Survey Methodology, December 1985
Vol. 11, No. 2, pp. 203–210
Statistics Canada

Estimating the Age/Sex Distribution of Small Area Populations[1]

DAVID S. O'NEIL and CHRIS D. McINTOSH[2]

ABSTRACT

This paper describes a method of producing current age/sex specific population estimates for small areas utilizing as inputs total population estimates, birth and death data and estimates of historical residual net migration. An evaluation based on the 1981 Census counts for census divisions and school districts in British Columbia is presented.

KEY WORDS: Age/sex population estimates; Small area; Residual net migration.

1. INTRODUCTION

The Central Statistics Bureau currently produces post-censal population estimates for a variety of sub-provincial areas using a regression approach (Central Statistics Bureau 1982). In addition to estimates of the total population by small area, age/sex specific estimates are also produced.

This paper outlines the method by which age/sex specific population estimates are derived for subprovincial areas of British Columbia, given an estimate of the total population.

2. OVERVIEW

The methodology used to derive the small area populations by sex and single years of age is divided into two parts.

The first part consists of examining historical residual net migration data compiled from censuses to derive a number of migration distributions by sex and single year of age for each small area (Shryock and Siegal 1980).

The second part of the methodology consists of aging the base population for each sex and adding births and subtracting deaths to yield a new population distribution for each area. This is referred to as the "natural base" population. The difference between the estimated total population by sex and the natural base population yields a residual term, which is equal to net migration by sex if the population and vital events for the two periods are exact. This small area sex specific residual term is distributed by single years of age according to a historical distribution, then added to the natural base population giving an age/sex specific population estimate for the area in the next time period.

Due to the timeliness of the input data, estimates of the total populations can be produced four months after the reference date of June 1, and the age/sex breakdowns one to two months later.

---

[1] Abridged version of the paper presented at the meeting of the Federal-Provincial Committee on Demography, Ottawa, November 28-29, 1985.

[2] D.S. O'Neil, SRL Sociometrics Resources Ltd., and C.D. McIntosh, Intersoft Resources Ltd., Central Statistics Bureau, Ministry of Industry and Small Business Department, Government of British Columbia, 2nd Floor, 1405 Douglas Street, Victoria, British Columbia, Canada V8W 3C1.
The views expressed in this paper are those of the authors and do not necessarily represent the views of the Government of British Columbia.

3. HISTORICAL NET MIGRATION DISTRIBUTIONS

Age/sex specific residual estimates of net migration were compiled for the census periods 1961/66, 1966/71 and 1971/76 for each of the 74 British Columbia school districts. These are referred to as the Historical Small Area Distributions.

Examination of these net migration distributions by small area showed them to be extremely unstable over time. In order to minimize the effects of this instability, a number of steps were taken.

First, migration distributions by small area were separated according to whether they occurred during a time of positive or negative total net migration. It was found that residual migration age distributions for many areas differed depending on whether net migration was positive or negative.

A further step taken to reduce the effects of unstable migration distributions was to group small areas of similar proportional migration distributions together, then calculate the positive and negative net migration distributions for each group of areas. These were called the Historical Grouped Distributions. Cluster analysis (using the SPSS/PC procedure) across selected age groups was used to group the historical small area migration distributions. Examination of cluster memberships from different periods resulted in the placing of the majority of areas into three clusters, while eight areas were maintained as unique independent clusters. Once areas had been arranged into groups, positive and negative migration distributions were calculated from the most recent periods of positive or negative net migration.

4. SMALL AREA POPULATION ESTIMATES BY SEX AND SINGLE YEAR OF AGE

As noted in Section 3, some areas showed considerable time-series variation in the residually calculated net migration distributions. This was likely the result of two factors. First, many of the areas under study possess small resource based economies subject to wide fluctuations, with consequent swings in migration levels. Second, a certain amount of instability is introduced when calculating a percentile distribution for a concept such as net migration, which may have either positive, negative, or zero values.

In order to guard against adopting a historical net migration distribution that may not be a representative distribution for the estimating year, five different historical sex-specific distributions were calculated, then distributed by single year of age. A description of these five different net migration distributions is given below.

1) The Historical Small Area Distribution for each small area having the same sign as the net migration to that small area was the first migration distribution.
2) The Historical Group Distribution for the group the small area belongs to, having the same sign as the net migration to that small area, was the second migration distribution.
3) The third migration distribution was calculated by separately totaling the migration from the most recent time period for all small areas with a positive and negative net migration, then calculating the age distributions.
4) The fourth distribution was the distribution of the natural base population for each small area.
5) The fifth and final distribution was the age distribution of migrants to British Columbia as a whole. For all the years under consideration, migration to B.C. has been positive, hence this is a positive distribution. Nevertheless, it was used as the fifth distribution regardless of whether the migration to a small area was positive or negative.

