Survey Methodology, December 1985
Vol. 11, No. 2, pp. 211–218
Statistics Canada

Estimating Population by Age and Sex for Census Divisions and Census Metropolitan Areas[1]

RAVI B.P. VERMA, K.G. BASAVARAJAPPA and ROSEMARY K. BENDER[2]

ABSTRACT

A methodology has been developed for producing population estimates by single years of age and sex for small areas (census divisions and census metropolitan areas). To assure reliability, the estimates by single years of age are grouped into five years and only these grouped data are recomended for dissemination. They are based on the age-sex composition of population from the last census, births by sex, deaths by single years of age and sex, estimates of migration by age and sex, and counts of family allowance recipients in the age group 1-14 years.

KEY WORDS: Cohort-component method; Mean absolute error; Index of dissimilarity; Separation factor.

1. INTRODUCTION

The objective of this paper is to describe the methodology for estimating population by age and sex for small areas (census divisions and census metropolitan areas), present findings of the evaluation of estimation methods, and finally to discuss the factors affecting the quality of estimates. According to the 1981 Census, the 266 Census divisions ranged in population from 2,000 to 2,000,000, and the 24 census metropolitan areas, from 100,000 to 3,000,000. The description of the estimation methods and principal data sources are presented in section 2. The results of the evaluation of migration and population estimates are given in section 3.

2. METHODOLOGY

The descriptions of the estimation methods, as well as the preparation of the basic input data are presented below.

2.1 Cohort-Component Method

For each census division (CD) and census metropolitan area (CMA), the cohort-component method is used to produce population estimates by age. The equations are as follows:

For the age 0, $$P_0^{t+1} = B - f_0 D_0 + \tfrac{1}{2} M_0 \quad (1)$$

For the age 1, $$P_1^{t+1} = P_0^t - \left[(1-f_0)D_0 + \tfrac{1}{2} D_1\right] + \tfrac{1}{2}(M_0 + M_1) \quad (2)$$

For ages 2 to 84, $$P_{a+1}^{t+1} = P_a^t - \tfrac{1}{2}(D_a + D_{a+1}) + \tfrac{1}{2}(M_a + M_{a+1}) \quad (3)$$

For ages 85+, $$P_{85+}^{t+1} = P_{84+}^t - \tfrac{1}{2} D_{84} - D_{85+} + \tfrac{1}{2} M_{84} + M_{85+} \quad (4)$$

---

[1] Revised version of the paper presented at the Federal-Provincial Committee on Demography meetings held on November 28-29, 1985 at Statistics Canada, Ottawa, Canada. This research was undertaken with support from the Small Area Data Program of Statistics Canada.

[2] Ravi B.P. Verma, K.G. Basavarajappa and Rosemary K. Bender, Demography Division, Statistics Canada,, 4th floor, Jean Talon Building, Tunney's Pasture, Ottawa, Ontario, Canada K1A 0T6.

where f_0 = Separation factor of deaths at age 0

M_a = Net migrants aged a between time t and $t+1$

B = Births between time t and $t+1$

D_a = Deaths at age a between time t and $t+1$

P_a^t = Population aged a at time t.

The cohort-component method is also used at the provincial level by Statistics Canada (Statistics Canada, Catalogue No. 91-210), and by the province of British Columbia for producing population estimates by age at the census division, school and health district levels (Central Statistics Bureau 1980).

2.2 Preparation of Basic Input Data

Since we are proposing to produce preliminary postcensal population estimates within eight months after the reference date, final data on components of population change cannot be used because they do not become available until after 18 to 24 months. Consequently, estimates would have to be used for each component.

