THE UTILIZATION OF ADMINISTRATIVE RECORDS
FOR STATISTICAL PURPOSES

L.E. Rowebottom

This paper advances the case that administrative records are a powerful source of statistics and in support of this conclusion provides an overview of the extensive utilization in Canada of administrative records for statistical purposes. The paper discusses recent developments and the changing environment which are seen as major determinants of both the creation of administrative data bases as well as their utilization. The capabilities of the computer, combined with the extensive demand for statistics and the limited financial resources available to meet that demand, are seen as combining to lead to more extensive use of administrative records. A variety of problems associated with the use of administrative records is specified and the development of strategies to meet these problems and permit utilization of administrative records is described. Recent developments in Canada intended to support the use of administrative records are indicated.

1. INTRODUCTION

This paper, on the utilization of administrative records in Canada for statistical purposes, comprises a survey of:

- a history of such utilization;
- reasons why administrative records are a powerful source of statistics;
- recent important developments and the changed environment within which such utilization must take place;
- problems accounting for the relatively slow pace of developments;
- prerequisites of effective and extensive utilization;
- ways being pursued to more fully utilize administrative records.

1 Adapted from a paper prepared for the XIII Session of the Committee on Improvement of National Statistics (COINS) of the Inter-American Statistical Institute, November 1977.

Administrative records are defined for purposes of this paper as those generated as a result of carrying out a wide range of functions such as registering, insuring, educating, taxing, and other benefits or penalties so that decisions which affect individuals may be taken by departments, agencies, institutions and other organizations of government. Administrative records contain information not only about individuals, e.g., admission to a hospital, but also about organizations and institutions of government, e.g., amount spent on construction or about businesses, e.g., gross business income, employment, etc. In many respects, they are the public sector equivalent of the accounting records of businesses which form the basis of various economic statistics. In much the same way, administrative records are created for non-statistical purposes but may be used for statistical purposes.

2. STATISTICAL USES OF ADMINISTRATIVE RECORDS IN CANADA

Many countries have long and effective records of utilizing administrative files for purposes of compiling both economic and social statistics. Thus, in Canada, registrations of birth, death, and marriages have long formed the basis of vital statistics, and custom invoices have similarly constituted the basis of external trade statistics. Examples of other administrative files and the statistics derived from them are as follows. This listing is not exhaustive but makes clear that substantial portions of the national statistical system are in large part dependent upon administrative records.

<table>
<thead>
<tr>
<th>The Administrative Records</th>
<th>The Statistics Derived From Them</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital admission and discharge records</td>
<td>Statistics on hospitalized illness, cause of hospitalization, days of hospital care, by age and sex. They comprise the only national measure of morbidity.</td>
</tr>
<tr>
<td>The Administrative Records</td>
<td>The Statistics Derived From Them</td>
</tr>
<tr>
<td>------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Student enrolment</td>
<td>Numbers of students by age, sex and grade, as well as level of education attained and &quot;drop-outs&quot; before graduation. For university students, measures of enrolment by discipline.</td>
</tr>
<tr>
<td>Personnel records of teachers</td>
<td>Salaries of teachers by qualification, years of experience, discipline, and other characteristics.</td>
</tr>
<tr>
<td>Accounting records of boards, commissions, departments and agencies of government</td>
<td>For federal, provincial and municipal governments, statistics of income by source, expenditure by object and function, assets and debts.</td>
</tr>
<tr>
<td>Offenses reported to police and persons charged by police</td>
<td>Statistics of crime by type of crime, by age and sex of persons charged, for Canada, the provinces and major municipalities.</td>
</tr>
<tr>
<td>Personal income tax returns</td>
<td>Statistics of income by geographic area, occupation, age sex, level of income, marital status and number of dependents. Statistics concerning businesses of which the individual may be a partner or the proprietor.</td>
</tr>
<tr>
<td>Corporate tax returns</td>
<td>Statistics of incorporated businesses.</td>
</tr>
<tr>
<td>Payroll Deduction Accounts</td>
<td>Statistics of payroll and employment.</td>
</tr>
</tbody>
</table>

