ISSUES IN ESTIMATING THE POPULATION OF CANADIAN MUNICIPALITIES

Rosemary Bender and Mario Bédard
Statistics Canada

ABSTRACT

Interest in current population data at the municipality level has increased significantly in recent years. The demand comes from a variety of sources, including individual municipalities, regional governments, and provincial as well as federal authorities. While the Canadian census has served in the past as the primary source of demographic data, users are now interested in obtaining estimates and projections based on the most current figures available.

This shift in focus is attributable to two main factors. First, in the current context of fiscal restraint, the planning of municipal programs and services, including their funding by the various levels of government, has increased the need for more up-to-date demographic data. The second factor is related to the adjustment for net census undercoverage (persons missed in the census or counted more than once) that is now built into Statistics Canada's population estimation program. Though the program does not extend to municipal level estimates, users require that population data for these areas be consistent with the adjusted population figures available for provinces and territories, census divisions and census metropolitan areas. This paper reviews the potential data sources and methods for estimating municipal populations in Canada. Issues discussed include the limitations and possible improvements of the data currently available, as well as the strengths and weaknesses of the assumptions underlying the models.

KEYWORDS

Custom Areas, Municipality, Population, Rate of Growth Method

1. INTRODUCTION

Interest in current population data at the level of the municipality has increased significantly over the recent past in Canada. The demand comes from a variety of sources. Individual municipalities request population data, often to fill one-time needs for particular projects. Other requests come from provincial government departments and agencies, who want current municipal population estimates for policy and planning purposes in areas such as health and municipal services.[1] The New Brunswick Statistics Agency is one such provincial organization with which Statistics Canada has recently signed an agreement to supply current population data by single year of age and sex for the province's 287 municipalities on an on-going basis.

[1] Certain provincial and territorial agencies produce population projections at the subprovincial level. The methods vary (component, regression and econometric) and take advantage of local area data.

Federal departments also are increasingly interested in using small area data as building blocks to generate the demographic estimates they require for their own administrative areas. This approach, with municipalities defined as the building blocks, could replace the present methodology where the population estimates for the administrative areas in question are approximated from census division (county) data, which overlap them. This new approach could be beneficially applied to generate the population estimates for areas such as Statistics Canada Labour Force Survey urban centres, Immigration and Citizenship Canada employment centres and Revenue Canada regional districts, to name but a few.

In the past, the census was relied upon as the primary source of population data for 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 their funding by the various levels of government. 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.

Thus, the demand for small area population estimates and projections is quite high, and the future demand is potentially even higher. However, while the accuracy of the estimates available for the larger geographic entities covered by PEP is relatively well established, the same does not apply to present day municipality level estimates. In this paper, we examine some issues pertaining to the estimation of the population of Canadian municipalities. They include: changing municipal boundaries (section 2.1); updates to census counts (3.1.1); geographic misallocation (3.1.5); lack of current data at the municipality level (3.2 and 3.3); extreme growth rates (4.2.2 and 4.3.2); and rate of growth as the best estimation method (6). To this end, we examine the data sources and their constraints, and we look at the methodology currently applied to generate the estimates. The methodology consists of a two-stage approach where population totals are generated first followed by the production of the age/sex estimates. We also examine some empirical results, and discuss future developments in the area.

2. MUNICIPALITIES AND OTHER SUBPROVINCIAL REGIONS DEFINED

2.1 Municipalities / Census Subdivisions (CSDs)

As a rule, Canadian municipalities are determined by provincial legislation.[2] In Statistics Canada, municipalities belong to the type of legislative/administrative areas known as *census subdivisions*

[2] 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. These geographic areas include subdivisions of county municipalities, of regional districts, and of unorganized territory.

(CSD), which also include Indian reserves, Indian settlements, and unorganized territories.[3] Municipality and other CSD types include boroughs, cities, hamlets, parishes, resort villages, towns and villages (municipalities are henceforth referred to as CSDs). At the time of the 1991 census, there were 6,006 CSDs across Canada (CSD frequencies by province and territory are shown in Table 1).[4] CSDs aggregate to census divisions, census metropolitan areas and, from our experience, to most user-defined geographic areas.

