

Issues in the Use of Administrative Records for Statistical Purposes

G.J. BRACKSTONE¹

ABSTRACT

Demands for statistics on all aspects of our lives, our society and our economy continue to grow. At the same time statistical agencies share with many respondents a growing concern over the mounting burden of response to surveys. One result of the search for alternative methods of satisfying statistical demands has been an increased emphasis on the use of administrative records for statistical purposes. This paper reviews recent experience at Statistics Canada in this area and discusses obstacles to the greater use of administrative records. Approaches to rendering administrative systems more useful for statistical purposes are reviewed, together with some important concerns related to information protection and record linkage.

KEY WORDS: Indirect estimation; Survey frames; Survey evaluation; Access; Confidentiality.

1. INTRODUCTION

Demands for statistics on many aspects of our lives, our society, our economy and our environment continue to grow. This may be due in part to our increased ability to handle and manipulate large sets of data as we move into the so-called information age, and it may also be a reflection of the increasing complexity of our social and economic systems and our desire to understand them better. Whatever their cause we face these demands in a climate of tight budgetary constraint for government statistical agencies. At the same time, statistical agencies are sensitive to the increased burden that would be imposed on respondents by an increase in survey-taking activity to meet these demands.

These factors have led to the exploration of other means of satisfying these statistical demands. Prominent among these alternative means is the increased use of existing administrative systems as sources of statistical data. This is not a new idea. For many years, statistical data have been a by-product of administrative processes in domains such as vital statistics, imports and exports, health care, and education. We will describe later how this usage of administrative data has spread more recently to statistics on businesses and on families and individuals.

The first sections of the paper describe the variety of types and uses of administrative records, illustrating some of their uses in Statistics Canada's program. The heavy dependency of Canada's statistical system on administrative records will be apparent. Section 6 discusses issues of accessing administrative sources and making them more appropriate for statistical use. Finally, a brief review of privacy concerns related to administrative record use is provided.

2. TYPES OF ADMINISTRATIVE RECORD

Administrative records come in many shapes and sizes. An important distinction is between those administered nationally (usually by the Federal Government) and those

¹ G.J. Brackstone, Assistant Chief Statistician, Statistics Canada, 26-J R.H. Coats Building, Tunney's Pasture, Ottawa, Ontario K1A 0T6.

administered sub-nationally (e.g., by provinces or municipalities). For the latter to be useful nationally, agreement between jurisdictions is required on items such as definitions, standards, record formats, and procedures. Such agreement is not always easy to achieve, particularly in domains that are constitutionally within provincial jurisdictions.

Administrative records vary in terms of their purpose, and their purpose is a prime determinant of their coverage and quality, and therefore of their statistical usefulness. Six broad categories of purpose can be distinguished.

(1) *Records maintained to regulate the flow of goods and people across borders.*

These include records of imports, exports, immigration and emigration. The coverage and content of the resulting administrative records depend on the particular laws and regulations to be enforced, and on the success of their enforcement. Typically such laws are well enforced. Immigration records, by definition, exclude illegal immigrants but otherwise are complete. However, since emigration from Canada is not controlled, no direct administrative emigration records exist. Administrative records on Canadian imports tend to be more accurate than those on exports since the former require more detailed documentation in order to assess their liability for duty.

(2) *Records resulting from legal requirements to register particular events.*

Examples include births, deaths, marriages, divorces, business incorporations or amalgamations, licensing, etc. Typically coverage and quality of records collected for this purpose are very high in Canada, since evidence of this type of registration is necessary to obtain rights or benefits.

(3) *Records needed to administer benefits or obligations.*

Examples include taxation, unemployment insurance, pensions, health insurance, and family allowances. The coverage and content of these records are highly program dependent. The population to which they apply may be very well covered, but for political or administrative reasons the definition of this population may not be the most useful definition analytically.

(4) *Records needed to administer public institutions.*

These include, for example, records related to schools, universities, health institutions, courts and prisons. Such records tend to focus on the institutional caseload rather than on the individuals passing through the institution. On the other hand, they usually provide very complete aggregate statistics on the population using these institutions. In Canada, many administrative records in this category fall within provincial jurisdiction.

(5) *Records arising from the government regulation of industry.*

Examples include records in the areas of transportation, banking, broadcasting and telecommunications. They also include records arising from the management of the supply or the price of some commodities, especially in the agriculture area.

(6) *Records arising from the provision of utilities.*

These include electricity, phone and water services. Their coverage of subscribers and the quality of information associated with services and billing are normally good. Many of these services are administered at the provincial or municipal levels.

