]In 1991 the universe of the Canadian census was expanded to include temporary residents (i.e. persons holding employment authorizations, student authorizations, visitor permits and ministerial permits).

- workers from another country staying in Canada temporarily and holding an employment authorization, and their families;
- students attending school in Canada whose usual residence is outside Canada and hold a student authorization, and their families;
- residents of another country visiting Canada temporarily.

Not included in the census counts are the following persons:

- govemment representatives of other countries and their families attached to the legation, embassy or other diplomatic body, residing in Canada;
- members of the armed forces of other countries and their families, stationed in Canada.

Adjustment for Net Census Undercoverage

After each census, Statistics Canada conducts studies to measure the quality of the data collected. The most imortant of these is the Reverse Record Check (RRC), designed to estimate the number of persons not enumerated in the census but who were part of the census universe. Undercoverage is considered to be the largest source of error. The RRC, first undertaken in 1966, consists of a number of tracing procedures applied to a sample of persons, all of whom should have been enumerated. This sample is drawn from five sources: persons enumerated in the previous census; persons born since the previous census; landed immigrants who entered Canada since the previous census; non-permanent residents at the time of the census (for 1991 only); and persons not enumerated in the last census. Undercoverage in 1991 was estimated at about 950,000 (3.4% of the adjusted population).

In 1991, Statistics Canada introduced the Overcoverage Study, which checked samples of persons enumerated in 1991 to determine if they were counted more than once, or if they were not part of the census universe. Overcoverage in the 1991 Census was estimated at about 150,000 persons (0.6% of the adjusted population). Net census undercoverage is the difference between undercoverage and overcoverage.

2.2.2 Vital Events: births, deaths, marriages and divorces

Vital events have been recorded on a regular basis by provincial and territorial governments since 1921. Vital statistics are therefore a rich source of information on births, deaths, marriages and divorces. Furthermore, because of the recording of vital events is required by law, the coverage and quality of the data are very high.

Births and deaths are tabulated at the census subdivision level by age (deaths only), sex and marital status (deaths only). From these statistics, numbers of births and deaths can be derived for any area comprised of whole census subdivisions including Canada, provinces, territories, census divisions and census metropolitan areas.

Marriages and divorces, on the other hand, are compiled according to the court, or place of occurrence. For many types of subprovincial areas this may not correspond to the place of residence. Thus the data on marriages and divorces are reliable only at the national provincial and territorial levels. These data are available by age, sex and previous marital status.

Table 2.1 Level of Geographic and Demographic Detail for Births and Deaths from Canadian Centre for Health Information, Statistics Canada

Area	Total	Age*	Sex	Marital Status*
Canada	X	X	X	X
Provinces	X	X	X	X
Territories	X	X	X	X
Census Divisions	X	X	X	S
Census Metropolitan Areas	X	X	X	S
Census Subdivisions	S	S	S	S

*Deaths only.
X = available.
S = available upon special request.

Table 2.2 Level of Geographic and Demographic Detail for Marriages and Divorces from Canadian Centre for Health Information, Statistics Canada

		For Each Spouse		
Area	Total	Age	Sex	Previous Marital Status*
Canada	X	X	X	X
Provinces	X	X	X	X
Territories	X	X	X	X
Census Divisions				
Census Metropolitan Areas				
Census Subdivisions				

*Marriages only.
X = available.

2.2.3 Immigration

Immigration statistics are available from the Employment and Immigration Canada Landed Immigrant file. All immigrants are required to submit an "Immigrant Visa and Record of Landing" form upon arrival in Canada.

Immigration statistics include only those immigrants who are lawfully admitted into Canada to establish permanent residence. They do not include immigrants entering Canada illegally; returning Canadian citizens who worked, studied or visited abroad; or visitors, boarders, commuters, students, workers, diplomatic and consular representatives and their families from other countries.

Table 2.3 Level of Geographic and Demographic Detail for Immigration from Employment and Immigration Canada

Area	Total	Age	Sex	Marital Status*
Canada	X	X	X	X
Provinces	X	X	X	X
Territories	X	X	X	X
Census Divisions	S	S	S	S
Census Metropolitan Areas	S	S	S	S
Locality				

*Data extracted from Employment and Immigration Canada files.
X = available.
S = available upon special request.

As the registration of immigrants is required by law, the coverage and quality of these data, are very high, at least at the national level. The distribution of immigrants by sub-national area is obtained through the intended place of destination, identified by locality codes. The statistics at the provincial and territorial level are of very good quality, while those for subprovincial areas are somewhat more problematic. First, the intended local area of destination indicated by the immigrants is often different from the actual destination. This is in large part due to the fact that many immigrants coming to Canada are not really certain of their destination and only know the names of a few (and often major urban) centres. Second, assigning locality codes to user-specified aras (e.g. census divisions) is very difficult. See Module 5 for details.

2.2.4 Non-permanent Residents

The census universe was expanded in 1991 to include non-permanent residents, who numbered about 223,000. Estimates of non-permanent residents for other years were developed from the Visitors Data System (VIDS) obtained from Citizenship and Immigration. The VIDS file provides information since 1981 on the number of permits authorizing temporary residence in Canada. A methodology was developed to derive the number of persons this represents and their demographic characteristics. In particular, it accounts for persons holding more than one permit and estimates the number of dependants who do not have permits (such as preschoolers and seniors). For years before 1981, estimates of the total number of non-permanent residents were derived from information on student authorizations and work permits available from the Education, Culture and Tourism Division of Statistics Canada.

2.2.5 Returning Canadians

As there is no direct data source on the number of Canadians returning to the country, their estimates were derived from a model based on information on the length of stay abroad using a form filled out by persons entering Canada by air (Customs and Excise data) and estimates of the number of emigrants by province and territory produced by Demography Division spanning the 1984-1985 to 1994-1995 period (Statistics Canada, Catalogue no. 91-537).

The total provincial/territorial estimates were distributed by age, sex and marital status according to the distribution observed in the 1991 Census data on the one-year mobility question.

2.2.6 Emigration and Internal Migration

Unlike immigration, there is no provision in Canadian law for people leaving the country or moving from one region to another within Canada to record this fact. Thus, estimates of their number and characteristics must be derived through secondary sources such as Canadian administrative files or immigration statistics of recipient countries.

The principal data sources for emigration and internal migration used at Statistics Canada are the Revenue Canada Income Tax file and the Child Tax Benefit file. Secondary sources include the Census, which provides important data on the demographic characteristics of migrants, and the United States immigration file, which contains information on immigrants to the United States whose last country of permanent residence was Canada. Details on the methodology used to estimate emigration and internal migration are given in **Annual Demographic Statistics** (91-213).

Table 2.4 Level of Geographic and Demographic Detail for Emigration and Internal Migration from Various Files

Area	Total	Age	Sex	Marital Status
1. Census*				
Canada	X	X	X	X
Provinces	X	X	X	X
Territories	X	X	X	X
Census Divisions	X	X	X	X
Census Metropolitan Areas	X	X	X	X
Census Subdivisions	X	X	X	X
2. Revenue Canada - Income Tax				
Canada	X	X[1]	X	
Provinces	X	X[1]	X	
Territories	X	X[1]	X	
Census Divisions	X	X[1]	X	
Census Metropolitan Areas				
Census Subdivisions	S	S[1]	S	
3. Revenue Canada (child tax benefit)				
Canada		X[2]		
Provinces		X[2]		
Territories		X[2]		
Census Divisions				
Census Metropolitan Areas				
Census Subdivisions		S[2]		

* Internal migration only.
1 Broad age group.
2 Ages 0-17.
X = available.
S = available upon special request.

MODULE 3
OTHER METHODS

Other Methods

3.1 The Regression Method

3.1.1 Methodology

The method uses multiple regression techniques to establish the mathematical relationship between the changes in series of indicators and changes in population figures. More specifically, we derive a regression equation expressing the relationship between (1) the change over an intercensal period in the proportion that a small area represents in a large area for several series of indicators and (2) the change in the proportion that the population of the small area represents in the population of the large area.

The regression method can produce results far more quickly than the component method. However, it estimates only total population and provides no information on the components of population growth. Furthermore, the regression method rests on the assumption, sometimes problematic (especially for small areas), that the relationship between changes in symptomatic indicators and demographic change remains constant. It also assumes that the quality of indicator data is constant.

At Statistics Canada, the regression method, employing the ratio-correlation or the difference-correlation technique, is used to estimate any event (population) using a set of symptomatic indicators as predictors (independent variables). The equation expressing this relationship for the j^{th} small area is given by:

Basic equation:

$$\Delta y_j = \alpha + \beta_1 \Delta x_{1j} + \ldots + \beta_n \Delta x_{nj} + \varepsilon \qquad (3.1)$$

where Δy = the vector of change in proportional values of the dependent variable between census years t and t-5;

Δx = the vector of change in proportional value for the specific independent variable (symptomatic indicator);

α = the regression intercept;

β = the regression coefficient for the specific independent variable;

ε = the vector of stochastic errors, such that $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2$.

In the ratio-correlation regression method,

$$\Delta y_j = \frac{P_{j,t}}{P_t} + \frac{P_{j,t-5}}{P_{t-5}} \qquad (3.2)$$

and in the difference-correlation method,

$$\Delta y_j = \frac{P_{j,t}}{P_t} - \frac{P_{j,t-5}}{P_{t-5}} \tag{3.3}$$

where $y_j = P_{j,t} / P_t$

and	t and t-5	=	two consecutive censal years;
	P	=	the population of a province (or larger area);
	P_j	=	the population of the jth small area, such that $\Sigma P_j = P$;

and the derivation of the Δxs can be inferred from that given for the Δys.

Equation 1 is fitted by the Least Squares method. For additional details on the method and a summary of the adjustments deemed necessary and the precautions taken in applying the method to control specific problems (multicollinearity, homogeneity of variance of the error term (ε) and heteroscedasticity), refer to Chapter VI in Population Estimation Methods, Canada, Catalogue No. 91-528.

Knowing the values of α, the βs, of P_{t+i}, and the value of the Δxs at time t+i, the population of a small area P_j at time t+i can be calculated, for any i^{th} year, where i ranges from 1 to 5.

In ratio-correlation, the population estimates are given by:

$$\hat{P}_{j,t+i} = [\ y_{j,t} \bullet \hat{\Delta} y_j\ ] \bullet \hat{P}_{t+i} \tag{3.4}$$

In difference-correlation, the population estimates are given by:

$$\hat{P}_{j,t+i} = [\ y_{j,t} + \hat{\Delta} y_j\ ] \bullet \hat{P}_{t+i} \tag{3.5}$$

where	$\hat{P}_{j,t+i}$	=	estimated population for small area j at time t+i;
	$\hat{P}_{t+i}$	=	independent estimate of population for the larger area containing all the small areas at time t+i;
	$y_{j,t}$	=	proportion of the population of small area j with respect to the larger area, according to the census counts at time t;
	$\hat{\Delta} y_j$	=	estimate of Δy_j, the change in proportion of the population of small area j with respect to the larger area, between time t and t-h as derived by the regression model.

Neither the ratio-correlation nor the difference-correlation method uniformly or routinely outperforms the other. A choice between the two is thus contingent upon a thorough evaluation of the results to date. Estimation procedure depends, in certain respects, on whether it is applied to census divisions or census metropolitan areas, as discussed below in sections 3.3 and 3.4.

3.1.2 Regression-Nested Procedure

Regression-based population estimates are subject to fluctuations over the intercensal years (due to temporal instability in the regression coefficients, inconsistency of input data quality, mis-specification of model construction). Consequently, annual growth rates based on regression estimates of population are sometimes not comparable to those from component data. To deal with this issue, and to ensure that the most valid estimates are taken into account, Statistics Canada has developed a technique called the *Regression-nested method* which resolves the inconsistency in the annual rate of growth based on the preliminary (regression-based) estimates of population and the final (component-based) estimates (Statistics Canada, Catalogue 91-538E, 1987; c. VI).

Table 3.1 Methodology for Preliminary Population Estimates (Regression-nested) for Census Divisions

Time	Regression Estimate	Component Estimate[1]	Preliminary Estimate
t+1	$\hat{P}_{t+1}$	$\hat{P}'_{t+1}$	$\hat{P}'_{t+1}$
t+2	$\hat{P}_{t+2}$	$\hat{P}'_{t+2}$	$\hat{P}'_{t+1} + [\hat{P}_{t+2} - \hat{P}_{t+1}]$
t+3	$\hat{P}_{t+3}$	$\hat{P}'_{t+3}$	$\hat{P}'_{t+2} + [\hat{P}_{t+3} - \hat{P}_{t+2}]$
t+4	$\hat{P}_{t+4}$	$\hat{P}'_{t+4}$	$\hat{P}'_{t+3} + [\hat{P}_{t+4} - \hat{P}_{t+3}]$
t+5	$\hat{P}_{t+5}$	$\hat{P}'_{t+5}$	$\hat{P}'_{t+4} + [\hat{P}_{t+5} - \hat{P}_{t+4}]$

1 The method uses census counts adjusted for net census undercoverage as a base population plus births and deaths from Vital Statistics Records and migration data derived from Revenue Canada Taxation Files.

Source: Statistics Canada, **Population Estimation Methods, Canada**, Catalogue No. 91-528, 1987, p. 79.

3.1.3 Census Divisions (CDs)

The specifications for the regression models for each province are based on census data for census divisions.

For producing population estimates, the regression models for each province utilize the best available symptomatic indicators of population change: the number of family allowance recipients aged 1-14 (most provinces); the population registered in provincial health insurance programs (Saskatchewan and Alberta), and the number of hydro accounts (British Columbia). Model specifications and coefficients are usually drawn from data on the intercensal period preceding the current postcensal estimate. These elements vary from one time period to another as a function of the periodic evaluation of their effectiveness. *In addition, in 1993 the Family Allowance program was replaced by the Child Tax Benefit program. Recipient of the new program were thus used as of 1993.*

The main reason for using different regression models for different provinces is to maximize the accuracy of population estimates by taking advantage of administrative files containing data on specific local areas.

The regression method and the variables selected are the ones which give the lowest mean absolute error.[6] The sum of the regression estimates for census divisions within a specific province is adjusted to the corresponding provincial total obtained by the component method. The accuracy of the regression-based estimates for census divisions was found to be a little weaker than that of the regression-nested and component-based methods.

