

*THE USE OF INDIVIDUAL ADMINISTRATIVE RECORDS
FOR SOCIAL STATISTICAL PURPOSES IN CANADA*

John Leyes, Statistics Canada

I. INTRODUCTION

Statistics Canada has exploited administrative records for their statistical potential for a long time. Customs records have been used for external trade data. Institutional records have been the basis of data derived on births, deaths, marriages, public safety, education, and hospitals. A variety of administrative data on the economy has also been exploited for statistical applications.

In the domain of social statistics, Statistics Canada launched a new initiative in 1979 to exploit the statistical potential of administrative records that are maintained on individuals. Since then, Statistics Canada has begun to tap the statistical potential of these records on an experimental basis. (Social statistics are defined to be limited to data on households, families, and individuals.) In particular, this initiative has been directed at the development of new data for (a) small areas that conform to standard statistical areas and (b) non-standard geographical areas specified by users of small area data. (For a discussion of the problems, considerations and potential of using mailing addresses, see Kopustas (1983); Statistics Canada, Section 6.0, 1983.)

This paper, then, has been written to describe some administrative data development work undertaken by Statistics Canada in the domain of social statistics over the last four years.

II. ADMINISTRATIVE RECORDS IN CURRENT USE

Three basic sets of records have been used. Each of these is described briefly below. (A more complete explanation of each is contained in Statistics Canada, 1983.)

1. Tax Records

Statistics Canada has dedicated a substantial portion of its developmental effort to the manipulation of the 100 percent personal income tax records.

These records include an array of data: postal code, demographic variables, sources of income variables, deductions from income, and taxes paid. The files received from Revenue Canada-Taxation have a record length of approximately 500 bytes. This record is reduced to about 325 bytes for processing purposes.

2. Family Allowance Records

A second source of administrative data is the family allowance files. At the present time, Statistics Canada has access to only a portion of the entire family allowance file. In particular, the data in this file are restricted to the postal code, month and year of birth of each child, language of correspondence (english or french), amount of monthly payment to each family, and reason for addition or deletion from the file (including birth, death and inter-provincial migration).

The family allowance program is a universal program for all eligible children. Children not eligible include the children of illegal immigrants and the children of foreign government officials. Well in excess of 99 percent of the children under age 15 are covered by this program. For 16 and 17 year olds, the coverage is lower since some of these teenagers earn income.

3. Unemployment Insurance Records

A third source of data has been the unemployment insurance records. These records include an array of data on reason for eligibility for unemployment insurance benefits, limited demographic detail, a geographic identifier (the postal code), and beneficiary detail (eligibility, earnings status, occupation code). (For additional information, see Statistics Canada, Section 6.0, 1983.)

III. SUBJECT MATTER APPLICATIONS

In general, data derived from the three sets of administrative records described in Section II can be summarized into four subject matter areas.

1. Labour force

The personal income tax data have considerable potential for estimating employment income and gross labour force. Also, the unemployment insurance files hold the potential for the development of small area indicators of unemployment as well as source data on reasons for unemployment along with a limited amount of information on occupation. (For additional information, see Norris and Mussely, 1983; and Leyes, et al., 1982).

2. Income data

The personal income tax file contains a wealth of data on the sources of income of individual Canadian residents. These records have been used primarily to produce data on sources of income by variables such as age, sex, marital status and geographical location.

3. Demographic data

All three sets of administrative records possess some demographic detail.

A. The tax records can be used for data on age, sex and marital status. Also, by using the dollar value of exemptions from the tax file, it is possible to impute the number of dependents.

B. The family allowance file contains accurate data on the birthdate of children eligible for this benefit. There is not, however, any information on the sex of the children nor is there any information on parents.

C. The unemployment insurance file includes a limited amount of data on beneficiaries -- age, sex.

4. Migration data

It has been demonstrated that both the personal income tax records and the family allowance records can be used for estimating inter-area migration. And in the case of the latter, interprovincial migration data have been derived from these records since the 1950's for estimating the provincial population in the years following each census. (See Norris et al., 1982; Hache and Norris, 1983).

IV. CONSTRAINTS ON THE USE OF INDIVIDUAL ADMINISTRATIVE RECORDS

In this section, a brief overview is provided of the legal and other restrictions on the use of administrative records for social data.

First, Canada has legislation on privacy. Under the terms and conditions of the "Privacy Act," administrative records are to be used in processes that directly affect individuals or for other purposes approved by an Act of Parliament. Since Statistics Canada has specifically been given authority to obtain administrative records for statistical purposes, the Privacy Act has not become a major obstacle in Statistics Canada's use of administrative records for social or other statistical purposes.

Secondly, there are restrictions on the use of administrative records because of the commonly held view that the use of administrative records, for purposes other than those for which they were originally intended, represents an unwarranted intrusion into personal and private matters. From a somewhat pragmatic perspective, perhaps the conventional understanding will evolve in the future and Canadian residents will come to

accept the use of administrative records for statistical purposes. This will, of course, require that these people must accept in principle that statistical usage of data is different from administrative use of data.

