11-617

no.86-09

katak s Canada Sarsta r Sarsta

c. 2



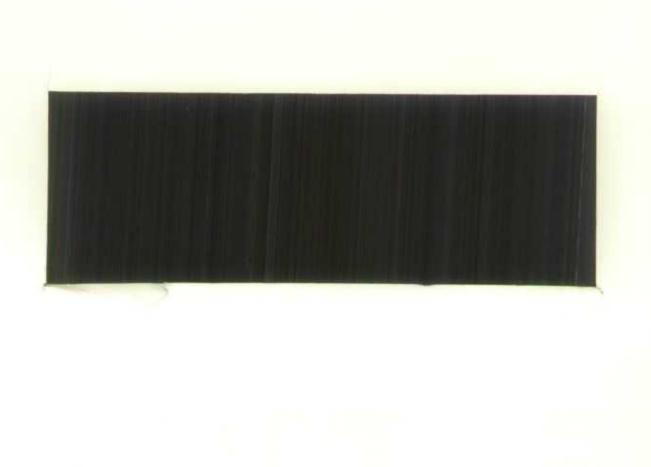
## Methodology Branch

Business Survey Methods Division

## Direction de la méthodologie

Division des méthodes d'enquetesentreprises

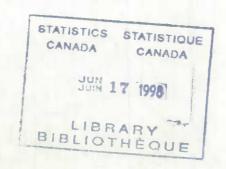




\*

WORKING PAPER NO. METH-86-009E/F

METHODOLOGY BRANCH



2-1668

# SOME STATISTICS CANADA EXPERIENCES IN USING ADMINISTRATIVE DATA TO COMPLEMENT SURVEY DATA

Gordon Brackstone, Michael Colledge and John Leyes February 1986

<sup>\*</sup> This is a preliminary version.

Do not quote without author's permission. Comments are welcome.

CARADA CARADA

#### INTRODUCTION

Administrative records have long been an important source of statistical data. In areas such as vital statistics, imports and exports, and many aspects of health care and education, statistical data have traditionally been the by-product of an administrative process. More recently, concerns over respondent burden and the mounting costs of direct data collection have led to the investigation and utilization of administrative records as a source of statistical data in subject-matter areas previously dependent entirely on censuses and surveys.

The quality of statistical data is a major concern of survey and census takers. Equally, quality is a primary concern in the development of administrative data for statistical purposes, particularly when the design of the administrative data collection process has not had the benefit of a statistician's input.

In this paper, attention is focussed on two particular areas of administrative data use in Statistics Canada, with emphasis on quality and comparability with traditional sources. In Part 1, the use of business tax data, in conjunction with business surveys, as a source of economic data is discussed; in Part 2, the utilization of the personal income tax file to produce demographic data is addressed.

## PART 1. ECONOMIC DATA. EXPERIENCE IN THE USE OF INCOME TAX DATA

#### 1.0 OPENING REMARKS

The basic reasons for using income tax information to augment or replace economic survey data are that it can result in a net decrease in data collection costs, an increase in the range of data available, and a reduction in response burden. The potential for these benefits was explicitly recognized when Statistics Canada was granted access to Revenue Canada income tax files under the Statistics Act in 1971.

In Canada there are two basic types of tax returns: Tl and T2. The Tl return is filed by an individual. It includes financial statements for all unincorporated businesses owned entirely or in partnership by the individual. Revenue Canada does not require that these financial statements be in a fixed format. Each Tl return is identified by the individual's social insurance number (SIN), or, for individuals with no SIN, by a unique number defined by Revenue Canada. To returns are filed

by corporations and by certain other legal entities required to report for income tax purposes. As for T1 returns, the accompanying financial statements are in free format. Revenue Canada assigns to each filer a T2 number which is, in principle, unique and unchanging over time. The T2 number is in no way related to any other identification number possessed by the corporation, e.g., the number associated with its charter, or with its payroll deduction account(s) (also maintained by Revenue Canada).

Tax data are made available to Statistics Canada in two forms: machine readable files; and the basic Tl and T2 documents. For economic survey purposes most of the data items of interest, including wages and salaries, depreciation, net profit, inventories, purchases, etc., are to be found in the financial statements attached to the return but not on the Tl and T2 forms themselves. Such data are not captured by Revenue Canada and hence not available in machine readable form. Also not captured are business descriptions provided by tax filers which are of utility in assigning standard industrial classification (SIC) codes to the reported economic activity. Thus Statistics Canada has an extensive program for sampling tax returns as they are being processed at Revenue Canada, assigning SIC codes, taking copies and subsequently capturing the additional data items of interest.

Income tax data can be used for two separate but interrelated purposes. Firstly they are of assistance in constructing and maintaining the frames which form the starting point for economic survey activity. Secondly they can be utilized to supplement or to replace financial data which would otherwise have to be gathered by annual surveys. In this context, there is an important distinction to be made between surveys for which the tax filing unit is also the survey reporting unit and those for which it is not. There is obviously more scope for application of tax information when tax and survey units coincide.

All these points are further discussed in the following sections. The uses presently made of tax data are outlined in Section 1.1; the results of some studies comparing tax and survey values are summarized in Section 1.2; experiences with a specific industrial survey which best exploits the full potential of tax data are described in Section 1.3; and some plans for the future are outlined in Section 1.4.

#### 1.1 USES OF INCOME TAX DATA

There are two surveys for which T2 (incorporated) tax data are the principal source of financial information, not just a supplement. Firstly, there is the annual Survey of Corporations which is based entirely upon information from T2 returns. The survey reporting unit is defined, with some exceptions, to coincide with the T2 tax filer. Published data include details of assets, equities, revenues, expenditures and profits, etc. Secondly, the Corporation and Labour Union Relations Act (CALURA) requires Statistics Canada to collect intercorporate ownership, financial and technological data from large "enterprises", defined in terms of ownership and control. The financial data for this survey have in the past, and will in the future, be obtained from T2 returns.

Tax data are exploited in varying measures to replace direct data collection for annual surveys of industrial statistics covering manufacturing, trade, services, transportation and construction. The lack of uniformity in use between industrial sectors is partly inherent to the sectors, partly related to the types of data considered most important, and partly historical. For all these surveys the target statistical unit is defined by Statistics Canada and termed the "establishment." The extent to which tax data can be used depends upon the degree of coincidence between the establishment and the tax filer. In general, it may be assumed that, whereas the activities of a small, independent establishment are likely to be covered by a single tax return on a one-to-one basis, the relationships between large establishments and tax returns will be more complex. Data for establishments which are jointly operated may be consolidated within a single tax return or may be spread over several tax returns. Thus there is more scope for use of tax data to supplement survey data in industries with a high proportion of small units, such as construction trade contracting, than in industries with predominantly large units, like manufacturing. The following paragraphs provide examples of tax data usage and some remarks on why it differs by sector.

Of all the annual industrial surveys, the major user of tax information is the Census of Construction. Since the 1975 reference year, tax records have been employed to classify the construction universe into larger and smaller units, and to provide SIC codes and financial information for the latter group. In 1983, 93% of the units accounting for 28% of the total construction industry revenue were covered by tax data. The survey is described in more detail in Section 1.3.

The Census of Manufactures has also made use of tax information since the 1975 reference year. In this case, tax data are collected for smaller units covering about 30% of the universe and accounting for 3% of total revenue. The data are

not used to delineate the target population nor to provide industrial classification. In this context it should be noted that there are three times more SIC codes for manufacturing than for construction. Industrial classification is determined by means of a short questionnaire especially designed for small units. The short questionnaire is also the basis for collection of information from small units for which the appropriate tax identifiers (SIN and T2 numbers) are unknown, and for gathering non-financial information which can not be obtained from a tax return.

Tax data were used to provide coverage of smaller units for the 1975 Motor Carrier Freight Survey but concerns about reliability lead to the approach being abandoned in subsequent years. These concerns reflected the attitude of survey takers towards tax data as much as scientific reasoning. A consequence is that survey data collection is presently restricted to medium and large sized units, with no coverage below \$100,000 total revenue.

Beginning with the 1981 reference year, the Annual Survey of Wholesale Trade has utilized tax data in place of survey data for smaller wholesale merchants, 70% of the units accounting for 7% of the revenue. There is no equivalent of the Census of Manufactures' short questionnaire. SIC codes derived from tax data are validated by direct contact with the corresponding business units in cases of uncertainty. A similar arrangement for production of annual retail statistics is being investigated but has not yet resulted in a publication. Likewise, annual statistics for selected service industries based on tax data for the smaller units have been proposed, but have not passed the experimental stage. This situation reflects, in part, the fact that there is no immediate prospect of taking another census of wholesale, retail and other service trades. (The last one was in 1971.)

Tax data have been used to supplement agricultural surveys, beginning on a small scale for the 1981 reference year. They have also provided the basis for a periodic small business publication and their use for a new small area program is being investigated.

#### 1.2 TAX AND SURVEY DATA: COMPARABILITY

The degree of comparability between tax and survey data is an important factor in determining the extent to which the former can replace the latter. There are several possible reasons for a lack of comparability. Firstly, there may be conceptual differences between the requirements of the Income Tax Act and of the survey for (nominally) the same data item. As an example, tax and survey definitions of "depreciation"

actually do not coincide. Secondly, the incidence and types of respondent errors may be quite different. A small business will usually arrange for an accountant to complete its tax return but it may respond to a survey questionnaire with rough estimates, or by reference to previously prepared accounts, including those prepared for tax purposes, or by sending the questionnaire to its accountant for completion. Thirdly, the incidence and types of data capture errors may vary according to the processing procedures. For certain data items, Statistics Canada may have tighter quality control than Revenue Canada, or conversely. Extraction of some items from a tax return for survey purposes may require considerable judgement, making such items more prone to data capture error than items taken directly from a questionnaire (where the respondent has had to exercise the judgement).

A number of tax/survey comparison studies were conducted when tax data were first being introduced into the industrial survey programs ten years ago. The studies covered the major industries - construction, manufacturing, wholesale, retail and surface transport. In each study, sample data for comparison purposes were created by linking tax and survey files at unit level. But there were a number of limitations. Linkages were carried out, in part, through automatic matching and were not comprehensively validated. This factor will have contributed to an increase of unknown size in the measured differences between tax and survey values. Furthermore, the matched records were not the result of a selection process with known probabilities.