3.1.4 Census Metropolitan Areas (CMAs)

The CMA population estimates are based on aggregation of the census division regression-nested estimates for all "CMA convergent CDs" (i.e., all CDs located entirely within CMAs, or whose boundaries intersect CMA boundaries). Based on the CD and CMA adjusted enumerated population for the previous census year, the population of a CMA is expressed as a ratio of the sum of the populations of the "convergent" CDs. This ratio (assumed to remain constant), is then applied to the sum of CD population estimates at the reference date, providing an estimate of population for the CMA at the same reference date. This operation is summarized in the following equation:

$$\hat{P}_{t+i}^{CMA} = \frac{P_t^{CMA}}{\sum_{CD} P_t^{CD}} \bullet \sum_{CD} P_{t+i}^{CD} \qquad (3.6)$$

where $\hat{P}_{t+i}^{CMA}$ = estimated population of CMA in year t+i;

P_t^{CMA} = adjusted enumerated CMA population in year t;

P_{t+i}^{CD} = estimated population in year t+i for a CMA convergent census division, (estimate, obtained by the regression-nested procedure);

P_t^{CD} = estimated or adjusted enumerated population in year t for a CMA convergent census division.

---

[6]In the weighted regression methods, the data are adjusted using weighting factors to ensure constant variance of the random error term.

At the census metropolitan area level, this method has been found to be superior to the regression method.

Example 3.1

Using the following information, estimate the population of Saint John, New Brunswick in 1990, using the regression method (ratio correlation). We are using an example from the 1986-1991 period to maintain a single source for family benefits:

	Population (thousands)		Family Allowance	
	1986	1990	1986	1990
Saint John	84.5	?	16.1	144.5
New Brunswick	727.7	743.0	16.7	152.4

Regression coefficients: $\alpha = 0.3202$
$\beta = 0.6834$

Solution: i) Independent variable x_i

$$x_i = \frac{AF_{90,SJ}}{AF_{90,NB}} + \frac{AF_{86,SJ}}{AF_{86,NB}}$$

$$= \frac{16.7}{152.4} + \frac{16.1}{144.5}$$

$$= 0.9835$$

ii) ~~In~~dependent variable Y_i

$$Y_i = \alpha + \beta x_i$$

$$= 0.3202 + (0.6834)\ (0.9835)$$

$$= 0.9923$$

iii) Population estimate

$$P_{90,SJ} = Y_i \bullet \frac{POP_{86,SJ}}{POP_{86,NB}} \bullet POP_{90,NB}$$

$$= 0.9923 \bullet \frac{84.5}{727.7} \bullet 743.0$$

$$= 85.6 \textit{ thousand}$$

Example 3.2

Estimate the population of Saint John, New Brunswick as of June 1, 1990, using the regression method (difference-correlation), using the same information on family allowances and population as in Example 3.1, and β coefficient = 0.6574.

Solution: i) Independent variable x_i

$$x_i = \frac{AF_{90,SJ}}{AF_{90,NB}} + \frac{AF_{86,SJ}}{AF_{86,NB}}$$

$$= \frac{16.7}{152.4} + \frac{16.1}{144.5}$$

$$= -0.0018$$

ii) Independent variable Y_i

$$Y_i = \beta + \beta x_i$$

$$= 0.6574 \; (\; -0.0018 \;)$$

$$= 0.0012$$

iii) Population estimate

$$P_{90,SJ} = Y_i \bullet \frac{POP_{86,SJ}}{POP_{86,NB}} \bullet POP_{90,NB}$$

$$= 0.0012 \bullet \frac{84.5}{727.7} \bullet 743.0$$

$$= 86.3 \textit{ thousand}$$

3.2 Rate of Growth Method for Census Subdivisions

In the past, the census was relied upon as the primary source of population data for census subdivisions (CSDs) or municipalities. However, users are now increasingly concerned with obtaining estimates and projections based on the most current demographic information available. This shift in focus is attributable to two main factors. First, in these time of fiscal restraint, all levels of government have to deal with reduced if not continuously diminishing budgets. Population data thus play an increasingly important role in the planning and costing of municipal programs and services and in their funding by the various levels of government. In particular, certain provincial and territorial agencies produce population estimates at the subprovincial level, taking advantage of local area data. The second factor relates to the adjustment for net census undercoverage, that is, persons missed or counted more than once in the census, which is now built into Statistics Canada's population estimation program (PEP). Although this program does not currently extend to municipalities, users require that population data for these areas be consistent with the adjusted population figures available for the provinces and territories, census divisions and census metropolitan areas, all of which are covered by PEP.

3.2.1 Census Subdivisions Defined

As a rule, municipalities are determined by provincial legislation.[7] In Statistics Canada, municipalities belong to the type of legislative/administrative areas known as the CSD, which also include Indian reserves, Indian settlements, and unorganized territories.[8] Municipality and other CSD types include boroughs, cities, hamlets, parishes, resort villages, towns and villages. At the time of the 1991 Census, there were 6,006 CSDs across Canada. CSD frequencies by province and territory are shown in Table 3.2 below. CSDs aggregate to census divisions, census metropolitan areas and, from our experience, to most user-defined geographic areas.

Table 3.2 Number of Census Divisions[9] and Census Subdivisions by Province and Territory, 1991 Census

	Total	Province / Territory											
		Nfld.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.
Adjusted Population July 6, 1996 (in thousands)	29,964	571	137	943	763	7,389	11,252	1,144	1,023	2,790	3,855	31	67
Census division	290	10	3	18	15	99	49	23	18	19	30	1	5
Census subdivision	6,006	404	126	118	287	1,637	951	293	953	438	691	36	72

One of the issues involved in estimating the population of CSDs revolves around the potential changes in boundaries over time.[10] Under the present methodology, where estimation is based on two censuses, the problem is somewhat lessened through the overlaying of the previous census data with the most current census geography. For instance, both the 1986 and the 1991 Census data bases are mapped according to the 1991 Census geography. Keeping constant the geographical definition of the CSDs greatly simplifies the population estimation procedures. However, the overlaying can result in slight misallocations in small

---

[7] In certain regions in Newfoundland, Nova Scotia and British Columbia, the term also describes geographic areas that have been created by Statistics Canada in co-operation with the provinces as equivalents for municipalities. See Table A1 in the Appendix.

[8] The reference date for designating CSDs for the 1991 Census is January 1, 1991. Indian reserves that had been populated at the time of the three previous censuses (i.e., 1976, 1981 and 1986) were recognized as CSDs. Also recognized were the Indian reserves identified by the department of Indian and Northern Affairs Canada as having a population of at least 10 inhabitants between the 1986 Census and January 1, 1991. Statistics Canada has recognized Indian settlements (generally located in unorganized territory) as CSDs with the co-operation of the provincial or territorial authorities.

[9] In Newfoundland, Manitoba, Saskatchewan and Alberta, provincial law does not provide for census division geographical areas. These have therefore been created by Statistics Canada in co-operation with the respective provinces. In New Brunswick, in order to maintain the integrity of component census subdivisions, census divisions do not respect the legal county limits. In Québec a few census divisions are combined to bring the number to less than 100.

[10] The extent of boundary changes varies over any given time period. The changes that occurred between the 1986 and 1991 Censuses are documented in *1991 Standard Geographical Classification: SGC 1991, Vol. III, Changes 1986 to 1991*, Statistics Canada (1992).

geographical areas. These must be taken into account when producing the estimates. Postcensal estimates are according to the 1991 Census geography and do not take into account boundary changes since then.

3.2.2 Data Sources for Census Subdivisions and their Constraints

The quinquennial census remains by far the primary source for CSD population data through its high quality detailed demographic information. The most important obstacle for producing accurate annual population estimates for CSDs is the lack of current information at this level of geography. There exist very few other sources of demographic information that can be readily used for this purpose. Vital statistics, e.g., birth and death registrations, constitute a high potential data source. Other data sources, as defined below, could eventually be used.

Census Data

The quinquennial census provides population counts as well as population characteristics according to a wide range of demographic and socio-economic variables for all CSDs in Canada. These data constitute the basis from which the CSD population estimates are produced. Before they can be used, however, some adjustments are necessary.

First, although the accuracy of the census is generally very high, some errors inevitably occur that are not identified and corrected before the data base is frozen. Errors usually involve two adjacent CSDs with a shift in population from one to the other. As a rule, the differences between the initial and the revised counts usually cancel out for the larger geographical entities incorporating the two CSDs (e.g., the census division). However, the impact of the update on an individual CSD can be significant and must be taken into account.

Second, the census counts are adjusted for net census undercoverage. The sample size for the Reverse Record Check and the Overcoverage Studies are too small to produce results for geographic units below the province and territory. However, differences in undercoverage rates tend to be greater among age/sex groups than among geographic regions. For example, the census misses more men than women and more young adults than older adults (Statistics Canada 1996). Thus, the approach that was adopted by Statistics Canada for small areas, including census divisions and CSDs, is to apply the provincial or territorial age/sex- specific rates of net census undercoverage (excluding partially enumerated Indian reserves and settlements) to the enumerated population of the small areas by age and sex.

Third, in either 1986 and/or 1991 a number of Indian reserves and settlements (which are defined as individual CSDs) throughout Canada did not participate fully in the census. Their population was estimated separately by Statistics Canada (see Lebrasseur D, 1992) and is incorporated directly into the corresponding CSD estimates for 1986 and 1991.

Fourth, the 1991 Census universe was expanded to include non-permanent residents. The 1986 Census data thus also require adjustment. The estimates of the number of non-permanent residents by province and territory for 1986 were based on information from the Citizenship and Immigration *Visitors Data System*, while the distribution by subprovincial region was the same as in the 1991 Census.

Finally, the 1986 Census counts for CSDs must be adjusted for 1991 geography. The total population is available from the C91 data base where 1986 microdata are used to reconstruct 1991 boundaries. For details by age and sex, a mapping procedure superimposes the 1991 geography onto the 1986 counts, resulting in slight misallocations for geographically small areas such as CSDs. As a rule, such misallocations are corrected by adjusting to the C91-based total.

Vital Statistics

Few data are available on the change in population at the level of the CSD. The best potential source for all CSDs in Canada is the vital statistics registration files on births and deaths. Because Canadian citizens have a legal obligation to register these vital events, coverage is virtually complete and the overall quality of the files is quite high. However, they currently present some constraints for use at the level of the municipality.

The major limitation is the misallocation of events among neighbouring areas. That is, births and deaths occurring within a given municipality are sometimes coded to a neighbouring area. In some extreme cases, all events are erroneously coded, which results in the municipalities registering no births or deaths at all, while others are given more than their share of events. The current coding procedures do not consistently adhere to 1991 boundaries. While such misallocations tend to cancel each other out for the larger geographic regions such as the census division, they represent a major impediment to the use of the data for municipalities. Ongoing research in this area includes examining using postal codes to better pinpoint each event's geographical coordinates.[11] Good quality vital statistics would allow consideration of alternative methodologies (e.g., the application of fertility and mortality rates or the vital rates method) for the estimation of the population of small areas. Alternatively, births and deaths could simply enter as natural growth into the standard equation for estimating population using the components approach (data sources would also have to be developed for estimating the other components of growth).

Other Data Sources

Other potential data sources include the use of administrative data available for selected areas, such as electric utilities connections and housing building permits. From an individual municipality's point of view, the best source of information is often the municipality itself. Though the local administrators may not have exact counts, they often have a good idea whether their population is increasing or decreasing and what is the order of magnitude of these increases or decreases. The obvious challenge is to find how to take advantage of this valuable information.[12] Use of data from other federal programs such as Revenue Canada's Child Tax Benefit or that same department's personal income tax file also remains a possibility. (At present, final annual migration estimates for census divisions and census metropolitan areas are derived from the latter.) In each of these cases, a thorough assessment would have to be made of the differences in the concepts and

---

[11] This research is being conducted by the Geocodes Committee, which involves the participation of Statistics Canada's Demography Division, Health Statistics Division, and Operations and Integration Division, as well as Health Canada's Laboratory Centre for Disease Control.

[12] The "best guesses" of municipalities were incorporated into a decision table based model by the Bureau de la statistique du Québec in 1989 for providing population estimates for the municipalities of that province.

definitions associated with municipality population estimation, of how well the data cover the population universe, of the quality of the data, its accessibility and its cost.

3.2.3 Rate of Growth Methodology for Estimating the Population of CSDs

At present, given the lack of comprehensive, factual data on population change, the preferred methodology for estimating municipal populations is based on each area's growth during the period spanning the last two quinquennial censuses (presently, the 1986 and 1991 Censuses), controlled to the next higher-up geographical entities for which factual data are available, i.e., census divisions. The approach entails three major steps. The first involves adjusting the census base populations as described in the previous section. The second step consists of generating *total* population estimates for the CSDs. Finally, estimates are produced by *single year of age and sex.*[13] In each step procedures are applied to ensure that the sum of the CSD estimates coincide with the corresponding census division to which they aggregate. As CSDs also aggregate to census metropolitan areas, these two sets of estimates should be made consistent as well. However, the methodology is not designed to accommodate this added dimension at this point.

Total Population

Although the methodology is essentially the same, slight variations are necessary when generating the intercensal estimates (period of time between the two quinquennial censuses) as opposed to the postcensal estimates (following the last census).

Intercensal Estimates

The intercensal total population estimates for CSDs are interpolated from the adjusted 1986 and 1991 population bases through the average annual growth rate r derived using formula (1) below. This annual rate is applied to the 1986 figure to obtain an estimate of the total population for 1987. The rate is then reapplied successively to obtain the populations for 1988 to 1990.

$$r = \sqrt[5]{\frac{P_{91}}{P_{86}}} - 1$$

The above applies to all CSDs where the total population is non zero at both the 1986 and 1991 end points. Where the total population of a CSD is nil at either one of the end points, the intercensal population is estimated through linear interpolation between the two ends.

Postcensal Estimates

The postcensal total population estimates are generated through the same approach, but with two control procedures. The first relates to high growth rates observed for the intercensal period, which are defined purely by the end points. It is assumed that extremely high growth (either positive or negative) is not sustained

---

[13] The single year of age estimates include the individual ages 0 to 89 and the 90+ age group.

for protracted periods of time. Thus, the annual growth of the total population is constrained to a maximum of 10% (absolute) for the postcensal period.[14]

A second set of control procedures is put into effect for the CSDs with zero population at one of the end points. For those with no population in 1991, a zero population is maintained in the postcensal period. When there is population observed in 1991 but none in 1986, the growth rate of the corresponding census division is applied.

