Estimating the Age/Gender Distribution of Small Area Populations in British Columbia

David O'Neil, Chris McIntosh
Population Section,
Central Statistics Branch,
Ministry of Government Services
Government of British Columbia

April, 1994

ESTIMATING THE AGE/GENDER DISTRIBUTION OF SMALL AREA POPULATIONS

ABSTRACT

This paper describes a method of producing current age/gender specific population estimates for small areas utilizing as inputs total population estimates, birth and death data, family allowance and old age security administrative records, and estimates of historical residual net migration. An evaluation based on the 1991 Census counts (adjusted for net census undercount) for census divisions and Local Health Areas in British Columbia is presented.

1. Introduction

The Central Statistics Branch currently produces post-censal estimates of the total population for a variety of sub-provincial areas using a regression approach (Central Statistics Branch 1993). In addition to estimates of the total population by small area, age/gender specific estimates are also produced.

This paper outlines the method by which age/gender 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 gender 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 gender and single year of age for each small area. [1]

The second part of the methodology consists of aging the base population for each gender 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 gender and the natural base population yields a residual term, which is equal to net migration by gender if the population and vital events for the two periods are exact. This small area gender 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/gender 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 July 1, and the age/gender breakdowns one to two months later.

[1] For more information, see Shryock, H.S. and Siegal, J.S. (1980) The Methods and Materials of Demography, Vol. 2. U.S. Bureau of the Census, pp 628-630.

3. Historical Net Migration Distributions

Age/gender specific residual estimates of net migration were compiled for the census periods 1961/66, 1966/71, 1971/76, 1976/81, 1981/86, and 1986/91 for each of the 79 British Columbia Local Health Areas. 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 Gender 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 gender-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 be 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, 1986 and 1991 Censuses 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 Local Health Areas 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, 1986 and 1991 Censuses 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 gender and single year of age in the next time period. These age/gender 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 was 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, three final adjustments were made. First, Family Allowance data was substituted for

the age groups 0-14. Second, Old Age Security counts were substituted for age groups above 64 years of age. In both cases of substitution, adjustment factors derived from the previous Census were applied to the Family Allowance and Old Age Security total counts in order to be consistent with the Census. The populations for the rest of the age groups were pro-rated to keep the total population of each area unchanged. The third 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 July 1, 1991 population estimates by five year age group to 90+, for 79 British Columbia Local Health Areas and 30 census divisions. The census division age/gender specific population estimates were derived by aggregating Local Health Area population estimates.

The accuracy of the small area age/gender specific population estimates derived from the previously described methodology was conducted by producing 1991 population estimates by gender and 5 year age groups to 90+ for 79 Local Health Areas, then comparing these results to the 1991 Census (adjusted for net census undercount).

As noted in Section 1, two distinctly different methodologies are used to estimate the small area total, and the age/gender specific populations. This evaluation is intended to measure only the accuracy of methodology that generates the age/gender estimates; hence, it is assumed that the estimates of total population at the provincial and regional levels are free from error. As a result, the following evaluation does not give the expected error associated with the overall small area estimation program, but rather the small area age/gender component only. [2]

[2] For an overall evaluation, see: *Estimating the Age/Gender Distribution of Small Area Populations in British Columbia -- Overall Evaluation* , unpublished report, Central Statistics Branch, Ministry of Government Services, Government of British Columbia, April 1994.

Two summary measures were used to evaluate the effectiveness of the age/gender specific population estimates. These were Average Absolute Percent Error (AAPE) and Index of Misallocation (IM). The AAPE is defined as:

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

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:

$$IM = 100 \times \tfrac{1}{2} \left[\sum_{i=1}^{N} \left| \left(P_{A_i} - P_{E_i} \right) \right| \right] \Big/ \sum_{i=1}^{N} P_{A_i}$$

where P_{Ai} is the actual cell population for age group *i*, and P_{Ei} is the estimated cell population for age group *i*, and it is assumed that $\sum P_{Ai} = \sum P_{Ei}$.