In some cases it was not possible to calculate all five distributions. This was the case if a small area never had a negative net migration in the past, but one is indicated for the estimating year under consideration. In situations such as this only distributions that can be calculated were used to distribute the small area net migration.

Empirical testing based on the 1981 Census indicated that of the five net migration distributions described above, number 1 (the Historical Small Area Distribution) produced the lowest average absolute percent error over all school districts and age groups, followed by number 2 (Historical Grouped Distribution), then number 3, etc. However, despite the fact that distribution number 1 produced the lowest error on average, it did not produce the lowest error in each case. Hence, a selection procedure was designed to substitute the population distribution produced by number 1, with either 2, 3, 4, or 5 in only those cases where the population distribution produced by number 1 was considered unrepresentative of the estimating year population distribution.

Empirical testing based on the 1981 Census resulted in the following selection procedure to be adopted.

First, all migration distributions possible were calculated and added to the natural base population, resulting in up to five possibilities for the small area estimated population by sex and single year of age in the next time period. These age/sex specific population estimates were then examined to determine which one produced the least change in the small area age structure from the previous year. This was done by first calculating the unweighted average percent difference between the age structures for each of the five possible populations in time t + 1 to the population in time t. Next, the standard deviations about these averages were calculated, and the distribution with the lowest standard deviation is flagged. If the standard deviation produced by using the Historical Small Area Distribution was significantly greater than the smallest standard deviation (i.e. of the flagged distribution), then the Historical Small Area Distribution was rejected. This procedure was repeated with the Historical Grouped Distribution, and so on until one of the five possible populations was selected.

Once the "best" population in time $t+1$ was calculated for all small areas, two final adjustments were made. First, family allowance data was substituted for the age groups 0-14, and the populations for the rest of the age groups were pro-rated to keep the total population of each small area constant. The second adjustment was to pro-rate the population to ensure the age distribution of the sum of the small area population estimates was consistent with the British Columbia age distribution estimated by Statistics Canada.

5. EVALUATION OF THE CURRENT METHODOLOGY

The following tables summarize the error associated with the June 1, 1981 population estimates by five year age group to 70+, for 74 British Columbia school districts and 29 census divisions. The census division age/sex specific population estimates were derived by aggregating school district population estimates.

The accuracy of the small area age/sex specific population estimates derived from the previously described methodology was evaluated by producing 1981 population estimates by sex and 5 year age groups to 70+ for 74 school districts, then comparing these results to the 1981 Census. Two summary measures were used to evaluate the effectiveness of the age/sex specific population estimates. These were Average Absolute Percent Error (AAPE) and Index of Misallocation (IM). The AAPE is defined as:

$$AAPE = 100 \times \left[\sum_{i=1}^{N} \left| (P_{Ei} - P_{Ai})/P_{Ai}) \right| \right] / N$$

where P_{Ei} is the estimated cell population for age group i, P_{Ai} is Census cell population for age group i, and N the number of cells. The IM is defined as:

$$IM = 100 \times \frac{1}{2}\left[\sum_{i=1}^{N} \left(\left| P_{Ai} - P_{Ei} \right| \right) \right] / \sum_{i=1}^{N} P_{Ai}$$

where P_{Ai} is the actual cell population for age group i, and P_{Ei} is the estimated cell population for age group i.

As seen in Table 1, relative to the 1981 Census the average absolute percent error over all age groups and regions is 6.20%, and the IM is 1.95%. The average percent errors for male and female are quite similar (AAPE's of 7.00% for both, and IM's of 2.15% for males and 2.08% for females).

By age, the highest errors occur in the 20-29 and 60-69 age groups. It should also be noted that there is some difference in the age distribution of errors between males and females. Males appear to have higher error in the upper age groups, while females have higher error in the very mobile 20-29 age groups.