V. THE ROLE OF USERS IN THE DEVELOPMENTAL PROCESS

Administrative records, practically speaking, contain a large amount of data. These data, moreover, can be tabulated for an almost infinite number of small areas, using the existing 600,000+ postal codes. Since a considerable amount of local knowledge and expertise are required to assess the reasonableness of these data, experimental tabulations were produced for distribution to "expert" users on an *ad hoc* basis.

In other words, users were actively sought out and asked to provide their "expert" judgement on the data for their geographical areas of interest.

The potential user community for administrative social data was defined to be similar to the community that uses data derived from the census of population. That is, since administrative records hold the potential for developing small area data, and since the census is the main source of small area data, it seemed reasonable to conclude that the two sets of users would be similar.

With these basic notions identified, data were produced and made available. The data were described as *experimental* and users were asked to comment on (1) the general appropriateness of the tabulations; (2) the extent to which the data seemed reasonable compared to their knowledge of the respective areas in which they were working and about which they would be expected to have local knowledge; and (3) whether they would be interested in obtaining these kinds of data in the future.

In general, and not surprisingly, this strategy was not very successful. Many users find themselves pulled from many different directions simultaneously. Although they have data needs, the data required vary from one day to another. In many cases, they are not generally aware of what their small area data needs may happen to be. Furthermore, users of small area data have come to consider small area data to be identical with census of population data. In many cases, they are reluctant to provide a serious review of data derived from an unfamiliar source such as administrative records.

Another important discovery in approaching users of small area data was their general unwillingness to provide any consideration to data described as *experimental*. Indeed there is irony in this finding since many of these data users previously expressed a desire for more frequent small area data. They seemed unwilling, nevertheless, to accept a role in evaluating at least annual small area data from administrative records.

In spite of the negative results in stimulating a dialogue with users, some feedback was obtained. Some reflected the basic difficulties arising from a geographical system based on mailing addresses rather than residential locations. A further finding resulted from the recognition that many users find themselves in a position of preparing scenarios, or developing alternatives for decision on the basis of no statistics, or on statistics that are out of date, or on the basis of statistics for geographical areas that are inappropriate for their requirements. These users welcomed the opportunity to use administrative social data for their small areas of concern, even though these data were based on new concepts and definitions.

Thus, and in summary, it has become clear that the role of users in evaluating administrative social data has been, and will likely continue to be marginal at best. In most instances, they have neither the time nor the inclination to invest resources in data that are essentially *experimental*. Although they often possess the local knowledge for providing critical judgements about the general utility of these data, it is diffi-

cult to obtain their cooperation in obtaining these insights.

VI. SOME USES OF ADMINISTRATIVE RECORDS

There is a growing number of uses of individual administrative records for statistical applications. The following examples are illustrative and do not represent a comprehensive listing of the many applications to date.

A. Individual tax records have been used to evaluate government programs (including the income of fishermen and the elderly).

B. Administrative social data are also being used for demographic studies. The migration of children, as defined by family allowance data, have been used since the 1950's to estimate interprovincial migration for the Canadian population. At this time, the principal application of tax data has been for the estimation of inter-census division migration for the 265 census divisions in Canada, units that are similar to counties in the United States. Furthermore, these migration data are being used in the population estimation methodology in the post-1981 census period.

C. The personal income tax file and the unemployment insurance files are useful sources of data on the labour force -- including employment income, gross labour force and unemployment.

D. In Canada a number of transfer payments are subject to taxation -- unemployment insurance benefits, family allowance benefits, pension benefits. Although these sources of income fall short of the total transfer payments received by individual Canadians, these sources of income have been used to estimate economic dependency on transfer payments for small areas. (Also, both the family allowance and unemployment insurance files offer the potential to tabulate transfer payments by small geographical areas, although this has not yet been attempted.)

E. The individual tax records have also been investigated with respect to their potential utility for conducting marketing studies and developing marketing strategies geared to direct mail advertising. And although this would seem to hold considerable long-term promise, it has not been possible to obtain demonstrable evidence that individual administrative records are useful for direct mail marketing. (It has not been proposed that administrative records be used directly to generate mailing lists. This, quite clearly, would be a violation of the secrecy provisions of the Statistics Act. Rather, it has been proposed that administrative social data be produced for small areas and that these data be then used to target direct mail advertising programs.)

F. A second type of marketing application is under development at this time. It involves the generation of aggregated data for a federal agency engaged in the provision of management advice and loans to small businesses. A pilot project is currently underway to develop small area data from the tax files, to mount the small area data on diskettes for micro computers, and to enable either management consultants or prospective small businessmen to access small area data for basic small business decisions. The data would, in this environment, be designed for decisions related to the market for small businesses that might include boutiques, shoe stores, restaurants, and similar kinds of retail services that are typically found in urbanized areas.

G. The tax file has also been used to define a sample frame, a frame that was developed for a survey designed to assess the incomes and pension plans to which Canadians have been contributing over time.