Subject to the above limitations, some valuable results, guiding the development of procedures for use of tax data, were For example "gross business income" from T1 tax obtained. returns was compared with "total sales" reported to manufacturing survey. The values were found to be within +1% for about 65% of the sample. For the remainder, they differed, often substantially, being divergent by a factor of 2 or more for 7% of the total sample. The high proportion of virtually coincident values leads to the conjecture that, in these cases, the respondent was reporting the same figure for both survey and tax purposes regardless of slight differences in concept between gross business income and sales. In other words, response to the survey was based on information prepared for tax purposes. On the other hand, the very wide range of different values suggests that when the survey respondent did not refer to his tax accounts, the discrepancies between survey and tax figures could be large. Bearing in mind that survey data are more prone to random reporting error and tax data to systematic bias, it is an open question of which data are likely to be closest to the "true" (target) values. This question is presently being addressed.

Comparison of other matched pairs of tax and manufacturing

survey data items such as net profits, wages and salaries, purchases, etc., produced even greater tax/survey differences than for the gross business income - sales case. In particular it was apparent that tax "purchases" and survey "purchases" were not comparable and that redefinition of this tax item was necessary in order for it to be considered as a survey replacement. Similar results were obtained from comparison of data from T2 tax returns and the 1975 survey of wholesale merchants, and were confirmed in a subsequent study based on data for the 1982 reference year.

Further insight regarding the incidence of errors in extracting and capturing tax data has been provided by a recent study comparing values obtained independently by Statistics Canada and Revenue Canada for the same data items from the same returns. Some substantial differences were detected. It was found for example that the figures for gross business income were within +1% for 91% of the units but differed by a factor a 2 or more for 2% of the units. For depreciation the corresponding figures were 84% and 5%, and for net profit 74% and 7%, respectively. Industrial classification of units to 80 major industry groups based on the business description provided by the tax filer (supplemented by reference to earnings and balance sheet statements) resulted in only 68% agreement.

#### 1.3 EXPERIENCE WITH THE CENSUS OF CONSTRUCTION

Nothwithstanding the limitations of tax data there are substantial benefits to be derived from their use, as illustrated by reference to the Census of Construction. This program, which, despite its name, is actually a sample survey, has made substantial usage of tax data since 1975 reference year. The universe is partitioned on the basis of revenue into "larger" and "smaller" units. The larger units are mailed a full scale survey questionnaire. The remainder of the universe is covered by a sample of tax returns, complemented by a short supplementary questionnaire mailed to a sub-sample of the tax sample. The questionnaire requests percentage breakdowns of the type of work undertaken by each respondent, from which SIC codes can be derived, and it requests average hourly wage rates and annual salaries. For the 1983 reference year, there were about 8,000 larger units, defined as having revenues of at least \$750,000; there were some 100,000 smaller units covered by a sample of 40,000 tax returns and by a sub-sample of short questionnaires eliciting about 15,000 responses.

Over a ten year period, the procedures for combining tax and survey data have been modified and fine-tuned to maximize use of tax information whilst preserving and improving the quality of the output. A bank of tax derived business descriptions has

been built up to facilitate assignment of an SIC code to a tax return. Comparisons between tax and supplementary survey data have lead to increased use of the former to identify inactive units and to provide wage and salary estimates. The following paragraphs elaborate upon these experiences.

When an inconsistency is detected between tax and survey, data as regards to whether a unit is active, or whether there are paid employees, then the tax information is preferred. This is a reversal of the initial policy under which survey data were given priority, and it indicates growth of confidence in the tax source. It is now considered that, with respect to activity and employment indicators at least, a respondent is less likely to give an erroneous response on a tax return than on a survey questionnaire. In particular, it is belived that survey respondents sometimes show a tendency to provide current year rather than reference year values for these items.

One of the problems which sets a limit to the use of tax data stems from the fact that Revenue Canada does not require a fixed format for the financial statements associated with the tax return. This gives rise to difficulty in assignment of values to items which do not appear explicitly in a particular set of financial statements but which could be expected to do so. For example, "employee benefits" or "sub-contracting expenses" may be included implicitly in direct cost items or may appear separately, under administrative costs, according to the accounting practice of the respondent.

In the case of a difference between tax and survey based \( \)
SIC codes, the latter is preferred. It is believed that, for most units in the construction sector, the survey "type of work" breakdown provides a more reliable SIC than does the tax filer's free format description of his/her activity. (This belief has not been put to a formal test.) However the bank of business descriptions and the experience of assigning SIC codes over the years has lead to improvements in both the proportion of tax returns which can be properly coded and the accuracy of the codes. For example, the proportion of returns which were assigned SIC codes rose from 77% in 1977 to 92% in 1982. Over the same period, the proportion of these codes in agreement with the corresponding survey based code went from 78% to 83%.

#### 1.4 CONCLUDING REMARKS

Over the past ten years the benefits of tax data in reducing operating costs and response burden have been clearly demonstrated in practice. Though studies have indicated some potential pitfalls in terms of quality, there is no doubt that the more tax data are used, the more effective usage will

become.

An opportunity to consolidate experience to date, to undertake further studies of quality and to rationalize the somewhat uneven growth in tax data usage is now arising through the new Business Survey Redesign Project. This undertaking, initiated in the latter half of 1984 will have had a substantial effect on virtually all economic surveys by the time it is completed in 1988. The twin foci of the project are the enhancement of tax data usage, and the complete overhaul of the procedures for provision of frames for economic surveys.

Maintenance strategy are the quality and costs of SIC codes which can be obtained from tax returns. Thus a study has begun to determine the respondent errors in providing business descriptions, the clerical errors in assigning codes based on such descriptions and the potential for introduction of automated coding procedures. In a second study, the current methods of sampling tax returns are being revised in the light of new technology and of new requirements, e.g., for longitudinal data. The prospect of increased coordination of Revenue Canada and Statistics Canada tax processing activities is being explored, and a general strategy for use of tax data by all annual industrial surveys has been formulated along the lines of the present Census of Construction usage. By the end of the project the knowledge and use of tax data for economic surveys will have increased significantly.

#### PART 2. SOCIAL DATA. AN EXAMPLE OF THE PERSONAL TAX FILES

#### 2.0 OPENING REMARKS

In the case of social data, a systematic effort to develop data on individuals, families and households from administrative records was initiated in the late 1970's. This work had the explicit objective of developing inter-censal small area social data, and concentrated initially on the personal income tax file. This file offered opportunities for small area analysis that were previously only possible for national, provincial, and large metropolitan areas of Canada, except in census years.

In comparison with the business data work described in Part 1, this work was focussed on the utilization of administrative data to supplement survey (or census) data, rather than on the construction or maintenance of frames. However, an important element in the evaluation of these data has been the extent to which they cover the population and its sub-groups as defined for census purposes. Since the coverage of administrative records is determined by administrative regulations and considerations, they rarely conform exactly to a desired statistical population. They may entirely omit certain sub-groups, under-represent others, and provide almost complete coverage of yet other sub-groups. In this part of the paper, an assessment is made of the coverage of the personal income tax file by making selected comparisons with data from alternative sources.

Although a variety of possibilities exist for the direct comparison of data, a generic formulation of <u>coverage</u> rates is defined as follows:

Measurement for Tax-Data Variable Y X 100
Measurement for Similar Variable,
but from an Alternative Data Source

While Y could include many different demographic counts and dollar amounts, the discussion and findings in this paper have been limited to demographic comparisons and to a few comments on Statistics Canada's success in deriving annual migration data from administrative records. Other uses of the personal income tax file not covered by this paper are the regular derivation of small area income and gross labour force data.

#### 2.1 DEMOGRAPHIC COVERAGE RATES

The coverage rates for demographic variables can be derived for the basic population, and for age and sex attributes.

## 2.1.1 Basic Population versus Taxfiler Counts, 1981

The 1980 personal income tax file (T1) is for the period of about March/April 1981, and it can be compared to the data collected in June, 1981 for the census program. Initially, it is possible to compare the count of taxfilers\* to the count of population from the Census. The result is as follows:

## TABLE 2.1. TAXFILER TO CENSUS OF POPULATION COMPARISON, 1981

No. of Taxfilers, Spring 1981	14,525,984
Census Population, June 1981	24,343,175
Percent Coverage	59.7%

This direct comparison can also be made for sub-national data such as provinces and cities. Table A.1 is appended with similar coverage rates for the provinces and territories as shown in column 5.

## 2.1.2 Basic Population versus Imputed Taxfiler Counts, 1981

While each tax filer can be considered independently, many taxfilers have dependents that can be imputed by using the dollar value of the personal exemptions. This imputation was undertaken for 1981 with the following result for Canada in Table 2.2 below.

<sup>\*</sup> The taxfiler counts for Tables 2.1, 2.2, 2.3 and 2.4 are after edit and imputation (i.e., records with missing values or illegitimate values may have values imputed.

TABLE 2.2 IMPUTED TAXFILER POPULATION, SPRING 1981, COMPARED TO CENSUS POPULATION, JUNE 1981 (000'S)

		POPULATION	
	TOTAL	MALES	FEMALES
Imputed Taxfiler Population	23,957	11,327	12,631
Census Population	24,343	12,068	12,275
Percent Coverage	98.48	93.88	102.9%

Once the taxfiler population is adjusted for the population of reported dependents, the coverage rates change considerably. For example, while the coverage appears to be 59.7 percent compared to the Census of Population in Table 2.1, the coverage of the Total Population by the imputed taxfiler population rises to 98.4, column 1 of Table 2.2.

There is, nevertheless, some double-counting and over-counting in the imputations. Since some taxfilers are both taxfilers and dependents, the over-counting phenomenon can be noted in column three of Table 2.2, where the estimated coverage of females in 1981 was 102.9 percent. While efforts are made to eliminate double-counting, it is not possible to eliminate it entirely. As for over-counting, there is no known method of eliminating the count of individuals who are legally claimed as dependents, but who are not residents of Canada. (Column 6 of Table 4.2) shows the coverage of the imputed population for provinces and territories by sex.)