Administrative records also vary in terms of the processes by which they are assembled. Most administrative processes with wide coverage are now automated, but differences in hardware and data formats (both between jurisdictions, and between the administrative agency and the statistical agency) have to be faced. Increased automation also leads to an increasing amount of modification to the originally reported records by the administrative agency before they are received by the statistical agency. While enhanced control of the quality of incoming forms may be beneficial to the final quality of the administrative file, additional work is required by the statistical agency to understand and evaluate the effects of any preliminary processing carried out by the administrative agency. In some administrative systems, the individual records remain at their local source and only aggregates are assembled centrally. This practice restricts the statistical agency's ability to evaluate the quality of the data and limits flexibility in statistical analysis of the data.

Finally, records differ in terms of their accessibility. Legal and regulatory provisions often govern access to, and use of, administrative records for secondary, including statistical, purposes. This topic is addressed further in Section 6.

3. USES OF ADMINISTRATIVE RECORDS

The statistical uses of administrative records may be categorized into four main areas. Most statistical applications of administrative records fall into one of these four categories or represent combinations or variations of these uses.

(1) *Direct Tabulation*

This includes the counting of units in files, cross-classification by attribute, and the aggregation of quantitative variables associated with each unit. Statistics on vital events and on external trade are important examples. Other examples include the publication of monthly counts of unemployment insurance claimants, and of beneficiaries by province, age, sex and length and type of benefit, and annual summaries of income distributions for each county based on the personal income tax file.

(2) *Indirect Estimation*

This category includes cases where data from administrative records comprise one of the inputs into an estimation process. For example, individual tax returns for the same taxfiler are linked from one year to the next in order to produce partial estimates of migration which can be weighted up with reference to census-based benchmarks. These estimates of migration then feed into Statistics Canada's population estimation program (which also makes use of administrative data on births, deaths and immigration). A second example is the use of taxation data for small businesses in lieu of seeking survey data from them. These tax-based data, adjusted if necessary, are combined with survey-based data for large businesses to provide industry aggregates.

Also within this category are uses that involve the linkage of different administrative or statistical files to produce estimates. For example, the linkage of the death register with files of individuals exposed to particular hazards in order to estimate differential mortality rates, or the linkage of records from tax files, unemployment insurance files, and manpower training files in order to analyse labour market attachment and adjustment.

(3) *Survey Frames*

In this category we include the use of administrative records to create, supplement or update frames to be used for censuses or surveys. A primary example is the use of payroll deduction information submitted by employers to Revenue Canada. The questionnaire which has to be completed by new payroll deduction account holders is a valuable means of identifying new businesses or changes in the structure of existing ones. Although in Canada we do not have a register of housing units, a second example would be the use of building permits or new telephone or electricity connections as signals of possible new housing units.

(4) *Survey Evaluation*

This category covers the use of administrative records for checking, validating or evaluating survey-derived data. This may be done either at the individual unit level, or at an aggregate level. Several census evaluation studies in the past have used immigration and taxation records to evaluate census questions on immigration and income, respectively, while family allowance records have been used in checking the census coverage of children.

An important determinant of how a particular administrative source will be used is the perceived quality of the administrative records compared to corresponding survey information. In some instances administrative records are used to evaluate survey responses, while in others survey-based data provide a means of benchmarking administrative-based estimates. The quality of administrative records has to be assessed in each individual case. In general, their quality for statistical purposes depends upon at least three factors:

- (i) the definitions used within the administrative system;
- (ii) the intended coverage of the administrative system;
- (iii) the quality with which data are reported and processed in the administrative system.

Weaknesses in any of these three factors can affect the statistical usefulness of the administrative records. The timeliness with which they are available is also an important consideration. Some of the potential limitations that need to be considered when deciding on the statistical use of administrative records have been described elsewhere (e.g., see Brackstone 1984). The strengths and weaknesses of administrative records compared to those of censuses and surveys are summarized in Table 1.

To illustrate the utilization of administrative records in Canada we will describe two areas of application within Statistics Canada. The first deals with the production of business statistics; the second addresses the production of statistics on individuals and families.