Population by Age and Sex

Intercensal Estimates

The intercensal populations for CSDs by age and sex are interpolated for each cohort from the 1986 and 1991 Census counts. For example, the change in the population for the cohort aged 20 in 1986 is calculated between 1986 and 1991 by comparing the population aged 20 in 1986 to the population aged 25 in 1991. The average annual growth rate for this cohort is given by:

$$r_a = \sqrt[5]{\frac{P_{91\ (age\ a+5)}}{P_{86\ (age\ a)}}} - 1$$

This average annual growth rate is then applied to the 1986 population aged 20 to obtain an estimate of the population aged 21 in 1987. The rate is then reapplied successively to obtain the population aged 22 in 1988, 23 in 1989 and 24 in 1990. Special procedures are used for the youngest and oldest cohorts.

Postcensal Estimates

The postcensal population for CSDs by age and sex are extrapolated for each 1991 cohort by reapplying successively the average annual rates observed for the 1986 cohorts. For example, the rates applied to the cohort aged 20 in 1986 for the 1986-1991 intercensal estimates are applied to the cohort aged 20 in 1991 for the 1991-1996 postcensal estimates. A threshold rate of 20% is incorporated for the age/sex postcensal estimates. Again, it is felt that extreme growth patterns observed between 1986 and 1991 are less likely to persist over a ten year period.

Before the growth rates matrix can actually be used to generate the individual age/sex estimates, procedures must first be applied to replace empty cells with rates calculated for the corresponding census division. Then, to ensure that rates are relatively smooth across ages, each age/sex rate is recalculated by applying a three-year moving average.

---

[14] While a threshold of 10% for total population (the threshold is set at 20% for the age/sex estimates) may be considered arbitrary, the examination of the data to date indicates that these levels will minimize spurious outliers while at the same time allow for some strong continued population growth or decline. These levels will be reassessed following analysis of the 1996 adjusted Census counts. For areas showing very high population change, local area information is sought to assess the current status of population change.

Ensuring Census Subdivision and Census Division Estimates are Consistent

The intercensal and postcensal population estimates for CSDs as calculated above use only the adjusted 1986 and 1991 Census counts. As CSDs aggregate to census divisions, the last step involves adjusting them through raking to the current census division population estimates available by single year of age and sex.

3.2.4 Empirical Observations

The production of CSD population estimates is being done on a cost recovery basis as part of an experimental program. Presently complete sets of CSD population estimates have been produced for six provinces and territories. We hope to have data for all twelve provinces and territories by early 1997. Empirical analysis played a crucial role in the fine tuning of the current methodology, including the development of many of the detailed adjustment procedures required to improve the quality of the estimates. In order to provide some insight into the empirical issues that were addressed during the development stages, we present some of them in light of the production of the age/sex population estimates for the CSDs of New Brunswick and Ontario.

According to 1991 Census geography, the province of New Brunswick had 287 CSDs. The largest of these is the city of Saint John, where close to 75,000 individuals were enumerated. At the other end of the scale, many CSDs of New Brunwick (and of all provinces and territories) had a population count of less than one hundred. Five of the Indian reserves in New Brunswick were incompletely enumerated, and were incorporated separately into the data base. Finally, three of the CSDs had a zero population count in 1991. Ontario on the other hand had 951 CSDs in 1991. The city of Toronto is the largest CSD with an enumerated population of 635,395. Adjustments were necessary for 34 incompletely enumerated Indian reserves in Ontario, and 13 CSDs had a zero population count in 1991.

The methodology for estimating the population by single year of age (0 to 90+) and sex for all the CSDs in New Brunswick and Ontario entails adjusting all the cells in the 1986 and 1991 Census population bases, calculating the average annual growth rate for each of these cells, and applying the rates to the individual age/sex cells in the base population.[15] The growth rates matrix is applied as is for generating estimates within the intercensal period (1986-91). As discussed above, however, further adjustments are necessary for producing postcensal estimates (post-91 period). One of these involves limiting the average annual growth of any given age/sex category to a maximum value of 20% (absolute).

As is shown in Table 3.3 approximately 2% of the age/sex cells for both New Brunswick and Ontario had growth rates exceeding the maximum. Extreme values (either negative or positive) plotted in Figure 3.1 are observed in relatively high proportions for the cohorts aged 15 to 25 years old (in 1986). As can be expected, relatively high proportions of extreme values (these ones negative) are also observed for the cohorts aged 70 and over.

[15] There are 52,234 cells per year for New Brunswick and 173,082 for Ontario. The figures correspond to the number of CSDs (287 and 951, respectively), multiplied by the number of age/sex groups (91x2).

Another adjustment to the postcensal rates matrix entails imputing a value to the empty cells. From Table 3.3, about 11% (5,997) of the cells in the growth rates matrix for New Brunswick were empty, slightly lower than the 14% (24,126) for Ontario. Figure 3.2 shows the percentage of empty cells by age across all CSDs for each province. Cohorts up to age 40 have relatively few empty cells (around 5%), while the frequency gradually increases with age thereafter. As can be expected, CSDs with very small populations account for a relatively high proportions of empty cells.

Table 3.3 Size of Age/Sex Growth Rates Matrix, Rates Exceeding |20%|, and Empty Cells

	N.B.	Ont.
Total number of elements in the age/sex rates matrix	**52234**	**173,082**
Total number of rates exceeding |20%|	**1,048**	**4,772**
.. Number of rates greater than 20%	594	3,094
.. Number of rates less than -20%	454	1,678
Total number of empty element	**5,997**	**24,126**

Figure 3.1a Percentage of Age/Sex Cells Exceeding |20%| By Age, New Brunswick CSDs

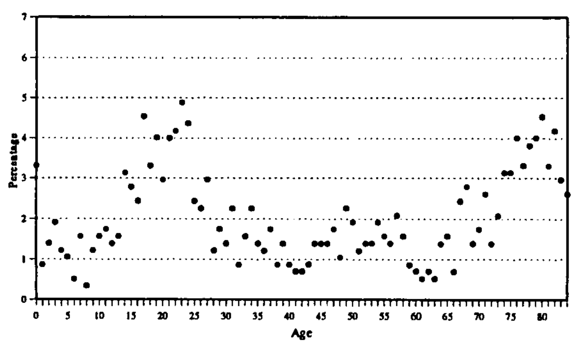


Figure 3.1b Percentage of Age/Sex Cells Exceeding |20%| By Age, Ontario CSDs

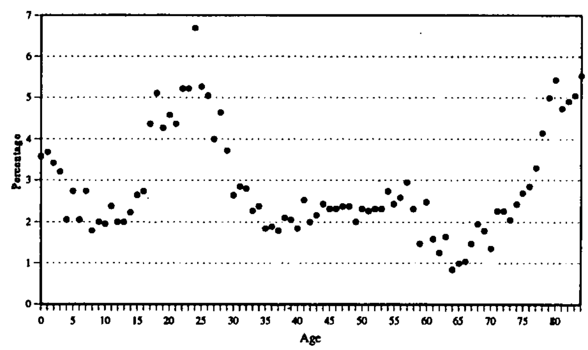


Figure 3.2a Percentage of Empty Age/Sex Cells By Age, New Brunswick CSDs

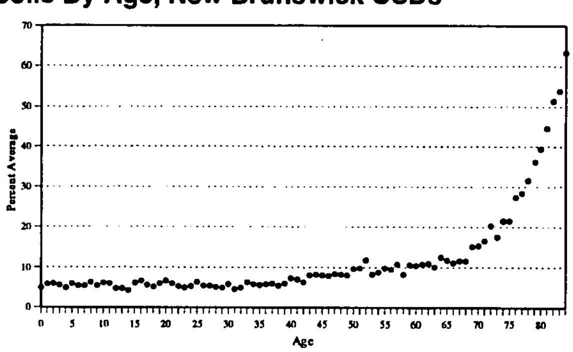
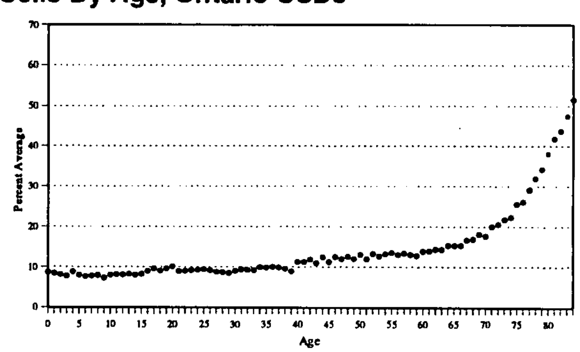


Figure 3.2b Percentage of Empty Age/Sex Cells By Age, Ontario CSDs



3.2.5 Alternative Methodologies

The current growth rate method for estimating the population of CSDs is one of various methods for population estimation. It was adopted and fine tuned following a study in which the performance and adequacy of a number of methods were examined. Census divisions were part of the study as they provide information on how the component method compares to other methods. Among the approaches that were evaluated are the component method (for census divisions only), and methods based on the rate of growth, proportional allocation, vital rates, and simply keeping the base population constant.

The performance of the various methods was evaluated by comparing total population estimates for 1991 (based on the 1986 adjusted census counts) to the 1991 adjusted census counts. This involved examining the results for each of Canada's 290 census divisions, where complete information is available for all methods considered. In a second phase, the rate of growth, proportional allocation, vital rates and constant methods were evaluated for selected CSDs. Details of the evaluation are presented in "Issues in Estimating the Population of Canadian Municipalities" by Bender and Bédard, published in *Annual Research Conference and Technology Interchange*, 1996. In summary, for census divisions the component method was found far superior to all other methods, and the rate of growth method ranked second. For CSDs, where no components are available, the rate of growth method had the lowest errors. In both cases however, the difference between the rate of growth and the next-best method was not as marked.

Table 3.4 Average Absolute Error of Closure (%) for 1991 according to Method of Estimation, Census Divisions and Census Subdivisions

	Census Divisions (290 CDS)	Selected Census Subdivisions (108 CSDs)
Component	2.3 %	n.a.
Rate of growth	4.8	6.6 %
Proportional Allocation	6.1	7.2
Constant Population	9.5	9.6
Vital Rates	5.7	n.a.

Example 3.3

a) Estimate the population of the municipality of West Hants of Hants County, Nova Scotia as of July 1, 1993 given the 1986 and 1991 adjusted census counts in the following table:

Answer: Population estimates for all census subdivisions of Hants County are required in order to control to the census division total.

	1986 adjusted census count	1991 adjusted census count	Average annual growth rate	Extrapolated estimate as of July 1, 1993	Population prorated to CD estimate as of July 1, 1993
	Pop 86	Pop 91	See formula (1)	$Pop91*(1+r)^2$	
Hants County	*18684*	*19354*		*19634*	*20106*
West Hants	**6909**	**6983**	**0.0021**	**7013**	**7182**
Windsor	1692	1700	0.0009	1703	1744
Hantsport	681	652	-0.0087	641	656
East Hants	9073	9644	0.0123	9882	10120
Indian Brook 14	329	375	0.0265	395	404

b) Estimate the number of men aged 32 in West Hants as of July 1, 1993 given the 1986 and 1991 adjusted census counts in the following table:

Answer: Population estimates for all census subdivisions of Hants County are required in order to control to the census division total.

	1986 adjusted census count aged 30	1991 adjusted census count aged 35	Average annual growth rate*	1991 adjusted census count aged 30	Extrapolated estimate aged 32 as of July 1, 1993	Population aged 32 prorated to CD estimate as of July 1, 1993
	Pop 86	Pop 91	See formula (2)		$Pop91*(1+r)^2$	
Hants County	*324*	*346*		*358*	*365*	*386*
West Hants	**104**	**111**	**0.0131**	**138**	**140**	**148**
Windsor	27	23	-0.0316	22	21	22
Hantsport	12	16	0.0592	10	11	12
East Hants	175	191	0.0177	183	188	199
Indian Brook 14	6	5	-0.0358	5	5	5

* Statistics Canada further smoothS these rates using a three-year moving average.

Quality of Population Estimates

This module describes the method used at Statistics Canada to evaluate quality, which depends on the accuracy of source data and varies with details sought and methods used.

4.1 Methods of Evaluating Estimate Quality

Strictly speaking, there are no objective criteria for validating postcensal population estimates. The adjusted Census, however, is a reliable benchmark. The difference between the two, called the "error of closure", provides a measure of accuracy for the postcensal estimates. Mean absolute percent error is used to measure estimate quality for a set of small areas.

4.1.1 Error of Closure

Error of closure is the difference between the census count and the postcensal estimates on a given date. Error of closure is calculated as follows:

$$\varepsilon = \hat{P} - P \tag{4.1}$$

where ε = error of closure;

$\hat{P}$ = estimated population;

P = adjusted census count.

4.1.2 Mean~~s~~ Absolute Percent Error (MAPE)

The quality of any estimate may be expressed as an error of closure, whatever the geographical detail or characteristics considered. However when ratio-regression or difference-regression is used to produce estimates for a set of sub-areas, a technique that consolidates error size for the entire set of small areas is preferable. Mean absolute percentage error (MAPE) is the technique used at Statistics Canada to evaluate result quality by CD or CMA. MAPE may be defined as the average of the relative differences (either positive or negative) between population estimates and census counts.

$$MAPE = \frac{1}{n} \Sigma \left| \frac{\hat{P}_j - P_j}{P_j} \right| x\ 100 \tag{4.2}$$

where $\hat{P}_j$ = estimated population of census division j.

P_j = adjusted census population of census division j.

n = number of census divisions in a given province.

4.1.3 Error of Closure - Sources

Error of closure has two main sources: errors in the components of population growth and census under/overcoverage.

4.2 Quality of Data on Components

Each component of population growth (births, deaths, immigrants, emigrants and interprovincial migrants) may contain a degree of bias and error. However, the data on births, deaths and immigration can be regarded as more accurate than the estimates of emigrants and interprovincial migrants (see Module 2, section 2.2.2) and returning Canadians. Statistics Canada is currently evaluating the quality of this last component, introduced in 1993.

4.2.1 Emigration Data

Table 4.1 presents the estimates of emigrants from Canada produced using different methods and different data sources, for 1981-1986 and 1986-1991. The Table shows that figures produced by different methods can vary widely, and that the deviations in 1981-1986 figures are much more greater than in the 1986-1991 figures. The emigrant population figures for the 1981-1986 period estimated by the residual method are far higher than those derived using the Family Allowance method.

Residual Method

This is a variation of the component method. Emigration is simply the difference between intercensal population growth and the sum of the population growth components other than emigration (see Table 4.1). As the data on births, deaths, immigration and temporary residents are assumed to be accurate, the higher value of the residual-base emigration estimate may be attributed to returning Canadians and the difference between undercoverage rates in two successive censuses.