More specifically, and with regard to double-counting for the female population, a major source of over-coverage can be attributed to the Child Tax Credit (CTC). When the CTC was introduced for the 1978 tax year, the coverage of imputed female taxfilers increased by more than 10%. Women who had previously appeared on the tax file only as dependents on their husbands'

tax returns began filing their own tax returns as well as being claimed as dependents. These women are therefore being counted twice: as taxfilers and as imputed dependents.

### 2.1.3 Coverage Rates by Age, 1981

Comparisons of the coverage rates by age reflect the lower rates of taxfiling in both the higher and lower age groups,

those less likely to be receiving income for which they have a tax liability (Table 2.3).

TABLE 2.3. IMPUTED TAXFILER POPULATION BY AGE, SPRING 1981, COMPARED TO CENSUS POPULATION, JUNE 1981

(1) AGE	(2) CENSUS	(3) NUMBER OF	(4) TAXFILERS+	(5) COVERA	(6) GE RATES
GROUPS	POPULATION	TAXFILERS	DEPENDENTS	(3)/(2)	(4)/(2)
<18	6,845,150	291,272	6,636,631	4.2	96.9
18-24	3,294,655	2,823,707	3,622,795	85.7	110.0 +
25-44	7,184,345	6,511,222	7,748,510	90.6	107.8
45-64	4,658,070	3,640,849	4,488,168	78.2	96.4
65+	2,360,970	1,258,934	1,461,164	53.3	61.9

Again, most of the women who file for the CTC fall into either the 18-24 or 25-44 age group, leading to double-counting in these groups. Another source of double-counting in the 18-24 age group for both males and females is children who earn some money and file their own tax return (usually to obtain a tax refund for taxes withheld by their respective employers), but are still claimed as dependents by their parents.

#### 2.2 MISSING DATA BY AGE, SEX AND MARITAL STATUS

A final demographic consideration pertains to the extent to which the basic demographic attributes are absent from the tax returns. Thus, looking solely at the population of taxfilers for a few recent years, it is possible to note the frequency (and percentage) with which these data items are not reported by taxfilers.

From the following table, it can be observed that the frequency of missing data is, on average, fairly low. This table does not indicate, however, to what extent the data are accurate (i.e., respondent reporting errors).

TABLE 2.4. PROFILE OF MISSING DATA BY AGE, SEX AND MARITAL STATUS DATA, PERSONAL INCOME TAX FILE: 1981, 1982, 1983

		FREQUENCY OF SEX	MISSING AGE	DATA FOR MARITAL STATUS	NUMBER OF TAXFILERS BEFORE E&I
1981 Missing	90	40,678	2,854	35,629 0.23	15,231,185
1982 Missing	do	42,043	13,477	122,029	15,238,257
1983 Missing	용	38,090 0.25	20,305	115,379	15,336,355

#### 2.3 MIGRATION DATA

Statistics Canada is currently using the annual personal income tax records as a source of annual migration data between geographical areas (between the ten provinces and two territories, and between the 266 Census Divisions).

The migration data are obtained by assigning a Census Division code to each individual record. For the majority of records, this assignment is based on the postal code attached to the taxfiler's address. Two successive years of tax data are compared and if there is a change in codes for the two years, it is assumed that the taxfiler migrated. A separate migrant file is created and the demographic information and tax exemptions information are used to estimate the total number of persons to be counted as migrants. In the final stage, since the tax file coverage is somewhat less than the coverage of the total population, the data are given a small adjustment to estimate the population of migrants from year to year.

After careful review in the early 1980's, a number of independent analyses indicated that no other source of migration data produced estimates as accurate as those from the taxfiler migrant file. As a consequence, it was concluded that in the post-1981 population estimates program, the final estimates would be based on the taxfiler migrant data file.

#### 2.4 FURTHER DEVELOPMENTS

In reviewing the coverage of the population in this part of the paper, it appears that the taxfiler counts plus dependents provide a reasonably high coverage of the population compared to the 1981 Census. When these comparisons are made by age and sex, it is apparent that the coverage rates depart from the broad population coverage. In particular, there is over-coverage for females (primarily in the 18-24 and the 25-44 age ranges), and undercoverage for the 65+ population.

Two separate developmental activities are underway at this time to address the coverage problems. One project is directed at the linkage of spouses as a means of reducing the overcounting of the female population. Additionally, one part of this project is directed at the linkage of children to their respective parents through address linkages as a means of reducing the over-counting of dependent children who file tax returns.

Secondly, while at least one third of the 65+ population does not file a tax return, virtually 100 percent of the population aged 65+ is eligible for and receives a monthly Old Age Security (OAS) benefit. It is probable that the tax file and OAS files can be merged on a geographical building block basis to create a geographically-linked data set that will ensure that the enhanced file will have a much improved coverage of the 65+ population.

Thus, while the overall coverage of the population is encouraging at 98.4 percent, the over- and under-counting dimensions are uneven. With the enhancements that are currently being investigated, it is anticipated that these sub-population coverage problems will be reduced over the next one to two years.

#### 2.5 CONCLUDING REMARKS

The principal objective in preparing Part 2 of this paper was to make a series of demographic comparisons between data derived from the personal income tax file and data derived from the 1981 Census of Population. Some general conclusions appear noteworthy:

o The imputed taxfiler population for 1981 (i.e., the sum of actual taxfilers plus imputed dependents) closely approaches the 1981 Census count (i.e., 24.0 million imputed taxfiler population versus 24.3 million counted in the census).

- The analysis of the 1981 imputed taxfiler population indicates that the coverage of the female portion of the population is higher than the male portion (102.9% versus 93.8%, respectively). The over-coverage of the female population is largely a result of double-counting, a problem that is being addressed independently at this time.
- o The imputed taxfiler population comparison also indicates that over-counting occurs in the 18-24 and 25-44 age groups.
- Overall, data for the variables sex, age and marital status are rarely missing.
- No other source of post-censal migration data provided better estimates overall than those derived from the annual personal income tax records.

#### CONCLUSIONS

The applications described in this paper are just two among several examples of administrative data use within Statistics Canada. Other notable areas of developmental work include the use of unemployment insurance files in conjunction with Labour Force Survey data to produce small area labour market measures, the utilization of old age security records alluded to in Part 2, and the development of family income estimation from the personal income tax file.

The usefulness of administrative data for statistical purposes is often limited either because the underlying concepts or definitions differ from the required statistical concepts, or because key variables required for analytical purposes are unavailable or of poor quality. In many cases a relatively minor change to an administrative procedure or form could greatly enhance the statistical usefulness of the administrative file without jeopardizing the primary administrative function. There is a need to develop mechanisms that would allow the statistical agency to influence the design of administrative systems so as to optimize their statistical value.

Tax data have become an essential element in Statistics Canada's business data programs, both as a source of frame data and as a source of response data primarily for small businesses. In the case of data on individuals, the tax data cannot be considered an alternative to traditional data derived from cen-

suses or surveys, but they constitute an important source of information that can be used to fill in inter-censal or small area detail. Thus, on both the economic and social sides, tax data complement data from surveys and censuses. The challenge for the statistical agency is to determine a mixture of these three data sources that takes advantage of the strengths of each at a minimum cost.

## APPENDIX

TABLE A.1. Population Comparison, by Province, 1981

(1) PROV	(2) CENSUS POPULATION	(3) NUMBER OF TAXFILERS	(4) TAXFILERS+ DEPENDENTS	COVERA	(6) GE RATES (4)/(2)
WELL D	567 600	204 745	562 752	£7 09	99.18
NFLD	567,680	294,145	562,752	51.8%	93.88
PEI	122,505	65,874	114,896	53.8%	
NS	847,445	478,039	829,739	56.48	97.98
NB	696,405	387,951	678,017	55.78	97.48
QUE	6,438,405	3,669,789	6,128,488	57.0%	95-28
ONT	8,625,110	5,370,843	8,761,173	62.3%	101.68
MAN	1,026,240	661,312	1,071,698	64.48	104.48
SASK	968,315	543,259	925,356	56.1%	95.68
ALTA	2,237,725	1,326,725	2,141,267	59.38	95.78
BC	2,744,465	1,690,495	2,677,875	61.68	97.68
YUKON	23,150	14,169	22,278	61.28	96.28
NWT				50.98	95.68
IAM T.	45,745	23,280	43,729	30.98	33.00
CANADA	24,343,175	14,525,984	23,957,268	59.78	98.48