4. ADMINISTRATIVE DATA AND BUSINESS SURVEYS

Statistics Canada is currently in the throes of a complete redesign of the infrastructure and strategy on which its business surveys program is based. In particular this involves the redesign of the business register (the frame for business surveys), the re-thinking of the role

Table 1
Comparison of Censuses, Surveys and Administrative Records as Sources of Statistical Data

Factors	Censuses	Surveys	Administrative records
1. Coverage	Aim at complete coverage of the population	Some surveys exclude certain sectors of the population (e.g., Indian reserves, remote areas)	Target populations are defined by administrative requirements
2. Content	Wide range of data items allows extensive cross-classification	Usually covers a narrow range of topics but in more depth than a census	Restricted to variables required for administrative purposes
3. Concepts/definitions	Can be based on the requirements of social and economic analysis	Can be based on the requirements of social and economic analysis	Defined by administrative requirements
4. Small area estimates	Available as a result of aim at complete coverage	Unavailable in most cases	Available, provided individual records are geographically coded to small areas
5. Quality control	Can be designed to minimize errors	Smaller size allows for even tighter control than in censuses	Under the control of the administrative agency and may not receive attention except for key variables
6. Cost	Expensive	Relatively low cost per survey, although the cumulative cost of a regular survey over a 5-year inter-censal period may be large	Relatively inexpensive if initial collection costs attributed to the administrative program
7. Frequency	Every 5 or 10 years (depending on topic)	May be annual, quarterly or monthly depending on topic	May be annual or monthly depending on administrative program
8. Timeliness	Data available six months to 2½ years after Census Day	Repeated regular surveys produce results in a few weeks. Ad hoc surveys may require several months	Dependent upon the administrative process. An annual file may not be available in a clean form until well into the following year
9. Stability	Changes are under the control of statisticians who respond to user needs	In repeated surveys, changes are infrequent to allow comparisons over time	Changes may occur due to legislative or regulatory change, or due to changes in administrative practice
10. Respondent burden	Heavy but infrequent. Reduced through the use of sampling	Light on average, though heavy for those selected	No additional burden

and use of tax data within the program, and the development of a consistent strategy for the design of both annual and sub-annual business surveys. This redesign was motivated by needs to:

- (a) overcome some noticeable data quality weaknesses in the current program;
- (b) better integrate data from different surveys;
- (c) minimize respondent burden by making maximum use of tax data;
- (d) reduce resources required for maintaining survey frames.

A more detailed description of this project can be found in Colledge (1987).

Income tax and payroll deduction data play a prominent role in the conduct of business surveys. Annual tax returns submitted by corporations (T2) and by individuals (T1) are available to Statistics Canada under the Statistics Act. The payroll deductions of income taxes by employers are also available. Statistics Canada makes use of these data from business for two distinct purposes:

- (i) maintenance of its frame of businesses;
- (ii) substituting income tax data for survey data.

4.1 Frame Maintenance

The maintenance of a frame of businesses is a complex task. This complexity stems primarily from the complex structure and inter-relationships of many businesses, particularly large ones, and from the difficulty of keeping track of the very large number of births and deaths occurring among small business. The term “business” itself needs careful definition. In fact a distinction must be made between legal structures (incorporated companies, etc.), operating structures (the way companies organize and operate themselves), and statistical structures (the units for which data are required for analytic purposes). A hierarchy of units can be defined within each of these structures. In the case of the statistical structure, Statistics Canada has defined a hierarchy comprising, from top down, enterprises, statistical companies, establishments and locations. The task of frame maintenance thus involves not only updating for births and deaths but also keeping track of changes in the relationships between the various units within complex businesses, including the relationships between the statistical and operating hierarchies.

The proposed frame strategy calls for the continuous maintenance of the current corporate structure of all companies above a certain threshold size (which varies with industry), including the relationship of this structure to tax reporting units. Companies updated in this way will account for at least 70% of economic activity in each industry.

An activity known as “profiling” is used to determine the internal structure of complex businesses. This involves interviewing officers of the business to understand their operating structure and identify the appropriate statistical units. An important source of information on changes to business (births and restructuring) is Revenue Canada’s payroll deduction (PD) system. The activation of a new PD account by an employer is treated as a signal that something has happened. Such signals are followed up with the business to identify whether a frame update is required. Other signals will be obtained from annual tax returns, from responses to regular surveys, and from routine profiling.

In the case of smaller companies, where the structure is usually simpler but the turnover is faster, no attempt is made to define the various types of unit and their inter-relationships.

Instead, administrative data are used directly. Two alternative lists of businesses are made available as a basis for surveying – one is the most recent set of annual tax returns; the second is the current set of PD accounts. In both cases, all units above the threshold are removed. These two lists overlap and the most appropriate one is used in each particular survey. The PD-based list, which is more current since PD accounts may be opened or closed at any time during the year, is preferred for sub-annual surveys. It has the disadvantage of excluding non-employers.