For the 1981-1986 period, residual emigration unadjusted for net undercoverage is very high (653,000). After correction of the population figures, by 1.6% (1981 Census) and by 2.7% (1986 Census), the number of emigrants estimated by the residual method is found to be 360,000, closer to the result of the Family Allowance (taxation) method (278,000). Unadjusted residual emigration in the 1986-1991 period is weaker than in the preceding period, due to a smaller difference in undercoverage rates between the 1986 (2.7%) and 1991 (2.8%) Censuses. When these figures are adjusted, residual emigration falls from 365,000 to 269,000, again closer to the result obtained using the Family Allowance method (208,000).

Table 4.1 Estimates of emigrants by different methods, Canada, 1981-1986 and 1986-1991

Method	1981-86	1986-91
1. Residual(1)		
(a) Unadjusted	653,000	365,000
(b) Adjusted for Net Undercoverage	360,000	269,000
2. Reverse Record Check	288,000	241,000
3. Revenue Canada Tax File	165,300	191,000
4. Family Allowance Method		
(a) f_c from tax file constant	278,000	208,000
(b) f_c from EIC file	281,000	194,000
(c) f_c from US immigration file	308,000	225,000

(1) Residual Method: Emigrants = ({Births - Deaths)} + [Immigrants] - Intercensal growth of population between June 1 of year t and May 31 of yeat t + 5.

Source: Demography Division, Statistics Canada.

Family Benefit Method

Emigration to the United States is derived from that country's immigration file. For other countries, the method entailes estimating emigration from the changes of address recorded in the Child Tax Benefit file for families with eligible children. Prior to 1993 the Family Allowance file was used as a source for child emigration data. A problem arises when moving from emigration of children eligible for benefits to emigration of the total population. To solve it, Statistics Canada calculates the ratio of the emigration rate for adults to the emigration rate for children. The accuracy of this ratio obviously has an impact on the quality of the resultant estimate. Known as the **f_c factor**, the ratio used in estimating the number of emigrants 18 years and over is based on either emigration data in the Revenue Canada tax file, Employment and Immigration Canada data on immigrants, or data on Canadian emigrants to the United States (Table 4.2).

The f_c factors derived from emigration data in the Revenue Canada tax file are available from 1989-1990, the first year of the revised system for the processing of taxation data. Although the emigration level in this file is considered too low, it nevertheless provides reliable information on the relationship between the rates of adult and child migration (i.e. f_c), which is equal to 1.10 for 1989-1990 and 1990-1991. This value was used in the official emigration estimates for the periods between 1981-1986 (278,000) and 1986-1991 (208,000).

It should be pointed out that, for most other sources, the value of f_c varies from year to year. Similarly, the factor's capacity for producing a good estimate of the number of emigrants varies by its source. For example, large numbers of Canadian retirees are attracted by the southern United States and emigrate there. As a result, the value of f_c established from data on emigration to the United States may not be usable for estimating emigrants to other destinations. The estimates of emigration using this source for 1981-1986 (308,000) and 1986-1991 (225,000) are higher than the estimates based on the tax file.

Table 4.2 Estimates of Emigrants by Family Allowance Method Using Different Values of f_c (Adult-Child Emigrant Ratios) 1986-1991

	Value of f^c					Number of Emigrants
Data Source of f^c	1986-87	1987-88	1988-89	1989-90	1990-91	
1. Revenue Canada Tax Files	1.10	1.10	1.10	1.10	1.10	2.080e+17
2. EIC Immigration Data	0.87	0.85	0.87	0.98	1.16	
3. Canadian Emigrants to the U.S.A.	1.29	1.41	1.31	1.38	1.15	

Source: Demography Division, Statistics Canada.

According to some authors (Beaujot and Rappak 1988), emigrant and immigrant flow data are associated, making it possible to compute f_c from the Employment and Immigration Canada (EIC) immigration file. For 1986-1991, f_c is in general less than 1, unlike the f_cs from the other two sources (i.e. tax file and United States immigration file). This means that they produce a lower estimate of the number of emigrants over the period (194,000). For 1981-1986, the factors are very close to the 1.10 value based on the taxation file and, as a result, emigration estimate is almost identical (281,000).

The change from the Family Allowance file to the Child Tax Benefit file has resulted in loss of child emigration data. Child emigration rates must now be adjusted for the non-universality of the new program. This is done with information on eligibility to Child Tax Benefits available from the income tax file. When data are available for the complete 1991-1996 period, the above evaluation will be repeated with special attention paid to the quality of the Child Tax Benefit data.

4.2.2 Interprovincial Migration Data

To test the accuracy of estimates of interprovincial migration obtained from the Revenue Canada tax file, two evaluations were conducted for the period 1986-1991: (i) comparison of sets of interprovincial migration data derived from the Revenue Canada tax files and Family Allowance files; and (ii) comparison of the errors of closure of population estimates for two sets of internal migration data. When data become available for the entire 1991-1996 period, this exercise will be repeated. Of particular interest will be the evaluation of the new migration data based on Child Tax Benefit introduced in 1993. As for emigration, they are adjusted for non-universality.

Table 4.3 compares net interprovincial migration estimates derived from four sources or methods:

the question on mobility asked in the 1991 Census;
the Family Allowance file;
the Revenue Canada tax file; and
the residual method.

For all provinces, estimates of internal migration derived from these different sources or methods are consistent inasmuch as the results are either positive or negative. For each source, net migration is positive for Ontario, British Columbia and Yukon and negative for all other provinces and territories.

In some respects, the estimates of net interprovincial migration obtained using the residual method are not strictly comparable to the results obtained using the other methods. By definition, the sum of net interprovincial migration in Canada should be zero; the residual method sets it at -57,690.

Table 4.3 Estimates of Net Interprovincial Migration from 1991 Census Question on Mobility, Family Allowance Files, Income Tax Files, and Residual Method, Canada, Provinces and Territories, 1986-1991

Geographic Area	1991 Census(1)	Family Allowance Files	Income Tax Files	Residual Method(2)
CANADA	0	0	0	-57,690
Newfoundland	-13,960	-16,894	-13,468	-21,139
Prince Edward Island	-855	-666	-122	-2,940
Nova Scotia	-4,870	-4,015	-1,672	-6,464
New Brunswick	-6,070	-4,423	-3,693	-2,441
Quebec	-25,550	-41,216	-39,366	-46,578
Ontario	46,955	65,174	70,543	76,673
Manitoba	-35,245	-36,232	-35,823	-45,806
Sakatchewan	-60,350	-66,300	-65,941	-78,978
Alberta	-25,015	-36,703	-40,237	-55,473
British Columbia	125,880	143,969	132,373	124,950
Yukon	780	1,201	1,094	1,999
Northwest Territories	-1,700	-3,895	-3,688	-1,493

(1) Population 5 years of age and over.
(2) The residual method for estimating net interprovincial migration is:
Net Migration = Growth of Census Population between time t and t + 5 - [(Births-Deaths) + (Immigration - Emigration)] - net growth of temporary residents - returning Canadians.
Source: Demography Division, Statistics Canada.

Errors of closure were used to measure the relative accuracy of the internal migration estimates. The other estimates of the components of population change were assumed to be accurate (Table 4.4).

Overall, there is very little difference between the comparable estimates based on errors of closure in the estimates of net interprovincial migration based on the Income Tax File and the Family Allowance file. As the Family Allowance program is no longer universal, the Income Tax File has become a far better choice, especially as the methods of producing good quality estimates have been refined.

Table 4.4 Error of Closure Between Alternative Population Estimates and Census Counts by Province and Territory, July 1, 1981, 1986 and 1991

Geographic Area	Error of Closure(1)(%)					
	1981		1986		1991	
	Income Tax	FA	Income Tax	FA	Income Tax	FA
Newfoundland	0.87	2.72	1.84	1.88	1.32	0.73
Prince Edward Island	-0.31	1.42	0.30	-0.06	2.15	1.74
Nova Scotia	0.74	1.57	0.35	0.16	0.52	0.27
New Brunswick	0.77	2.77	0.96	0.66	-0.17	-0.26
Quebec	0.73	0.72	0.75	0.83	0.10	0.08
Ontario	0.75	0.74	-0.05	-0.12	-0.06	-0.11
Manitoba	1.62	0.93	-0.16	-0.26	0.90	0.86
Saskatchewan	0.23	2.11	-0.03	0.19	1.29	1.26
Alberta	-2.70	-4.25	1.04	0.94	0.59	0.72
British Columbia	0.17	-0.09	-0.10	0.10	0.22	0.56
Yukon	-8.76	-11.50	-5.99	-4.39	-3.12	-2.75
Northwest Territories	-3.70	-3.86	-3.81	-3.77	-3.59	-3.92
Average Absolute Error						
10 provinces	0.89	1.73	0.56	0.52	0.73	0.66
Provinces and Territories	1.78	2.72	1.28	1.11	1.17	1.11

(1) Error of closure is calculated using the following equation:

$$\text{Error of closure} = \left(\frac{\text{Estimate} - \text{Census count}}{\text{Census count}} \right) \times 100$$

Income Tax: Revenue Canada Income Tax File.
FA: Family Allowance File.
Census: Census count, adjusted for net undercoverage.
Source: Estimates of interprovincial migration based on Family Allowance file, Demography Division, Statistics Canada.
Estimates of interprovincial migration based on tax data, Small Area and Administrative Data Division, Statistics Canada.

4.3 Quality of Census (Base Population) Data

The postcensal estimates (total population, population by age and sex, and population by age, sex and marital status) contain certain inaccuracies stemming from adjusted census data (the base population on which these estimates were constructed).

The errors attributable to census data include (a) errors in adjusting for coverage, (b) response and processing errors and (c) errors due to differences in concept definitions.

(a) In 1993, Demography Division introduced a new series of population estimates beginning in 1971 which correct for coverage error by adding to the base population an estimate of net census undercoverage. This adjustment is made for total population, by age and sex, and by age, sex and marital status.

Coverage errors are primarily the result of undercoverage (overcoverage is generally approximately 20% of gross undercoverage). The undercoverage rate varies not only by province but also by sex, age and marital status, as well as from census to census (see Tables 4.5, 4.6 and 4.7). Such differential undercoverage inevitably leads to bias in the estimates. For example, to estimate the male population aged 20-24 on June 1, 1991, the population aged 15-19 in 1986 was used as a base. All the demographic events affecting these cohorts between 1986 and 1991 were then factored in. In 1986, the undercoverage rate for the 15-19 age group was about 3%; in 1991, the same cohorts (who were then between the ages of 20 and 24) had an undercoverage rate of 7%. Consequently, even if the component data accurately reflected the events occurring to that cohort during the 5-year interim, the 1991 estimate of the number of persons aged 20-24 would be 4% higher than the number counted in the 1991 Census.

Estimates of net census undercoverage for the total population of the provinces and the territories and for the population of Canada by age group and sex were derived directly from two coverage studies: the Reverse Record Check and the overcoverage study. Statistical techniques were used to distribute these estimates by finer demographic and geographical detail. While these models improve the quality of the population estimates, they are subject to error. For censuses befor 1991, information on undercoverage was limited to the results of the RRC. Net census undercoverage for 1971 to 1986 was estimated by applying to the RRC results for those years the ratio of overcoverage and undercoverage observed in 1991. With very few exceptions, undercoverage exceeds overcoverage at all levels of demographic and geographic disaggregation.

Table 4.5 Estimated Net Population Undercoverage (number and rate), by Province and Territory, 1981, 1986 and 1991 Censuses

Province	1981		1986(1)		1991(2)(3)	
	Number	Rate	Number	Rate	Number	Rate
		Percent				
Canada	403,401	1.63	696,036	2.68	791,412	2.82
Nfld.	7,948	1.38	9,114	1.58	11,380	1.96
P.E.I.	1,127	0.91	2,258	1.75	1,221	0.93
N.S.	6,859	0.80	15,625	1.76	17,192	1.87
N.B.	10,345	1.46	16,351	2.25	24,157	3.23
Que.	100,490	1.54	160,581	2.40	179,793	2.54
Ont.	138,877	1.58	268,670	2.87	373,990	3.58
Man.	7,698	0.74	26,625	2.44	20,338	1.83
Sask.	6,871	0.70	19,871	1.93	17,720	1.76
Alb.	46,538	2.04	60,175	2.48	51,262	1.97
B.C.	74,192	2.63	112,497	3.76	90,019	2.67
Yukon	615	2.59	1,068	4.35	1,148	3.97
NWT	1,841	3.87	3,201	5.77	3,192	5.23

(1) Included as overcoverage is an estimate of 6,000 temporary residents enumerated in 1986 but not part of the universe.
(2) Undercoverage for the 0-4 age group (females) is adjusted for an overestimate of undercoverage in the coverage studies.
(3) Includes temporary residents.

Source: Statistics Canada, Demography Division. Revised Postcensal and Intercensal Estimates: Canada, Provinces and Territories 1971-1993. Methodology for Derivation of Annual Estimates of Population by Age and Sex, M. Michalowski, October 1993.