TABLE A.2. Population Comparison, By Sex and Province, 1981

MALES					
(1) PROV	(2) CENSUS POPULATION	(3) NUMBER OF TAXFILERS	(4) TAXFILERS+ DEPENDENTS	(5) COVERAGE (3)/(2)	RATES
NFLD	285,690	158,476	263,713	55.5%	92.38
PEI	60,940	34,690	54,403	56.98	89.38
NS	419,575	256,023	388,812	61.08	92.79
NB	346,005	205,514	317,151	59.48	91.79
QUE	3,172,200	1,945,320	2,860,212	61.3%	90.29
ONT	4,246,790	2,816,990	4,139,461	66.38	97.5
MAN	506,510	341,816	502,569	67.5%	99.2
SASK	486,075	297,457	442,055	61.28	90.9
ALTA	1,143,220	725,338	1,044,845	63.48	91.4
BC	1,365,155	900,999	1,280,808	66.08	93.8
YUKON	12,175	7,741	11,163	63.68	91.7
NWT	23,965	12,604	21,558	52.68	89.9
CANADA	12,068,285	7,702,968	11,326,750	63.8%	93.89
FEMALE.	<u>s</u>				
FEMALE.	(2)	(3)	(4)	(5)	(6)
(1)	(2) CENSUS	NUMBER OF	TAXFILERS+	COVERAGI	E RATES
(1)	(2)				E RATE
(1) PROV	(2) CENSUS	NUMBER OF	TAXFILERS+	COVERAGI	106.0
(1) PROV NFLD	(2) CENSUS POPULATION	NUMBER OF TAXFILERS	TAXFILERS+ DEPENDENTS 299,039 60,493	COVERAGE (3)/(2)	E RATE: (4)/(2)
(1) PROV NFLD PEI	(2) CENSUS POPULATION 281,990	NUMBER OF TAXFILERS	TAXFILERS+ DEPENDENTS 299,039	COVERAGE (3)/(2) 48.1%	106.09 98.2
(1) PROV NFLD PEI NS	(2) CENSUS POPULATION 281,990 61,565	NUMBER OF TAXFILERS 135,669 31,184	TAXFILERS+ DEPENDENTS 299,039 60,493	COVERAGE (3)/(2) 48.1% 50.6%	106.09 98.29 103.09
(1) PROV NFLD PEI NS NB	(2) CENSUS POPULATION 281,990 61,565 427,870	NUMBER OF TAXFILERS 135,669 31,184 222,016	TAXFILERS+ DEPENDENTS  299,039 60,493 440,927 360,866 3,268,276	COVERAGE (3)/(2) 48.1% 50.6% 51.9%	106.09 98.29 103.09
(1) PROV NFLD PEI NS NB QUE	(2) CENSUS POPULATION 281,990 61,565 427,870 350,400	NUMBER OF TAXFILERS 135,669 31,184 222,016 182,437	TAXFILERS+ DEPENDENTS  299,039 60,493 440,927 360,866	COVERAGE (3)/(2) 48.1% 50.6% 51.9% 52.1%	106.00 98.20 103.00 100.10
(1) PROV  NFLD PEI NS NB QUE ONT	(2) CENSUS POPULATION 281,990 61,565 427,870 350,400 3,266,205	135,669 31,184 222,016 182,437 1,724,469	TAXFILERS+ DEPENDENTS  299,039 60,493 440,927 360,866 3,268,276	COVERAGE (3)/(2) 48.1% 50.6% 51.9% 52.1% 52.8%	106.0 98.2 103.0 100.1 105.6
(1) PROV  NFLD PEI NS NB QUE ONT MAN	(2) CENSUS POPULATION 281,990 61,565 427,870 350,400 3,266,205 4,378,320	NUMBER OF TAXFILERS 135,669 31,184 222,016 182,437 1,724,469 2,553,853	299,039 60,493 440,927 360,866 3,268,276 4,621,712	COVERAGE (3)/(2) 48.1% 50.6% 51.9% 52.1% 52.8% 58.3%	106.0: 98.2: 103.0: 100.1: 105.6: 109.5:
(1) PROV  NFLD PEI NS NB QUE ONT MAN SASK	(2) CENSUS POPULATION 281,990 61,565 427,870 350,400 3,266,205 4,378,320 519,730	NUMBER OF TAXFILERS 135,669 31,184 222,016 182,437 1,724,469 2,553,853 319,496	TAXFILERS+ DEPENDENTS  299,039 60,493 440,927 360,866 3,268,276 4,621,712 569,129	COVERAGE (3)/(2) 48.1% 50.6% 51.9% 52.1% 52.8% 58.3% 61.5%	106.0: 98.2: 103.0: 100.1: 105.6: 109.5:
	(2) CENSUS POPULATION 281,990 61,565 427,870 350,400 3,266,205 4,378,320 519,730 482,240	NUMBER OF TAXFILERS 135,669 31,184 222,016 182,437 1,724,469 2,553,853 319,496 245,802	TAXFILERS+ DEPENDENTS  299,039 60,493 440,927 360,866 3,268,276 4,621,712 569,129 483,301	COVERAGE (3)/(2) 48.1% 50.6% 51.9% 52.1% 52.8% 58.3% 61.5% 51.0%	106.0: 98.2: 103.0: 103.0: 100.1: 105.6: 109.5: 100.2:
(1) PROV  NFLD PEI NS NB QUE ONT MAN SASK ALTA BC	(2) CENSUS POPULATION 281,990 61,565 427,870 350,400 3,266,205 4,378,320 519,730 482,240 1,094,505	NUMBER OF TAXFILERS 135,669 31,184 222,016 182,437 1,724,469 2,553,853 319,496 245,802 601,490	TAXFILERS+ DEPENDENTS  299,039 60,493 440,927 360,866 3,268,276 4,621,712 569,129 483,301 1,096,422	COVERAGE (3)/(2) 48.1% 50.6% 51.9% 52.1% 52.8% 58.3% 61.5% 51.0% 54.9%	106.0: 98.2: 103.0: 103.0: 100.1: 105.6: 109.5: 100.2: 100.2:
(1) PROV NFLD PEI NS NB QUE ONT MAN SASK ALTA	(2) CENSUS POPULATION 281,990 61,565 427,870 350,400 3,266,205 4,378,320 519,730 482,240 1,094,505 1,379,310	NUMBER OF TAXFILERS  135,669 31,184 222,016 182,437 1,724,469 2,553,853 319,496 245,802 601,490 789,496	299,039 60,493 440,927 360,866 3,268,276 4,621,712 569,129 483,301 1,096,422 1,397,067	COVERAGE (3)/(2) 48.1% 50.6% 51.9% 52.1% 52.8% 58.3% 61.5% 51.0% 54.9% 57.2%	E RATES

CAHIER DE TRAVAIL NO. BSMD-86-009F/E
MÉTHODOLOGIE

LES DONNÉES ADMINISTRATIVES COMME COMPLÉMENT DES DONNÉES D'ENQUÊTE: EXPÉRIENCES À STATISTIQUE CANADA

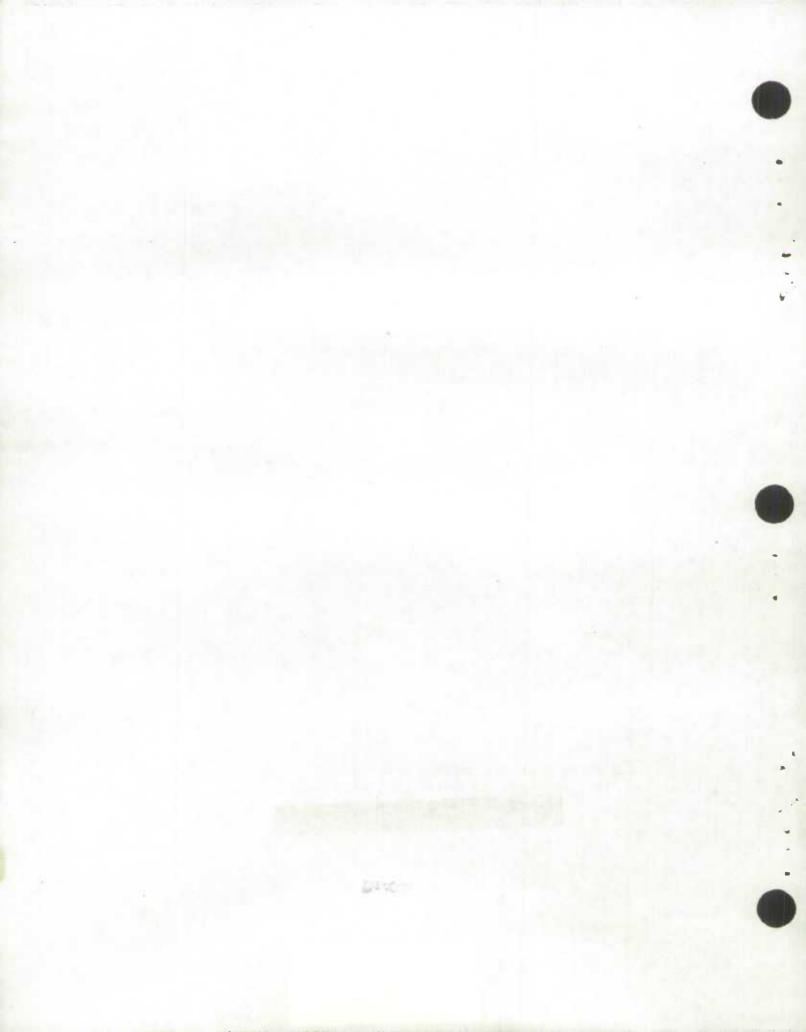
Document rédigé par Gordon Brackstone, Michael Colledge et John Leyes, Direction de la méthodologie, Statistique Canada.

200740

AGAMAD SUDITICIS CAMADA (IBRARY)
AGAMAD SUDITICITICS CAMADA
AGAMAD SUDITICITICS CAMADA

7.2

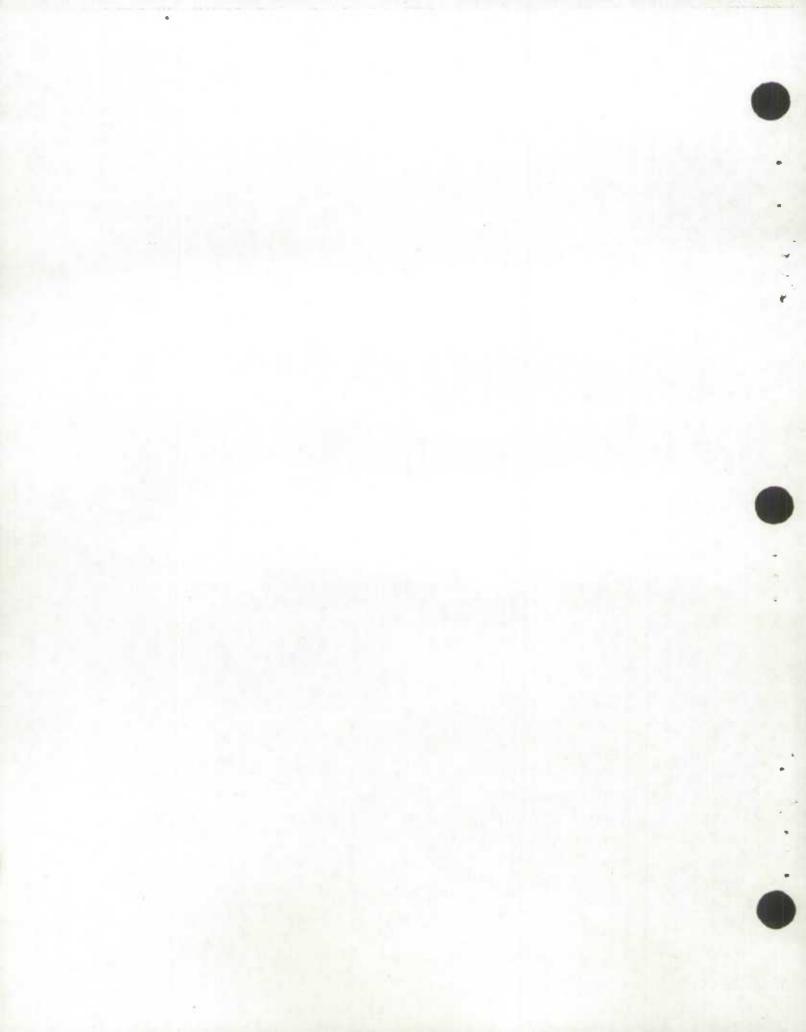
<sup>\*</sup> Il s'agit d'une version préliminaire. Ne pas citer sans la permission des auteurs. Nous aimerions recevoir vos commentaires.