4.2 Substituting for Survey Data

In the interests of minimizing both response burden and costs, tax data are used to replace survey data where feasible. The concepts and definitions underlying tax data do not uniformly coincide with the survey definitions required to assure consistency in the System of National Accounts or for other analytic purposes. Therefore care has to be taken in selecting from tax returns the data items that come closest to the required survey definitions. Furthermore, tax data do not contain the full range of variables required by many annual business surveys. In particular, they lack production statistics.

A further problem in utilizing tax data lies in establishing the relationship between the unit for which a tax return is submitted and the unit(s) to be surveyed. This is a problem particularly for the large complex businesses referred to earlier.

The strategy that has been developed for annual surveys is to make use of tax data primarily for small businesses where there is usually a one-to-one relationship between the taxfiler and the business. This approach significantly reduces the response burden on small businesses, without unduly affecting the quality of final data, since the bulk of economic activity is reported through the survey returns of larger companies.

It is clear from this brief overview of the new business survey strategy and infrastructure that there is a fundamental dependence on tax data for the continuing functioning of the program. This requires a very close working relationship between Statistics Canada and Revenue Canada so that the impact of administrative and procedural changes in the tax system can be assessed and prepared for in advance.

5. SOCIO-ECONOMIC DATA FROM ADMINISTRATIVE SOURCES

A systematic effort to develop data on individuals, families and households from administrative records was initiated in the late 1970s. The original motivation for this work was the rising costs of census-taking and the search for cheaper alternatives. It quickly became apparent that the statistical potential of administrative records on individuals in Canada lay in supplementing the quinquennial census through the provision of data for small areas inter-censally, rather than in replacing the census. It is not possible to achieve the coverage, geographic specificity, and range of individual, family and household characteristics required from a census with the existing administrative record systems. Nevertheless, the emulation of census coverage using a combination of administrative record systems is being pursued, together with the study of the possibility of replacing some census questions with data derived from administrative sources.

This section will concentrate on the use of administrative records to supplement census data inter-censally. The focus of the developmental work has been on administrative record systems that are national in scope (e.g., income tax, unemployment insurance (UI), family allowance, old age security) rather than systems that are administered at provincial or lower

levels (e.g., health insurance, driver's licences, municipal assessments). In the latter case the problem of standardization across jurisdictions is added to the other problems inherent in the statistical use of administrative records.

The annual individual tax file (T1) has proven to be the principal source of statistical data on individuals. The first use of this file was its direct tabulation to produce statistics on income and labour force participants by age and sex for provincial and sub-provincial areas. Identification of geographic location of taxfilers is based on the postal code indicated on the record. A file that provides a conversion from postal code to the various levels of census geography (province, county, municipality, electoral district, etc.) has been developed. Special tabulations can also be produced for user-defined areas described in terms of postal codes.

Data derived in this way are, of course, based on the concepts, definitions and regulations implicit in the Income Tax Act. These may not conform to definitions desired for analytic purposes (e.g., some forms of social assistance which are not taxable may be excluded). Income can be broken down by source – in particular, employment income can be separated. Variables available for cross-classification are limited (e.g., age, sex and marital status). Occupation, though asked on the tax form, is not reported nor coded with sufficient quality to be statistically useful. The coverage of these data is limited by the need to file a tax return. Low income individuals and dependents are therefore under-represented. Over time, changes to tax law can have a significant impact on coverage; e.g., the introduction of the Child Tax Credit, that required low income earners to file a tax return in order to claim the credit, led to a marked increase in coverage in 1978 compared to the previous year.

Despite these reservations, data produced by direct tabulation from income tax files provide a useful inter-censal source of small area income data. A recent publication from Statistics Canada made use of this source to produce data for Forward Sortation Areas, i.e., the first three characters of the postal code (Statistics Canada 1987). Since a prime concern in the publication of data for small areas is to ensure that no individual data can be deduced from aggregate totals for small areas, data are not provided for areas with less than 100 taxfilers.

A second use of the individual tax file is for estimating annual migration. This is achieved by matching individuals on tax files for two successive years and comparing the Census Division (or county) code assignment for each year. If there has been a change in code, it is assumed that the taxfiler has migrated. Demographic and tax exemption information are used to estimate the total number of persons who have migrated with the taxfiler. In a final stage, since the tax file does not cover the whole population, an adjustment is made to estimate the total number of migrants from year to year. Since 1981, tax-based migration estimates have been used in Statistics Canada's population estimates program. A full description of the methodology for estimating migration from tax records can be found in Norris and Standish (1983).

While data on individual incomes can be derived from tax data as described earlier, more analytic and policy interest focuses on family income. To derive family income from the individual tax file requires the capacity to identify and match records of individuals belonging to the same family. Development of family income data in this way has been proceeding with encouraging results. A description of methodology and results can be found in Auger (1987).

