

The Development of Alberta Health Care Records and Their Application to Small-Area Population Estimates¹

**F. AHMAD, R. CHOW, O. DEVRIES,
A. HASHMI, and M. MARCOGLIESE²**

ABSTRACT

This paper examines the use of administrative files from Alberta's Health Care Insurance Plans combined with Vital Statistics data as inputs for estimating population. Results, which are presented and compared with Census data, indicate that Health Care data can be used to produce accurate population estimates at the provincial level and for smaller areas such as census divisions and municipalities.

KEY WORDS: Administrative files; Component method; Small areas; Residual net migration.

1. BACKGROUND

During the mid to late 1970's, the Province of Alberta experienced rapid economic growth led by activity in the oil and gas industry, which generated high population growth. Governments, in order to effectively provide goods and services for the influx of people into various regions, required timely data on where and by how much population was growing. With the need for up-to-date population data, it was felt that the federal quinquennial census was not sufficiently frequent nor current (census data are released about twelve to eighteen months after the reference year). Consequently, provincial agencies, and in particular, the Alberta Bureau of Statistics, began investigating alternative sources of timely population data.

After examining a number of potential sources, the Bureau began assessing administrative health care insurance data from the Alberta Health Care Insurance Plan (AHCIP) files to develop population statistics. The remainder of this paper highlights work undertaken by the Bureau to develop the AHCIP records and to use the data in estimating small-area population.

2. DEVELOPMENT OF AHCIP RECORDS INTO HEALTH CARE COUNTS

This section describes briefly the nature of the AHCIP records and evaluates the counts developed.

2.1 Developing Health Care Counts Data

The Bureau receives selected registration records via computer tape, on a quarterly basis, from the AHCIP registration-billing system. (The tape contains only a partial listing, in particular, all names, identifiers, etc. have been stripped such that the confidentiality of all individuals is strictly preserved.) The file contains information such as addresses, postal codes, registration and cancellation dates, age and sex for every registrant. (A detailed description of the record layout is available upon request.)

¹ Abridged version of the paper presented at the Federal-Provincial Committee on Demography meeting held on November 26-27, 1985, Ottawa, Canada.

² F. Ahmad, R. Chow, O. DeVries, A. Hashmi and M. Marcogliese, Alberta Bureau of Statistics, Alberta Treasury, Sir Frederik W. Haultain Building, 9811-109th Street, Edmonton, Alberta, Canada T5K 0C8.

The reporting unit of the AHCIP file is the registration. Each registration may contain up to twenty-five individuals; one registrant (usually the person who pays the premiums) and up to twenty-four dependents. There are currently about 1.7 million active registrations accounting for roughly 2.6 million individuals. In addition, the file is historical and includes all individuals ever covered under AHCIP since its inception in 1969.

The file is processed through four phases.

- a) Edit-notes and/or corrects errors according to edit check criteria.
- b) Purge-uses the edited raw data file and selects active individuals.
- c) Consolidation-matches postal codes between the purged file and the Bureau's Postal Code Translator File (PCTF) and attaches the geographic reference information to the AHCIP records.
- d) Aggregation-takes the consolidated file and aggregates males and females by single years of age for each postal code. This reduces the number of records/individuals from approximately 2.6 million to fewer than 120,000 and significantly reduces the subsequent systems processing costs.

The aggregated file is used for the production of age and sex counts by any geographic area definable through the 60,000 PCTF Alberta codes.

2.2 Evaluation of the Counts Data

To evaluate the health care counts data, Census of Canada population figures for 1976 and 1981 were used for comparison. The 1981 AHCIP records were considered to be more accurate than the 1976 file, therefore, the evaluation relied more heavily upon the 1981 census comparisons. Also used as a second basis of comparison were municipal censuses data, even though these data generally were not considered to be as reliable as Canada Census figures. The municipal censuses, however, provided insight into the magnitude of the variations as well as the relative distributions of age, sex and trends (growth or decline) over time. An additional source of comparison was intercensal population estimates prepared by the Bureau and by Statistics Canada.