LES DONNÉES ADMINISTRATIVES COMME COMPLÉMENT DES DONNÉES D'ENQUÊTE : EXPÉRIENCES À STATISTIQUE CANADA

Document rédigé en prévision de la Rencontre de la Conférence des statisticiens européens sur les méthodes statistiques, février 1986

Document rédigé par Gordon Brackstone, Michael Colledge et John Leyes, Direction de la méthodologie, Statistique Canada.



#### INTRODUCTION

Les documents administratifs constituent depuis longtemps une source importante de données statistiques. En effet, dans des domaines comme la statistique de l'état civil, les importations et les exportations, pour de nombreux aspects des soins de santé et de l'éducation, les données statistiques sont en régle générale un sous-produit du processus administratif. Par ailleurs, les préoccupations exprimées ces derniers temps au sujet du fardeau de réponse et de l'augmentation des coûts de collecte des données ont pour résultat de favoriser l'étude et l'utilisation de documents administratifs comme source de données statistiques dans des domaines spécialisés pour lesquels on devait auparavant s'en remettre entièrement aux résultats de recensements et d'enquêtes.

La qualité des données statistiques est une source de préoccupation constante pour les responsables des recensements et des enquêtes. Pareillement, le souci de produire des données administratives de qualité pour des besoins statistiques est grand, en particulier lorsque la conception du processus de collecte de ce genre de données n'a pas bénéficié des connaissances d'un statisticien.

Dans le présent exposé, on examine deux domaines particuliers d'application des données administratives à Statistique Canada et on met l'accent sur la qualité et la comparabilité de cette information et des sources de renseignements conventionnelles. Dans la partie 1, on analyse l'utilisation des données fiscales des entreprises, de même que des résultats d'enquêtes auprès des entreprises, comme source de données économiques; dans la partie 2, on traite de l'utilisation des dossiers de l'impôt sur le revenu des particuliers pour produire des données démographiques.

## PARTIE 1. DONNÉES ÉCONOMIQUES UTILISATION DES DONNÉES SUR L'IMPÔT SUR LE REVENU

## 1.0 CONSIDÉRATIONS PRÉLIMINAIRES

Les raisons fondamentales pour lesquelles les données sur l'impôt sur le revenu sont utilisées comme complément ou en remplacement des données d'enquêtes économiques, c'est qu'elles peuvent se traduire par une diminution nette des coûts de collecte des données par un accroissement de la variété des données produites et par une diminution du fardeau de réponse. Aussi, en accordant à Statistique Canada l'accès aux dossiers fiscaux de Revenu Canada, aux termes de la Loi sur la statistique de 1971, on reconnaissait explicitement l'importance de ces avantages.

Au Canada, il y a deux genres de déclaration fiscale, la formule T1 et la formule T2. La formule T1 est remplie par un particulier. Ce particulier doit inclure les états financiers de toutes les entreprises individuelles dont il est l'unique propriétaire ou l'un des associés. Revenu Canada n'exige pas que ces états financiers soient produits selon une présentation uniforme. Chaque formule T1 est identifiée par le numéro d'assurance sociale (NAS) du déclarant ou, dans le cas des déclarants qui n'ont pas de NAS, par un numéro unique attribué par Revenu Canada. Les formules T2 sont remplies par les sociétés constituées et par certaines autres personnes morales qui doivent faire une déclaration pour le calcul de l'impôt sur le revenu. Comme dans le

cas des formules T1, les états financiers joints aux déclarations peuvent être présentés sous n'importe quelle forme. Revenu Canada attribue à chaque déclarant un numéro T2 qui, en principe, est unique et ne peut être modifié avec le temps. Ce numéro ne peut être en aucune façon apparenté à un autre numéro d'identification d'une société, par exemple le numéro de sa charte ou les numéros de ses comptes de retenue sur la paye (ces documents sont également tenus par Revenu Canada).

Les données fiscales sont présentées à Statistique Canada sous deux formes, soit des fichiers ordinolingues et les documents de base T1 et T2. Pour les enquêtes économiques, la plupart des éléments d'information, notamment les traitements et salaires, l'amortissement, le bénéfice net, les stocks, les achats, etc., peuvent être tirés des états financiers joints aux déclarations mais pas des formules T1 et T2. Comme cette information n'est pas saisie par Revenu Canada, elle n'est pas produite sous forme de fichiers ordinolingues. Revenu Canada ne traite pas également la description des entreprises fournie par les déclarants, laquelle sert à l'attribution des codes de la classification des activités économiques (CAÉ) relativement à l'activité économique déclarée. Par conséquent, Statistique Canada administre un vaste programme d'échantillonnage des déclarations fiscales, qui consiste à sélectionner des déclarations au moment de leur traitement à Revenu Canada, à attribuer les codes CAÉ, à tirer des copies et, ultérieurement, à saisir les autres renseignements d'intérêt.

Les données de l'impôt sur le revenu peuvent servir à deux fins distinctes quoique étroitement liées. Tout d'abord, elles facilitent la construction et la mise à jour des bases utilisées pour les enquêtes économiques. Ensuite, elles peuvent servir à compléter ou remplacer les données financières qui, autrement, devraient être recueillies lors d'enquêtes annuelles. Dans ce contexte, il convient de faire une distinction entre les enquêtes pour lesquelles l'unité de déclaration fiscale est également l'unité déclarante, et celles pour lesquelles ces unités ne le sont pas. De toute évidence, les enquêtes pour lesquelles ces unités sont semblables offrent beaucoup plus de possibilités d'application des données fiscales.

Toutes ces questions sont traitées plus en détail dans les sections suivantes: la section 1.1 décrit les applications actuelles des données fiscales; la section 1.2 résume les résultats de certaines études de comparaison des données fiscales et des données d'enquête; la section 1.3 décrit l'expérience acquise pour une enquête particulière qui tire profit au maximum des données fiscales; enfin, la section 1.4 décrit certains projets futurs.

#### 1.1 APPLICATIONS DES DONNÉES SUR L'IMPÔT SUR LE REVENU

Il y a deux enquêtes pour lesquelles l'information fournie dans les déclarations T2 (sociétés constituées en corporation) est la principale source de données financières, et non simplement une information complémentaire. Il s'agit, premièrement, de l'enquête annuelle sur les sociétés, qui est fondée entièrement sur l'information fournie dans les déclarations T2. Dans cette enquête, l'unité de déclaration est définie, sauf exception, de façon à correspondre aux déclarants qui produisent une formule T2. Les données publiées fournissent des renseignements sur l'actif, l'avoir des actionnaires, les recettes, les dépenses, les bénéfices, etc. Deuxièmement, conformément à

la Loi sur les déclarations des corporations et des syndicats ouvriers (CALURA), Statistique Canada doit recueillir des données techniques et financières, de même que de l'information sur la propriété des sociétés, auprès de grandes "entreprises" classées selon la propriété et le contrôle. Les données financières produites dans cette enquête continueront, comme dans le passé, à être tirées des déclarations T2.

Les données fiscales sont utilisées à des degrés différents à suppléer la collecte directe d'information dans les enquêtes annuelles sur des secteurs d'activité tels que la fabrication, le commerce, les services, les transports et la construction. Le manque d'uniformité des méthodes de collecte dans ces enquêtes est en partie inhérent à ces secteurs d'activité, en partie attribuable à la variété des données jugées les plus importantes et, enfin, il tient à des facteurs historiques. Dans toutes ces enquêtes, l'unité statistique visée est définie par Statistique Canada et est désignée par le terme "établissement". Le degré d'utilisation des données fiscales dépend du degré de concordance de l'établissement et du déclarant fiscal. De façon qénérale, on peut considérer que les activités d'un petit établissement indépendant sont susceptibles d'être prises en compte dans une seule déclaration fiscale, alors que les liens entre les grands établissements et leur déclaration fiscale sont beaucoup plus complexes. Les données relatives aux établissements qui sont exploités par une même entreprise peuvent être consolidées dans la même déclaration fiscale ou peuvent être fournies sur plusieurs déclarations. Aussi, il y a plus de possibilités d'utilisation des données fiscales pour compléter les données d'enquête dans le cas des secteurs d'activité qui comptent une forte proportion de petites unités, par exemple les sous-traitants dans les métiers de la construction, que dans le cas des branches d'activité comprenant principalement de grandes unités, comme par exemple dans la fabrication. Des exemples d'utilisation des données fiscales sont présentés ci-après, de même que certaines observations sur les raisons pour lesquelles ces utilisations diffèrent entre les secteurs d'activité.

Parmi toutes les enquêtes annuelles sur les secteurs d'activité, c'est le recensement de la construction qui utilise le plus les données fiscales. Depuis l'année de référence 1975, les dossiers fiscaux servent à répartir l'univers du secteur de la construction en petites et grandes unités, et à attribuer les codes CAÉ, de même qu'à recueillir des renseignements financiers sur les petites unités. En 1983, les données fiscales portaient sur 93% des unités, représentant 28% du total des recettes du secteur de la construction. Cette enquête est décrite en détail dans la section 1.3.

On utilise les données fiscales également pour le recensement des manufactures, depuis l'année de référence 1975. Dans ce cas, les données sont utilisées pour représenter les petites unités qui couvrent 30% de l'univers et interviennent pour 3% des recettes globales. Cette information ne sert pas à définir l'univers ni à fournir une classification par branche d'activité. Il convient de souligner ici qu'il y a trois fois plus de codes CAÉ pour le secteur de la fabrication que pour celui de la construction. Le classement par branche d'activité est déterminé au moyen d'un questionnaire abrégé conçu spécialement pour les petites unités. Ce questionnaire sert également à la collecte de données auprès des petites unités dont on ne connaît pas les codes d'identification de dossier fiscal (NAS et numéro de T2), et à la collecte de données de nature non financière qui ne figurent pas sur une déclaration fiscale.