A characteristic of many of the files referred to above is that they are generated by the provinces as a result of their responsibility for the administration of such important areas as health, justice, education, the registration of vital events and a wide variety of other functions. Provinces and municipalities also account for a large portion of the total income and expenditure of government. Given such division of responsibilities and a corresponding creation by the provinces of important administrative records, Statistics Canada as the national statistical agency has devoted a great deal of effort over many years to the co-ordination of administrative records so that they would be comparable from province to province and with the federal systems, and usable to
produce significant national statistics. Careful painstaking and continuing effort to obtain and maintain agreement on definitions, classifications, codes, and wording of questionnaires and administrative forms, has been the dominating activity required to derive good statistics from administrative records. Progress has been rewarding although slow. In many areas substantial progress has been made, in others, such as welfare, much remains to be done.

3. WHY USE ADMINISTRATIVE RECORDS FOR STATISTICS

It should not be taken for granted that there is a widespread agreement that administrative records provide significant sources of data for the derivation of useful statistics. Before further discussing the problems and possibilities of using administrative records for statistical purposes, it is important to describe why they should be so used. Skepticism is not uncommon and as a basis for giving priority to their exploitation it is important to provide a wider basis of understanding to both those who hold and those who use administrative files.

For a number of administrative files, the principal historic reason for using them, and still the overwhelming reason, is that they provide almost the sole source of essential statistics. Examples are import and export statistics based on customs declarations, and mortality and demographic statistics based on birth, marriage and death registration. Alternative sources of trade statistics, for example, could be developed but at inordinate cost in resources and response burden, as well as reductions in the statistics available. Even where administrative records do not compromise the sole source, they are frequently powerful complements to other data sources. Thus, income statistics derived from tax returns and those derived from household surveys complement each other and the statistical system is enriched by the use of both sources. For example, tax returns provide geographic detail not otherwise available on an annual basis and survey data provides more socio-demographic information about those in receipt of various levels of income. Similarly,
personal tax returns can supply financial data relating to unincorporated businesses which complements data from business surveys and can often reduce significantly the number of respondents needed for those surveys.

Another important reason for the use of administrative records is that they have been generated for other purposes and can be utilized for statistical purposes at marginal cost. Of course, even the marginal cost is not zero and can be quite high in absolute terms; and the usefulness of statistics derived from administrative records must justify the cost of their production.

A combination of these reasons - the uniqueness of the source and benefits which justify marginal costs - account for the historic derivation of many statistics from administrative records. In the areas of health, education and crime, it has been possible to study both the institutions and the population involved with them, e.g., patients and students, by utilizing the administrative records of the institutions.

At the same time, an important limitation of such files for social statistical purposes is that they relate to particular populations and not to the population at large, and it is this limitation which has given emphasis to the development of household surveys to measure such characteristics of the whole population as level of education attained and health status. However, as with the income statistics referred to above, statistics of education and health, for example, derived from the two sources - administrative records and household surveys - complement each other and one can seldom be replaced with the other. For example, information about the delivery of health care based on hospital records is greatly enriched, not replaced, by information about the health/illness conditions of the whole population derived from household surveys.
In the case of business statistics, the entire population of corporations, employers and unincorporated businesses are covered by the corporate, Payroll Deduction Account, and personal taxation systems. The bureau has access to these files for statistical purposes and they form the basis for the Business Register - a major tool in the integration of business survey-taking activity.

As an example of the way in which different data sources complement each other, household sample surveys can be taken frequently and can explore social issues in depth, but they can seldom provide information for small local areas where many public decisions are made. These characteristics of household surveys highlight a contrasting characteristic of some important administrative data files, namely, the fact that they are continuously updated and also relate to very large populations. Files which have such characteristics have great potential for the production of small area data, as well as longitudinal data derived from cohorts of individuals for lengthy periods. Both of these kinds of statistics are expensive to obtain from surveys. Some of the administrative records which relate to large portions of the population and are continuously maintained are those derived from the administration of unemployment insurance, family allowances, the collection of personal income taxes and health insurance.