A second important administrative source of data on individuals is the unemployment insurance (UI) system. Files of both claimants and beneficiaries are available to Statistics Canada. The UI claimant and beneficiary files contain individuals who, for a variety of reasons, may be entitled to UI benefits. Not all of these individuals are considered to be unemployed according to the standard international definition of unemployment as incorporated in the Labour Force Survey (LFS), the source of published unemployment rates.

If a closely corresponding category in the UI system can be found, these files can be used to tabulate counts of "unemployed" for small areas. However, since even the best choice of category in the UI system does not correspond exactly with the definition of "unemployed" used in the LFS, attention has to be focused on how to integrate or reconcile these two sources of data. For example, monthly counts for small areas from the UI system might be used as indicators of changes in unemployment at the local level which could be calibrated to reliable LFS estimates at a higher geographic level (e.g., the province). Various methods of estimation along these lines have been investigated (e.g., regression estimation, SPREE – structure preserving ratio estimation), though without as yet any final conclusion as to the most appropriate method. A description of this work can be found in Trottier and Choudhry (1985) while Feeney (1987) describes a similar approach in the Australian context. A time series modelling approach which exploits the correlated structure of the error over time appears very promising (Choudhry and Hidioglou 1987).

These examples have illustrated that, in the case of statistics on individuals, the primary uses of administrative records are for direct tabulation and as input into estimation processes. This contrasts with the examples from the business side where frame maintenance and substitution for survey responses were the main uses.

While these two examples represent two important developing areas of administrative record use in Statistics Canada, they cover only a small fraction of the administrative files used by the Agency. There is, for example, a widespread and long-standing use of administrative records in the social institutions area (education, health, justice) both for creating survey frames and for obtaining statistical data. Current developmental work on telephone surveying and on address registers is using administrative records to develop frames of dwellings or households. A recent internal survey identified more than 50 administrative systems being used for statistical purposes. These covered the full range of types and uses described in Sections 2 and 3, and included examples from areas as varied as disease registries, motor vehicle licences, aircraft landings, milk marketing boards, fuel sales tax, municipal construction records, and customs and excise.

6. ACCESSING AND INFLUENCING ADMINISTRATIVE SYSTEMS

It is clear from this review of the use of administrative records for statistical purposes, that administrative records are a vital input to many of Statistics Canada's programs. This leads to a consideration of measures the Agency can take to protect the supply of data from administrative sources, and perhaps to make them more useful for statistical purposes. In this section we will deal with the two primary issues of obtaining access to administrative records, and influencing their content, design or associated procedures.

6.1 Access

The legal authority for access to administrative records is provided by Section 12 of the Statistics Act (1971):

"A person having the custody or charge of any documents or records that are maintained in any department or in any municipal office, corporation, business or organization, from which information sought in respect of the objects of this Act can be obtained or that would aid in the completion or correction thereof, shall grant access thereto for those purposes to a person authorized by the Chief Statistician to obtain such information or such aid in the completion or correction of such information."

While this provision appears to give fairly broad access rights, it is not without limitations. In some cases, legislation governing the administrative process places restrictions on access or secondary use of the administrative data. This leads to a confrontation of legislation that will at best delay the negotiation of access. In some cases, access for statistical purposes is specifically permitted.

Enabling legislation is a necessary but not sufficient condition for the productive utilization of administrative records. A co-operative approach to the development and utilization of administrative records for statistical purposes is likely to be far more effective in obtaining access to administrative records than an approach involving legal arguments and sanctions. Indeed, once access is obtained, the subsequent step of influencing design or procedures is only achievable if there is a spirit of co-operation between the administrative and statistical agencies.

Access to administrative records by Statistics Canada is strictly a one-way street. Individual micro-data are provided from the administrative agency to the statistical agency, but only confidentiality-protected aggregate data can flow back. The only exception to this rule is the case where the administrative agency depends on the statistical agency to organize, format, edit, process, or restructure its records, and a version of the original micro-data is passed back to the supplying agency.

6.2 Influencing Change

We have already alluded to the potential impact of changes in administrative regulations or practices on resulting statistics. Discontinuities in time series based on administrative records can be caused by simple changes in the coverage of a program, the introduction of an incentive to join or leave a program, or procedural changes that affect quality or completeness of records. Thus the statistical agency has to guard against, and react to, externally imposed changes.

There are other kinds of changes that the statistical agency might like to see implemented. A frequent frustration of the statistician trying to use administrative records is the feeling that the administrative records could be so much more useful if only relatively minor changes were made. For example, the addition of an extra question, the use of a different concept, the coverage of an additional subgroup, or the introduction of a quality check might significantly enhance the statistical value of the records. On the other hand, why should the administrative agency contemplate changes not required for the primary administrative process, changes which would probably in some measure add to the cost and complexity of the administrative process?