Les dossiers fiscaux ont été utilisés pour prendre en compte les petites unités dans l'enquête auprès des transporteurs routiers de 1975, mais cette mesure a été supprimée à cause des réserves exprimées au sujet de la fiabilité des données recueillies. Ces réserves étaient liées à la fois à l'attitude des responsables des enquêtes vis-à-vis de ce genre de données et à des arguments de nature technique. Par voie de conséquence, le processus de collecte des données de cette enquête vise actuellement les unités de taille moyenne et grande, ce qui exclut les unités dont les recettes globales sont inférieures à \$100,000.

Pour l'enquête annuelle sur le commerce de gros, on a utilisé, à partir de l'année de référence 1981, des données fiscales au lieu des données d'enquête dans le cas des petits grossistes, soit pour 70% des unités qui représentaient 7% des recettes. Il n'existe pas d'équivalent au questionnaire abrégé du recensement des manufactures. Les codes CAÉ tirés des dossiers fiscaux sont confirmés, en cas d'incertitude, auprès des unités commerciales Des dispositions semblables concernant la production des correspondantes. statistiques annuelles sur le commerce de détail sont en cours d'examen, mais non pas encore donné lieu à une publication. Pareillement, un projet de production de statistiques annuelles sur certains services fondées sur les données fiscales relatives à des petites unités a été présenté et en est encore au stade expérimental. Cette situation traduit en partie le fait qu'il n'existe pas de projet immédiat de tenir un autre recensement du commerce de gros, du commerce de détail et d'autres services. (Le dernier recensement a eu lieu en 1971.)

Des données fiscales ont servi à compléter les enquêtes sur l'agriculture, sur une petite échelle d'abord, à compter de l'année de référence 1981. Des données fiscales ont également constitué l'information de base d'une publication périodique sur les petites entreprises, et on examine finalement la possibilité de les utiliser dans un programme de données régionales.

## 1.2 COMPARABILITÉ DES DONNÉES FISCALES ET DES DONNÉES D'ENQUÊTE

Le degré de comparabilité des données fiscales et des données d'enquête est un facteur important qui permet de déterminer dans quelle mesure les données fiscales peuvent remplacer les données d'enquête. Il y a plusieurs causes qui peuvent rendre ces données non comparables. Premièrement, il peut y avoir des différences sur le plan des concepts entre les exigences de la Loi sur l'impôt sur le revenu et celles des enquêtes recueillant les éléments d'information ayant les mêmes noms. Par exemple, les définitions d'"amortissement" établies pour les déclarations fiscales et pour les enquêtes ne sont pas semblables. Deuxièmement, la nature et l'incidence des erreurs dues aux répondants peuvent varier beaucoup. Ainsi, une petite entreprise confie habituellement à un comptable la production de sa déclaration d'impôt, mais pourra remplir un questionnaire d'enquête en fournissant des estimations approximatives ou en consultant des comptes établis antérieurement, notamment ceux produits pour l'impôt, ou encore confiera le questionnaire à son comptable. Troisièmement, la nature et l'incidence des erreurs de saisie des données peuvent varier selon la méthode de traitement utilisée. En effet, il peut arriver que, pour certains éléments d'information, Statistique Canada exerce un contrôle qualitatif plus sévère que Revenu Canada, et inversement.

L'extraction de certains renseignements des déclarations fiscales, pour les besoins d'une enquête, peut nécessiter un élément de jugement considérable et risquer d'accroître les risques d'erreurs de saisie des données, par opposition à la collecte directe de renseignements à l'aide d'un questionnaire (dans quel cas le répondant est celui qui a du faire preuve de jugement).

Des études comparatives des données d'enquête et des données fiscales ont été effectuées après que des données fiscales ont été utilisées pour la première fois dans les programmes d'enquête sur les secteurs d'activité, il y a une dizaine d'années. Ces études portaient sur les grands secteurs d'activité, notamment la construction, la fabrication, le commerce de gros, le commerce de détail et le transport terrestre. Dans chacune de ces études, un échantillon de données a été tiré au moyen d'une comparaison des dossiers fiscaux et des dossiers d'enquête au niveau des unités. Ce travail comportait toutefois des limites. En effet les comparaisons étaient effectuées, du moins en partie, par appariement automatique des dossiers et, de ce fait, n'étaient pas vérifiées complètement. Ce facteur a favorisé une augmentation, d'importance inconnue, des différences mesurées entre les chiffres des enquêtes et ceux tirés des déclarations fiscales. En outre, le choix des dossiers appariés n'a pas été fait selon un procédé de sélection avec probabilités connues.

Sous réserve des limites décrites ci-dessus, on a obtenu des résultats précieux qui servent à l'élaboration de méthodes d'application des données fiscales. Par exemple, les chiffres du "revenu d'entreprise brut" déclarés dans les formules I1 ont été comparés aux chiffres des "ventes totales" déclarés dans les questionnaires d'enquête sur les industries manufacturières. Les chiffres étaient à ±1% comparables pour près de 65% des unités de l'échantillon. Dans le cas des autres unités, les chiffres comparés diverqeaient, souvent beaucoup, selon un facteur de 2 ou plus pour 7% des unités de La forte proportion de valeurs pratiquement concordantes l'échantillon. laisse supposer que, dans ces cas, les répondants ont déclaré les mêmes chiffres dans les questionnaires d'enquête et dans la déclaration fiscale, peu importe que les définitions du revenu brut et des ventes aient été légèrement différentes. En d'autres termes, la réponse fournie aux enquêtes était fondée sur l'information établie pour la déclaration d'impôt. Par contre, la très grande variété de valeurs divergentes donne à penser que, lorsque le répondant à une enquête ne consultait pas ses documents fiscaux, les écarts entre les chiffres des enquêtes et ceux des dossiers fiscaux pouvaient être importants. Si l'on garde à l'esprit que les données d'enquête sont susceptibles de comporter davantage d'erreurs de réponse aléatoires et que les données fiscales, des biais systématiques, la question reste posée de savoir quelles données seront probablement les plus proches des valeurs vraies (valeurs visées). Cette question fait actuellement l'objet d'une étude.

La comparaison d'autres paires de données fiscales et de données d'enquête sur les industries manufacturières, par exemple les bénéfices nets, les salaires et traitements, les achats, etc., a produit des écarts encore plus importants entre les données fiscales et les données d'enquête que ceux obtenus dans le cas de la comparaison des chiffres relatifs au revenu tiré d'une entreprise brut et aux ventes totales. En particulier, les "achats" déclarés pour l'impôt sur le revenu et les "achats" déclarés dans une enquête n'étaient pas comparables, d'où la nécessité de redéfinir cet élément d'information fiscale de façon qu'il puisse remplacer les données d'enquêtes

correspondantes. Des résultats semblables ont été obtenus dans la comparaison des données fournies sur les déclarations T2 et sur le questionnaire de l'enquête sur le commerce de gros de 1975; ces résultats ont été confirmés par ceux d'une enquête subséquente fondée sur les données de l'année de référence 1982.

L'incidence des erreurs d'extraction et de saisie de données fiscales a été examinée également dans une étude récente qui comportait la comparaison de valeurs obtenues séparément par Statistique Canada et Revenu Canada pour les mêmes éléments d'information tirés des mêmes déclarations. Cette étude a révélé des écarts marqués. On a observé, par exemple, que les chiffres relatifs au revenu tiré d'une entreprise brut concordaient à ±1% pour 91% des unités étudiées, mais divergeaient selon un facteur de 2 ou plus pour 2% des unités. Dans le cas du poste "amortissement", les chiffres correspondants ont été de 84% et 5%, et dans le cas du profit net, de 74% et 7% respectivement. La classification industrielle des unités selon 80 grands groupes, à partir de la description de l'entreprise fournie par le déclarant fiscal (étayée par les états des résultats et les bilans joints) a produit une concordance dans 68% des cas seulement.

#### 1.3 CAS DU RECENSEMENT DE LA CONSTRUCTION

En dépit des limites des données fiscales, il y a des avantages considérables à tirer de leur utilisation, comme on a pu le constater avec le recensement de la construction. Ce programme, qui, contrairement à ce que son titre laisse supposer, est en fait une enquête par échantillon, comporte l'utilisation d'une grande proportion de données fiscales depuis l'année de référence 1975. Son univers est réparti, selon les recettes déclarées, en grandes et petites unités. Les grandes unités reçoivent par la poste un questionnaire complet et les autres unités de l'univers sont incluses dans un échantillon de déclarations fiscales, lequel est divisé en un sous-échantillon auquel on envoie par la poste un questionnaire abrégé complémentaire. Dans ce questionnaire, on demande de faire des répartitions en pourcentage du genre de travaux exécutés par chaque répondant (cette information permet d'attribuer les codes CAÉ), et de déclarer les taux horaires moyens et les traitements et salaires annuels. Pour l'année de référence 1983, cette enquête comprenait 8,000 grandes unités qui, selon la définition, avaient des recettes d'au moins \$750,000. Il y avait quelque 100,000 petites unités couvertes par un échantillon de 40,000 déclarations fiscales et par un sous-échantillon de questionnaires abrégés permettant d'obtenir environ 15,000 réponses.

Sur une période de dix années, les méthodes de regroupement des données fiscales et des données d'enquête ont été modifiées et affinées de façon à maximiser l'utilisation des données fiscales tout en préservant et améliorant la qualité du produit final. Une base de descriptions d'entreprise tirées des déclarations fiscales a été établie afin de faciliter l'attribution des codes CAÉ aux déclarations fiscales. La comparaison des données fiscales et des données d'enquête supplémentaires a eu pour résultat d'accroître l'utilisation des données fiscales pour le repérage des unités inactives et la production d'estimations des salaires et traitements. Les travaux menés dans ce domains sont décrits ci-après.