One of the issues involved in estimating the population of CSDs revolves around the potential changes in CSD boundaries over time.[5] Under the present methodology, where estimation is achieved by inter/extrapolation from two census population bases, 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 process. However, the overlaying can result in slight misallocations in small geographical areas.

2.2 Census Divisions (CDs)

Census divisions are intermediate geographic areas between CSDs and the province or territory. Often generally referred to as "counties", they are also, as a rule, established by provincial law.[6] Census divisions are made up of exact groupings of CSDs. In turn, they aggregate to a province or a territory. CD types include communautés urbaines, counties, district municipalities, districts, metropolitan municipalities, municipalités régionales de comté, regional districts, regional municipalities, regions and united counties. At the time of the 1991 census, there were 290 census divisions across Canada (CD frequencies by province and territory are shown in Table 1.)[7]

3. DATA SOURCES FOR CSDs AND THEIR CONSTRAINTS

---

[3] For the 1991 census, 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.

[4] Statistics Canada (1992a) Table 23B provides CSD types and their frequencies by province and territory.

[5] The extent of boundary changes varies over any given time period. The changes that occurred between the 1986 and 1991 censuses are documented in Statistics Canada (1992b).

[6] In Newfoundland, Manitoba, Saskatchewan and Alberta, provincial law does not provide for these administrative geographic areas. Therefore, census divisions have been created by Statistics Canada in co-operation with these provinces. In New Brunswick, in order to maintain the integrity of component census subdivisions, census divisions do not respect the legal county limits.

[7] Statistics Canada (1992a) Table 23A provides CD types and their frequencies by province and territory.

The most important obstacle for producing accurate population estimates for CSDs is the lack of current information at this particular level of geography. There exist very few sources of demographic information that are readily accessible. The quinquennial census, which provides high quality detailed information, remains by far the primary source. Vital statistics, e.g., births and deaths registrations, constitute a high potential data source. Other data sources, as defined below, could eventually become usable.

Table 1. Adjusted Provincial and Territorial Population, July 1, 1991 -- Number of Census Divisions 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 1, 1991 (in thousands)	**28,120**	580	131	918	749	7,081	10,472	1,113	1,006	2,601	3,380	29	61
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

* In the province of Quebec, a few CDs are combined by Statistics Canada to bring the number to less than 100.

3.1 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, presently the 1986 and 1991 census data, constitute the basis from which the current CSD population estimates are produced. Before they can be used, however, census data need to be verified for any changes (updates) starting from the time since the data base was frozen. They must also be adjusted for net census undercoverage, and incompletely enumerated Indian reserves and settlements (which are recognized as individual CSDs). As well, starting in 1991, the census universe was expanded to include non-permanent residents. Thus, the 1986 census counts must also be adjusted accordingly. Finally, the 1986 data must also be verified and corrected for any misallocation resulting from their overlaying with the 1991 census geography.

3.1.1 Census Data Updates

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. Though relatively few in number, these errors must be taken into account. Indeed, for very small areas, the differences between the data base figures and the updated population counts are at times exceedingly high. For example, the 1991 census data base figure for the total population of Dufferin Parish in the province of New Brunswick is 861, but the updated count reads slightly more than half as much, i.e., 480.

Errors of this type usually involve two CSDs, with a shift in population from the one to the other. The CSD involved along with Dufferin Parish in the preceding example is Saint Stephen Parish.

Here, the data base count was revised from 1,437 to 1,818. As a rule, the differences between the initial and the revised counts usually cancel out, and are thus transparent for the larger geographical entities incorporating the two CSDs (e.g., the census division). However, as in this particular case, the impact of the update on each individual CSD is often far from negligible. Finally, the updates are usually made available for the total population, but they do not extend to the population characteristics such as age and sex. When required, these must be generated, usually by applying the corresponding distribution from the original data base to the updated total population count.

3.1.2 Adjustment for Net Census Undercoverage

The adjustment for net census undercoverage, which is now built into the official population estimation program, is based on the results of two quality studies. Both of these surveys are conducted by Statistics Canada following a census: the Reverse Record Check, which provides information on persons missed in the census who should have been enumerated; and the Overcoverage Study, which estimates the number of persons counted more than once, or counted when they were not part of the census universe.

The sample size for these studies is 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. Thus, the approach 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 to the enumerated population of the small areas by age and sex.