Basic findings:

- a) On a provincial basis, AHCIP counts overestimate both Canada Census and total municipal censuses figures by about 3.5% to 4.5%. Age and sex distributions are more accurate and the correlation coefficients indicate consistency of trends (over/under estimates) over time.
- b) At the census division (CD) level, AHCIP counts varied from Canada Census figures from -2.6% to 9.7% (see Table 1). Comparisons with intercensal population estimates indicated a similar variance. As with the provincial level data, age and sex distributions and the trend consistency proved highly reliable.
- c) At the census consolidated subdivision (CCSD) level, for fifty of the seventy-one CCSDs, health care data were within $\pm 10\%$ of the Census counts. The largest discrepancy was -56.5% (Municipal District 135).

Most problem areas had major urban centres located close to the county, municipal district and improvement district boundaries. No specific anomalies were found when testing the age and sex distributions, although relationships were not as strong as with the province and the census division levels.

- d) At the census subdivision (CSD) level, preliminary figures showed discrepancies between the AHCIP counts and 1981 Census data ranged from -100% to +955%.

Consequently, the twenty-eight largest areas of over 5,000 in population were used at the CSD level. The six largest CSDs (Edmonton, Calgary, Lethbridge, Medicine Hat, Red Deer and St. Albert) displayed overcounts ranging from 3% to 9%. Eight other CSDs differed up to $\pm 20\%$, while sixteen showed somewhat greater than $\pm 20\%$ variation. Again, no specific age and sex distribution anomalies were detected, although discrepancies were greater than those at more aggregated levels. As well, twenty-seven of the twenty-eight CSDs indicated high trend consistency.

As the geographic area decreases in size, AHCIP counts become less reliable; age and sex distributions, although less accurate, still remain strong; and trend consistency (counts over time) remain highly correlated with a few notable exceptions. The limitations of AHCIP counts as population indicators primarily can be attributed to one of two main sources: a) the AHCIP administrative procedures/inaccuracies; or b) use of postal codes.

a) *AHCIP Administrative Procedures:*

- 1) As an insurance programme, a chief concern is to supply coverage. Therefore, efforts are directed to getting people onto the system to ensure universal coverage with less effort placed on getting individuals off the system. This has resulted in more people being registered than are actually in the province.

Table 1
Comparisons of Alberta Health Care Counts and Canada Census Data
for Alberta Census Divisions

Census Division	Year							
	1976				1981			
	Census Count	AHCIP Count	Percent Difference Count	Actual Difference Count	Census Count	AHCIP count	Percent Difference Count	Actual Difference Count
1	46,990	45,789	-2.56	-1,201	55,375	55,748	0.67	373
2	96,995	97,229	0.24	234	110,477	111,567	0.99	1,090
3	32,898	33,884	3.00	986	35,652	36,463	2.27	811
4	12,130	12,101	-0.24	-29	12,119	12,038	-0.67	-81
5	35,424	35,656	0.65	232	38,382	38,457	0.20	75
6	524,554	538,432	2.65	13,878	668,682	699,999	4.68	31,317
7	37,866	38,235	0.97	369	40,071	40,359	0.72	288
8	95,384	95,063	-0.34	-321	123,642	124,666	0.83	1,024
9	19,903	21,832	9.69	1,929	21,670	23,338	7.70	1,668
10	67,171	67,168	0.00	-3	78,417	78,532	0.15	115
11	632,909	646,799	2.19	13,890	762,041	796,884	4.57	34,843
12	63,129	62,011	-1.77	-1,118	84,221	86,183	2.33	1,962
13	46,305	47,258	2.06	953	53,701	54,282	1.08	581
14	19,386	21,039	8.53	1,653	24,635	25,991	5.50	1,356
15	106,993	111,678	4.38	4,685	128,639	134,451	4.52	5,812
Unknown ^a		48,462				19,279		
Alberta	1,838,037	1,922,636	4.60	84,599	2,237,724	2,338,237	4.49	100,513

^a Unknown, represent counts without address identifiers.

Source: Statistics Canada 1976 and 1981 Censuses; Alberta Health Care Insurance Plan data, prepared by Alberta Bureau of Statistics, Alberta Treasury.