De façon générale, lorsqu'il y a incohérence entre les données fiscales et les données d'enquête relativement à l'activité d'une unité ou au nombre d'employés rémunérés, on se fie davantage à l'information tirée des déclarations fiscales. Cette préférence est un renversement de la ligne de conduite adoptée au début, selon laquelle on donnait priorité aux données d'enquête, et elle constitue une marque de confiance à l'égard de l'information fiscale. Il est généralement admis maintenant que, pour ce qui a trait aux indicateurs d'activité et d'emploi du moins, un répondant est moins enclin à donner une réponse inexacte dans sa déclaration d'impôt que dans un questionnaire d'enquête. Tout particulièrement, on estime que les répondants à une enquête sont portés à fournir, pour ces éléments d'information, des chiffres correspondants à l'année courante plutôt qu'à l'année de référence.

Un des problèmes qui limite l'utilisation des données fiscales est le fait que Revenu Canada n'exige pas une présentation uniforme des états financiers joints à la déclaration fiscale. Cela complique l'attribution de valeurs à des postes qui ne figurent pas explicitement dans un ensemble d'états financiers mais qui seraient censées y être. Par exemple, selon la méthode comptable du répondant, les "prestations aux employés" ou les "dépenses de sous-traitance" peuvent être incluses implicitement dans les coûts directs ou déclarées séparément dans le poste des coûts administratifs.

S'il y a divergence entre les codes CAÉ déterminés selon les données fiscales ou selon les données d'enquête, on se fie habituellement aux données d'enquête. En effet, on estime que, pour la plupart des unités incluses dans le secteur de la construction, la répartition du genre de travaux exécutés établie dans le questionnaire d'enquête fournit une information plus fiable que la description de l'activité du déclarant fournie dans les documents fiscaux. (Cette conviction n'a toutefois pas été vérifiée.) La base de descriptions des entreprises et l'expérience acquise en l'attribuant des codes CAÉ au cours des années ont contribué à augmenter la proportion de déclarations fiscales qui peuvent être codées correctement et à améliorer l'exactitude des codes. Par exemple, la proportion de déclarations auxquelles on a pu attribuer des codes CAÉ est passée de 77% en 1977 à 92% en 1982. Durant la même période, la proportion de ces codes qui correspondaient aux codes attribués selon les données d'enquête a grimpé de 78% à 83%.

#### 1.4 CONCLUSIONS

Au cours des dix dernières années, les avantages de l'utilisation des données fiscales sur le plan de la réduction des coûts d'exploitation et du fardeau de réponse ont été clairement prouvés dans la pratique. Quoique certaines études aient révélé des lacunes éventuelles sur le plan de la qualité, il ne fait aucun doute que plus les données fiscales seront utilisées, plus elles seront efficaces.

Le programme de remaniement des enquêtes entreprises offre l'occasion de tirer le meilleur parti possible de l'expérience acquise, d'effectuer d'autres études sur la qualité des données et de rationaliser l'application croissante, quoique inégale, des données fiscales. Ce programme, qui a été lancé dans la seconde moitié de l'année 1984, aura une incidence considérable sur presque toutes les enquêtes économiques, d'ici à ce qu'il soit terminé en 1988. Les deux objectifs principaux de ce projet sont l'intensification de l'utilisation des données fiscales et la refonte complète des méthodes de construction des bases de sondage des enquêtes économiques.

La fiabilité et les coûts d'attribution des codes CAÉ à partir des déclarations fiscales sont d'une importance capitale pour l'élaboration d'une stratégie de mise à jour des bases de sondage d'enquête. Aussi, on a entrepris une étude pour déterminer les erreurs commises par les répondants dans la description de leur entreprise, les erreurs dues aux préposés à l'attribution des codes fondée sur ces descriptions et la possibilité d'appliquer des méthodes de codage automatique. Dans une autre étude, on révise les méthodes actuelles d'échantillonnage des déclarations fiscales en fonction de nouvelles techniques et de nouvelles exigences, par exemple. la production de données longitudinales. On examine la possibilité de coordonner davantage les programmes de traitement des données à Revenu Canada et à Statistique Canada, et un programme général d'application des données fiscales dans toutes les enquêtes annuelles des secteurs d'activité a été élaboré en fonction de ce qui a été expérimenté avec le recensement de la construction. Les aspects théoriques et pratiques de l'utilisation des données fiscales dans les enquêtes économiques seront mieux connus, une fois ces projets achevés.

#### PARTIE 2. DONNÉES SOCIALES EXEMPLE D'UTILISATION DE LA DÉCLARATION D'IMPÔT SUR LE REVENU DES PARTICULIERS

### 2.0 CONSIDÉRATIONS PRÉLIMINAIRES

Dans le cas des données sociales, un programme systématique de production de données sur les particuliers, les familles et les ménages à partir des documents administratifs a été lancé à la fin des années 1970. Ce travail avait pour objectif explicite de produire des données sociales intercensales sur les petites régions et a porté au début sur les déclarations d'impôt sur le revenu des particuliers. Cette déclaration rendait possible l'analyse régionale, qui jusque-là, sauf les années de recensement, pouvait être faite seulement à l'échelle nationale et provinciale et au niveau des grandes régions métropolitaines.

Si on le compare au programme des données sur les entreprises décrit dans la première partie, ce programme visait l'utilisation de données administratives pour compléter les données d'enquête (ou de recensement) plutôt que la construction ou l'alimentation de bases. Cependant, un des éléments importants de l'évaluation de ces données était leur taux de couverture de la population et de ses sous-groupes définis pour les besoins du recensement. Comme le champ d'observation offert par les documents administratifs est déterminé par des considérations et des règlements administratifs, il est rare que ces documents correspondent exactement à l'univers statistique souhaité. En effet, ils peuvent exclure entièrement certains sous-groupes et en sous-représenter d'autres tout en assurant le dénombrement presque exhaustif d'autres sous-groupes. Dans la présente partie de l'exposé, nous évaluons le taux de couverture des déclarations d'impôt sur le revenu des particuliers en comparant certaines données qu'elles contiennent à des données d'autres sources.

Bien qu'il y ait diverses possibilités de comparaison directe des données, on peut définir une formule générique des taux de couverture de la façon suivante:

Mesure de la variable Y de données fiscales Mesure d'une variable semblable, mais provenant d'une autre source de données X 100

La variable Y pourrait inclure beaucoup de chiffres de population et de sommes différents mais l'analyse et les conclusions du présent exposé se limitent à des comparaisons démographiques et à quelques commentaires sur la réussite de Statistique Canada dans le domaine de l'estimation de données annuelles sur les mouvements migratoires à partir de documents administratifs. La production de données régionales sur le revenu et de données brutes sur la population active est une autre utilisation des déclarations d'impôt sur le revenu des particuliers, mais elle n'est pas examinée dans le présent exposé.

#### 2.1 TAUX DE COUVERTURE DÉMOGRAPHIQUE

Les taux de couverture pour les variables démographiques peuvent être calculés pour la population en général et pour des caractéristiques selon l'âge et le sexe.

## 2.1.1 Chiffres de base de la population et nombre de déclarants, 1981

La déclaration d'impôt sur le revenu des particuliers (T1) de 1980 couvre à peu près la période mars-avril 1981 et les données qu'elle contient peuvent être comparées à celles qui ont été recueillies au recensement de juin 1981. On peut commencer par comparer le nombre de déclarants\* au chiffre de population du recensement. Voici le résultat.

TABLEAU 2.1 COMPARAISON DU NOMBRE DE DÉ CHIFFRE DE POPULATION DU RECENSEME	
Nombre de déclarants, printemps 1981	14,525,984
Population du recensement, juin 1981	24,343,175
Taux de couverture	59.7%

Cette comparaison directe peut aussi être faite pour des sous-groupes de la population nationale, par exemple pour la population de provinces et de villes. La colonne 5 du tableau A.1 de l'annexe donne des taux de couverture analogues pour les provinces et les territoires.

## 2.1.2 Chiffres de population et nombre de déclarants, 1981

Chaque déclarant peut être considéré isolément, mais beaucoup de déclarants ont des personnes à charge qui peuvent faire l'objet d'une imputation en fonction de la valeur monétaire des exemptions personnelles. Cette imputation a été effectuée pour l'année 1981, et le résultat pour le Canada apparaît au tableau 2.2 qui suit.

<sup>\*</sup> Les totaux des déclarants qui figurent aux tableaux 2.1, 2.2, 2.3 et 2.4 sont des chiffres obtenus après vérification et imputation (c'est-à-dire que des valeurs ont pu être imputées dans le cas des dossiers où il manquait des valeurs ou qui contenaient des valeurs fausses).

## TABLEAU 2.2 POPULATION DE DÉCLARANTS APRÈS IMPUTATION, PRINTEMPS 1981, PAR COMPARAISON À LA POPULATION DU RECENSEMENT, JUIN 1981 (en milliers)

		POPULATION	
	TOTAL	HOMMES	FEMMES
Population des déclarants après imputation	23,957	11,327	12,631
Population du recensement :	24,343	12,068	12,275
Taux de couverture	.: 98.4%	93.8%	102.9%

Une fois la population des déclarants ajustée pour tenir compte des personnes à charge déclarées, les taux de couverture sont considérablement modifiés. Par exemple, le taux de couverture de la population des déclarants par rapport à la population du recensement est de 59.7% au tableau 2.1, mais quand on considère la population des déclarants après imputation, ce taux atteint 98.4%, comme le montre la première colonne du tableau 2.2.

Il y a cependant un certain élément de double compte dans les imputations, car certains déclarants sont en même temps des personnes à charge. On peut observer le surdénombrement dans la troisième colonne du tableau 2.2, où le taux de couverture estimé des femmes était de 102.9% en 1981. Des efforts sont faits pour réduire le double compte, mais il n'est pas possible de l'éliminer complètement. Quant au surdénombrement, il n'y a pas de moyen connu d'exclure les personnes légalement déclarées personnes à charge mais qui ne résident pas au Canada. (La colonne 6 du tableau A.2 donne, par province et territoire et par sexe, les taux de couverture de la population après imputation.)

Plus précisément, en ce qui concerne le double compte dans la population des femmes, on peut attribuer une bonne par surdénombrement au crédit d'impôt pour enfants (C.I.E.). Quand le C.I.E. est apparu pour l'année fiscale 1978, le taux de couverture de la population des déclarantes après imputation a augmenté de plus de 10%. Des femmes qui jusque-là figuraient seulement comme personnes à charge dans les déclarations d'impôt de leur maris ont commencé à produire leurs propres déclarations sans toutefois cesser d'être déclarées personnes à charge. Ces femmes sont donc comptées deux fois: comme déclarantes et comme personnes à charge.