H. Statistics Canada also has a 10 percent longitudinal file of taxfilers for the period 1967-80. This file contains an array of data on the geographical location of individual taxfilers, their incomes, their deductions and exemptions, and their taxes paid over this period. The file has been used for several applications in recent years. In two instances, users derived data to assess the return migration phenomenon (i.e., a taxfiler who would leave, for example,

one province and return in a subsequent year would be classified as a return migrant). Secondly, data from the longitudinal file have been used for the development of limited time-series data on segments of the population. For example, it has been possible to produce annual estimates of the income of taxfilers by demographic characteristics such as age, sex and marital status. Finally, limited studies have been undertaken with the longitudinal data to assess the income patterns of migrants with non-migrants.

I. The family allowance data have been used by individual school boards to obtain data on the ages of children for the administration of their school districts. These data have been used for planning school consolidations, instructional staff requirements, and new school buildings; preparing long-range projections of the pre-school age population; and for defining alternative school boundary arrangements. In addition, data from the family allowance records have been used for day care planning.

J. Statistics Canada has access to a monthly file of unemployment insurance beneficiaries. It has been possible to develop an experimental small area unemployment indicator based on counts of unemployment insurance beneficiaries. These data have been reviewed by users as a potential source of data. No formal uses, however, of these data have been identified thus far.

VII. USERS' PERCEPTIONS OF ADMINISTRATIVE SOCIAL DATA

There has been remarkably little feedback from users on their uses of administrative social data. On the basis of some sample enquiries, it has been concluded that the potential market for administrative social data are circumscribed by several notable considerations:

A. Some users do not look favourably on the use of data for which there is a long list of caveats or detailed statements of limitations, many of which are differences with traditional data sources rather than because the data have inherent limitations. (For example, administrative data may often understate the absolute magnitude without distorting the relative relationships among the data themselves.)

B. Some users have expressed considerable unwillingness to use data that have been designated as experimental and for which there exist no objective criteria for determining whether and if this designation will ever be eliminated. The word experimental has been too strong and the caveats too explicit for them to feel comfortable.

C. One of the more interesting objections to small area administrative data has arisen because these data provide insights that were previously unavailable. In the absence of these data, it was possible to formulate policies without small area data. In fact, it was suggested that small area data might be politically undesirable since it would increase the information of interested parties, to the detriment of officials in their planning and policy-making activities.

D. Users have also expressed general dissatisfaction with the lack of conformity of the definitions and concepts of data derived from administrative records with the definitions and concepts of data derived through direct household collection programs. For example, in using the personal income tax records, it is not possible to estimate directly a concept that parallels family income.

VIII. THE FUTURE OF USING ADMINISTRATIVE SOCIAL DATA

In the short to medium term, the future of administrative social data is limited. The data, in general, have not been well received by the user community. There is, moreover, a highly pervasive and inculcated belief that small area data and census data are synonymous — a belief that has evolved over many generations of users of small area data. Unless, therefore, there is a major, unforeseen and exogenous force to alter this

widely held belief, it is unlikely that administrative social data will become a widely used and acceptable source of small area social data.

REFERENCES

1. Canada, Privacy Act.
2. Canada, The Statistics Act. 1981.
3. Hache, Jean-Guy, Norris, D., "New Sources of Data on Canadian International Migration," American Statistical Association Proceedings 1983, Section on Survey Research Methods.
4. Kopustas, Nelson, Norris, D., Leyes, J., "Geocoding Administrative Data," American Statistical Association Proceedings 1983, Section on Survey Research Methods.
5. Leyes, John, "Developing Administrative Data: Research and Development Prototypes Underway in Statistics Canada." Statistics Canada (unpublished paper, mimeo), August 1980.
6. Leyes, John, Bobet, E., Radley, L., "The Use of Unemployment Insurance Records to Derive an Unemployment Indicator," American Statistical Association Proceedings 1982, Section on Survey Research Methods.
7. MacDonald, W. Stephen, Wood, K.S., Coffey, W.J., Analysis of Small Area Data: Concepts and Methodology, Institute of Public Affairs, Dalhousie University Halifax, N.S. (mimeo). Paper presented to Canadian Regional Science Association, June 1983.
8. Norris, Douglas, Britton, M., Verma, R., "The Use of Administrative Records for Estimating Migration and Population," American Statistical Association Proceedings 1982, Section on Survey Research Methods.
9. Norris, Douglas, Leyes, J., Kopustas, N., "The Development of Social Data from Administrative Records: The Canadian Experience," American Statistical Association Proceedings 1981, Section on Survey Research Methods.
10. Norris, Douglas, Mussely, A., "A Measure of Annual Work Experience Using Individual Income Tax Records," American Statistical Association Proceedings 1983, Section on Survey Research Methods.
11. Statistics Canada, "Overview of and Guide to the Use of Administrative Social Data," Administrative Data Development Division, Statistics Canada (mimeo), March, 1983.
12. Yakimishyn, M., "Are You Planning for Children," Education Manitoba. Department of Education, Province of Manitoba, pp. 7-9, October 1982.