- 2) Mailing addresses are used rather than residential addresses, which has created difficulties in assigning geographic locations. Discrepancies occur in areas where significant rural populations surround an urban centre and the rural populace pick up their mail in the urban centre. Consequently, most urban areas are overcounted while rural areas are undercounted.
- 3) Incomplete and inaccurate data, especially related to postal codes, make it difficult to produce small-area statistics due to undercounting.
- 4) Time lags in reporting and recording of the data influence counts. Generally speaking, it takes three to six months to get an individual onto the system (birth, in-migrant) but it requires usually much longer to be removed from the active system (death, out-migrant). The lags, however, are difficult to follow and differ substantially depending on the circumstances.

b) *Postal Codes:*

- 1) Postal codes define delivery service areas (where a person gets his mail), not necessarily a residence. This factor limits the accuracy of assigning AHCIP registrations to appropriate geographic areas. In particular, it creates urban-rural split problems, as discussed.
- 2) A six-digit postal code, by itself, is not always enough to determine the service delivery area. A rural route, suburban service, or box number may be required to further specify a more exact location.
- 3) Postal codes have been insufficient, especially in rural areas, to aggregate to appropriate levels. For example, there are approximately 363 census subdivisions in Alberta, but the Bureau's PCTF can derive only 324 of these.

The problems outlined above have precluded the release of AHCIP counts as approximations of actual population. Although the counts were quite good in some areas, in others, they were poor or inconsistent. With the strong relationships between health care, age and sex distributions and those of Canada Census, as well as the consistency of trends over time, the counts have been used in conjunction with the Bureau's population estimation methodology (as discussed in the next section).

3. APPLICATION OF HEALTH CARE COUNTS TO SMALL-AREA POPULATION ESTIMATES

The Bureau has produced intercensal population estimates for Alberta and provincial census divisions for nearly a decade. During this period, various methodologies and data sources have been examined and used to improve the quality of these estimates. To date, significant success has been achieved with the component method using health care counts as input data. These data have been used to derive the age and sex structure of the Alberta population at the provincial and census division level and to produce provincial and census division population estimates. Also, recently, the data have been used to test the applicability in preparing census subdivision population estimates.

3.1 Estimation Methodology

The estimation methodology employed by the Bureau to produce subprovincial population estimates is comprised of two parts. Part one presents the method of estimating migrant population. Part two outlines the method used to develop population estimates.

a) *Estimating Migrant Population Using Health Care Counts*

The Bureau developed data from three administrative files: counts from AHCIP records; births from data supplied by Alberta Vital Statistics; and deaths, also supplied by Alberta Vital Statistics. These sources were used to calculate net migration. Basically for any small area, the growth of health care counts is obtained from the differences in counts between time t and time $t-1$. This residual less the area's natural increase (births minus deaths) calculates the inflow (or outflow) of individuals, i.e., net migration. This procedure is mathematically expressed as:

$$HMIG = [(HC_t - HC_{t-1}) - (B - D)]$$

Where:

$HMIG$ = health care net migration counts between time t and $t-1$

HC_t = total health care counts at time t

HC_{t-1} = total health care counts at time $t-1$

B = total births during time interval t to $t-1$

D = total deaths during time interval t to $t-1$.

This health care migrant population estimate, however, is subject to the same over and under counting difficulties discussed in Section 2. As a result, although this approach would prepare estimates for small areas at the provincial level, these estimates would be less reliable than the provincial migration estimates currently derived using interprovincial flows to family allowance recipients. (The family allowance files are also used by Statistics Canada, which ensures provincial estimates generally are consistent with those produced at the federal level.)

To further improve the small-area migration estimates and to ensure consistency with estimates at the provincial level, should the small areas be aggregated to a provincial total, an adjustment using a ratio distribution was encompassed. With this approach, the ratio of net migration from health care counts for an area over the net migration from health care counts for the province is multiplied by the provincial net migration calculated in connection with the Bureau's quarterly population estimates. Mathematically, the equation is:

$$AMIG_i = \frac{HMIG_i}{HMIG_a} \times PMIG$$

Where:

$AMIG_i$ = adjusted net migration in area i

$HMIG_i$ = health care net migration of counts for area i

$HMIG_a$ = health care net migration of counts for Alberta

$PMIG$ = estimated provincial net migration from Alberta's quarterly population estimates.