Table 1

Error by Age Group Across School District
1981 Estimated Versus Census
Absolute Average Percent Error (AAPE) and Index of Misallocation (IM)

Age	Total		Male		Female	
	AAPE (%)	IM (%)	AAPE (%)	IM (%)	AAPE (%)	IM (%)
0-4	3.33	0.96	3.94	1.21	3.62	1.04
5-9	2.80	0.76	3.28	0.88	3.62	1.02
10-14	2.33	0.64	3.54	0.84	2.88	0.87
15-19	5.20	2.01	5.68	2.01	6.18	2.24
20-24	13.32	4.77	13.50	4.62	14.54	5.12
25-29	8.31	4.07	8.42	3.70	9.41	4.65
30-34	5.02	2.12	5.42	2.45	5.72	2.06
35-39	4.88	1.33	5.73	1.62	5.38	1.34
40-44	4.52	1.33	5.84	1.51	4.67	1.52
45-49	3.60	1.22	4.47	1.37	4.78	1.49
50-54	5.66	1.33	5.86	1.48	6.68	1.54
55-59	6.11	1.72	6.19	1.78	7.82	1.97
60-64	8.86	2.44	10.35	2.95	8.91	2.17
65-69	10.60	2.66	12.53	3.52	11.44	2.30
70+	8.49	1.95	10.19	2.35	9.33	1.94
Average	6.20	1.95	7.00	2.15	7.00	2.08

As seen in Table 2 , on average higher percent errors are associated with areas of small population size. The higher percent errors in smaller areas may be associated with the instability of the smaller (resource based) economies, and associated instabilities in net migration distributions.

By census division, similar error patterns are observed. As seen in Table 3, the average absolute percent error across all regions and age groups is 4.83%, 5.19% for males and 5.60% for females. The IM is 1.27% for the total, 1.41% for males and 1.35% for females. Again, the error is bimodal, with peaks at 20- 29 and 60-69. In addition, the females have higher errors than males in the 20-29 age groups, while the reverse is true in the 60-69 age groups.

Table 2

School District Error by Population Size

Population Grouping	Total		Male		Female	
	AAPE (%)	IM (%)	AAPE (%)	IM (%)	AAPE (%)	IM (%)
0-9,999	8.87	3.16	10.14	3.89	10.27	3.65
10,000-24,999	6.07	2.47	6.92	2.96	6.62	2.58
25,000+	3.66	1.67	3.92	1.78	4.09	1.78
School District Average	6.20	1.95	7.00	2.15	7.00	2.08

Table 3

Error by Age Group Across Census Division
1981 Estimated Versus Census

Age Group	Total		Male		Female	
	AAPE (%)	IM (%)	AAPE (%)	IM (%)	AAPE (%)	IM (%)
0-4	2.37	0.54	3.20	0.76	2.28	0.58
5-9	1.52	0.50	1.71	0.55	2.13	0.68
10-14	1.69	0.39	2.75	0.57	2.50	0.60
15-19	3.81	1.39	3.79	1.30	4.68	1.63
20-24	9.83	3.07	9.30	2.91	10.90	3.41
25-29	7.02	3.04	7.30	2.87	8.09	3.37
30-34	3.28	1.29	3.31	1.43	3.85	1.25
35-39	3.34	0.66	3.06	0.57	4.21	0.88
40-44	3.86	0.88	4.29	1.01	4.16	0.90
45-49	2.91	0.70	3.20	0.75	3.75	0.83
50-54	4.82	0.64	4.41	0.75	6.10	0.86
55-59	5.49	1.34	5.36	1.55	6.94	1.30
60-64	7.88	1.95	8.37	2.29	7.94	1.74
65-69	8.48	1.89	10.30	2.67	9.79	1.43
70+	6.16	0.81	7.46	1.20	6.73	0.71
Avg	4.83	1.27	5.19	1.41	5.60	1.35

Table 4 (Census Division Error By Population Size) shows the improvement in error levels resulting from aggregating to larger sub-provincial areas. Table 7 illustrates the negative relationship between error levels and population size on a Census Division level.

A comparison of Tables 5 and 6 again demonstrates the improvement in error levels when aggregating to larger age/sex cell sizes. Although this does indicate that some precautions should be observed when utilizing age/sex estimates for some small areas, we do not believe it should preclude use of the estimates for these areas.

Table 4

Census Division Error by Population Size

Population Grouping	Total		Male		Female		N
	AAPE (%)	IM (%)	AAPE (%)	IM (%)	AAPE (%)	IM (%)	
0-39,000	7.22	1.94	7.55	2.13	8.79	2.29	10
40,000-59,999	4.32	1.82	5.03	2.14	4.91	1.83	10
60,000+	2.51	0.87	2.73	.98	2.84	0.90	9
Census Division Average	4.83	1.27	5.19	1.41	5.60	1.35	29

Table 5

- School District -
Number of Estimates by Error Range

	Average Absolute Percent Error Range				
	< 5	5 to 10	10 to 15	15+	Total
No. of Cells	674	239	101	96	1110
Percent	61%	22%	9%	9%	100%

Table 6

- Census Division -
Number of Estimates by Error Range

	Average Absolute Percent Error Range				
	< 5	5 to 10	10 to 15	15+	Total
No. of Cells	306	77	25	27	435
Percent	70%	18%	6%	6%	100%

6. FINAL REMARKS

The procedure outlined above has particular advantages for use in a region with well developed sources of historical small area population and vital statistics data. It is felt that a procedure utilizing net-migration estimates is relatively straightforward, produces acceptable error levels, and can produce age/sex estimates soon after the reference date. Although the optimal situation would be to have in- and out-migration estimates, currently little information is available on small area migration flows within British Columbia. One further improvement to the system being considered is the incorporation of Old Age Security counts to increase the stability and accuracy of estimates in the older age groups.