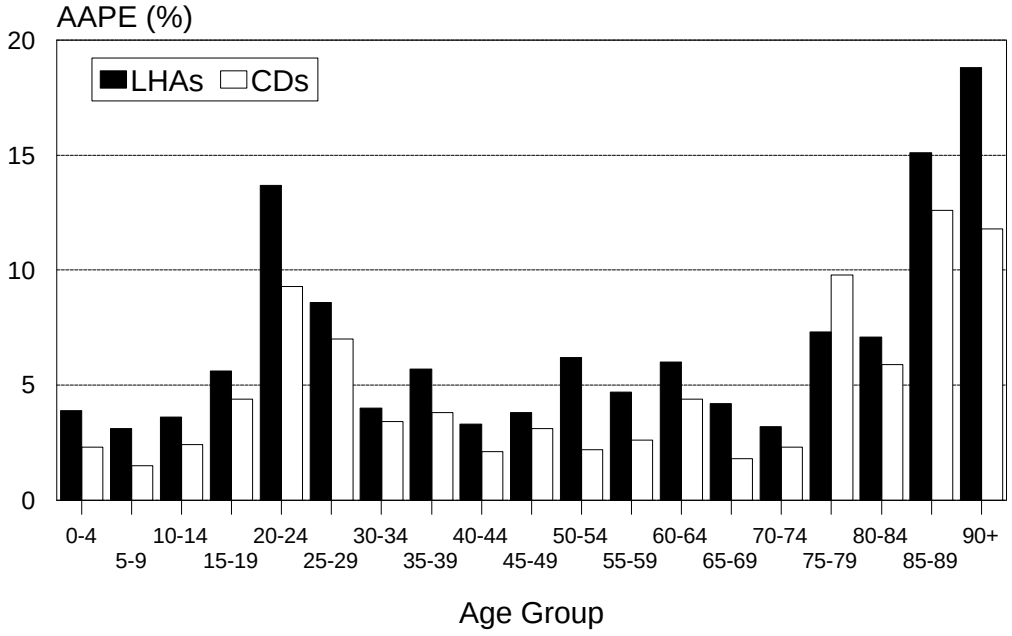
As seen in Table 1, relative to the 1991 adjusted Census, the average absolute percent error over all age groups and regions is 6.7%, and the IM is 1.67%. The average percent errors are 8.0% for males and 7.2% for females, and IM's are 1.86% for males and 1.85% for females).

Table 1
Error by Age Group Across Local Health Areas
1991 Estimated Versus Adjusted Census
Average Absolute Percent Error (AAPE)

	Total	Male	Female
Age Group	AAPE (%)	AAPE (%)	AAPE (%)
0-4	3.9	5.5	5.3
5-9	3.1	5.0	5.8
10-14	3.6	5.3	5.9
15-19	5.6	5.9	7.0
20-24	13.7	14.9	14.1
25-29	8.6	9.0	10.8
30-34	4.0	5.6	5.1
35-39	5.7	7.1	5.3
40-44	3.3	3.6	4.1
45-49	3.8	4.0	5.5
50-54	6.2	6.1	4.2
55-59	4.7	5.4	5.9
60-64	6.0	7.4	5.9
65-69	4.2	3.3	4.1
70-74	3.2	6.1	3.6
75-79	7.3	5.7	3.7
80-84	7.1	11.7	9.0
85-89	15.1	18.1	12.7
90+	18.8	22.9	18.7
Average	**6.7**	**8.0**	**7.2**
IM (%)	**1.67**	**1.86**	**1.85**

By age, the highest errors occur in the 20-29 and 85 plus age groups. For British Columbia, this is an understandable outcome of the high mobility of young adults, and the large error in the older age groups is mainly attributable to small numbers. It should also be noted that there is some difference in the age distribution of errors between males and females.

Figure 1
Error by Age Group Across LHAs and CDs
1991 Estimated Versus Adjusted Census
Average Absolute Percent Error (AAPE)



Central Statistics Branch, Ministry of Government Services, Government of B.C.

As seen in Table 2, higher percent errors are associated, on average, with areas of small population size. This is also seen in Figure 2 (Average Absolute Percent Error and Total Population). 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.

Table 2
Local Health Area Error by Population Size

	Total	Male	Female	
Population Grouping	**AAPE (%)**	**AAPE (%)**	**AAPE (%)**	**Number Of Cases**
0-9,999	11.5	13.4	11.2	27
10,000-24,999	5.3	6.2	6.9	19
25,000+	3.7	4.6	4.1	33
Average	**6.7**	**8.0**	**7.2**	**79**

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 4.9%, 6.0% for males and 5.1% for females. The IM is 1.10% for the total, 1.21% for males and 1.15% for females. Again, the error peaks at the 20-29 and 85 plus age groups.