The challenge for a statistical agency is to persuade the administrators that the benefits from such a change outweigh any additional administrative costs. This is made harder to the extent that the benefits do not accrue to the department responsible for the administrative system, but to separate policy-making departments and other statistical users.

It is usually easier to build statistical requirements into a system from its inception than to make changes to a system that is already operational. Therefore, a mechanism that would allow statistical requirements to be considered during the design, or the major redesign, of an administrative system is preferable to one that only tries to adjust existing systems. A topical case in Canada is in the area of tax reform, currently under consideration by the government. This could significantly change the collection of business data in Canada. Involvement of statisticians in the design of such a system could greatly enhance the statistical benefits derived from the system. Of course, the institution of a new administrative system is a relatively rare occurrence, so that adjustment to existing systems is also necessary if statistical benefits

are to be obtained in the short run. On the other hand, the comparative rarity of design or redesign of major administrative systems strengthens the argument for not missing opportunities to influence such exercises when they do arise.

6.3 Mechanisms

A variety of measures or mechanisms, some bilateral involving the statistical agency and a specific administrative department, others of a broad government-wide nature, can assist the statistical agency in accessing and influencing administrative systems. These include:

- (i) bilateral committees at a senior level to review and discuss issues of mutual interest, including problems related to the supply of administrative data;
- (ii) feedback of statistical data to the administrative agency to demonstrate both usefulness of the data and, perhaps, weaknesses arising from administrative practices;
- (iii) provision of technical advice or services in support of the administrative agency's own statistical activities;
- (iv) a government information collection policy that requires, for example, any data collection activity plan (statistical or administrative) to be reviewed by a central agency;
- (v) statistical planning in the form of a requirement that each new program proposal include a plan for acquiring the statistical information needed to monitor and evaluate the program;
- (vi) promotion of the use of standard statistical definitions (e.g., family, business establishment, unemployed) in administrative systems;
- (vii) audits that identify the use of administrative records as a cost-efficient alternative to other means of acquiring information;
- (viii) political instruction to make greater use of particular administrative systems or seek alternatives to survey-taking;
- (ix) removal of legislative impediments to access or use of administrative records for statistical purposes.

Statistics Canada's experience in dealing with other federal government departments has been most successful in cases where close bilateral arrangements have been developed. The introduction of senior bilateral committees in the early 1980s was supportive of such arrangements, and in some cases instrumental in creating them. Government-wide measures such as information management and statistical planning have been less successful in facilitating administrative record use. Government audits and cabinet directives have provided impetus to activities aimed at increasing administrative data use, but the increased use itself is again dependent upon close working relationships with particular departments. While it is convenient to characterize the statistical agency as the progressive agency trying to break down unreasonable barriers to administrative data use, it must also be recognized that there may be inertia to the associated changes within the statistical agency itself. Staff whose careers have been based on survey design and survey-taking may need convincing that budgetary restrictions and data needs now necessitate combining these with other approaches.

Since the above comments have focused on federally administered systems, we will add a few words about provincial records. While some of the above measures apply equally to provincially administered records, the fundamental problem in dealing with subnational

jurisdictions is that of adherence to common standards. Differing provincial needs and priorities, facilitated by increasing technological capacity, will lead to divergent administrative systems in the absence of any centralizing force. Statistics Canada has used a variety of mechanisms in the past in attempts to encourage conformity, but with only mixed success. As with federal government custodians of administrative records, mutual benefit has to be the major incentive to conformity. Federal-provincial committees exist in several subject areas. The Vital Statistics Council, consisting of provincial registrars of vital events and representatives of Statistics Canada, is a successful and long-standing example. Such committees have developed and monitored conventions for reporting certain data items in the past. For example, the framework for municipal finance reporting was developed as a result of federal-provincial meetings on municipal financial statistics.

7. CONFIDENTIALITY, PRIVACY AND PUBLIC RELATIONS ISSUES

Even with the legal authority to exploit administrative records and co-operative administrative agencies to supply them, careful consideration has to be given to the public perception of the use of administrative records beyond their original purpose. Since the effectiveness, if not the survival, of a statistical agency depends critically upon the continuing co-operation and trust of respondents, it must take extreme care before embarking on any activity with the potential to undermine that co-operation or trust.

