An Overview of the Strengths and Weaknesses of the Selected Administrative Data Files[1]

RAVI B.P. VERMA and PIERRE PARENT[2]

ABSTRACT

Twelve administrative data files are reviewed to determine if some of them could be used to derive migration data, in case the universality of the currently used family allowance files be limited, as a result of federal legislation.

It is found that none of the twelve files have strengths and weaknesses strictly comparable to those of the family allowance files. Further developments of the Health Care, and to a lesser extent the Old Age Security files are highly recommended.

KEY WORDS: Administrative files; migration; qualitative evaluation.

1. INTRODUCTION

In Canada, both family allowance and income tax files have a wide range of utility in producing the migration and population estimates for the different geographic areas (see Statistics Canada Catalogue Nos. 91-001, 91-210, 91-211 and 91-212). Data from the family allowance files are made available within 2 to 3 months after the reference date. In contrast, income tax data are available within 12 to 15 months after the reference date. However, income tax data provide the estimates of migration flows for the census divisions, and also by age and sex.

In terms of accuracy of population estimates, both family allowance and income tax files are good and they are comparable (see Norris and Standish 1983; Norris 1983; Verma *et al.* 1984; Verma and Basavarajappa 1985). One of the special features of the family allowance and income tax files is the fact that they are national in character. Another feature is that the records contain addresses with the postal codes. Thus, this could provide the migration information for local areas. However, in recent years, there seems to be some possibility that family allowance could cease to be universal as a result of government legislation. For example, coverage might be limited to the lower- and middle-sectors of the population. If this file ceased to be universal, its utility as a migration data source would be very severely limited. Hence, our population estimation activities would be jeopardized which in turn would affect other programs as revenue sharing, involving the annual distribution of $20 billion among provinces.

For this reason, alternate sources need to be explored. An attempt is made here to assess the strengths and weaknesses of some of the selected administrative data files for estimating migration and population for provinces and territories, census divisions, census metropolitan areas and other regions in Canada.

The twelve administrative files are qualitatively evaluated as an alternative to family allowance files. On the basis of their strengths and weaknesses, they are divided into the following three groups:

Major potential files for estimating migration flows
- i) Health Insurance Files
- ii) Old Age Security File

Major potential files used as a symptomatic indicator of population change and net migration
- iii) Hydro Connections
- iv) Telephone Customers
- v) School Enrollments

Other files with limited or uncertain potential for estimating migration flow/net migration
- vi) Driver's License
- vii Building Permits
- viii) Unemployment Insurance Beneficiaries
- ix) Labor Force Survey
- x) Voters' List
- xi) Retail Sales
- xii) Trucking Statistics

1.1 Criteria for Evaluating Administrative Data Files

The assessment of the usefulness of the various administrative data sources for estimating interprovincial and intraprovincial migration is done with respect to ten criteria: universe, coverage, method of determining migration information, types of migration, characteristics of records, reference date/period (and monthly availability), time-lag, historical availability, consistency and computerization (Almond 1982).

The new data source would have high potential if it contains features of the family allowance files, as described in Table 1. The most important criteria are: coverage, timeliness, consistency, monthly or quarterly availability, disaggregation using the postal code or other geocodes. The file or set of files that can meet these standards would probably qualify as replacement source to family allowance.

2. MAJOR POTENTIAL FILES FOR ESTIMATING MIGRATION FLOWS

Health Insurance and Old Age Security files are major potential files for estimating migration flows among provinces, territories and census divisions. Strengths and weaknesses of each of these two files are presented below.

2.1 Health Insurance File

Health Insurance is a provincial responsibility. Each province thus keeps a file of people eligible for the program. All residents in the province (including newly arrived immigrants and foreign students) are covered by the provincial insurance, except for RCMP and Armed Forces personel, and for the federal penitentiary inmates, covered by the federal government. Everybody who establishes its residence in a province must fill out a proper application, from which data on in-migrants, by province of origin, and on international immigrants can be compiled. Virtually complete coverage, monthly availability, minimal time lags and information usually detailed by age, sex and family composition of the migrants are the main strengths of the files. There should also be a very strong incentive for interprovincial migrants to apply to the program. Consequently, migration data should be reliable.