3.1.3 Adjustment for Incompletely Enumerated Indian Reserves

In both 1986 and 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 and is incorporated directly into the corresponding CSD estimates for 1986 and 1991.

3.1.4 Adjustment for Non-Permanent Residents in 1986

The 1991 census universe was expanded to include non-permanent residents. This group includes five categories of persons residing in Canada: individuals claiming refugee status, persons holding a student visa (for all types of educational institutions beginning with primary schools), individuals holding a work permit, persons holding a special permit issued by the minister of Citizenship and Immigration, and all non-Canadian-born dependants of all of these individuals. Although they have permission to reside in Canada only temporarily, non-permanent residents are active participants in the socio-economic fabric of Canadian society. They are eligible for many social benefits including medicare and child tax benefits, their children attend schools, they are consumers, etc. The estimates of the number of non-permanent residents by province and territory for 1986 were obtained from the Citizenship and Immigration Visitors Immigration Data System, while the distribution by subprovincial region were derived from the actual non-permanent counts in the 1991 census.

3.1.5 Geographic Misallocation in 1986

The centroid method, which is applied for overlaying the earlier census characteristics (such as age and sex) data base with current census geography, can result in slight misallocations in geographically small areas such as CSDs. These can be detected by summing across the characteristics for each CSD and comparing the resulting sum against the CSD total, which is derived independently and is considered correct. As a rule, such misallocations are corrected by adjusting the characteristics counts through raking to coincide with the independent total.

3.2 Vital Statistics

Few data are available on the change in population during the intercensal period 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, the files currently present some constraints for use at the geographical scale of the CSD.

The major limitation is the misallocation of events among neighbouring areas. That is, births and deaths occurring within a given CSD are sometimes coded to a neighbouring area. In some extreme cases, all events are erroneously coded, which results in a particular CSD registering no births or deaths at all, while others are given more than their share of events. The current coding procedures are also inconsistent in tracking boundary changes over time. While such misallocations tend to cancel each other out for the larger geographic regions such as the CD, they represent a major impediment to the use of the data for CSDs. Ongoing research in this area includes examining usage of postal codes to better pinpoint each event's geographical coordinates.[8] Good quality vital statistics would allow the consideration of solid alternative methodologies (e.g., 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 (while other data sources would have to be developed for estimating the other components of growth).

3.3 Other Data Sources

Other potential data sources include administrative data available for selected areas, such as electric utilities connections and housing building permits. 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, the final postcensal migration estimates entering into the estimation of the population of CDs via the components method are derived from the latter. In each case, however, a thorough assessment would have to be made of the differences in the concepts and definitions associated with municipality population estimation, of how well the data cover the population universe, of the quality of the data and of their accessibility and cost.

---

[8] 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.

Finally, from an individual CSD'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.[9] Also, these data reflect current boundaries of the CSD and would need to be adjusted to 1991 boundaries.

4. CURRENT METHODOLOGY FOR ESTIMATING THE POPULATION OF CSDs

At present, given the lack of comprehensive, factual data on population change, the preferred methodology for estimating CSD 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., the CDs.

The approach entails three major steps. The first involves adjusting the base populations, i.e., the census counts, for data updates, net undercoverage, incompletely enumerated Indian reserves and settlements, and non-permanent residents (1986 only). The second step consists in generating the *total* population estimates for the CSDs. The third step is optional and involves producing the *age/sex* estimates when requested by the user.[10] In each step, raking procedures are applied to ensure that the sum of the CSD estimates coincide with the corresponding CD to which they aggregate.

4.1 Adjusting the 1986 and 1991 Census Counts

The major adjustments that are applied to the census counts in order to generate the adjusted population base are discussed in Section 3.1. The last step in generating the adjusted population base involves ensuring that estimates for CSDs are consistent with corresponding CD data (see section 4.4). Work is still in progress in setting up the adjusted population bases for all provinces and territories. To date, comprehensive adjusted base year data have been completed for three provinces (New Brunswick, Nova Scotia and Ontario).

4.2 Estimating the Total Population of CSDs

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).

---

[9] 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.

[10] Estimates can be produced by single year of age for ages 0 to 89 and age group 90+.

4.2.1 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). 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.