This adjusted migration estimate (AMIG) is then used as input into estimating population.

b) *Estimation of Population For Small Areas*

The adjusted estimated net migration (AMIG) for each area is used in an equation using the components of population growth (births, deaths and migration):

$$P_{i_t} = P_{i_{t-1}} + (B_i - D_i) + AMIG_i$$

Where:

P_{i_t} = estimated population in area i at time t

$P_{i_{t-1}}$ = population in area i at time $t-1$

3.2 Evaluation of Small-Area Estimates

Using the above approach, the Bureau has developed population estimates for Alberta's fifteen census divisions and twenty-eight municipalities with populations over 5000. The results, so far, have been promising.

The results of a comparison between 1981 census data and estimates for 1981 prepared with 1976 census figures as a base population using the above described methodology, at the census division level, are presented in Table 2. For thirteen of the fifteen divisions the estimates were within $\pm 2.0\%$ variation compared to the 1981 census. Only the two smallest CDs (9 and 14) showed a five-year deviation greater than 2.0%. The average absolute deviations (i.e., average annual deviations) were no greater than 0.5% for all census divisions.

The twenty-eight population estimates for municipalities were compared to the 1981 census counts, as well as available data from municipal censuses conducted from 1982 to 1984 (Tables 3 and 4). Federal census comparisons showed nineteen estimates of the twenty-eight municipalities had an average absolute deviation of less than $\pm 1.0\%$. Only six municipalities had annual differences greater than 2.0%. Comparisons with municipal censuses conducted between 1982 and 1984, yielded twenty-two instances of deviations within $\pm 1.0\%$, fourteen ranging between $\pm 1.0\%$ and $\pm 3.0\%$, while nine had deviations greater than $\pm 3.0\%$.

In general, the estimation results have been satisfactory and encouraging. The development of AHCIP registrant counts and the component approach employed to estimate population have improved the accuracy of the population estimates produced and opened up possibilities for deriving estimates for user-defined small geographic areas. The Bureau will continue to investigate ways to improve the AHCIP counts (some of which are related to new administrative procedures being incorporated for the AHCIP). Also, the population estimation methodology will be further refined as new data techniques become available.

3.3 Summary of Advantages and Disadvantages of Using AHCIP

Using health care counts in deriving small-area population estimates has a number of advantages and disadvantages.

Table 2
 Comparisons of Canada Census Counts and Alberta Bureau of Statistics
 Population Estimates for Alberta Census Divisions

Census Division	Census 1976	Bureau Estimates ^a			
		Natural Increase ^b 1976-81	Net Migration 1976-81	Growth 1976-81	Population 1981
1	47,000	2,730	6,080	8,810	55,810
2	96,980	6,120	7,190	13,310	110,290
3	32,870	2,310	100	2,410	35,280
4	12,140	490	- 520	- 30	12,110
5	35,460	1,820	790	2,610	38,070
6	524,570	33,860	107,540	141,400	665,970
7	37,820	2,010	- 10	2,000	39,820
8	95,400	6,140	20,860	27,000	122,400
9	19,850	1,040	200	1,240	21,090
10	67,230	1,650	8,550	10,200	77,430
11	632,830	43,880	90,880	134,760	767,590
12	63,130	6,470	16,130	22,600	85,730
13	46,300	2,040	4,320	6,360	52,660
14	19,450	2,200	2,430	4,630	24,080
15	107,010	10,260	10,040	20,300	127,310
Alberta	1,838,040	123,020	274,580	397,600	2,235,640

Census Division	Census 1981	Difference		Average Absolute Deviation
		Number	%	
1	55,360	450	0.81	0.16
2	110,470	- 180	- 0.16	0.03
3	35,640	- 360	- 1.01	0.20
4	12,120	- 10	- 0.08	0.02
5	38,430	- 360	- 0.94	0.19
6	668,680	- 2,710	- 0.41	0.08
7	40,030	- 210	- 0.52	0.10
8	123,690	- 1,290	- 1.04	0.21
9	21,630	- 540	- 2.50	0.50
10	78,390	- 960	- 1.22	0.24
11	762,080	5,510	0.72	0.14
12	84,220	1,510	1.79	0.36
13	53,690	- 1,030	- 1.92	0.38
14	24,650	- 570	- 2.31	0.46
15	128,640	- 1,330	- 1.03	0.21
Alberta	2,237,720	- 2,080	- 0.09	0.02

^a Data are experimental.