ACKNOWLEDGEMENTS

The authors would like to thank Don McRae, Steve Miller, Ravi Verma, Garnett Picot, and Paul Knapp whose input to and support for the development of the estimation breakdown system should not go unrecognized.

Table 7

Error by Census Division Across Age Groups
1981 Estimated Versus Census

Census Division	Total Population	Total		Male		Female	
		AAPE (%)	IM (%)	AAPE (%)	IM (%)	AAPE (%)	IM (%)
1000 East Kootenay	53,725	4.24	2.04	5.24	2.29	3.88	2.15
3000 Central Kootenay	52,045	4.00	2.18	4.03	2.13	5.06	1.69
5000 Kootenay-Boundary	33,235	2.32	1.23	2.34	1.18	3.21	1.68
7000 Okanagan-Similkameen	57,185	5.04	2.64	6.02	3.08	4.72	2.49
9000 Fraser-Cheem	56,930	3.12	1.60	3.33	1.78	4.15	2.08
11000 Central Fraser Valley	115,015	3.14	1.43	3.46	1.52	3.65	1.81
13000 Dowdney-Alouette	62,000	2.10	1.15	2.56	1.23	2.26	1.32
15000 Greater Vancouver	1,168,700	1.63	0.94	1.68	0.93	1.67	0.98
17000 Capital	249,475	1.64	0.87	2.31	1.21	1.18	0.61
19000 Cowichen Valley	45,315	3.09	1.66	3.36	1.69	3.85	2.08
21000 Nanaimo	84,815	3.07	1.58	3.40	1.74	3.22	1.66
23000 Alberni-Clayoquot	32,560	2.75	1.36	2.88	1.27	3.27	1.68
25000 Comox-Strathcona	68,620	1.44	0.80	1.85	0.87	2.85	1.50
27000 Powell River	19,050	5.36	2.58	5.06	2.44	6.18	3.03
29000 Sunshine Coast	16,625	4.84	2.57	6.79	3.58	5.65	2.81
31000 Squamish-Liltooet	18,925	1.82	0.99	2.56	1.37	3.10	1.58
33000 Thompson-Nicola	102,430	2.13	1.10	2.07	0.10	2.65	1.37
35000 Central Okanagan	85,235	3.96	1.93	3.91	1.88	4.32	2.14
37000 North Okanagan	69,033	5.26	2.52	6.44	3.06	5.05	2.50
39000 Columbia-Shuswap	45,425	3.04	1.63	3.56	1.84	2.99	1.66
41000 Cariboo	58,810	3.18	1.93	3.90	2.18	3.42	2.06
43000 Mount Waddington	14,675	8.96	3.04	5.13	1.59	17.77	5.49
45000 Central Coast	3,050	17.99	7.62	21.62	8.86	14.92	7.34
47000 Skeena-Queen Charlotte	24,030	4.82	2.09	5.70	2.58	4.61	1.84
49000 Kitimat-Stikina	41,790	6.26	1.99	4.99	1.66	8.59	2.78
51000 Bulkley-Nechako	38,310	6.23	2.31	5.76	2.10	6.83	2.57
53000 Fraser-Fort George	89,430	3.50	1.41	3.39	1.25	3.72	1.68
55000 Peace River-Liard	55,340	8.00	2.95	9.43	3.65	7.34	2.83
57000 Stikine	2,685	17.15	6.89	17.89	6.88	22.39	8.35
Average Error		4.83	2.17	5.19	2.31	5.60	2.51

REFERENCES

CENTRAL STATISTICS BUREAU (1982). British Columbia municipal population estimation methodology. Unpublished report. Victoria: British Columbia Ministry of Industry and Small Business Development.

SPSS INC. (1984). *Statistical Package for the Social Sciences/PC.* Chicago, B265-B280.

SHRYOCK, H.S., and SIEGAL, J.S. (1980). The methods and materials of demography, 2. U.S. Bureau of the Census, 628-630.