Births and Deaths

Preliminary estimates of births by sex for year (t) are obtained by multiplying the proportional distribution by small areas of provincial total births by sex for year ($t-1$) with the provincial preliminary total births for year (t). Similarly, preliminary estimates of deaths by age and sex for year (t) are obtained by multiplying the proportional distribution by small areas of provincial total deaths by age and sex for year ($t-1$) with the provincial preliminary total deaths for year (t). Finally, they are converted into cohort deaths on the assumption that dates of birth of those who die and the number of deaths are uniformly distributed over a 12 month period except for deaths of age 0. The formulae are as follows:

For age 0,
Cohort deaths (0) = deaths (0) × 0.89
For age 1,
Cohort deaths (1) = [deaths (0) × 0.11] + [deaths (1) × 0.5]
For ages 2 to 84,
Cohort deaths (age) = [deaths (age-1) × 0.5] + [deaths (age) × 0.5]
Cohort deaths (85+) = deaths (84) × 0.5 + deaths (85+).

In the above formulae, the separation factors (f) are 0.89 for age 0, 0.11 for age 1 and 0.5 for all other ages.

Residual Net Migration

First, the estimates of total population for the postcensal years for CDs and CMAs prepared by the regression-nested procedure are split by sex using the sex composition from the latest census. The regression-nested procedure is described elsewhere (Statistics Canada, Catalogue No. 91-211). For males and females, residual total net migration is computed by taking the difference between the population change and the natural increase. For each area, this is distributed by five year age groups using migration data by age from three sources: residual

net migration from the 1976 and 1981 censuses, migration data from income tax files and the 1981 mobility question. The mobility question referred to is "Where were you on June 1, 1976?" in the 1981 Census. From the responses obtained for this question, in-migrants to and out-migrants from each small area can be tabulated. The five year age groups are split into single years of age using SPRAGUE multipliers. Before applying Sprague multipliers, the residual net migration is first split into in and out migration. Using in and out tax migration data as a reference, this calculation is done individually for each five-year age group.

$$\text{Residual In-Migration} = \frac{\text{Tax Data In-Migration}}{\text{Tax Data Net Migration}} \times \text{Residual Net Migration}$$

$$\text{Residual Out-Migration} = \text{Residual In-Migration} - \text{Residual Net Migration}$$

Using the preceding ratios, major problems occur when the split net migration is not of the same sign as the reference tax data on net migration. In this case, the sign of the split net migration is kept, but the resulting in and out migration are exchanged to yield the appropriate sign. This is based on the assumption of equal magnitude of a reversal of the migration flow.

2.3 Counts From The Family Allowances File, Ages 1-14 years

Estimates of population produced by the cohort-component method for the age groups 1-4, 5-9, 10-14 are replaced by counts of family allowance recipients at these ages which are readily available for CDs and CMAs, within 3 to 4 months after the reference date. Family allowances are paid universally in Canada and hence the counts are considered to be complete for all practical purposes. The data on the family allowance recipients are not provided by sex. Hence they are split into males and females using the sex composition from the latest census.

2.4 Adjustments for Consistency with Provincial and Census Division Estimates

Postcensal regression-nested estimate of total population of each CD and CMA become available within six months after the reference date. In addition, provincial estimates of population also become available by age and sex about the same time. Estimates of population by age and sex prepared as described above for the CDs within each province are controlled with respect to the census division total population estimates, and to the provincial population estimates by age and sex on a pro rata basis. For the census metropolitan areas, the age and sex totals are adjusted only to the CMA total population estimate.

3. EVALUATION

The evaluation is done with respect to three criteria: (i) accuracy; (ii) timeliness and (iii) consistency. Each of these is discussed below.

3.1 Accuracy

The accuracy of population estimates by age and sex depends to a large extent on the accuracy of estimation of the age-sex distribution of migrants, as the data on deaths by age and sex are considered satisfactory. Thus an evaluation of population estimates by age and sex indirectly throws light on the accuracy of migration estimates by age and sex. The accuracy is examined by comparing the estimates with the corresponding census counts.