^b Natural increase refers to the number of births minus the number of deaths.

Note: Components may not add to total due to rounding.

Source: Statistics Canada 1976 and 1981 Censuses; Alberta Bureau of Statistics Estimates.

Table 3

Comparisons of the Canada Census Counts and Alberta Bureau of Statistics
Population Estimates for Selected Alberta Municipalities

Municipality	Census 1976	Bureau Estimates ^a			Census 1981	Difference %	Average Absolute Deviation
		Natural Increase ^b 1976-1981	Net Migration 1976-1981	Popu- lation 1981			
Airdrie	1,410	580	5,090	7,070	8,410	-15.9	3.2
Brooks	6,340	730	2,370	9,440	9,420	0.2	0.0
Calgary	469,920	30,310	93,760	593,990	592,740	0.2	0.0
Camrose	10,100	150	2,570	12,830	12,570	2.1	0.4
Crowsnest Pass	5,250	40	-410	4,880	7,310	-33.2	6.6
Drayton Valley	4,300	530	1,760	6,590	5,040	30.8	6.2
Drumheller	6,150	20	220	6,390	6,510	-1.8	0.4
Edmonton	461,360	27,900	51,240	540,510	532,250	1.6	0.3
Edson	4,040	510	2,490	7,040	5,840	20.5	4.1
Fort McMurray	15,420	2,900	14,140	32,460	31,000	4.7	0.9
Fort Saskatchewan	8,300	800	2,660	11,760	12,170	-3.4	0.7
Grande Prairie	17,630	1,970	6,300	25,900	24,260	6.8	1.4
Hinton	6,730	760	-820	6,670	8,340	-20.0	4.0
Innisfail	2,900	230	1,930	5,060	5,250	-3.6	0.7
Lacombe	3,890	150	1,210	5,240	5,590	-6.3	1.3
Leduc	8,580	920	3,430	12,930	12,470	3.7	0.7
Lethbridge	46,750	2,070	4,400	53,220	54,070	-1.6	0.3
Medicine Hat	32,810	1,770	6,010	40,590	40,380	0.5	0.1
Peace River	4,840	580	970	6,390	5,910	8.1	1.6
Ponoka	4,640	-10	530	5,160	5,220	-1.1	0.2
Red Deer	32,180	2,300	11,790	46,270	46,390	-0.3	0.1
Spruce Grove	6,910	1,110	4,710	12,730	10,330	23.2	4.6
St. Albert	24,130	2,360	6,670	33,160	32,000	3.6	0.7
Stettler	4,180	500	580	5,270	5,140	2.5	0.5
Taber	5,300	320	410	6,020	5,990	0.5	0.1
Vegreville	4,160	80	860	5,090	5,250	-3.0	0.6
Wetaskiwin	6,750	300	2,440	9,490	9,600	-1.1	0.2
Whitehorse	3,880	600	1,150	5,630	5,590	0.7	0.1
Alberta	1,838,040	123,020	274,580	2,235,630	2,237,720	-0.1	0.0

^a Data are experimental.

^b Natural increase refers to the number of births minus the number of deaths.

Note: Components may not add to total due to rounding.

Source: Statistics Canada 1976 and 1981 Censuses; Alberta Bureau of Statistics Estimates.