Table 3

Error by Age Group Across Census Divisions
1991 Estimated Versus Adjusted Census
Average Absolute Percent Error (AAPE)

	Total	Male	Female
Age Group	AAPE (%)	AAPE (%)	AAPE (%)
0-4	2.3	2.8	3.1
5-9	1.5	3.2	3.8
10-14	2.4	2.6	3.5
15-19	4.4	5.5	4.3
20-24	9.3	10.2	9.6
25-29	7.0	6.5	8.2
30-34	3.4	5.0	4.2
35-39	3.8	4.5	3.8
40-44	2.1	1.9	2.8
45-49	3.1	2.8	3.9
50-54	2.2	3.2	3.1
55-59	2.6	3.9	3.7
60-64	4.4	5.3	4.6
65-69	1.8	2.4	2.9
70-74	2.3	5.3	4.2
75-79	9.8	5.3	3.7
80-84	5.9	12.1	9.1
85-89	12.6	13.4	6.0
90+	11.8	17.4	13.1
Average	**4.9**	**6.0**	**5.1**
IM (%)	**1.10**	**1.21**	**1.15**

Table 4 (Census Division Error By Population Size) shows the improvement in error levels resulting from aggregating to larger sub-provincial areas.

Table 4

Census Division Error by Population Size

Population Grouping	Total AAPE (%)	Male AAPE (%)	Female AAPE (%)	Number Of Cases
0-24,999	10.4	12.3	10.1	7
25,000-74,999	3.6	4.6	4.3	14
75,000+	2.6	3.1	2.5	9
Average	**4.9**	**6.0**	**5.1**	**30**

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

Table 5

Local Health Area

Number of Estimates by Error Range

	<5	5 to 10	10 to 15	15+	Total
No. of Cells	989	269	109	134	1501
Percent	65.9%	17.9%	7.3%	8.9%	100.0%

Table 6

Census Division

Number of Estimates by Error Range

	<5	5 to 10	10 to 15	15+	Total
No. of Cells	434	78	33	25	570
Percent	76.1%	13.7%	5.8%	4.4%	100.0%

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/gender 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.

Table 7

Error by Census Division Across Age Groups
1991 Estimated Versus Adjusted Census

	1991 Census*	Total	Male	Female
		AAPE (%)		
Alberni - Clayoquot	32,101	3.50	5.38	4.15
Bulkley - Nechako	39,421	3.18	3.21	4.64
Capital	308,162	2.20	2.61	1.82
Cariboo	62,971	4.01	3.20	5.67
Central Coast	3,640	15.66	13.61	15.86
Central Fraser Valley	89,817	2.25	2.85	2.55
Central Kootenay	52,509	2.93	2.88	4.25
Central Okanagan	114,991	2.18	1.94	2.50
Columbia - Shuswap	42,837	4.07	4.83	3.97
Comox - Strathcona	85,217	2.90	4.25	2.98
Cowichan Valley	62,389	3.61	4.30	3.63
Dewdney - Alouette	92,497	2.88	3.30	2.67
East Kootenay	53,841	2.98	4.13	3.36
Fort Nelson - Liard	5,332	27.20	20.88	15.62
Fraser - Cheam	70,612	3.38	3.84	3.26
Fraser - Fort George	93,290	2.82	5.14	2.46
Greater Vancouver	1,586,126	1.97	2.06	1.91
Kitimat - Stikine	43,237	4.96	7.71	6.63
Kootenay Boundary	32,070	4.04	4.17	4.62
Mount Waddington	14,286	3.18	5.36	7.73
Nanaimo	104,921	2.32	1.75	2.60
North Okanagan	63,482	2.54	3.39	2.33
Okanagan - Similkameen	68,577	3.27	3.88	3.26
Peace River	54,814	3.22	3.54	5.05
Powell River	19,434	6.46	7.16	6.81
Skeena - Queen Charlotte	24,438	3.69	6.85	4.72
Squamish - Lillooet	25,070	4.20	8.61	7.30
Stikine	2,061	13.64	28.93	16.24
Sunshine Coast	21,417	3.31	3.21	4.00
Thompson - Nicola	107,340	3.47	4.13	3.25
Average Error	**3,376,900**	**4.9**	**6.0**	**5.1**
IM (%)		**1.10**	**1.21**	**1.15**