Finally, the use of such files for statistical purposes does not involve significant additional risk related to privacy or confidentiality if the records are handled appropriately so that the statistical output contains no identifiable information. However, their use to produce statistics may not be perceived as involving minimal risk. In fact, an important task confronting statisticians is to demonstrate and persuade that the risk is so low, and that files are so carefully safeguarded, that the small risk which is inevitably involved is more than justified by the benefits derived.
4. RECENT DEVELOPMENTS AND THE CHANGING ENVIRONMENT

While the historic use of administrative records provides a perspective from which to view current developments, it does not provide an adequate understanding of recent developments which have radically changed the environment within which administrative records are used (or not used) for statistical purposes. Probably the most important of these developments are computers which have eliminated the monopoly, or near monopoly, on data processing capability formerly held by large statistical offices. Effective ability to process data is now available to many agencies of government, computer stored data bases now provide administrators with ready access to their records for management purposes, and the development of data base technology involving random access to computer stored data is rapidly extending that ability further than even the tremendous improvements which sequential batch computer processing made possible. The process of creating, storing and retrieving administrative records is usually referred to as a management information system but the product of such systems is typically statistics. The essential points are that where there was historically only one or a few producers of statistics derived from administrative records there can now be many, and that senior managers of administrative programs want access to information derived from their data bases.

A second change of major consequence has been the recent development by governments of large socio-economic programs such as unemployment insurance and old age pensions which have generated large data bases typically covering significant proportions of the population. These new data files, along with long established files, processed by the new computer technology, have resulted in the statistical system rather suddenly acquiring new producers of significant statistics in various subject-matter areas and a new major need to co-ordinate the data in this enlarged system.
While this paper has described these developments as "new" they have, of course, been underway for some while and were recognized and articulated at least four years ago (see [1] and [2]). Since that time there have been large increases in the number of machine-readable data files.

We have entered a still further phase in the utilization of administrative records for statistical purposes reflecting the following additional changes in the environment:

- financial constraint on the part of all levels of government which are seeking to reduce their commitments;
- reflecting such constraints, a concern about the costs of statistics;
- frequent complaints on the part of the business community, politicians and the media about the burden of completing statistical questionnaires;
- widespread concern about privacy and confidentiality;
- elements (known or assumed) of duplication within the statistical system, including duplication between administrative and other data bases;
- a conviction on the part of both suppliers of data and users of resource;
- record linkage technology which is now relatively advanced;
- an insistent demand by sophisticated users for access to micro-data bases reflecting the development of new analytical tools;
- statisticians' concern about the quality and comparability of the numbers being generated by the many more players who are now part of the system;
- in certain cases the simple fact that sufficient data to satisfy the demands for statistics cannot be obtained by traditional survey-taking methods without massive increases in expenditure and respondent burden;
machinery for protection of confidentiality and security of data files is greatly improved;

· developments in methodology which enable us to use financial data from business financial statements to generate pseudo-survey data usable at the macro-level, when processed in conjunction with data derived from small sample surveys.

All of the above have led to the conviction that not using administrative records to the full is a luxury which can no longer be afforded, and that priority (one amongst several) must go to their fuller effective use as an important part of the statistical base.