Table 4
Comparisons of Alberta Municipal Censuses and Alberta Bureau of Statistics
Population Estimates for Selected Municipalities

Municipality	1982			1983			1984		
	Bureau Esti- mate ^a	Muni- cipal Census	Devia- tion %	Bureau Esti- mate ^a	Muni- cipal Census	Devia- tion %	Bureau Esti- mate ^a	Muni- cipal Census	Devia- tion %
Airdrie	9,450	9,980	-5.3	9,830	10,430	-5.8	10,080	--	--
Brooks	9,640	--	--	9,790	--	--	9,510	--	--
Calgary	614,930	623,130	-1.3	622,510	620,690	0.3	615,140	619,810	-0.8
Camrose	12,880	12,810	0.6	12,970	--	--	13,070	12,750	2.5
Crowsnest Pass	7,490	7,580	-1.1	7,530	--	--	7,350	--	--
Drayton Valley	5,120	4,870	5.2	5,200	--	--	5,310	4,920	7.9
Drumheller	6,660	--	--	6,700	6,670	0.4	6,620	--	--
Edmonton ^b	550,930	551,310	-0.1	557,400	560,090	-0.5	551,140	--	--
Edson ^b	6,110	6,290	-2.9	6,220	--	--	6,080	7,110	-14.5
Fort McMurray	32,930	33,580	-1.9	33,600	34,490	-2.6	35,150	35,350	-0.6
Fort Saskatchewan	12,530	12,460	0.6	12,650	12,470	1.4	12,620	--	--
Grande Prairie	24,650	--	--	24,910	24,080	3.5	25,370	24,410	3.9
Hinton	8,820	8,820	0.0	8,980	8,830	1.8	8,950	8,900	0.6
Innisfail	5,420	5,440	-0.4	5,460	--	--	5,440	5,440	0.0
Lacombe	5,810	5,720	1.5	5,850	5,850	5,950	-1.8	5,850	--
Leduc	12,880	--	--	13,010	--	--	13,290	--	--
Lethbridge ^b	55,440	56,500	-1.9	55,900	58,090	-3.8	57,500	--	--
Medicine Hat ^b	41,070	--	--	41,440	42,270	0.7	41,540	--	--
Peace River	6,080	--	--	6,150	--	--	6,250	--	--
Ponoka	5,310	--	--	5,310	--	--	5,280	--	--
Red Deer	48,450	48,560	-0.2	49,230	50,260	-2.0	50,860	51,070	-0.4
Spruce Grove	11,080	10,780	2.7	11,410	11,310	0.9	11,550	11,570	-0.1
St. Albert	33,170	32,980	0.6	33,740	35,030	-3.7	34,840	35,530	-1.9
Stettler	5,180	--	--	5,220	--	--	5,300	--	--
Taber	6,140	--	--	6,210	--	--	6,360	6,380	-0.4
Vegreville	5,280	5,250	0.6	5,290	--	--	5,390	--	--
Wetaskiwin	9,880	9,900	-0.2	9,990	10,020	-0.3	10,080	--	--
Whitecourt	5,710	--	--	5,840	--	--	5,710	--	--

^a Data are experimental.

^b Annexation took place between 1982 and 1984.

Note: "--" indicates that a municipal census is not available.

Source: Alberta Municipal Affairs, 1982-1984 Municipal Censuses; Alberta Bureau of Statistics Estimates.

Advantages:

- a) AHCIP registration data provides universal coverage of all individuals in Alberta;
- b) Registration lag appears to be random and does not adversely affect distributions or trends of the counts;
- c) Data are available on a timely/frequent basis; and
- d) The file contains some socio-economic information on registrants and dependents (e.g., age, sex and marital status) to enable the production of more than basic population estimates.

Disadvantages:

- a) Residency based on postal codes can lead to some inaccuracies;
- b) AHCIP registrants can leave the system, for example, death and out-migration, without notifying AHMC resulting in overcounts; and
- c) Administrative procedures may cause discrepancies/inaccuracies in the number of Alberta Health Care registrants.

4. CONCLUSION

Our experience with health care development has been very positive. The greatest potential is the use of the counts in a component model to produce estimates for small areas as well as the excellent age-sex distribution ratios and trend consistency. Costs of development of the demographic reporting systems were not considered excessive in light of these benefits. For other provincial agencies contemplating the development of provincial health care files, the Bureau would certainly be willing to discuss its experiences in more detail and make available additional information, such as record layouts and system processing costs.