Table 1

Distribution of Census Divisions/CMAs Showing the Accuracy of Population Estimates by Age, 1981

Provinces	Methods of Migration Estimation	Levels of Mean Absolute Error (%) by Sex							
		Males				Females			
		Under 3	3-5	5-10	10+	Under 3	3-5	5-10	10+
Newfoundland	R	8	2	0	0	10	0	0	0
	M	2	7	1	0	3	3	3	1
	T	0	0	5	5	1	1	3	5
Prince Edward Island	R	1	1	1	0	2	0	0	1
	M	1	2	0	0	1	2	0	0
	T	0	0	1	2	0	0	2	1
Nova Scotia	R	8	4	3	3	8	5	3	2
	M	3	6	4	5	5	6	2	5
	T	0	2	8	8	1	3	11	3
New Brunswick	R	10	1	2	2	7	4	3	1
	M	4	7	3	1	3	8	3	1
	T	0	2	5	8	0	4	5	6
Quebec	R	13	27	23	13	24	18	19	15
	M	12	26	26	12	17	23	21	15
	T	1	5	37	33	3	13	32	28
Ontario	R	30	8	8	7	37	5	4	7
	M	8	16	18	11	21	10	10	12
	T	0	8	34	11	4	10	31	8
Manitoba	R	4	6	8	5	4	6	8	5
	M	1	5	12	5	0	6	7	10
	T	0	1	8	14	0	1	4	18
Saskatchewan	R	10	5	1	2	9	5	2	2
	M	1	11	5	1	4	5	5	4
	T	1	1	10	6	1	1	12	4
Alberta	R	9	3	1	2	8	3	3	1
	M	5	5	3	2	5	4	5	1
	T	0	2	5	8	0	3	5	7
British Columbia	R	19	3	3	4	23	1	1	4
	M	9	13	2	5	14	8	3	4
	T	0	0	13	16	0	3	16	10
CMA	R	14	8	2	0	19	3	2	0
	M	2	17	5	0	10	9	4	1
	T	1	7	13	3	1	10	12	1

Note: R: Residual based age distribution of migrants, 1976-81.
M: Mobility based age distribution of migrants, 1981.
T: Annual tax migration data.
Source: Demography Division, Statistics Canada, 1985.

For each CD and CMA, three sets of population estimates by age and sex as of June 1, 1981 produced by using the age distribution of migrants from the three sources (residual (1976-81), mobility (1976-1981) and annual tax files) and counts from family allowance files as described in sections 2.1 to 2.4 were compared with the 1981 census counts. The differences were termed errors and for each small area, a summary index known as the "mean absolute error" (MAE) was computed by taking the arithmetic mean of percentage errors disregarding

the sign for 16 five year age groups. The smaller the value of this index, the more accurate are the estimates. In Table 1, a classification of CDs by provinces and of CMAs is presented for four levels of mean absolute error: under 3%, 3-5%, 5-10% and over 10%. Overall, it appears that the residual based age distribution of migrants gives better estimates. For males, about 66% of the total number of census divisions had an MAE under 5%. For females this percentage was slightly higher, at 69%. In contrast, lower percentages were observed for the mobility (55% and 57%) and tax migration data (9% and 19%), for males and females, respectively.

For CMAs too, the residual age distribution of migrants seems to give better estimates. The proportions of cases with MAE under 3% were 58% and 79% for males and females respectively. Mobility and tax based age distributions of migrants ranked second and third respectively, for both males and females.

With the exception of Prince Edward Island, the relative accuracy of the three sets of age distribution observed for Canada largely holds good for each province. This is true for both males and females. However, in some cases the residual based age distributions seem to give results similar to the mobility based distributions. Such similarity was observed for males in three provinces (Newfoundland, New Brunswick and British Columbia), whereas for females it was found only in New Brunswick.

It should be noted that the age distribution of migrants derived by the residual method uses the census age distributions of 1976 and 1981. Consequently, the population estimates as of June 3, 1981 prepared by using the migrant age distribution based on the residual method can be expected to be similar to the 1981 census age distribution. Hence, on the basis of this comparison we cannot conclude that the migrant age distribution derived by the residual method is better than the distribution derived from mobility question or from tax files.