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

(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. These cases occur rarely, and are observed essentially for very sparsely and possibly only intermittently inhabited Indian reserves and settlements.

4.2.2 Postcensal Estimates

The postcensal total population estimates are generated through the same approach, but for a few supplementary control procedures. Firstly, the growth rate for the intercensal period is given by the end points themselves, that is, it is purely data driven and can be quite high in some cases. It is assumed that extremely high growth (either positive or negative) is not sustained for protracted periods of time. Thus, the annual growth of the total population is constrained to a maximum of 10% (absolute) for the five-year postcensal period.

Another set of control procedures is put into effect for the CSDs with zero population at one of the end points. For those where zero population is observed in 1991, zero population is extrapolated into the postcensal period. Where zero population is observed in 1986 while non-zero population is observed in 1991, the growth rate of the corresponding CD is applied.

4.3 Estimating the Population of CSDs by Age and Sex

Here, again, slight variations are necessary when generating intercensal estimates as opposed to postcensal estimates. Involved are maximum thresholds, as well as procedures to ensure that the matrix of postcensal annual growth rates does not incorporate any empty elements and to ensure that neighbouring rates are relatively smooth.

4.3.1 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:

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

(2)

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. This procedure is used for the cohorts aged 5 to 84. For those born between 1986 and 1991, the average annual growth rate is estimated by comparing the transversal change in population for ages 0 to 4 between 1986 and 1991. This same approach is used for the 90+ age group. For the cohorts over age 85 in 1986 (and thus aged 90+ in 1991), the average annual growth rates are estimated from trends on declining populations for cohorts over 80 years of age.

4.3.2 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 for rates with absolute values over 20% for the population for a given age/sex is incorporated for the 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.[11]

Before the growth rates matrix can actually be used to generate the individual age/sex estimates, procedures must first be applied to ensure that each of its elements contains a valid non empty value, as well as to ensure that neighbouring rates present relatively smooth values. The first procedures result in empty elements being assigned rates calculated at the level of the corresponding census division. Once this is done, each age/sex rate is recalculated by applying a 3-year moving average smoothing technique.

4.4 Ensuring Census Subdivision and Census Division Estimates are Consistent

The intercensal and postcensal population estimates for CSDs as calculated above use the only information available at the present time on demographic change at that particular level of geography: the 1986 and 1991 censuses. It is only for CDs that current annual data on population change (births, deaths and migration) are available. The last crucial step in estimating the population

[11] For areas showing such high population change, local area information is to be sought to assess the current status of population change.

of CSDs, which aggregate into CDs, involves adjusting the interpolated and extrapolated estimates to the CD population estimates available by single year of age and sex.

First, the estimates of the total population for CSDs are adjusted proportionately to coincide with the estimates of the total population for CDs. Then, the estimates by age and sex are made to simultaneously respect two sets of marginal totals: (1) for each CSD, the sum across all ages and sexes must equal the CSD total; (2) for each age and sex category, the sum across all CSDs must equal that of the corresponding CD. This is accomplished through iterative prorating, where the rows and columns of the age-sex matrix are successively prorated to the margin totals. After a limited number of iterations, the process converges to a point where the rows and columns simultaneously agree with the marginal totals.

5. CURRENT METHODOLOGY: AN EVALUATION

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 ensure the quality of the estimates. In order to provide some insight into the empirical issues that were addressed during the development stages, we now examine some of these issues in light of the production of the age/sex population estimates for the province of New Brunswick's CSDs. The examination will also provide insight into the application of the methodology.

5.1 New Brunswick CSDs

Altogether, the province of New Brunswick had 287 CSDs at the time of the 1991 census. 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, the smallest populated CSD 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 287 CSDs recognized in 1991 had a zero population count.[12] The methodology for estimating the population by single year of age (0 to 90+) by sex for all the CSDs in the province of New Brunswick entails adjusting the 52,234 cells (287 CSDs *times* 91 ages *times* 2 sexes) in each of the 1986 and 1991 census population bases, calculating the average annual growth rate for each of these cells, and applying them to the individual age/sex cells in the base population.[13]

Table 2. Size of Age/Sex Growth Rates Matrix, Rates Exceeding |20%|, and Empty Cells

Total number of elements in New Brunswick age/sex rates matrix	**52,234**
Total number of rates exceeding |20%|	**1,048**
Number of rates greater than 20%	594

[12] The three CSDs are Red Bank 7, Tabusintac 9 and Saint Mary's 24. All three are also Indian reserves.