Table 4.6 Estimated Net Population Undercoverage (number and rate) by Sex and Age Group, Canada, 1981, 1986 and 1991 Censuses

Age Group	1981		1986(1)		1991(2)(3)	
	Number	Rate	Number	Rate	Number	Rate
	Both sexes					
All ages	403,401	1.63	696,036	2.68	791,412	2.82
0-4 years	21,145	1.17	32,955	1.79	43,043	2.21
5-9 "	18,675	1.04	34,457	1.88	44,674	2.29
10-14 "	14,231	0.74	26,847	1.48	33,661	1.76
15-19 "	53,312	2.25	55,332	2.79	56,975	2.96
20-24 "	115,543	4.70	181,397	7.45	147,909	7.01
25-29 "	52,571	2.36	112,820	4.60	158,774	6.27
30-34 "	36,178	1.74	78,341	3.46	104,817	4.04
35-39 "	28,120	1.70	31,674	1.54	54,847	2.34
40-44 "	27,903	2.04	40,206	2.43	51,105	2.39
45-49 "	7,965	0.63	13,039	0.98	23,671	1.42
50-54 "	4,597	0.37	15,413	1.24	10,353	0.78
55-59 "	7,988	0.67	25,301	2.06	15,867	1.28
60-64 "	4,421	0.45	14,010	1.23	12,363	1.04
65 years and over	10,752	0.45	34,244	1.25	33,353	1.04
	Male					
All ages	240,332	1.95	410,359	3.18	468,984	3.37
0-4 years	10,754	1.16	16,963	1.80	22,583	2.26
5-9 "	10,392	1.13	19,207	2.04	19,910	1.99
10-14 "	7,295	0.74	10,117	1.09	16,881	1.72
15-19 "	27,458	2.27	31,751	3.12	26,564	2.70
20-24 "	63,955	5.17	111,864	9.00	82,726	7.74
25-29 "	31,951	2.86	73,808	5.96	102,435	7.97
30-34 "	21,795	2.09	44,901	3.98	73,028	5.57
35-39 "	23,951	2.83	25,626	2.47	37,112	3.17
40-44 "	22,085	3.17	28,122	3.35	34,511	3.21
45-49 "	6,816	1.06	10,281	1.53	15,109	1.80
50-54 "	3,635	0.58	8,284	1.33	7,924	1.18
55-59 "	4,159	0.73	12,240	2.02	10,446	1.69
60-64 "	1,909	0.41	5,469	1.02	5,943	1.03
65 years and over	4,177	0.41	11,726	1.02	13,812	1.03
All ages	163,069	1.31	285,677	2.18	322,428	2.28
0-4 years	10,391	1.18	15,992	1.78	20,460	21.5
5-9 "	8,283	0.95	15,250	1.71	24,764	2.59
10-14 "	6,936	0.74	16,730	1.89	16,780	1.80
15-19 "	25,854	2.23	23,581	2.45	30,411	3.23
20-24 "	51,588	4.23	69,533	5.84	65,183	6.26
25-29 "	20,620	1.85	39,012	3.21	56,339	4.51
30-34 "	14,383	1.39	33,440	2.95	31,789	2.47
35-39 "	4,169	0.51	6,048	0.59	17,735	1.52
40-44 "	5,818	0.87	12,084	1.48	16,594	1.56
45-49 "	1,149	0.18	2,758	0.42	8,562	1.04
50-54 "	962	0.15	7,129	1.15	2,429	0.37
55-59 "	3,829	0.62	13,061	2.10	5,421	0.87
60-64 "	2,512	0.48	8,541	1.42	6,420	1.05
65 years and over	6,575	0.48	22,518	1.42	19,541	1.05

(1) Included as overcoverage an estimate of 6,000 temporary residents enumerate in 1986 but not part of the universe.
(2) Undercoverage for the 0-4 age group (females) is adjusted for an overestimate of undercoverage in the coverage studies.
(3) Includes temporary residents.
Source: Statistics Canada, Demography Division. Revised Postcensal and Intercensal Estimates: Canada, Provinces and Territories 1971-1993. Methodology for Derivation of Annual Estimates of Population by Age and Sex, M. Michalowski, October 1993.

Table 4.7 **Estimated Population Undercoverage (number and rate) by Sex and Marital Status, Canada, 1981, 1986 and 1991 Censuses**

	1981		1986(1)		1991(2)(3)	
Sex and marital status	Number	Rate	Number	Rate	Number	Rate
Both Sexes			696,036	2.68	791,412	2.82
Now married or separated			188,615	1.48	228,389	1.65
Divorced			41,320	1.96	44,588	2.12
Widowed			24,903	5.64	29,101	4.67
Never married			441,198	3.92	489,334	4.12
Age under 15			94,259	1.72	121,378	2.09
15 years and over			346,939	6.01	367,956	6.06
Male			410,359	3.18	468,984	3.37
Now married or separated			105,340	1.66	133,470	1.92
Divorced			23,069	1.71	39,077	3.56
Widowed			3,650	7.69	8,315	9.71
Never married			278,300	4.62	288,122	4.54
Age under 15			46,287	1.67	59,374	1.99
15 years and over			232,013	7.22	228,748	6.79
Female			285,677	2.18	322,428	2.28
Now married or separated			83,275	1.30	94,919	1.37
Divorced			18,251	2.01	5,511	1.82
Widowed			21,253	4.22	20,786	1.00
Never married			162,898	3.11	201,212	3.63
Age under 15			47,972	1.79	62,004	2.19
15 years and over			114,926	4.50	139,208	5.15

(1) Included as overcoverage is an estimate of 6,000 temporary residents enumerated in 1986 but not part of the universe.
(2) Undercoverage for the 0-4 age group (females) is adjusted for an overestimate of undercoverage in the coverage studies.
(3) Includes temporary residents.
Source: Statistics Canada, Demography Division. Revised Postcensal and Intercensal Estimates: Canada, Provinces and Territories 1971-1993. Methodology for Derivation of Annual Estimates of Population by Age and Sex, M. Michalowski, October 1993.

(b) The principal response and processing errors are non-response, respondent misinterpretation, coding errors and incorrect non-response imputation. For example, the response rate in the 1991 Census was 98.3% for the question on sex, 98.3% for the question on age and 97.7% for the question on marital status.

(c) The statistics on deaths, marriages, divorces, new widowhoods and immigration refer to legal marital status (married, widowed, divorced and never-married). The Census, on the other hand, uses a more practical definition, closer to the actual situation. Thus, in the Census, the "married" category includes people who are living together (e.g. common-law unions). The Census also regards all persons under 15 years of age as single (never-married). The difference in the concept of married persons, particularly the phenomenon of common-law partners, is probably responsible for the most significant deviations observed between Census figures and estimates of population by marital status. A study of common-law partners is currently being conducted. In future, it should be possible to correct the situation by incorporating into the model a reasonable estimate of the net formation or dissolution of common-law couples.

4.4 Errors in Other Data Sources and Methodology

Errors due to estimation methodologies and data sources other than censuses are difficult to quantify, although not insignificant. The more detailed the breakdown of the data, the larger the coefficient of error. The component totals contain a certain amount of initial error, and the methodology used to classify them by sex and age, and then by marital status, produces additional error in the figures at each stage.

Nevertheless, the components can be divided into two categories according to the quality of their data sources: births, deaths, immigration, marriages and divorces, for which the sources of final data may be considered very good; and emigration, interprovincial migration and new widowhoods (estimated using deaths of married persons) for which the indirect methods used may be a more substantial source of error. Lastly, the size of the error due to component estimation may vary by province, sex, age and marital status, and errors in some components (births, emigration, new widowhoods) may have a greater impact on a given age group, sex or marital status category.

4.5 Estimate Quality by Type

Error of closure, which measures estimate quality, varies by estimate type. Generally speaking, the more sophisticated the estimates, the greater the likelihood of different errors offsetting one another to some extent, at least in the overall results.

At Statistics Canada, estimates are distinguished on the basis of three criteria: frequency of release; definitive character of estimate; and degree of result breakdown. Table 4.8 shows the types of postcensal population estimates.

Intercensal estimates contain the same types of error as postcensal estimates, as well as errors resulting from the way in which errors of closure are distributed on the basis of the time elapsed since the reference census (see Module 1, section 1.3.3).

Table 4.8 Types of Postcensal Population Estimates Produced by Statistics Canada

Estimate	Geographic Unit	Reference Date	Status and Time Elapsed Since Reference Date	Publication (Catalogue No.)
1. Total population	Canada, provinces and territories	January 1	Preliminary (3 months)	Cat. No. 91-002 (Quarterly)
		April 1	Updated (6-15 months)	and
		July 1 (annual)		Cat. No. 91-213 (Annual)
		October 1	Final (24-33 months)	
	CDs and CMAs	July 1	Preliminary (8 months)	Cat. No. 91-213 (Annual)
			Final (32 months)	
2. Age, sex and marital status	Canada, provinces and territories	July 1	Preliminary (4 months)	Cat. No. 91-213 (Annual)
			Updated (8 months)	Available on request
			Final (20 months)	Available on request
3. Number of census families	Canada, provinces and territories	July 1	Preliminary (8 months)	Cat. No. 91-213 (Annual)
			Updated (10 months)	
			Final (24 months)	

4.5.1 Accuracy of Population Estimates by Total, Sex, Age and Marital Status for Canada, Provinces and Territories

If it is assumed that the quality of the basic data remains constant, existing postcensal estimates should have an acceptable degree of reliability, as indicated in Tables 4.9, 4.10 and 4.11. Errors of closure are sensitive to the quality of data from the censuses preceding and following the reference period. Over a five-year period, these errors compound the annual inaccuracies in data on components.

As of July 1 (the end of the reference period), postcensal estimates of total population deviate relatively little from census data. In 1991, the difference at the national level was 0.2%; at the provincial level, the largest deviation (Prince Edward Island) was 2.2% (Table 4.9).

Table 4.9 Error of Closure of the Estimates of Total Population, Canada, Provinces and Territories, July 1, 1991

Province or Territory	June 4, 1991 Census adjusted to July 1, 1991	Final Postcensal Estimates	Deviation	
			Number	Percent
CANADA	28,085,300	28,175,300	57,700	0.21
Nfld.	579,900	587,600	7,700	1.32
P.E.I.	130,800	133,700	2,800	2.15
N.S.	918,100	922,900	4,800	0.52
N.B.	748,500	747,300	-1,300	-0.17
Que.	7,081,200	7,088,400	7,200	0.10
Ont.	10,471,200	10,465,100	-6,100	-0.06
Man.	1,113,300	1,123,300	10,000	0.90
Sask.	1,007,000	1,020,000	13,000	1.29
Alb.	2,600,300	2,615,500	15,200	0.59
B.C.	3,376,900	3,384,300	7,400	0.22
Yukon	29,000	28,100	-900	-3.12
NWT	61,200	59,000	-2,200	-3.59

Note: The error of closure is equal to the July 1, 1991 final postcensal estimates minus the 1991 Census count adjusted to July 1 and for error of closure. The percentage error of closure is: error of closure, divided by 1991 Census count adjusted to July 1, multiplied by 100. Each figure is rounded independently to the nearest hundred.

There is very little difference between the errors of closure for the male and female populations. However, when the other demographic characteristics of the population are considered, the differences are greater. Specifically, the estimates for the population on the 25-29 age group shows a greater deviation from the census count than the estimate for other age groups, even after adjustment for net undercoverage (Table 4.10).

Table 4.10 Error of Closure of the Estimates of Population by Sex and Age Group, Canada, Provinces and Territories, July 1, 1991

	Both sexes		Male		Female	
Age group	Number	Percent	Number	Percent	Number	Percent
All ages	57,700	0.21	30,700	0.22	27,000	0.19
0-4 years	6,000	0.31	3,500	0.35	2,500	0.26
5-9 "	-33,100	-1.70	-12,900	-1.29	-20,300	-2.12
10-14 "	-19,100	-0.99	-7,400	-0.76	-11,600	-1.24
15-19 "	-30,300	-1.57	-16,400	-1.66	-13,900	-1.48
20-24 "	12,000	0.57	15,200	1.42	-3,100	-0.30
25-29 "	88,300	3.49	50,600	3.95	37,700	3.03
30-34 "	21,400	0.82	11,100	0.84	10,300	0.80
35-39 "	30,100	1.29	11,600	0.98	18,600	1.59
40-44 "	-24,400	-1.14	-12,800	-1.18	-11,700	-1.10
45-49 "	8,800	0.53	8,300	0.99	500	0.06
50-54 "	-3,900	-0.29	-2,600	-0.38	-1,300	-0.20
55-59 "	-5,800	-0.47	-7,200	-1.17	1,400	0.23
60-64 "	6,300	0.53	300	0.05	6,000	0.99
65 years and over	1,300	0.04	-10,500	-0.78	11,800	0.63

Note: The error of closure is equal to the July 1, 1991 final postcensal estimates minus the 1991 Census count adjusted to July 1, 1991. The percentage error of closure is: error of closure, divided by the 1991 Census count adjusted to July 1, multiplied by 100.

Table 4.11 Error of Closure of the Estimates of Population Aged 15 Years and Over, by Sex and Marital Status, Canada, July 1, 1991

Marital Status	Both sexes		Male		Female	
	Number	Percent	Number	Percent	Number	Percent
All statuses	103,843	0.47	47,451	.43	56,392	0.50
Single	405,564	6.68	214,428	6.36	191,136	7.07
Married	-527,193	-3.80	-276,259	-3.97	-250,934	-3.62
Widowed	17,670	1.28	7,175	3.06	10,495	0.92
Divorced	207,802	21.74	102,107	25.32	105,695	19.13

Note: The error of closure is equal to the July 1, 1991 updated postcensal estimates minus the 1991 Census count adjusted to July 1. The percentage error of closure is: error of closure, divided by the 1991 Census count adjusted to July 1, multiplied by 100.

4.5.2 Accuracy of Population Estimates for CDs and CMAs

Table 4.12 shows the accuracy of population estimates for CDs and CMAs for the 1986-1991 period, derived using the component method. For all provinces and territories, the average absolute error is 2.2%; for individual provinces, it ranges from 1.2% (Ontario) to 3.0% (Manitoba).

Table 4.12 Accuracy of the Regression Method for Estimating the Total Population for Census Divisions by Province and for Census Metropolitan Areas, 1991

Area	1991 AAE (%)
Newfoundland	2.40
Prince Edward Island	2.39
Nova Scotia	1.64
New Brunswick	1.82
Quebec	2.52
Ontario	1.20
Manitoba	3.04
Saskatchewan	2.70
Alberta	2.43
British Columbia	1.54
Yukon	3.13
Northwest Territories	2.94
Canada	2.16
Census Metropolitan Areas	0.69

Note: AAE: Average absolute error = $\frac{1}{N} \sum \left| \frac{E_i - P_i}{P_i} \right| x\ 100$

E_i: Estimated population for region i;

P_i: Census population total for region i;

N : Number of census metropolitan areas or number of census divisions in a given province or territory.

MODULE 5

COMMON PROBLEMS FOR LOCAL AREA ESTIMATES OF POPULATION

Common Problems for Local Area Estimates of Population

There are four common problems in estimating population for local areas. They are related to:

- converting the data compiled for one type of boundary specification to data for the selected boundary specification;
- boundary changes over time;
- appropriateness of data conversion files; and
- lack of data for local areas.

Conversion of Geographic Data

Converting data for one type of local area into data for a different type is often difficult due to partial overlap of the two types of areas. For example, the proportion of a Census Division (CD) that is included in a Census Metropolitan Area (CMA) ranges from 5% to 100%, making conversion of population estimates for CDs into estimates for CMAs difficult.

Another problem is a lack of the detailed information on boundary specifications needed to link two types of areas. Examples of this situation include: Manitoba Health Service Regions and CDs, Saskatchewan municipalities and CDs, and Canada Employment Centres and CDs.