Table 2 presents the percentage distribution of CD and CMA outliers. The outliers are those CDs with an MAE of over 10% and those CMAs over 5%. They are presented by sex and the three sources of migrant age distributions. As expected, both for males and females, the proportion of outliers is generally low for estimates using residual based age distribution. On the other hand, the percentage of outliers tends to be high for estimates using tax based migration distribution.

Temporal Stability of the Three Sets of Estimates During Postcensal Years, 1982-1984

For postcensal years, as there are no standard age distributions with which the estimates can be compared, the three population estimates by age and sex are compared with each other to learn of the temporal stability among them. A summary index knows as the "index of dissimilarity" calculated as half of the sum of absolute differences in two percentage age distributions is used for this purpose. The range of the index is from 0 to 100. The smaller the value, the greater is the similarity between the two distributions compared. The small areas are classified into three levels of dissimilarity: (i) the smallest level of difference with indices between 0% and 5%; (ii) the medium level of difference with indices between 5 to 10% and (iii) the outliers showing the index value of 10% and over. The classification of CDs is presented in Table 3 and that of CMAs in Table 4.

From Table 3, it appears that all the three population distributions tend to be similar and on average, a high percentage of cases, about 90%, are in the smallest category of differences (0%-5%) with only about 7% falling in the 5% to 10% category.

The percentage of cases with the extreme level of differences (index of dissimilarity exceeding 10%) were also examined for the ten provinces and their total. For males, the percentages of extreme cases were small, 3 to 5% between the residual and mobility based age distributions. For females, a relatively higher proportions of outliers were noticed. For other comparisons, residual vs tax based, and mobility vs tax based, slightly higher proportions of outliers were found. The results were similar for census metropolitan areas (see Table 4).

Table 2

Percentage of Outliers[a] Among Census Divisions by Province, and of CMA's 1981

Provinces	Males			Females		
	R	M	T	R	M	T
Newfoundland	0	0	50	0	10	50
Prince Edward Island	0	0	67	33	0	33
Nova Scotia	17	28	44	11	28	17
New Brunswick	13	7	53	7	7	40
Quebec	17	16	43	20	20	37
Ontario	13	21	21	13	23	15
Manitoba	22	22	61	22	43	78
Saskatchewan	11	6	33	11	22	22
Alberta	13	13	53	7	7	47
British Columbia	14	17	55	14	14	34
Total	15	16	43	15	20	35
CMA	8	21	67	8	21	54

Note: R: Residual based age distribution of migrants.
M: Mobility based age distribution of migrants.
T: Tax based age distribution of migrants.
[a] The outliers are those CDs with MAE of over 10% and those CMAs with MAE of over 5%.
Source: Table 1.

Table 3

Distribution of Census Divisions by Level of Index of Dissimilarity Obtained by Comparing the Age Distributions of Population Based on Residual, Mobility and Tax Migration Sources, 1982 to 1984

Year/ Index of Dissimilarity	Males			Females		
	Residual vs Mobility	Residual vs Tax	Mobility vs Tax	Residual vs Mobility	Residual vs Tax	Mobility vs Tax
YEAR 1982						
0-5	245	237	242	240	241	234
5-10	7	13	10	8	5	11
10+	8	10	8	12	14	15
Total	260	260	260	260	260	260
YEAR 1983						
0-5	235	221	223	230	229	223
5-10	11	18	21	10	13	13
10+	14	21	16	20	18	24
Total	260	260	260	260	260	260
YEAR 1984						
0-5	240	226	229	235	233	231
5-10	11	16	14	15	13	12
10+	9	18	17	10	14	17
Total	260	260	260	260	260	260

Source: Demography Division, Statistics Canada, October 1985.