5. PROBLEMS OF USING ADMINISTRATIVE RECORDS

There are a number of prerequisites for such effective use but before turning to them it is helpful to examine some of the problems associated with the use of administrative records which explain why progress has not been more rapid. Some problems referred to earlier are worth repeating in the following listing:

- administrative records are frequently held in multiple jurisdictions, for example in the provinces as already described;
- notwithstanding earlier comments about the importance of computers some of the technologies are still relatively new and not yet familiar to those responsible for data handling, e.g., computer utilization, data base technology, record linkage, micro-data bases;
- concern about possible negative public and political reaction has undoubtedly inhibited development, and properly so. Given the widespread concern about "big brother" the use of administrative records could be imperilled if statisticians were to move too quickly ahead of public and political opinion;
the problems of standards, comparability, consistency, coverage, and similar problems are intrinsically complex; for example, administrative records are subject to changes in procedures which may seriously affect comparability through time;

- in contrast to surveys of households and businesses, administrative records produced by program departments of government are not under the control of statisticians, and their use as a data source requires collaboration of administrators well beyond that required of respondents to most surveys;

- responsibility for the processes involved in using administrative files for statistical purposes is diffused, frequently overlapping between the central statistical office and other agencies;

- there are frequently uncertainties as to who has access to data files, under what authority, and with what rights and obligations;

- there is a fairly widespread perception amongst statisticians that their business is surveys, and that while the use of administrative records is not to be inhibited, neither is it a priority task to be supported with the dedication of significant resources responsible for researching, investigating, co-ordinating, and producing statistics derived from administrative records;

- frequently administrative records do not fit logically into one subject-matter area or one sphere of observation, and the old proverb "everybody's business is nobody's business" applies.

6. PREREQUISITES FOR UTILIZATION

The above represents a formidable array of problems. In addition, the following are considered to be almost prerequisites for extensive exploitation:

i) Co-ordination in terms of standards, concepts, definitions, classifications, to ensure comparability of files containing data relating to similar subject-matter and variables. However, even more important than the co-ordination of such technical matters is the development of a spirit of co-operation between administrators and statisticians, based on mutual understanding of each other's problems, objectives and contributions.
ii) The existence of a central authority to develop such standards in consultation with those acquiring and holding the administrative records, and to work towards implementation of the standards, either vested with legal authority or being persuasive as it largely is in the case of Statistics Canada.

iii) Some organization which will give sustained support to the use of administrative records as a priority.

iv) An organization which develops and maintains an inventory of administrative records, including good documentation of machine-readable (and other) data files. The organization responsible for compiling such an inventory, and disseminating information about its content, should have authority to acquire required documentation about each file maintained in the jurisdiction involved, e.g., the departments and agencies of the central government.

v) The central statistical agency should have a legal right of access to administrative records of program departments and agencies, probably with some neutral agency monitoring the effectiveness of the relationships between statisticians and administrators. This is now the case, for taxation records.

vi) A central statistical agency which does have right of access must operate under strong confidentiality laws, practices, and procedures. These are universally agreed upon as essential. In terms of modern technology they are also required to enable linkage of data files (from both administrative and survey sources); to develop and implement policies and procedures for the dissemination of micro-data files containing only unidentifiable records; and to constrain the unnecessary production of data bases containing the combined data from many separate systems.

vii) The critical accumulations of technical skills and computer capacity required to acquire, edit, link and process administrative records, as well as evaluate, analyze and disseminate the statistics derived from them.
viii) A willingness on the part of administrators, based on enlightened self-interest, to utilize their records in collaboration with statisticians for the purpose of publishing statistics based on them, or to make them available to statisticians for such purposes; and also to perhaps modify their own procedures somewhat for a recognized common good. Statisticians also need to equip themselves with the expertise to participate in the design of systems of administrative data files in such a way as to assist administrators and at the same time meet their own needs.

7. RECENT DEVELOPMENTS

Some of these prerequisites have existed in Canada for some while - most importantly the relatively strong central statistical agency with legal right of access to administrative records; laws and practices protecting confidentiality; much of the required skill and technology and cooperation between statisticians and administrators. (It is worth noting that the "right of access" referred to above is only exercised in collaboration with federal, provincial and municipal departments and agencies, and is not enforced by either confrontation or reference to the courts).