Boundary Changes over Time

Boundaries of local areas (e.g. census divisions and census metropolitan areas) are modified from census to census. These changes cause difficulty in estimating population over a long period of time. Examples of this type of change include the changes in the number of CDs in Ontario and Alberta between 1981 and 1986. In Ontario, the number of CDs fell from 53 in 1981 to 49 in 1986; in Alberta, the number rose from 15 in 1981 to 19 in 1986.

Between 1986 and 1991, there were changes in the number of CDs in two provinces: in British Columbia, the number of CDs rose from 29 in 1986 to 30 in 1991; and in Quebec, the number of CDs rose from 76 in 1986 to 99 in 1991.

Other types of changes affect CMAs. Sherbrooke was made a CMA in 1986. There were boundary changes in all but 3 CMAs between 1981 and 1986. Few changes of this kind were made over the 1986-1991 time period.

Boundary changes over time for census subdivisions are too numerous to list. Module 3 describes the procedure used to present annual estimates in constant boundaries.

Appropriateness of Data Conversion Files

Statistics Canada currently uses postal codes conversion files to break data down by CD. The usefulness of such files is incontestable. Postal codes specify very small land units, which may be consolidated to construct new boundaries.

However the use of postal codes to reorganize data is not entirely without problems. The postal codes system, although universal in Canada, is relatively unsophisticated in rural areas, and is constantly changing. When producing data by CD, the three following problems arise:

- one postal code may cover a number of CDs; (e.g. rural postal codes often cover areas far larger than the CDs);

- new postal codes are constantly being created; there may be a more or less significant time lag between creation of a postal code and its inclusion in the conversion files, and in its adoption by the households concerned; if the postal codes appear in the conversion files after data extraction, some indicators are not fully factored in;

- some postal codes refer to community mail boxes, and thus do not accurately reflect the location of the actual residence.

Data Availability

With the exception of non permanent residents and returning Canadians, data on the components are available for census divisions and census metropolitan areas. Current information for non-permanent residents and returning Canadians is only available at provincial and territorial levels. The most recent census is used to derive subprovincial distributions.

At the census subdivision level, however, this is not the case and extrapolation models must be used. The largest obstacle to the component method is migration (births and deaths could be derived from Vital Statistics files). The problems encountered at the census subdivision level with migration based on the income tax file is small numbers and problems of postal code allocation.

While comprehensive data sources for all subprovincial areas in Canada are hard to come by, many sources exist for selected areas. See section on Other Data Sources of Module 3 (3.2).

MODULE 6

UNDERSTANDING AND INTERPRETING POPULATION ESTIMATES

Understanding and Interpreting Population Estimates

The Basics of Demographic Analysis

The procedure most commonly used in making population estimates is highly demographic (component method) as it consists of adding to the base population the demographic events that alter population size and structure. It is thus useful to be aware of these events, know how the relevant data are presented and understand the adjustments that are sometimes necessary if the data are to meet the requirements of the estimation model selected.

Tracking over time provides insight into the data configuration requirements of the component method, as that configuration is often different from the one used to disseminate population statistics.

When the results are available, they must be interpreted as a function of the specific requirements that led to their consultation or production. A number of demographic tools (indices, tables and tabulations, graphs) have been designed to assist the analyst and make the results more comprehensible to the user.

This module discusses in brief:

- how data are used in preparing estimates;
- how to analyze a population broken down by age and sex;
- how estimated data are used in the computation of many indices that are good indicators of the demographic dynamics of the population estimated; and
- the precautions to be taken when comparing demographic behaviour over time or space (comparative indices), in order to avoid bias of interpretation linked to the influence of differences in composition by age and sex.

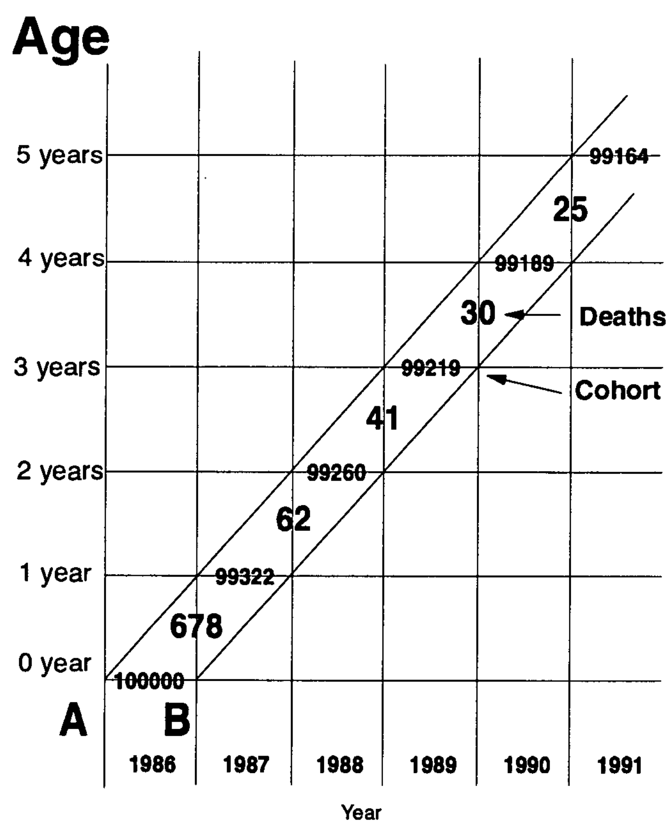
6.1 Tracking Over Time

We have seen that estimating or projecting population figures by age consists of aging the members of a population observed at a given point in time, factoring in the risk of death, migration, having children and arrival of immigrants (Module 2, section 2.1.2). ***Births, deaths*** and ***immigration*** are events that are systematically registered and counted, and serve as the basis for the publication of statistics or dissemination of computerized data.

The Lexis diagram is used to represent the change over time of a ***cohort***, (sometimes called a "birth cohort" when it is made up of all the individuals born in the same year, or simply "a cohort" when it is made up of individuals who experience the same event at a given point in time, e.g. marriage, first birth).

Passage from one age (or time period) to another is read on the ordinate, and passage from one year (or other time unit) to another on the abscissa. Thus, all the lines that mark the lives of members of a given cohort occur within a diagonal corridor (Figure 6.1) that shows the events that mark their progress (years of schooling, marriage, birth of children, ... and finally, death).

Figure 6.1 Change in cohort size over the first years of life

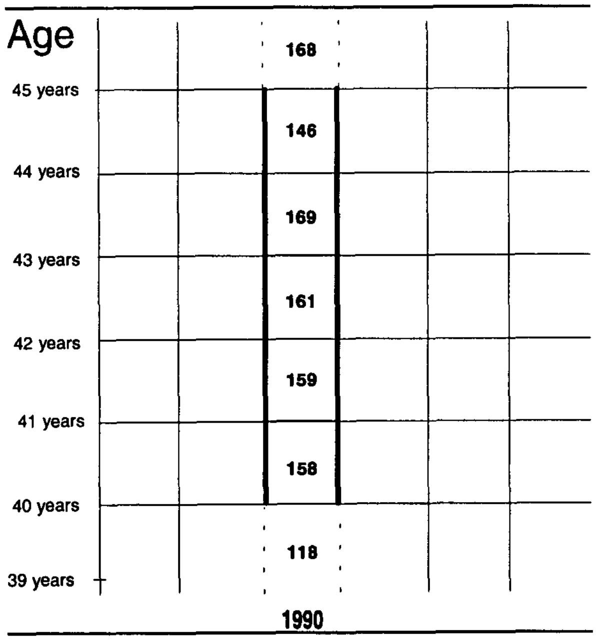


Source: Statistics Canada, Life Tables, Canada and Provinces, 1985-1987, Catalogue No. 84-532, Occasional.

Generally the records of demographic events occurring during a given year are disseminated in the form of more or less detailed statistical tabulations that display various characteristics of the persons who experience the events in question or the events themselves. For example, births are classified according to a range of characteristics of the mother, more rarely of the father (e.g. age, marital status, place of residence) and of the newborn (e.g. live birth, still birth, sex, weight).

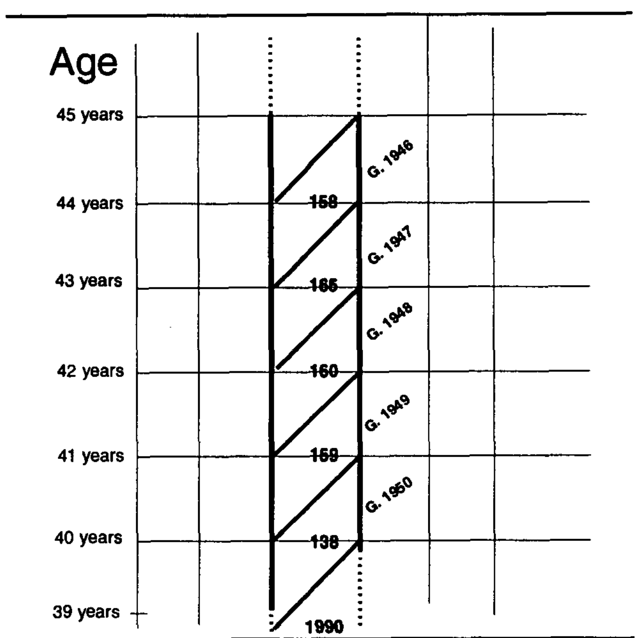
In most cases, the events in a single year are classified by the age (in whole years) of the persons who experience them. This type of classification does not allow for tracking cohorts (Figure 6.2 a). On the Lexis diagram, events related to a given age and experienced during the year in question are shown in a square that encloses a segment of the lifelines of two cohorts. The estimate of population by age requires that the data on components be classified by cohort rather than age (Figure 6.2 b).

Figure 6.2A **Classification of Population Data by Age, Number of Deaths, Males, Ontario, 1990**



Source: Statistics Canada, Canadian Centre for Health Information, Final Data.

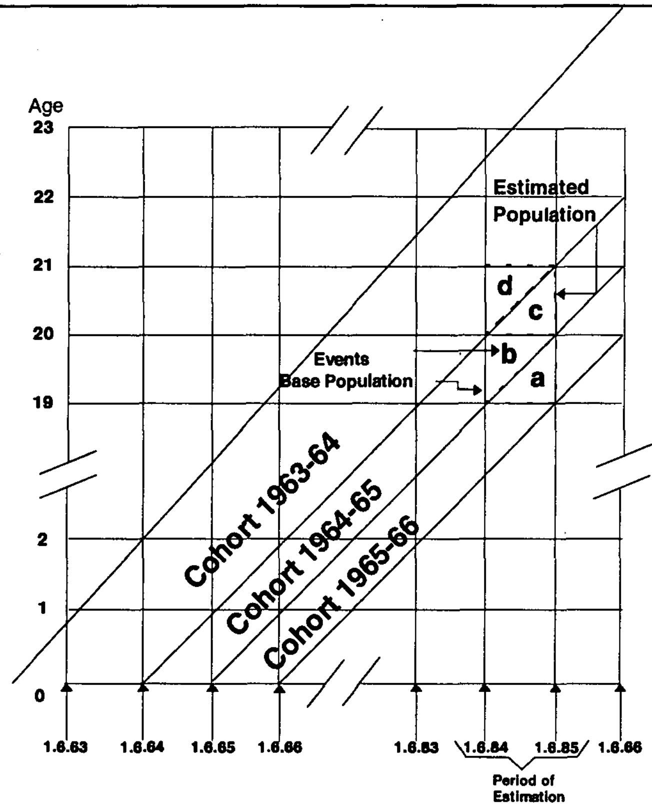
Figure 6.2B **Classification of Population Data by Birth Cohort, Number of Deaths, Males, Ontario, 1990**



Source: Statistics Canada, Canadian Center for Health Information, Final Data.

Some countries, now including Canada, establish statistics classified by both age and year of birth. Before Canada adopted this method, it was necessary to "reclassify" event data to make them meet the requirements of the component method. Figure 6.3 illustrates the process. The 1964-65 cohort is distributed from 1.6.1984 to 31.5.1985 (for the 19 and 20 years age group). Without access to classification by date of birth, it is generally assumed that the events related to a given age (triangles a and b for 20 years of age and triangles c and d for 21 years of age) are distributed equally over time. This means that when we take one half of the events experienced at 20 years of age (triangle b) and 21 years of age (triangle c), we obtain the events related to the 1964-65 cohort, for which we wish to estimate the population size as of June 1, 1985.

Figure 6.3 Dual Classification of Demographic Data



Source: Statistics Canada, Demography Division, Population Estimation Methods, Canada, Catalogue No. 91-528E, Ottawa, March 1987, p. 22.

6.2 Analysis of Population Growth and Structures

The purpose of population estimates is to gain insight into population size and demographic characteristics. The user must either be told how to interpret the results or be in a position to conduct a minimal analysis of the estimated data reflecting population growth and composition by age, sex and marital status, and any variations for the area or interval considered.

6.2.1 Measuring Population Growth

Population growth over a given time period is simply the difference between the size of that population at the beginning and the end of the time period.

$$G = P_t - P_o \tag{6.1}$$

where:
G	=	total growth of the population
P_t	=	population at end of time period
P_0	=	population at beginning of time period

If information on the components of growth is available, a distinction can be made between growth due to natural increase (difference between births and deaths) and growth due to migratory increase (difference between in-migration and out-migration).

These absolute numbers may be expressed as rates which, unless otherwise indicated, are annual.

$$\text{Rate of increase} = \frac{G}{P} \tag{6.2}$$

where P = average population = ½ ($P_O + P_t$)

Rate of natural growth or migratory growth may be calculated in the same way by replacing the numerator by the appropriate figure.

To obtain annual growth rate when working by five-year period, the numerator is simply divided by 5, assuming constant annual change in population size. If it is assumed that the rate is constant, it is preferable to use the equation given for calculating "r" in Module 1 (equation 1.3).