Table 1

Description of the Administrative Data Files Currently
Used to Derive Migration Data in Canada

Criteria	F.A. Monthly Statistics Report	F.A. M0024 File
Universe	Children in payment of F.A.	Children entitled to F.A. (as opposed to "in payment")
Coverage	25% of total population in 1984. Virtually 100% of children aged 0-17	Similar to F.A. Monthly Statistics
Method for determining status	Compilation of change of address notices	Compilation of change of address notices
Types of migration	Interprovincial migration, by province of origin and destination	Similar to F.A. Monthly Statistics, plus international migration
Characteristics	Origin-destination. Age: total 0-17 only. Family size: refers to the number of children in family	Origin-destination. Age: year and month of birth. Language (E or F). Type of account (regular, foster, foreign or agency)
Reference date/period	Month: refers to the amount of information processed during that time	Month: refers to month of real migration
Time-lags	Data processed a given month is available at the end of that month and refers to migration of approximately two months earlier	Data released semi-annually. Contains information on last six months' migration. Available approximately 3 months after end of semi-annual version
Historical availability	January 1974 onwards for children migration data. From 1947 to 1973, only information on family migration was available	December 1977 to present
Consistency	OVER TIME: change in 1974. Slight, problems since 1980. AMONG PROVINCES: good	OVER TIME: generally good. Slight problems since 1982. AMONG PROVINCES: problems with Ont. since 1982. Slight problems with Nfld. and N.S. in 1983
Level of computerization	In provincial offices, yes. But data are sent to Health and Welfare Canada central office on print-outs	Yes, well developed

Note: F.A. is an abbreviation for Family Allowance.

Table 1

Description of the Administrative Data Files Currently
Used to Derive Migration Data in Canada (Concluded)

Criteria	F55 Program	Revenue Canada File
Universe	Children entitled to F.A.	Tax filers (must have filed two consecutive years)
Coverage	Similar to F.A. Monthly Statistics	Filers matched two consecutive years total up to approximately 75% of population aged 18 and over
Method for determining status	Symptomatic indicator	Comparison of the return address of matched returns. Correction is brought for unmatched returns
Types of migration	Net migration	Intraprovincial, Interprovincial and International
Characteristics	Number of children by geographical area. Age	Origin-destination. Broad age-sex group
Reference date/period	Twice a year, as of June 1, and December 1 (refers to the number of children entitled to F.A. as of these dates)	Year: refers to the period between two consecutive filings, i.e. approximately the April-March period. Used as June-May data
Time-lags	Available approximately three months after reference date	Preliminary available 6-8 months after end of reference period. Final data, 10-12 months
Historical availability	December 1977 to present, with entitlement information. Available back to 1974 for children in payment	1966-67 to present
Consistency	Generally good	Changes in tax laws results in-change in coverage and in number of matched returns over time and provinces
Level of computerization	Yes, well developed	Yes, well developed

There are also, however, certain weaknesses. The fundamental limitation is that neither Ontario nor Quebec can provide migration data. In the latter case, however, new developments are promising, but for Ontario, nothing is expected. Unless a special source is derived for Ontario, this would compromise the high potential of this file. There could also be a consistency problem, since each province independently administers its file.

At the subprovincial level, migration could also be derived since the Provincial Health Care offices should be informed of any change of address. In the facts, however, all changes are not known.

Health Care files could also be used in regression estimates, especially in provinces that run periodic address checks to clean the file and count only the desired population.

2.2 Old Age Security Records

Health and Welfare Canada is responsible for the administration of the Old Age Security file. Canadian residents aged 65 and over who totalled a sufficient number of years of residence in the country are eligible. It represents approximately 10% of the total population. Coverage among eligible people is virtually universal. Also the financial incentive to report change of address is very strong. Another strength of the file is its timely availability. Information on people moving in a given month is compiled and received by Statistics Canada two or three months later. Finally, Old Age Security, being a federal program, provides comparable data for the provinces; even if the information is compiled by provincial regional offices, they all follow the same procedure.

The main shortcoming of this file for migration estimates purpose is the fact that it refers to a small portion of the population (varying from 7.3% in Alberta to 12.2% in P.E.I.), the elderly moreover showing a rather different migration pattern than the rest of the population. Unlike child migration, which can obviously be related to adult migration and then be blown up to estimate total migration, no similar efficient method could be developed to estimate total migration from the Old Age Security file. Although this could not be used as the main source for migration estimates, however, this file could provide a very interesting estimate of the elderly migration.

3. MAJOR FILES USED AS SYMPTOMATIC INDICATORS OF POPULATION CHANGE AND OF NET MIGRATION

Data from some administrative files could be useful for generating total population estimates. For example, School Enrolments, Hydro Connections or Telephone Residential Customers could be used in regression techniques as symptomatic indicators (see McRae 1985 for an application of Hydro Connections to population estimates). This method and the corresponding sources are generally used for producing small area population estimates, but if no other technique gives valuable estimates at the provincial level, these sources will be seriously considered.