[13] The figure 52,234 corresponds to the number of CSDs (287), multiplied by the number of age/sex groups (91x2). It does not include the number of total population estimates (one for each CSD, i.e., 287), which have to be generated first.

Number of rates less than -20%	454
Total number of empty values	**5,997**

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 2, there were 1,048 age/sex cells with a growth rate exceeding the maximum in New Brunswick. This represents approximately 2% of the total number of cells. As can be seen in Figure 1, extreme values (either negative or positive) 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.

Figure 1. Percentage of Age/Sex Cells Exceeding |20%| By Age for N.B. CSDs

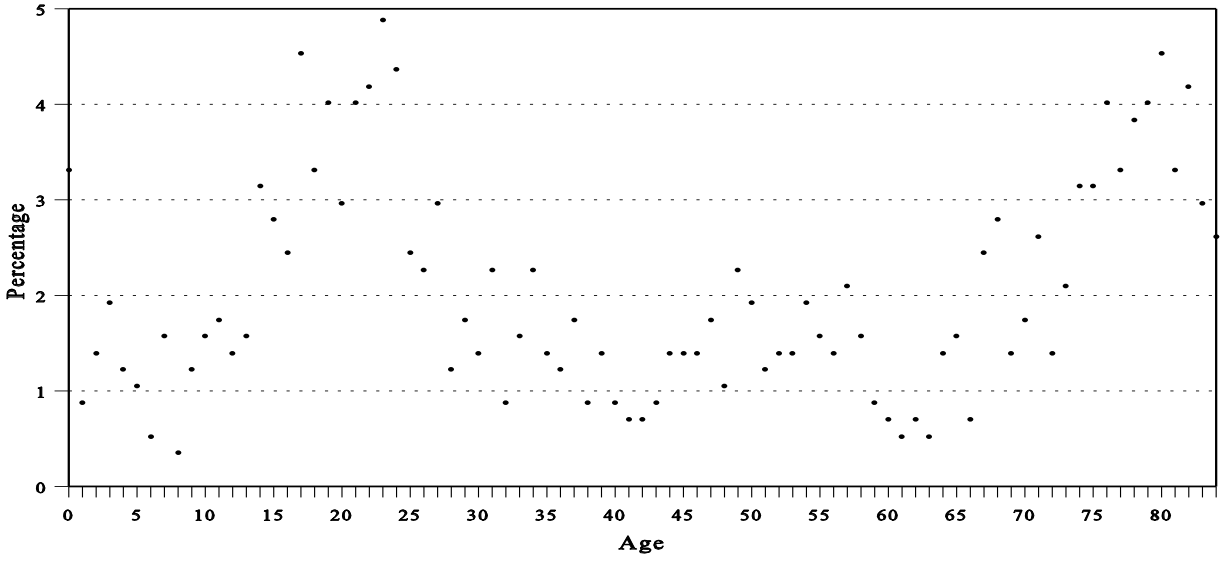
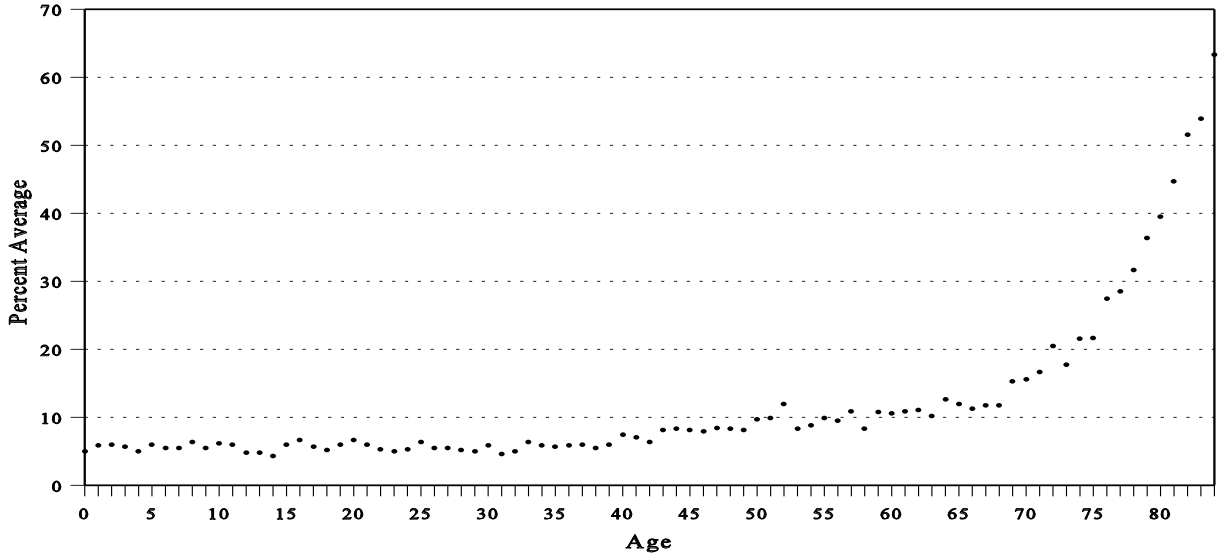