More recently a number of additional steps have been taken to improve the co-ordination of all statistical activities within the federal government and a number of them relate to the above prerequisites. A committee of senior government officials on federal statistical activities, chaired by the Chief Statistician, has been created. This committee is concerned with questions of access, priority-setting and planning for the statistical system. Under its aegis a federal statistical clearinghouse has been set up to prepare an inventory of machine-readable files, to promote their use, to facilitate assessment and cost evaluation, and to improve processing. It is intended that the clearinghouse continuously collect documentation about federal data files, prepare and distribute catalogues and disseminate information about classifications, codes and data handling.
Also, the federal government has passed human rights legislation within the past few months which is intended to co-ordinate the collection, retention, use and storage of information by government institutions. The legislation entitles individuals to ascertain what records concerning them are held within the control of a federal government institution and used for administrative purposes. To this end it calls for the annual publication of the name or other identification of each federal information bank and the types of records contained in it. The clearinghouse operation referred to above will be used to provide such a list. The individual's right to access to records about himself or herself relates only to files used for administrative purposes and not to files used for statistical purposes. However, the legislation does impact upon the statistical system, since it requires the responsible cabinet minister (who may or may not be the minister responsible for Statistics Canada) to cause to be kept under review the utilization of existing information banks and to approve proposals for the creation of new information banks or the substantial modification of existing ones. For a more extensive description of this legislation and its impact see [3] (Part 3, Privacy: Legal Basis in Canada). Statistics Canada is expected to participate in the review of proposals for new and substantially modified information banks, in much the same way that it has been responsible in recent years for reviewing new survey proposals for purposes of avoiding duplication, extending the use of existing survey data and co-ordinating survey activities of the federal government.

As previously indicated all the provinces produce extensive administrative files, and most of them have provincial statistical bureaus supported by legislation similar to that of Statistics Canada. A Consultative Council on Statistical Policy comprising directors of provincial statistical bureaus (or others where such bureaus do not exist) which is also chaired by the Chief Statistician, meets annually to review programs which involve the provinces as both users and producers of
statistics. This council recently created a federal-provincial committee on administrative records to initiate the development of inventories of administrative files held in the provinces, similar to that being undertaken by the federal clearinghouse referred to earlier in this paper.

In Statistics Canada an office of co-ordination of administrative records has recently been created as part of a restructuring of Statistics Canada. The co-ordinating activities to be carried by the office are now being developed.

Statistics Canada is also giving priority to the use of administrative records by assigning budgetary priority to such activities even in a period of severe financial constraint. How this budgetary priority can be made manifest also remains to be determined.

Statistics Canada has also developed a policy for the release of micro-data files which ensures that the observations contained in the files cannot be identified with any particular individual and a number of such files are now released routinely. Taxation-derived data, except where the release is provided for explicitly under existing legislation (e.g., CALURA) are, of course, fully confidential under the Income Tax Act. Finally, a limited number of files have been linked, e.g., cancer and death files, and a policy regarding record linkage is being developed by the office of co-ordination referred to above.

These recent developments provide important additions to a long established base which has permitted the utilization of administrative records as has been described in this paper. It is anticipated that, taken together, they will make possible new and much more extensive exploitation of the very substantial statistical potential that lies within administrative records.
RESUME

Cet article suggère que les dossiers administratifs sont une source puissante de statistiques. Afin d'appuyer cette conclusion on fournit une description générale de l'exploitation au Canada des dossiers administratifs à des fins statistiques. L'article traite de certains développements récents et de l'environnement changeant qui sont considérés comme les principaux déterminants de la création et de l'exploitation des bases de données administratifs. On estime que les capacités de l'ordinateur, avec la demande extensive pour les statistiques et les ressources financières limitées disponibles pour satisfaire cette demande, mènent conjointement vers une exploitation plus étendue des dossiers administratifs. On spécifie une variété de problèmes qui sont associés avec l'utilisation des dossiers administratifs et on décrit le développement de stratégies qui répondraient à ces problèmes et qui permettraient l'exploitation des dossiers administratifs. On indique certains développements récents au Canada qui devraient appuyer l'utilisation des dossiers administratifs.

REFERENCES