3.1 Hydro Connections

Electric companies keep files of their customers. Information on the type of account (residential, commercial, farm, ...) and the address and postal code of the customer are available. Coverage of residential households is virtually complete. Sometimes there is only one file for the province, but sometimes 2 companies (Manitoba and Newfoundland) or even more (B.C. and Ontario) provide the electric facilities within the province. In most provinces data can be produced for the entire territory, as of any date and within a short time-lag, but for a few provinces it can be hard to get the data. The main weaknesses of the files are of two kinds. In addition to the previously cited problem there may also be slight inconsistencies due to the difference between provincial definitions of residential households (since

it responds to administrative criteria), and even within one province, if more than one company is involved. Nevertheless, Hydro Connections could be a very good source for population estimates. As a matter of fact, they were tested in British Columbia, where population estimates for municipalities and school districts were produced. The results were good. This method could also be tested and eventually be extended to provincial level estimates, if need be.

3.2 Telephone Companies

In Canada, telephone services are insured by 14 major telephone companies. Information on customers with residental lines (address and postal code) is available. The situation is roughly similar to that of the Hydro Connections files. Data can usually be obtained for specified dates within a rather small delay and the coverage is fairly high. Here again, more than one company may serve a given province, and also, a company may serve more than one province. Despite the fact no estimate based on Telephone files has been tested in Statistics Canada, it is felt that they have the potential to produce good results.

3.3 School Enrollments

Each provincial government maintains a computerized file on students enrolled in its primary and secondary school system, containing information on school addresses with the postal code and on the number of students, by age and grade. Information on the number of students refers to September 30 and is available between 4 and 10 months after the reference date, the time-lag varying by province. The coverage of students is also very good.

There are some weaknesses associated with this file. For example, its annual character plays against its use for producing quarterly estimates. Also its date of reference (September 30 instead of June 1), along with the up to 10 months delay is another handicap. At the subprovincial level, finally, it often can be observed that some students reside in a given administrative region, but go to school in a different one. This also could affect the quality of the estimates. It should be pointed out here that the school enrolment data, at one time, were used in Statistics Canada (and also by the U.S. Bureau of the Census, using a component method developed by them. See U.S. Bureau of the Census 1973, Chap. 23, p. 51); the deviations associated with that method were much higher than those with other methods. In case no other file could provide adequate population estimates, regression estimates with that file could produce acceptable results, at the provincial level at least.

4. FILES WITH LIMITED POTENTIAL

4.1 Driver's License

Each province maintains a file listing persons aged 15 (or 16, or 17) and over licensed to operate a motor vehicle. Using the provincial files, migration could be estimated in two ways: 1) compilation of changes of driver's address for estimating flows of migration; and 2) as a symptomatic indicator of the population change, through the variation of the number of people licensed in a given region. Currently, Ontario uses drivers' licenses to estimate intraprovincial migration, but very few other provinces could provide migration flow information, especially at the subprovincial level. In order to do so, it would require too much work and consultation with the provincial ministry. Despite the fact that drivers are forced by the law to report their change of address, not all do so, and no sufficiently detailed statistics are available.

The driver's license file could also be used in regression techniques. Data available at any specified date in many cases and short delays are positive points. However, coverage and consistency concerns might affect the quality of the data. For example, 83% of adults in Saskatchewan own a driver's license, as against 73% in Manitoba, 85% of males and 62% of females accounting for the latter province's average proportion. In addition, the poor, comparatively recent immigrants, and Indians and residents of remote communities in the

North have below average rates for holding licenses (Stock 1981, p. 44). For estimate purposes, it is often preferable to have a 100% coverage of a small subpopulation (e.g. children) than an 80-85% coverage of a large subpopulation (e.g. adults), especially if the coverage is selective with respect to migration. Although it does not necessarily make a file inappropriate for estimate purposes, it affects its potential.

4.2 Building Permits

Statistics Canada collects new building permits for cities and rural areas in Canada. On average, the coverage rates vary between the urban (98.5%) and the rural (62.5%) areas. The building permit data are available on a monthly basis at the census division level. These data could be also used as a symptomatic indicator of the population change. However, one of the weaknesses of the building permit data is the fact that they refer to the date of permit. Due to this, it is not certain whether the building has been constructed and also, whether it has been occupied. Another weakness is the fact that the number of permits issued is not necessarily directly related to population change, especially in the case of a decreasing population.

Thus, the use of building permit data also seems to be limited in estimating population for the different geographic areas.

4.3 Unemployment Insurance Commission

The Unemployment Insurance Commission keeps a list of the beneficiaries of the program. A 10% sample of this file has been developed to produce statistics and it could provide migration information. However, this file could hardly be used to estimate migration in Canada. First, a 10% sample of unemployed corresponds to less than 1% of the population. From such a small subpopulation, flows of migrants between provinces could not be derived. Also the non-representativity of that sample (young adults representing a good part of non-employed) calls for suspicion concerning the migration data from that file.