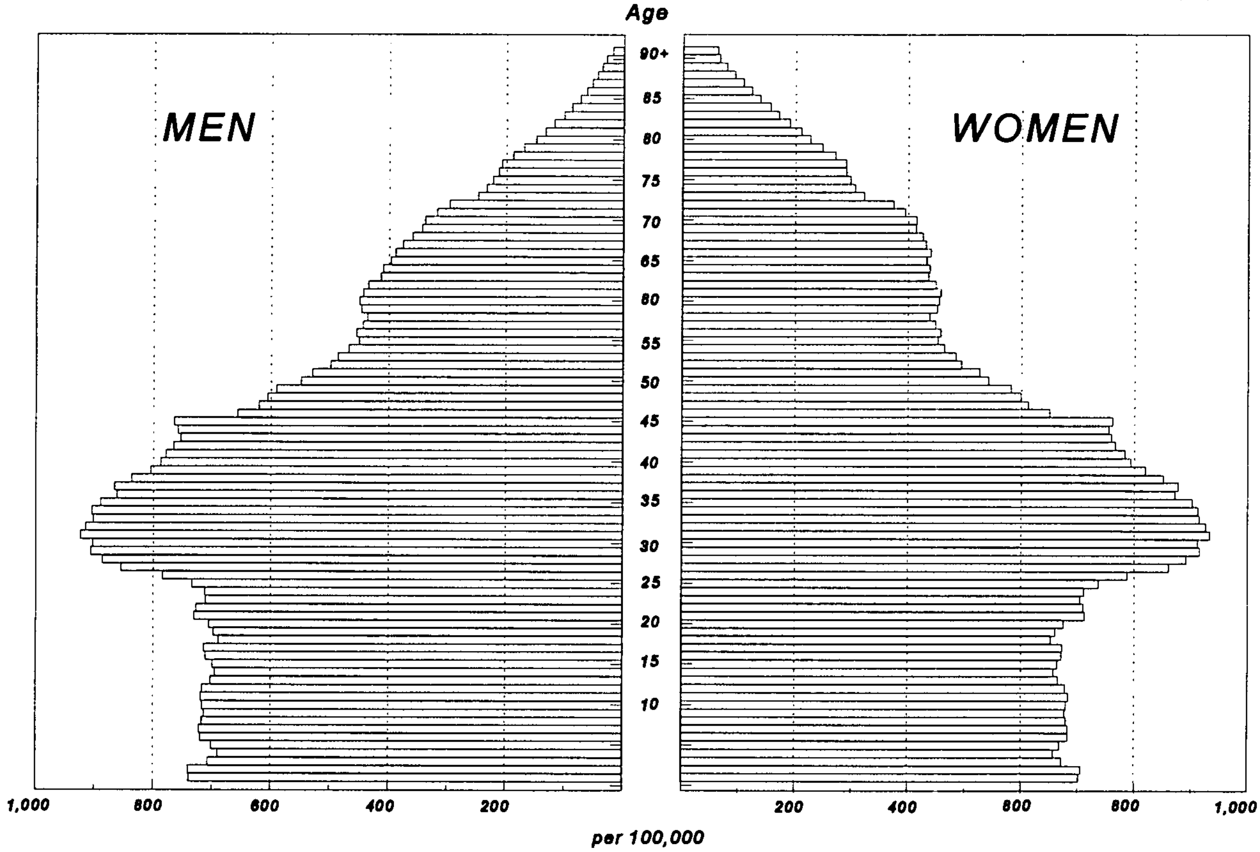
6.2.2 Analysing the Age/Sex Structure of the Population

It is not easy to deduce inter-regional differences or changes in distribution over time from tables that classify population by age and sex. The tools developed for analyzing structures are either graphs that permit visualization of the relationship between the sexes and between cohorts or indices that consolidate the information.

Age Pyramid

The age pyramid is a graphical representation of the age-sex structure of a population. It is essentially two sets of bar graphs laid back-to-back. Each set represents age (or age group) on the ordinate and population size on the abscissa. By convention, the male population comprises the left-hand side of the pyramid and the female population the right (Figure 6.4).

Figure 6.4 Age Pyramid of the Population of Canada, 1991



Source: Statistics Canada, Demography Division

Male/Female Ratio

This index, a simple ratio between the sexes, usually reduced to 100, may be calculated for total population (all ages) population, at birth, or any age or age group selected by the analyst according to the requirements of the study being conducted.

$$\text{Male/female ratio} = \frac{\textit{Males}}{\textit{Females}} * 100 \qquad (6.3)$$

$$\text{Male/female ratio at birth} = \frac{\textit{Male births}}{\textit{Female births}} * 100 \qquad (6.4)$$

Example 6.1

In 1992, the male/female ratio in Canada =

$$\frac{14{,}149{,}703 \ \textit{males}}{14{,}392{,}510 \ \textit{females}} * 100 \quad = \quad 98.3$$

Male/female ratio at birth for 1991-92 =

$$\frac{206{,}632 \ \textit{male births}}{196{,}475 \ \textit{female births}} * 100 \quad = \quad 105.2$$

Dependency Ratio

This is the ratio between persons not in the labour force (usually defined as the 0-19 and 65+ age categories) and persons in the labour force (20-64 years). The age limits for persons in the labour force and not in the labour force change according to the requirements expressed by estimate users.

$$\text{Dependency ratio} = \frac{\textit{Not in the labour force}}{\textit{In the labour force}} * 100 \qquad (6.5)$$

Example 6.2

In 1992, the dependency ratio in Canada =

$$\frac{11{,}108{,}942 \ \textit{individuals} \ 0\text{-}19 \ \textit{and} \ 65+ \ \textit{years of age}}{17{,}433{,}271 \ \textit{individuals} \ 20\text{-}64 \ \textit{years of age}} * 100 \ = 63.7$$

Example 6.3 **Structure Analysis: Change in Population Size by Broad Age Group, Dependency Ratio, Youth Index and Age Index, Halifax and Victoria Census Metropolitan Areas (CMAs), Both Sexes, 1986 and 1991**

Age Group	Population Size		Change from 1986 to 1991	
	1986	1991	Absolute	Percent
Halifax				
0-19 years	83,663	86,119	2,456	2.9
20-64 years	192,928	209,595	16,667	8.6
65 years and over	26,973	31,199	4,226	15.7
All Ages	303,564	326,913	23,349	7.7
Dependency Ratio	57.3	56.0		
Youth Index	43.4	41.1		
Age Index	14.0	14.9		
Victoria				
0-19 years	61,648	66,810	5,162	8.4
20-64 years	157,419	174,797	17,378	11.0
65 years and over	47,668	54,812	7,144	15.0
All ages	266,735	296,419	29,684	11.1
Dependency Ratio	69.4	69.6		
Youth Index	39.2	38.2		
Age Index	30.3	31.4		

Source: Statistics Canada, 1986 Census, Census Metropolitan Areas and Census Agglomerations: Part 1, Catalogue No. 94-127.
Statistics Canada, 1991 Census, Age, Sex and Marital Status, The Nation, Catalogue No. 93-310.

6.2.3 Rates and Ratios

There is a link between the size of a population and the number of demographic events that occur within it. A rate is an index used to establish the relative frequency of a demographic event as a ratio of the size of the population within which the event occurs. There is usually an annual dimension: the numerator is the ***annual*** number of events; the denominator is the average size of the population over the year considered. Population size is obtained by averaging the population at the beginning and end of the year or using the mid-year figure. The latter option is the one most often used in Canada where the population is known as of July 1 of each year. A rate is usually expressed as percent (%) or per thousand (‰) individuals in the population; it may also be expressed as a fraction.

The term ***rate*** is also used to mean the ratio of a sub-population to the total population. Rates of this kind are not appropriate for analyzing demographic phenomena as such; however, they are often used to study such aspects as labour force or school attendance.

Crude Rates and Specific Rates

Crude rates and specific rates belong to the first category. ***Crude rates***, the least detailed, simply relate number of events observed during a year to the average total population for the same year. As the phenomena studied by demography usually vary according to such factors as age and sex, it is preferable (when the data permit) to calculate ***specific rates***, which break the population down into more homogeneous groups according to the probability of the phenomenon in question. (e.g. birth rate is a crude rate, fertility rate by age a specific rate).

Mortality

Crude Death Rate (CDR)

$$CDR = \frac{D}{P} * 1\ 000 \tag{6.6}$$

where	D	=	Deaths in the population during the year
	P	=	Average population

Age-specific Death Rate (d_a)

$$d_a = \frac{D_a}{P_a} * 1\ 000 \tag{6.7}$$

where	D_a	=	Deaths of persons of age a
	P_a	=	Average population of age a

Infant Mortality Rate (IMR)

This is a specific rate. It relates the number of deaths of infants under one year to the number of live births during the year rather than the average population under one year

$$IMR = \frac{D_0}{B} * 1\ 000 \tag{6.8}$$

where	D_0	=	Number of deaths of infants under one year
	B	=	Number of live births over the year

Fertility and Birth Rate

Crude Birth Rate (CBR)

$$CBR = \frac{B}{P} * 1\ 000 \qquad (6.9)$$

where B = Live births in the year considered
P = Average population

General Fertility Rate (GFR)

$$GFR = \frac{B}{W_{15-44}} \qquad (6.10)$$

where B = Number of live births in the year
W_{15-44} = Average number of women in the childbearing age range (15 to 45 years)

Child-woman Ratio (CWR)

$$CWR = \frac{P_{0-4}}{W_{15-44}} \qquad (6.11)$$

where P_{0-4} = Children under five years of age
W_{15-44} = Women between the ages of 15 and 44

N.B. This index is used when vital status statistics on births are of poor quality or missing, and data on population distribution by age and sex are available. It is approximately five time greater than the overall fertility rate when infant and juvenile mortality are low and fertility is fairly constant.

Age-specific Fertility Rates (f_a)

$$f_a = \frac{B_a}{W_a} \qquad (6.12)$$

where B_a = Live births among mothers of age a
W_a = Women of age a

Fertility Table

Fertility by age or age group is usually presented in tabular form (Example 6.4). The age-specific fertility rate series provides insight into the fertility timetable, and the aggregate gives the strength of the phenomenon, i.e. the average number of children per woman.

Example 6.4 Age-Specific Fertility Rate, Canada, 1990	
Age	Rate (per 1,000)
15-19	26.6
20-24	85.5
25-29	132.2
30-34	88.1
35-39	28.8
40-44	4.0
45-49	0.1
TFR	**1.83**

Completed Fertility Rate (CFR) and Total Fertility Rate (TFR)

The strength of fertility within a real cohort is known as the ***completed fertility rate (CFR).*** The index corresponds to the sum of the age-specific fertility rates observed in a female cohort between the ages of 15 and 45. Calculating the index requires a compilation of data for a 30-year time period. If the rates observed over a single year are totalled, they simulate the fertility of a cohort of women (known here as a ***fictitious cohort***) who would have displayed (between the ages of 15 and 45) the fertility behaviour of the women observed that year. To avoid any confusion with the concept of completed fertility rate, the average number of children per woman in a fictitious cohort is referred to as the ***total fertility rate (TFR).*** CFR and TFR are rarely given per 1,000 women; the sum of the rates expressed per thousand is usually reduced to one.

$$\text{CFR or SFI} = \sum_{a=15}^{45} W_a \qquad (6.13)$$

Note: When fertility rates are calculated by five-year age groups, each rate applies to each of the five-year age groups and must be assigned a weight of 5 at the time of aggregation. An alternative is to multiply the sum of the rates by 5.

Gross Reproduction Rate (GRR) and Net Reproduction Rate (NRR)

Gross reproduction rate (GRR) is related to completed fertility rate or total fertility rate. GRR is the average number of female live births experienced by one woman during her fertile period, in the absence of death. It is calculated by multiplying the completed fertility rate or the total fertility rate by the female/male birth rate, or the ratio of female births to all births, i.e. by 0.488, (for every 100 female births, 105 male births are usually observed).

$$GRR = CFR \bullet 0.488 \text{ or } TFR \bullet 0.488 \qquad (6.14)$$

Net reproduction rate (NRR) factors in mortality in the female cohort whose reproduction is under study. Its calculation requires combining the fertility rate series with the probabilities of survival of women. A simplified computation method weights the gross reproduction rate using the proportion of women in the cohort who survive until the age of fertility (in this case, the probability of surviving until the average age of delivery, which varies around 30 years of age).

$$NRR = GRR \bullet S_{a'} \qquad (6.15)$$

where $S_{a'}$ = Probability of surviving from birth to average age of delivery a'

Migration

Because of the complexity of migration and the nature of the information available, analysis of the phenomenon is often indirect. Using the conventional demographic approach for a given area, the following rates may be calculated.

Rate of Internal Migration (IM)

This rate relates all changes of residence over one year in the area under study to the average population in the same year. In Canada, data on migrants are more readily available than data on migration; for this reason, the numerator reflects number of migrants rather than number of migrations. Note that, for intemal migration, the number of out-migrants equals the number of in-migrants.

$$IM = \frac{M}{P} \bullet 1\,000 \qquad (6.16)$$

where M = Number of out-migrants
P = Average population

Rate of Emigration (E)

This rate relates all the emigrations (in Canada, the emigrants) in one year to the average population for that year.

$$E = \frac{EM}{P} \bullet 1\,000 \qquad (6.17)$$

where EM = Number of emigrants
P = Average population

Rate of Immigration (I)

Logically, the number of immigrations (in Canada, immigrants) would be compared to the population at risk of emigration. This would be of little interest, as the population at risk is the total population outside the area in question. In fact, rate of immigration is calculated as a ratio of immigrations (immigrants) to average host population, for simple accounting considerations. The immigration rate, combined with the emigration rate, provide insight of the effect on the host population of international migration.

$$I = \frac{IM}{P} \tag{6.18}$$

where IM = Number of immigrants
P = Average population

Rate of Net Migration (NM)

This rate is the difference between the rate of immigration and the rate of emigration.

$$NM = I - E \tag{6.19}$$

6.3 Comparative Analysis

Much of demography is concerned with the comparison of populations and demographic behaviour in space and time. Some of the comparisons involve frequency distributions, in which case we are comparing two by two a great many classes or categories and seeking to express the result of the comparison by a single index. Consolidated indices, usually crude rates, are often compared. These rates reflect both the frequency of the phenomenon in the population under study and the population's composition. If, for example, effect of mortality is being compared for two or more populations, we must ask whether a difference in crude death rates is attributable, at least in part, to differences in age (mortality being highly correlated with age).

The demographer applies a number of methods or techniques that are in general use in social statistics in order to measure or offset the effects of differences in structure. The most common methods are given below.

Dissimilarity Index

This index measures the differences between the relative distributions of two populations on the basis of the categories or classifications proper to the phenomenon under study. The index is obtained as follows:

a) add the absolute values of the differences between the relative frequencies of each category; and

b) divide the result by two, so that the index assumes a value between 0 and 1 or between 0% and 100%.

The dissimilarity index corresponds to the minimum proportion by which one population or the other should change category to make both distributions identical.

The dissimilarity indices are used most frequently in situations where a large number of distributions are to be compared, either to each other or to a single standard distribution, or when large numbers of comparisons between pairs of distributions are to be made.