Figure 2. Percentage of Empty Age/Sex Cells By Age for N.B. CSDs



Another adjustment to the postcensal rates matrix entails imputing a value to the empty cells. As shown also in Table 2, there were 5,997 empty cells in the growth rates matrix, that is, over 11% of the total number of cells. Figure 2 shows the percentage of empty cells by age across all New Brunswick CSDs. Cohorts up to age 40 have relatively few empty cells (around 5%). The frequency of empty cells gradually increases with age thereafter. As can be expected, CSDs with very small populations present relatively high proportions of empty cells across ages.

6. ALTERNATIVE METHODOLOGIES

The current interpolation/extrapolation methodology for estimating the population of CSDs is one of various methods for population estimation. It was adopted and perfected following a study in which the performance and adequacy of a number of these methods were examined and evaluated. Among the approaches that were evaluated are the component method (for CDs), and methods based on proportional allocation, vital rates, and assuming zero population growth.

Component Method: the population estimate for any reference date for a given area is obtained by adding to the adjusted census base population an estimate of the change in population since that time. The components that make up this estimate of demographic change include births, deaths and migration. The migration in turn can be broken down into international, interprovincial and intraprovincial flows in and out of the region. Since these components are not currently available at the level of geography of the CSD, this method was only included for evaluation at the CD geographical scale.

Proportional Allocation: under this particular method, the population of a given CD is broken down by CSD according to the distribution observed in the last census, adjusted for net census undercoverage. This method can be applied to generate estimates for both the total and the age/sex population.

Vital Rates: when migration data are not available, birth and death rates can be used to estimate the population directly. First, these rates are derived as of the previous census. Then, by assuming that the rates are constant over time, estimates of the current population can be derived from the current number of births and deaths. This is done separately for each vital event, and the resulting populations are averaged. This approach can only be applied to estimate total population. Age/sex estimates must be generated using another method.

Zero Growth Method: here, the adjusted census year population estimates are held constant throughout the estimation period. This approach can be applied for both the total and the age/sex estimates.

6.1 Evaluation of Methods

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 CDs, where complete information is available for all methods considered. In a second phase, the proportional allocation, rate of growth and zero growth methods were evaluated for a number of particular CSDs.

The results of this analysis were presented in May 1995 at the conference of the *Canadian Population Society* and are summarized in Bender and Bédard (1995). The basic measure of accuracy used is the error of closure, defined as the percentage difference between the population estimate for 1991 and the 1991 adjusted census count. To summarize, **the component method is far superior to all other methods**. For CDs, this method produced the lowest average absolute error, with most errors below 2%. **The rate of growth method ranked second.** Like the vital rates, proportional allocation and zero growth methods, the modal interval for the error of closure was between 5 and 10%. However, the distribution of errors for the rate of growth method was more skewed toward the lower values. Similar results were observed at the CSD geographical scale for the three methods examined.

7. CONCLUSIONS AND FUTURE RESEARCH

The component method is by far the preferred method for the estimation of population. It produces few extreme errors and has the lowest average absolute error of closure, less than half that of the next best method. Comprehensive information on the components of population growth is available for CDs, but not, however, for smaller geographical areas such as CSDs. This lack of comprehensive information prevents this method from being applied at this particular geographical scale.