Public awareness and concern over privacy and related issues of information access and control have risen in many countries in recent years. In Canada, passage of the Privacy Act in 1982 bore witness to this mounting concern. The Privacy Act requires, *inter alia* and with some exceptions, that an index of all personal information banks under the control of federal government institutions be published periodically, that individuals have the right of access to information about themselves contained in such information banks, and that personal information be used only for purposes consistent with the purpose for which it was obtained. One of the exceptions to this last provision is that personal information may be disclosed

“... to any person or body for research or statistical purposes if ... the purpose for which the information is disclosed cannot reasonably be accomplished unless the information is provided in a form that would identify the individual to whom it relates, and ... a written undertaking (is obtained) that no subsequent disclosure of the information will be made in a form that could reasonably be expected to identify the individual to whom it relates.” (Privacy Act 1982 Section 8(2)(j)).

This provision covers the use of administrative records for statistical purposes as far as the Privacy Act is concerned. However, this Section is subject to any other Act of Parliament so that a clause forbidding such use in an Act governing an administrative process would have precedence.

While the Privacy Act and other Acts recognize statistical work as a legitimate secondary use of administrative records under certain conditions, this alone will not allay public concern over the existence of data banks that could be used to an individual's detriment. It is doubtful whether the average citizen appreciates the distinction between statistical use, where the identity of the individual record is of no lasting interest, and administrative use, where the essence of the individual record is the particular unit to which it relates. It would be easier to explain and utilize this distinction if we could state unequivocally that identifiers are never needed for statistical purposes. Unfortunately this is not the case. Several legitimate statistical

techniques do require identifiers in intermediate data manipulations. These techniques all involve some form of matching data from different files or different occasions, and identification is required to ensure that the correct records are matched. Once the matching has been accomplished the records can be anonymized provided no subsequent linkage is planned. Examples include the requirement for names in a population census to ensure coverage and permit coverage measurement, longitudinal studies using administrative records, epidemiological investigations, and evaluation studies to check survey responses against administrative sources. Explaining why identifiers are needed when identity is of no interest is an interesting challenge facing the statistical agency.

A further source of concern may relate to the undertaking of confidentiality itself. Despite Statistics Canada's record of confidentiality protection there are doubtless respondents who are skeptical about the protection their information enjoys. This concern may be heightened by the use of enumerators who are known to respondents, particularly in small communities. Some respondents seem to assume there is a high degree of information exchange actually taking place between federal departments, and in some cases do not distinguish between different departments of government.

An additional concern may relate, not to the trustworthiness of the present custodians of information banks, but to a fear that personal information cannot be protected against future violation, either illegally, or by a legitimate elected authority with different views on privacy. Protection against this possibility would require the removal of all identifying information from statistical data bases.

This public concern over privacy and the manipulation of personal information requires the statistical agency to consider measures it can take to prevent or minimize negative public reaction to its legitimate use of administrative records for statistical purposes. Since this is essentially an issue of public perception, it is important that the statistical agency be open about its practices, and that any of the following measures that are implemented are clearly visible to the interested public.

- (a) Public communications to respondents and users should continually stress the importance attached to confidentiality of all individual (micro) data acquired by the statistical agency.
- (b) The one-way nature of micro-data flow should be stressed. Micro-data flow into the statistical agency, but only confidentiality-protected aggregates or summaries flow out. This applies equally to survey or census data and data from administrative records.
- (c) The benefits of administrative record use in terms of reduced respondent burden and savings to the taxpayer should be emphasized. Such claims should be supportable by real measures of cost and respondent burden savings.
- (d) An explicit and public policy on record linkage stipulating the conditions under which the statistical agency will undertake such activities can be helpful both in demonstrating careful consideration and control of linkage activities, and in forestalling linkage requests that would violate the conditions.
- (e) The Privacy Act requires that individuals be informed of the purpose for which any personal information is being collected. Administrative agencies should be encouraged to ensure that statistical purposes are included in such statements. Even though statistical purposes may be a permissible secondary use of administrative records, their explicit mention on the collection form will serve to avoid subsequent surprise.

- (f) The physical security that surrounds the use of sensitive administrative records should be clearly visible, and perhaps even tighter than that in use generally within the Agency. For example, in Statistics Canada, the divisions having primary custody of tax data are housed in limited access areas within buildings that are themselves subject to security checks on entry.
- (g) Exemption of statistical files from examination by security or intelligence services is an important element in maintaining public trust in the absolute confidentiality of data provided to the statistical agency. An exemption for Statistics Canada data (the sole institutional exception within government) was provided when the new Canadian Security and Intelligence Service was formed in 1983.