* Adjusted for net census undercount

Figure 2
Local Health Area
Absolute Average Percent Error - 1991

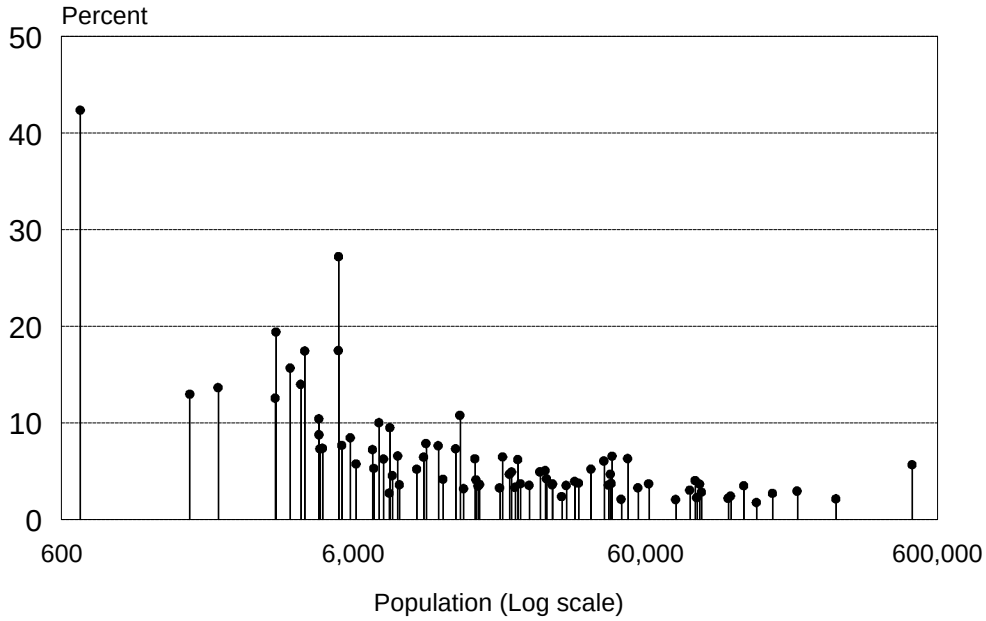
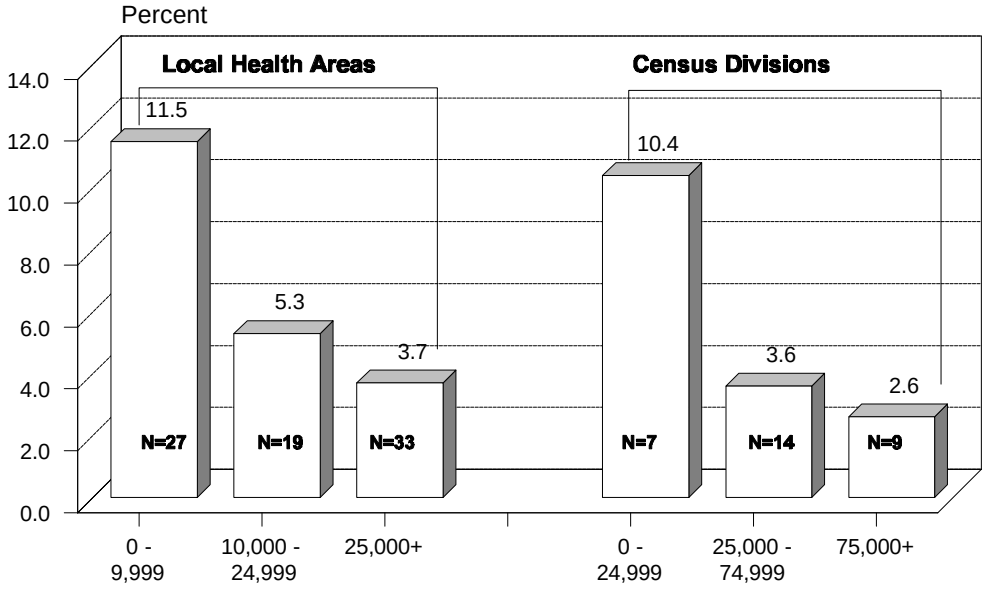


Figure 3
Local Health Area and Census Division
Absolute Average Percent Error - 1991



Central Statistics Branch, Ministry of Government Services, Government of B.C.