Efforts should be made to develop component data sources at this level. For births and deaths, procedures could be improved to better identify CSDs from other geographical information present on the registrations such as the locality code or postal code (zip code). The potential for obtaining good vital statistics at the CSD level in the near future is relatively high. Migration, however, remains the largest obstacle. International, interprovincial and internal migration components for CDs have for some time been derived by Statistics Canada from the Revenue Canada income tax files. This program, which generates the migration estimates from postal code-based geocodes, has

recently been extended to produce estimates for census metropolitan areas.[14] The postal code is an excellent allocation in urban areas, but in rural areas one postal code can span several CSDs. Producing migration estimates for CSDs requires detailed development, including much refinement to the geocoding procedures. The 1991 Census also provides migration data for CSDs covering the five-year period 1986-1991. Small numbers and the quinquennial migration period however render the data of little practical use.

While obtaining the number of migrants at the CSD level is not possible at the present time, information on the number of filers and their dependents could be derived to provide an indication of the magnitude of population change. The use of this and other potential symptomatic indicators such as hydro connections and housing building permits, or other information obtained directly from the CSDs should also be pursued.

Analysis of various, currently applicable, population estimation methods for selected CSDs has indicated that the rate of growth method performs best. It produces the lowest average absolute error of closure and has the fewest extreme errors. Thus, the rate of growth method forms the basis of our current methodology for population estimation for CSDs, using the most recent information on the population of these areas (census data) and annual estimates of demographic change available for the larger CDs. As *no information is available at the CSD level on annual population change,* two assumptions are implicit with the methodology. First, the general pattern of annual population growth for CSDs is assumed to equal the average annual growth between the 1986 and 1991 censuses. Second, the difference between the sum of the interpolated (extrapolated) estimates and a control CD count is assumed to be distributed proportionately across the CSDs of the region. To the extent that these assumptions are not realized, some of the population estimates for individual CSDs may not reflect actual changes in the population. Awareness of the data limitations is especially important when considering the population by age and sex. In fact, it is felt that the quality of the estimates increases with the size of the age groups. The estimates by single year of age should thus only be used to generate aggregate age groupings.

REFERENCES

BENDER, R. (1995) "Impact of New Population Estimates on Health and Vital Statistics," *Health Reports*, Ottawa: Statistics Canada, Cat. 82-003, vol. 7 no. 1.

BENDER, R and BÉDARD, M. (1995) Venturing into Population Estimates for Census Subdivisions, Ottawa: Statistics Canada, internal working paper.

LÉTOURNEAU, E. (1990) Estimation de la population des municipalités 1989," Quebec: Bureau de la statistique du Quebec, internal working paper.

LONG J.F. (1993) Postcensal Population Estimates: States, Counties, and Places, Washington: U.S. Bureau of the Census, Population Division Technical Working Paper No. 3.

---

[14] Canadian census metropolitan areas are either similar in size or larger than census divisions. Recent evaluation of the census metropolitan area migration estimates derived from the Revenue Canada income tax files indicates the data are of good quality.

PERREAULT, J. (1995) Who does What in regard to Subprovincial Population Projections?, Minutes of the Federal/Provincial Committee on Demography, Ottawa.

STATISTICS CANADA (1987) Population Estimation Methods, Canada. Ottawa: Statistics Canada, Cat.91-528, 1987.

STATISTICS CANADA (1992a) 1991 Census Dictionary. Ottawa: Statistics Canada, Cat. 92-301.

STATISTICS CANADA (1992b) 1991 Standard geographical classification: SGC 1991. Volume III, Changes - 1986 to 1991. Ottawa: Statistics Canada, Cat. 12-573.

STATISTICS CANADA (1994) Revised Intercensal Population and Family Estimates July 1, 1971-1991. Ottawa: Statistics Canada, Cat. 91-537.

STATISTICS CANADA (1995) Annual Demographic Statistics 1994. Ottawa: Statistics Canada, Cat. 91-213.

STATISTICS CANADA (1995) Compendium of Updates: 1991 Geographic Reference, Data and Digital Products. Ottawa: Statistics Canada.