Table 4

Distribution of Census Metropolitan Areas by Level of Index of Dissimilarity Obtained by Comparing the Age Distributions of Population Based on Residual, Mobility and Tax Migration Sources, 1982 to 1984

Year/	Males			Females		
Index of Dissimilarity	Residual vs Mobility	Residual vs Tax	Mobility vs Tax	Residual vs Mobility	Residual vs Tax	Mobility vs Tax
YEAR 1982						
0-3	24	24	24	23	22	23
3-5	0	0	0	0	1	0
5+	0	0	0	1	1	1
Total	24	24	24	24	24	24
YEAR 1983						
0-3	22	23	22	21	20	20
3-5	2	0	0	1	2	3
5+	0	1	2	2	2	1
Total	24	24	24	24	24	24
YEAR 1984						
0-3	21	21	21	21	20	20
3-5	2	2	2	0	1	0
5+	1	1	1	3	3	4
Total	24	24	24	24	24	24

Source: Demography Division, Statistics Canada, October 1985.

In conclusion, it may be said that although the three age distributions of migrants (residual, mobility and tax based) differed from each other, age distributions of population resulting from these were largely similar.

3.2 Timeliness

Timeliness refers to the availability of estimates within as short a time as possible after the reference date. Using the preliminary population totals (regression-nested estimates) which become available within six months from the reference date, the estimated numbers of births, deaths by age and net migrants by age as described in Sections 2.1 to 2.4, the population estimates by age and sex for CDs and CMAs could be prepared within eight months of the reference date.

3.3 Consistency

Consistency refers to the consistency in the sources of data sets used for estimation at various levels of administrative or other disaggregated areas and to the uniformity in the methods of estimation. While in certain cases, a different method may have to be used, it is highly desirable to use the same method throughout in order to ensure the methodological consistency of various levels of geographic disaggregation.

For provinces, CDs and CMAs, the sources of data are the same for births and deaths: the vital registration records. For migration data too, the sources are the same namely, tax files and mobility data from the census for all levels of geographic disaggregation. However, an additional data set, the residual age data derived from the two consecutive censuses is also used.

There is full methodological consistency between provinces and other levels as the cohort-component method is used in all cases.

4. CONCLUSION

By using the cohort-component method, three sets of estimates by age and sex have been prepared for CDs and CMAs. Each set uses a different migration component by age and sex: (i) tax file based; (ii) mobility data from the latest census and (iii) the residual derived from the two consecutive censuses.

Although the three age distributions of migrants differ from each other, the resulting estimates of population by age and sex were largely similar. Each set involves its own assumptions. Using a residual age distribution of migrants for postcensal estimation assumes that the age distribution remains constant for the period of estimation. A similar assumption is involved in using mobility data by age and sex for postcensal years. The data from tax files assume that the age-sex distribution remains the same for any two consecutive years. However, the type of movement measured by each of these sets is not the same. The residual measures only the net movement between the two consecutive censuses (e.g. 1976-81). The mobility question also measures five-year movements ranging from 0-4 years. The tax files, on the other hand show the movement during roughly a 12 month period. On the basis of the comparisons made in the paper, it cannot be concluded that one migrant data set giving rise to population estimates is better than the other. A more satisfactory evaluation of the three sets of estimates can be made only when the next census results become available.

REFERENCES

CENTRAL STATISTICS BUREAU (1980). Population estimates by age groups for school districts, 1977-80. Unpublished Technical Report, Government of B.C.

NORRIS, D., and STANDISH, L. (1983). A Technical report on the development of migration data from taxation records. Technical Report, Administrative Data Development Division, Statistics Canada.

STATISTICS CANADA (Annual). *Postcensal Annual Estimates of Population by Marital Status, Age, Sex and Components of Growth for Canada and the Provinces*, Vol. 2, 2nd issue. Catalogue 91-210, Ottawa: Ministry of Supply and Services.

STATISTICS CANADA (Annual). *Postcensal Annual Estimates of Population for Census Divisions and Census Metropolitan Areas (Regression-Method)*. Catalogue 91-211, Ottawa: Ministry of Supply and Services Canada.

STONE, Leroy O. (1980). Evaluating the relative accuracy and significance of net migration estimates. *Demography*, 4, 310-330.