Use of Specific Rates

The demographic behaviour of populations may be compared without structural bias using, rather than gross rates, appropriate specific rates (e.g. age, sex, marital status). The disadvantage of this approach for the analyst is that he or she is faced with an enormous volume of data but no overall summary comparison.

Standardization

The purpose of standardization is to render comparable the summary indices (i.e. crude rates) for different populations. The standard population method, the most common, reconstructs the crude rates for the two populations to be compared on the basis of a common structure. When we apply to this common structure the series of specific rates (e.g. by age, sex, marital status) of each population to be compared, we obtain the event figures "adjusted" for structural differences. These figures are used as the basis for calculating "adjusted" crude rates, often known as ***comparative rates*** or ***standardized rates***.

The method may be used to compare any number of populations, either populations in different areas or in a single area at different points in time.

The following steps are used to compare a number of populations for given demographic behaviour, using the standard population method.

1) Select a standard population. One possible option is the structure of one of the populations to be compared. The advantages of this option are (1) that the gross rate for the population that provides the standard structure does not have to be recalculated and (2), more importantly, it is not necessary to know the specific rates for that population (this characteristic is important in cases where it is impossible to know these rates as the observed population size is too small).

2) Apply to the standard population the rates specific to each of the populations to be compared. This operation gives, for each category (e.g. age, sex), the number of events that would have been observed if each population had the structure (e.g. by age, sex) of the standard population.

3) Add these "fictitious" events and calculate the adjusted gross rate, using the usual equation, and relate this sum to the total size of the standard population.

Examples 6.5 and 6.6 apply the method to Canadian data to compare different populations as to mortality (example 6.5) and populations of a single area at different points in time (example 6.6).

Example 6.5 **Compare the male mortality rates for 1991 of Canada and Quebec**

Crude male mortality rate, 1991:
Canada : 7.53 per thousand
Quebec : 7.65 per thousand
Difference 0.12

Age	Canada Rate 1991	Standard Population Quebec, 1991	Deaths based on standard population
	(Per thousand)		
	(1)	(2)	(1) x (2)
0 - 4	1.96	229,632	451
5 - 9	0.19	236,043	45
10 - 14	0.32	250,188	80
15 - 19	0.99	235,905	234
20 - 24	1.12	262,121	295
25 - 29	1.19	329,238	392
30 - 34	1.31	336,251	440
35 - 39	1.66	302,191	502
40 - 44	2.19	275,394	602
45 - 49	3.55	225,083	800
50 - 54	5.78	174,169	1,006
55 - 59	9.85	161,083	1,586
60 - 64	16.45	147,372	2,424
65 - 69	25.71	121,792	3,131
70 - 74	39.75	84,002	3,339
75 - 79	62.65	56,672	3,550
80 - 84	95.25	30,594	2,914
85 - 89	144.49	12,626	1,824
90+	208.99	4,984	1,041
Total	7.53	3,475,337	24,657

Standard rates:
Quebec, thus based on events: 26,600 / 3,475,337 • 1,000 = 7.65
Canada, assuming Quebec's age structure: 24,657 / 3,475,337 • 1,000 = 7.09
Adjusted difference: 0.56
Conclusion: The difference in mortality rates for men in Canada and Quebec is larger than is indicated by the crude rates. Actually Quebec's younger popoulation relative to that of Canada compensate somewhat for the province's higher mortality.

Example 6.6 Compared the male mortality rates of 1991 and 1981 for Canada			
Crude male mortality rates for Canada: 1991: 7.53 per thousand 1981: 7.87 per thousand Difference: 0.34			
Age	Rate 1991	Standard population Population 1981	Deaths based on standard population
	(per thousand)		
	(1)	(2)	(1) x (2)
0 - 4	1.96	923,877	1,813
5 - 9	0.19	928,147	175
10 - 14	0.32	1,002,558	321
15 - 19	0.99	1,228,051	1,217
20 - 24	1.12	1,248,577	1,403
25 - 29	1.19	1,124,341	1,337
30 - 34	1.31	1,039,645	1,361
35 - 39	1.66	838,976	1,395
40 - 44	2.19	690,958	1,511
45 - 49	3.55	644,540	2,290
50 - 54	5.78	623,695	3,603
55 - 59	9.85	569,995	5,613
60 - 64	16.45	458,957	7,548
65 - 69	25.71	389,564	10,014
70 - 74	39.75	278,818	11,084
75 - 79	62.65	179,234	11,228
80 - 84	95.25	94,340	8,986
85 - 89	144.49	43,782	6,326
90+	208.99	19,957	4,171
Total	7.53	12,328,008	81,398
1991 crude rates: Based on events: 96,964 / 12,328,008 • 1,000 = 7.87 Standardized to 1981 population age structure: 81,398 / 12,328,008 • 1,000 = 6.60 Adjusted difference: 1.27 Conclusion: The male mortality rate for Canada decrease by 1.27% between 1981 and 1991. This decrease was not as apparent with the crude 1991 rate due to the ageing of the population.			

1976-1981. Intercensal Annual Estimates of Population for Census Divisions, Statistics Canada, Catalogue no. 91-521, Occasional (Second issue), May 1984.

1971-1976. Revised Annual Estimates of Population for Census Divisions, Statistics Canada, Catalogue no. 91-521, Occasional (First issue), July 1979.

1966-1971. Revised Estimates of Population for Counties and Census Divisions, Statistics Canada, Catalogue no. 91-513, Occasional, May 1973.

1961-1966. Population Estimates for Counties and Census Divisions, Dominion Bureau of Statistics, Catalogue no. 91-206, Annual, February 1969.

Census Metropolitan Areas

1986-1991. Revised Intercensal Population and Family Estimates, July 1, 1971-1991, Statistics Canada, Catalogue no. 91-537, Occasional, April 1994.

1981-1986. Intercensal Annual Estimates of Population for Census Divisions and Census Metropolitan Areas, Statistics Canada, Catalogue no. 91-521, Occasional (Third issue), July 1989.

1976-1981. Estimates of Population for the Census Metropolitan Areas of Canada, Statistics Canada, Catalogue no. 91-207, Annual (Last issue), May 1984.

1971-1976. Estimates of Population for the Census Metropolitan Areas of Canada, Statistics Canada, Catalogue no. 91-207, Annual, June 1978.

1966-1971. Estimated Population of the Metropolitan Areas of Canada, Statistics Canada, Catalogue no. 91-207, Annual (First issue), May 1973.

1961-1966. Unpublished data, available on request.

Estimates of Census Families

1986-1991. Revised Intercensal Population and Family Estimates, July 1, 1971-1991, Statistics Canada, Catalogue no. 91-537, Occasional, April 1994.

1981-1986. Intercensal Estimates of Families, Canada and Provinces, Statistics Canada, Catalogue no. 91-529, Occasional, (First Issue), November 1988.

Estimates for intercensal periods prior to 1947-1980. No intercensal estimates have been published on the characteristics of census families. Nevertheless, some data are available upon request, for selected periods.

1947-1965. Unpublished data, available upon request.

REFERENCES

Estimates for intercensal periods prior to 1991 have been published in the following publications:

Total Population, 1867 to 1971

Postcensal Annual Estimates of Population by Marital Status, Age, Sex and Components of Growth for Canada, provinces and territories, June 1, Volume 1, Catalogue no. 91-210.

Marital Status, Age and Sex

1971-1991. Revised Intercensal Population and Family Estimates, July 1, 1971-1991, Statistics Canada, Catalogue no. 91-537, Occasional, April 1994.

1921-1971. Revised Annual Estimates of Population, by Sex and Age Group, Canada and the Provinces, Statistics Canada, Catalogue no. 91-512, Occasional, July 1973.

1921-1971. Revised Annual Estimates of Population by Sex and Single Year of Age, Canada, Provinces and Territories, unpublished data, available upon request.

1966-1971. Estimates of Population by Marital Status, Age and Sex, Canada and Provinces, Statistics Canada, Catalogue no. 91-203, June 1974.

1961-1966. Estimates of Population by Marital Status, Age and Sex, Canada and Provinces, Statistics Canada, Catalogue no. 91-203, June 1969.

1956. Census of Canada, Volume 1, Table 23, October 1957.

1952-1960. Population Estimates by Marital Status, Age and Sex, unpublished data, available upon request.

1941-1951. Population Estimates by Marital Status, Age and Sex for Canada and Provinces, Statistics Canada, Catalogue no. 6003-504, April 1954.

1931-1941. Estimated Population of Canada and Provinces, by Marital Status and Sex, Statistics Canada, Catalogue no. 7-9010, May 1947.

Census Divisions

1986-1991. Revised Intercensal Population and Family Estimates, July 1, 1971-1991, Statistics Canada, Catalogue no. 91-537, Occasional, April 1994.

1981-1986. Intercensal Annual Estimates of Population for Census Divisions and Census Metropolitan Areas, Statistics Canada, Catalogue no. 91-521, Occasional (Third issue), July 1989.

7.3.1 Adjustment for Net Census Undercoverage

In January 1993, Statistics Canada announced a decision to incorporate estimates of census error undercoverage into the population estimation program[12]. Measurement of net undercoverage is based mainly on Reverse Record Checks and overcoverage studies. Other sources (e.g. results of demographic analysis) are also used.

7.3.2 Temporary Residents and Returning Canadians

In 1991, the universe for the Census of Canada was extended to include temporary residents (i.e. every person holding an employment authorization, student authorization, visitor's permit or Ministerial permit (see 2.2.1)).

In order to factor in the expansion of the census universe, two new components of population increase have been introduced (in-migration and out-migration of temporary residents). The procedure uses the Employment and Immigration Canada visitor file as the source of data on the number and movement of these individuals.

From now on, population estimates will also include Canadians who leave the country and return at a later date. Data on their movements are taken from the Customs and Excise Canada file and the Health and Welfare Canada Family Allowance file. Statistics Canada is currently developing methods for making use of the Revenue Canada Child Tax Benefit file.

7.3.3 Replacement of Family Allowance Source

The Health and Welfare Canada Family Allowance file has proved an excellent data source for estimating migration (emigration and interprovincial migration) and has been used as the basis of a symptomatic indicator for estimating sub-provincial populations by the regression method. In January 1993, however, the Family Allowance program was replaced by the Child Tax Benefit program, with the data now administered by Revenue Canada. Although the modified Family Allowance program can continue to provide the same type of population data, it is no longer universal as high income families are no longer eligible. To deal with this change in coverage, adjustment factors must be developed using data on family propensity to migrate according to revenue. Such information is available from the Revenue Canada Tax file and from the Census.

---

[12]Only the estimation program is affected by this decision. The census data as such are not adjusted.

distinction is also made between estimates based on the census preceding the reference date (postcensal estimates) and based on the two censuses defining the intercensal period (intercensal estimates). Interested readers may obtain more information on the nature and timing of the various estimates from the estimation manual published by Statistics Canada in 1987[16].

7.2 Custom Services for Population Estimates Available under Demography Division's Estimation Program

Demography Division produces population estimates at various levels of geography. Population size totals, as well as age and sex distributions, are available at the national, provincial and territorial, census division and census metropolitan area levels. At the lower levels of geography, various sources of data are used in the production of estimates, including data from administrative files (e.g. taxation, family benefit, health care). Marital status estimates are available down to the provincial/territorial level.

In addition, the estimation program is sufficiently flexible in terms of its data sources and methods to allow for a wide variety of custom work on a cost-recovery basis. A great many user-defined areas can be built to meet the demands of our users for timely and accurate data down to relatively fine geographic levels.

Custom estimates may be used to describe current demographic and socio-economic conditions, and also to simulate future patterns in an effort to assess the impact of current and expected trends in population growth and distribution. Services such as hospitals, housing, schools, roads, facilities for the aged and the chronically ill, as well as services for target groups such as lone parents, aboriginals and the immigrant population are planned, implemented and evaluated using demographic estimates as an important part of their input. Private industry makes use of estimates to meet a broad range of consumer-related needs, including marketing and distribution of products and services. Examples of non-standard geographic areas for which custom estimates have been produced by Demography Division include: municipalities, urban centres and economic regions for the Statistics Canada Labour Force Survey, Revenue Canada regional districts, and Employment and Immigration employment centres (see 3.2).

7.3 Research and Development

The Statistics Canada Population Estimation Program is constantly changing. Current modifications to the Program stem from the conclusions of methodology research projects, changes in administrative files and changes in theoretical approach.

Significant developments affected the production of population estimates as of June 1, 1993. In the fall of the same year, the 1991 Census was adopted as the population frame.

A brief description of recent changes follows.

---

[16] Statistics Canada. 1987. Population Estimation Methods, Canada. Catalogue 91-528. Supply and Services Canada.

Statistics Canada Population Estimation Program: Products, Services, Research and Development

7.1 Population Estimates Produced by Statistics Canada

The Statistics Canada Population Estimation Program includes the production of population estimates at the national, provincial, territorial and subprovincial levels. For the larger areas, the estimates are produced more frequently and according to more detailed demographic characteristics. For subprovincial areas, estimates of the total population are published annually, with details by age and sex available upon request. Statistics Canada also estimates the number and characteristics of census families. More specifically, the population and family estimates and the corresponding components of demographic change produced by Statistics Canada include the following.

Canada, Provinces and Territories

1. Quarterly estimates of the total population as well as the components of demographic change (births, deaths, immigration, emigration and, for provinces and territories, interprovincial migration).

2. Annual estimates of the population by age, sex and marital status and the components of demographic change (those mentioned above in addition to marriages, divorces and new widowhoods).

3. Annual estimates of the number of census families by characteristic (type (husband-wife or lone-parent), size, broad age group of children, broad age group of husband, broad age group of wife, broad age group of lone parent).

Subprovincial Areas

1. Annual estimates of the total population of census divisions and census metropolitan areas based on the regression method.

2. Annual estimates of the total population of census divisions and census metropolitan areas based on the component method and the components of demographic change (births, deaths, immigration, emigration, interprovincial migration and intraprovincial migration).

3. Annual estimates of the population by broad age group and sex for census divisions and census metropolitan areas, available upon request.

4. Annual estimates of the population by broad age group and sex for user-specified areas.

In order to satisfy criteria of timeliness, accuracy and consistency, Statistics Canada releases various sets of estimates for a given reference date. Preliminary estimates are generally available within a few months of the reference date; the final estimates, based on final source data, involve a much longer time lag. A

MODULE 7

STATISTICS CANADA POPULATION ESTIMATION PROGRAM: PRODUCTS, SERVICES, RESEARCH AND DEVELOPMENT

STATISTICS CANADA LIBRARY