While the above points represent some specific measures that can be taken to avoid or respond to public reaction to the use of administrative records, ultimately the statistical agency must have strong political support for this kind of activity. The political credit to be gained from demonstrated reductions in costs and respondent burden, coupled with strong political assurances of the protection of individual data, provide a strong platform for politicians to dispel public concern over the use of administrative records for statistical purposes. At the same time they must immediately and unambiguously confront and correct any suggestion that statistical records be used for administrative purposes.

8. CONCLUSION

Administrative records are and will continue to be an increasingly important source of statistical data. The relative strengths and weaknesses of data derived from administrative systems, in terms of cost, coverage, quality, relevance and timeliness, in comparison to census- or survey-based data, dictate the manner in which these sources of data are most effectively used. Current uses of administrative records include direct tabulation, indirect estimation, substitution for survey responses, frame construction and maintenance, and data evaluation. These uses now permeate most statistical programs and can be expected to extend even further in the future.

In Canada, administrative records have become part of the fabric of our statistical system. Their use has been one of the means by which Statistics Canada has been able to maintain its programs in the face of declining budgets. In the process, respondent burden has been reduced and new, or more frequent, data series have become available. Since we do not have administrative registers as such, considerable attention has been paid to issues of coverage and the joint use of both administrative and survey-based data to ensure valid estimation of universe totals. The use of record linkage techniques, though requiring careful controls, has proven to be very valuable, particularly for business data, longitudinal labour market studies, and epidemiological work.

With the growing use of administrative records, statistical agencies are becoming increasingly dependent upon other agencies for the uninterrupted flow of input data to their statistical programs. Whatever the legislative and policy environment in which the statistical agency operates, the establishment of close co-operative arrangements with supplying agencies is crucial. The ability of the statistical agency to influence the design or redesign of administrative systems rests on a mutual understanding of the requirements of the two agencies. Establishment of a government-wide policy or principle that the statistical agency should have a voice in decisions regarding the design of administrative systems, or more generally, in proposals

for meeting the statistical needs of new programs, can help the statistical agency in this regard, but is no substitute for the fostering of close co-operation with administrative agencies.

A variety of mechanisms can be considered to assist the statistical agency in gaining the access and influence it requires within the government system. The applicability and effectiveness of each mechanism will depend upon the underlying legislative and political climate, and on the mandate and status of the statistical agency within the government apparatus. Statistics Canada's experience has been that close bilateral working relationships with administrative departments, based on a principle of mutual benefit, is the most effective approach. Political support for the use of administrative records is important and has been forthcoming through recent government decisions related to budget reductions.

ACKNOWLEDGEMENTS

This paper has drawn upon ideas and work of many people at Statistics Canada. In particular, parts of this paper incorporate material prepared in collaboration with Michael Colledge, Ivan Fellegi and John Leyes.

REFERENCES

- AUGER, E. (1987). Family data from the Canadian personal income tax file. Paper presented at 1987 American Statistical Association meetings.
- BRACKSTONE, G.J. (1984). The impact of technological change on census-taking, *Estadística*, 36, 43-60.
- CANADA, STATISTICS ACT (1971). *Statutes of Canada 1970-71-72*, c.15.
- CANADA, PRIVACY ACT (1982). *Statutes of Canada 1980-81-82*, c.11.
- CHOUHDRY, G.H., and HIDIROGLOU, M.A. (1987). Small area estimation: Some experiences at Statistics Canada. *Proceedings of the 46th Session of the International Statistical Institute*, (forthcoming).
- COLLEDGE, M.J. (1987). The Business Survey Redesign Project: Implementation of a new strategy at Statistics Canada. *Proceedings of the Third Annual Research Conference, U.S. Bureau of the Census*, (forthcoming).
- FEENEY, G.A. (1987). The estimation of the number of unemployed at the small area level. In *Small Area Statistics, An International Symposium*, (Eds. R. Platek, J.N.K. Rao, C.E. Sarndal and M.P. Singh), New York: John Wiley.
- NORRIS, D.A., and STANDISH, L.D. (1983). A technical report on the development of migration data from taxation records. Working Paper, Statistics Canada.
- ROWEBOTTOM, L.E. (1978). The utilization of administrative records for statistical purposes. *Survey Methodology*, 4, 1-15.
- STATISTICS CANADA (1987). *Urban FSA and rural postal code summary data*. Catalogue No. 17-602, Statistics Canada.
- TROTTIER, I., and CHOUHDRY, G.H. (1985). Model based unemployment estimates for small areas. In *Small Area Statistics, An International Symposium '85 (Contributed Papers)*, (Eds., R. Platek and M.P. Singh), Laboratory for Research in Statistics and Probability, Carleton University/University of Ottawa.