The Commission also maintains a file of wage earners who are contributing to the Unemployment Insurance program. However, no in depth analysis of this file has been done.

4.4 Labour Force Survey

In 1982, Statistics Canada conducted a sample survey of 56,000 households in Canada. The civilian non-institutionalized population aged 15 and over, included in the sample, residing in all provinces were asked a question on their migration history of the past 5 or 6 years. Other valuable information is also available. However, its very small sample (approx. ½% of the population) and the fact that the survey was conducted only once eliminates the Labour Force Survey as migration estimates source.

4.5 Voters' List

Data on voters are generally available in Canada. Federal and provincial election lists could easily be obtained while obtaining municipal lists would necessitate more work. Those lists give information on the number of canadian citizens aged 18+ (landed immigrants are included at the municipal level only). They cover an average 90-95% of the target population. The main shortcoming of that source is that it is not available at regular intervals. Federal and provincial lists are made for elections about every 4 years at dates that are not useful for estimation purposes. It thus seems pointless to consider voters lists.

4.6 Retail Sales

Data on retail sales are collected by Statistics Canada on the basis of sales figures from large stores and from a sample of smaller businesses. These data are collected on a monthly basis and they are made available 3 months after the reference date. These data could be used as a symptomatic indicator of the population change. However, the utility of this data

set seems to be limited in the case of population and migration estimations. This could be due to the fact that retail sales are heavily affected by the economic fluctuations which may not accurately reflect changes in the size of population.

4.7 Trucking Statistics (Moving Companies)

Statistics on a sample of five major moving companies are available in Canada. They cover about 90% of all moves. The interprovincial migration flow could be assessed by weighing the number of reported moves between two different provinces/territories. However, trucking statistics are seriously affected by a time-lag of two years or more.

5. CONCLUDING REMARKS

In this report, an overview of strengths and weaknesses of twelve administrative data files has been presented in order to make recommendations for selecting an alternative data source to the family allowance files. It has been found that there is no file with strengths and weaknesses strictly comparable to those of the family allowance files. However, if the family allowance files cease to be universal, one could suggest the following recommendations:

– Continue further developments in the use of the provincial health insurance file and the Old Age Security records of the federal government in order to produce the total population and migration estimates on a quarterly basis;

– Examine the quality of annual population estimates for the provinces and territories, produced by the Component Method II using the migration estimates from the provincial school enrollment data files; and

– Test the accuracy of the provincial administrative data files (health insurance files, hydro connections, telephone companies and driver's licence) as symptomatic indicators of the population change and the residual net migrants for sub-provincial areas (census divisions and census metropolitan areas in Canada).

REFERENCES

ALMOND, M.M. (1982). An inventory of sources of Canadian migration data. Working Paper, Demography Division, Statistics Canada.

McRAE, D.G. (1985). Use of hydro accounts in the regression population estimates model in British Columbia. Presented at the Federal-Provincial Committee on Demography, Ottawa, Canada.

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

NORRIS, D.A. (1983). New sources of Canadian small area migration data. *Review of Public Data Use*, 11-25.

STATISTICS CANADA (Quarterly). *Estimates of Population for Canada, Provinces and Territories*. Catalogue 91-001, Ottawa: Minister of Supply and Services Canada.

STATISTICS CANADA (Annual). *Postcensal Annual Estimates of Population by Marital Status, Age, Sex and Components of Growth for Canada, Provinces and Territories*. Catalogue 91-210, Ottawa: Minister of Supply and Services Canada.

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

STATISTICS CANADA (Annual). *Postcensal Annual Estimates of Population for Census Divisions and Census Metropolitan Areas (Component Method)*. Catalogue 91-212, Ottawa: Minister of Supply and Services Canada.

STOCK, R. (1981). *Migration Estimates from Current Administrative Files: Data Sources and Methodologies*. Canadian Plains Research Center, University of Regina.

U.S. BUREAU OF THE CENSUS (1973). *The Methods and Materials of Demography*. Washington, D.C.: U.S. Government Printing Office.

VERMA, R.B.P., BASAVARAJAPPA, K.G., BENDER, R.K. (1984). The regression estimates of population for sub-provincial areas in Canada. *Survey Methodology*, 9, 219-240.

VERMA, R.B.P., BASAVARAJAPPA, K.G. (1985). Recent developments in the estimation of population for small areas in Canada by regression. Presented at the International Symposium on Small Area Statistics, Ottawa, Canada.