## 2.1.3 Taux de couverture par groupe d'âge, 1981

La comparaison des taux de couverture par groupe d'âge indique des taux de déclaration plus bas pour des groupes d'âge supérieur et inférieur, c'est-à-dire pour les personnes moins susceptibles de toucher un revenu imposable. (Lableau 2.3)

TABLEAU 2.3 POPULATION DES DÉCLARANTS PAR GROUPE D'ÂGE, APRÈS IMPUTATION, PRINTEMPS 1981, PAR COMPARAISON À LA POPULATION DU RECENSEMENT, JUIN 1981

(1) GROUPES D'ÂGE	(2) POPULATION DU RECENSEMENT	(3) NOMBRE DE DÉCLARANTS	(4) DÉCLARANTS ET PERSONNES À CHARGE		OUVERTURE (4)/(2)
18	6,845,150	291,272	6,636,631	4.2	96.9
18-24	3,294,655	2,823,707	3,622,795	85.7	110.0
25-44	7,184,345	6,511,222	7,748,510	90.6	107.8
45-64	4,658,070	3,640,849	4,488,168	78.2	96.4
65+	2,360,970	1,258,934	1,461,164	53.3	61.9

Là encore la plupart des femmes qui réclament le crédit d'impôt pour enfants sont dans le groupe 18-24 ans ou 25-44 ans, d'où double compte dans ces groupes. Il y a une autre cause de double compte chez les 18-24 ans hommes et femmes: ce sont les enfants qui gagnent de l'argent et produisent leur propres déclarations (en général pour se faire rembourser l'impôt retenu à la source par leur employeur) mais continuent d'être déclarés personnes à charge par leur parents.

## 2.2 DONNÉES MANQUANTES SELON LE GROUPE D'ÂGE, LE SEXE ET L'ÉTAT MATRIMONIAL

Une dernière considération au plan démographique concerne la fréquence des cas où les caractéristiques démographiques de base ne figurent pas dans les déclarations d'impôt. En examinant la population des déclarants seulement pour les quelques dernières années, on peut observer la fréquence (et le pourcentage) de la non-déclaration de ces données par les déclarants.

Dans le tableau qui suit, on constatera l'assez faible fréquence des cas où des données manquent. Ce tableau n'indique toutefois pas le degré précision des données (c'est-à-dire la proportion d'erreurs commises au moment de remplir les formules de déclaration).

TABLEAU 2.4 PROFIL DES DONNÉES MANQUANTES SELON LES DONNÉES DE LA DÉCLARATION D'IMPÔT SUR LE REVENU DES PARTICULIERS CONCERNANT L'ÂGE, LE SEXE ET L'ÉTAT MATRIMONIAL, 1981, 1982, 1983

	FRÉQUENCE DES CA MANQUANTE LE SEXE L'ÂGE	S SUR	
1981	: 40,678 2,854	35,629	15,231,185
Pourcentage de données manquantes	· · · · · · · · · · · · · · · · · · ·		
1982	42,043 13,477	122,029	15,238,257
Pourcentage de données manquantes	0.28 0.09	0.80	
1983	38,090 20,305	115,379	15,336,355
Pourcentage de données manquantes		0.75	

#### 2.3 DONNÉES SUR LES MIGRATIONS

Statistique Canada utilise actuellement les déclarations annuelles d'impôt sur le revenu des particuliers comme source de données sur les migrations annuelles entre régions, c'est-à-dire entre les dix provinces et les deux territoires et entre les 266 divisions de recensement.

Les données sur les migrations sont obtenues par attribution d'un code de division de recensement à chaque dossier fiscal. Dans la majorité des cas, l'attribution de ce code se fait en fonction du code postal correspondant à l'adresse du déclarant. Si la comparaison des déclarations de deux années successives révèle un changement de code, on suppose alors qu'il y a eu migration. Un dossier distinct pour chaque migrant est constitué; l'information démographique et les données sur les exemptions personnelles à compter comme migrants. Enfin, comme la couverture représentée par les dossiers fiscaux est un peu inférieure à la couverture de la population totale, les données font l'objet d'un léger ajustement en vue de l'estimation de la population de migrants chaque année.

Après une étude approfondie effectuée au début des années 1980, un certain nombre d'analyses indépendantes ont montré qu'aucune source de données sur les migrations ne produisait des estimations aussi précises que les données provenant des dossiers de migration des déclarants. Aussi a-t-il été décidé qu'après 1981 les estimations finales effectuées dans le cadre du programme des estimations démographiques seraient faites à partir des données contenues dans les dossiers des déclarants migrants.

#### 2.4 PROJETS

L'étude de la couverture de la population dans cette partie de l'exposé semble montrer que les totaux de déclarants augmentés des personnes à charge donnent une assez bonne couverture de la population par comparaison au recensement de 1981. Quand les comparaison sont faites selon l'âge et le sexe, on constate que les taux de couverture diffèrent du taux de couverture de la population dans son ensemble. On peut noter en particulier un surdénombrement pour les femmes (principalement dans les groupes d'âge 18-24 ans et 25-44 ans) et un sous-dénombrement pour le groupe des 65 ans et plus.

Deux projets sont actuellement élaborés pour résoudre les problèmes de couverture. Un premier concerne l'appariement des déclarations des conjoints en vue de réduire le surdénombrement dans la population des femmes. Une autre partie de ce projet concerne l'appariement des enfants à leur parents, au moyen de l'adresse postale, en vue de réduire le double compte des enfants à charge qui produisent des déclarations.

Un deuxième projet a trait aux personnes âgées de 65 ans et plus. Le tiers au moins des personnes de ce groupe d'âge ne produisent pas de déclaration, mais presque toutes reçoivent des prestations mensuelles de sécurité de la vieillesse (SV). Sans doute peut-on apparier les dossiers fiscaux et les dossiers de SV selon des unités de base géographiques pour constituer un ensemble de données structuré géographiquement et ainsi obtenir une meilleure couverture de la population des 65 ans et plus.

Ainsi, le taux de couverture globale (98.4%) de la population est satisfaisant, mais le surdénombrement et le sous-dénombrement sont inégaux. Avec les améliorations actuellement envisagées, on prévoit que les problèmes de couverture de sous-groupes seront atténués dans un an ou deux.

#### 2.5 CONCLUSIONS

Le principal objectif de cette deuxième partie de l'exposé était de faire des comparaisons démographiques entre les données tirées des déclarations d'impôt sur le revenu des particuliers et les données du recensement de la population de 1981. Il convient ici de faire quelques conclusions générales.

- Le chiffre de la population des déclarants après imputation pour 1981 (c'est-à-dire le nombre de déclarants augmenté des personnes à charge imputées) est proche du chiffre de la population du recensement de 1981 (24.0 millions de déclarants, après imputation, et 24.3 millions de personnes dénombrées dans le recensement).
- L'analyse de la population des déclarants après imputation pour 1981 montre que le taux de couverture est plus élevé pour les femmes que pour les hommes (102.9% contre 93.8%). Le surdénombrement de la population des femmes est en grande partie attribuable au double compte, problème qui fait actuellement l'objet d'une étude.
- o Les comparaisons faites avec la population des déclarants après imputation montrent aussi qu'il y a surdénombrement dans les groupes des 18-24 ans et des 25-44 ans.

- o The analysis of the 1981 imputed taxfiler population indicates that the coverage of the female portion of the population is higher than the male portion (102.9% versus 93.8%, respectively). The over-coverage of the female population is largely a result of double-counting, a problem that is being addressed independently at this time.
- o The imputed taxfiler population comparison also indicates that over-counting occurs in the 18-24 and 25-44 age groups.
- Overall, data for the variables sex, age and marital status are rarely missing.
- o No other source of post-censal migration data provided better estimates overall than those derived from the annual personal income tax records.

#### CONCLUSIONS

The applications described in this paper are just two among several examples of administrative data use within Statistics Canada. Other notable areas of developmental work include the use of unemployment insurance files in conjunction with Labour Force Survey data to produce small area labour market measures, the utilization of old age security records alluded to in Part 2, and the development of family income estimation from the personal income tax file.

The usefulness of administrative data for statistical purposes is often limited either because the underlying concepts or definitions differ from the required statistical concepts, or because key variables required for analytical purposes are unavailable or of poor quality. In many cases a relatively minor change to an administrative procedure or form could greatly enhance the statistical usefulness of the administrative file without jeopardizing the primary administrative function. There is a need to develop mechanisms that would allow the statistical agency to influence the design of administrative systems so as to optimize their statistical value.

Tax data have become an essential element in Statistics Canada's business data programs, both as a source of frame data and as a source of response data primarily for small businesses. In the case of data on individuals, the tax data cannot be considered an alternative to traditional data derived from cen-

TABLEAU A.1 Comparaison des populations des provinces, 1981

(1) Provin-	(2) Population du	(3) Nombre de déclarants	(4) Déclarants et	Taux de	5) couverture (4)/(2)
	Recensement		personnes à charge		
TN.	567,680	294,145	562,752	51.8%	99.1%
1PÉ.	122,505	65,874	114,896	53.8%	93.8%
NE.	847,445	478,039	829,739	56.4%	97.9%
NB.	696,405	387,951	678,017	55.7%	97.4%
QUÉ.	6,438,405	3,669,789	6,128,488	57.0%	95.2%
ONT.	8,625,110	5,370,843	8,761,173	62.3%	101.6%
MAN.	1,026,240	661,312	1,071,698	64.4%	104.4%
SASK.	968,315	543,259	925,356	56.1%	95.6%
ALB.	2,237,725	1,326,725	2,141,267	59.3%	95.7%
CB.	2,744,465	1,690,495	2,677,875	61.6%	97.6%
YUKON	23,150	14,169	22,278	61.2%	96.2%
T. N0.	45,745	23,280	43,729	50.9%	95.6%
CANADA	24,343,175	14,525,984	23,957,268	59.7%	98.4%

TABLEAU A.2 Comparaison des populations, par sexe et par province, 1981

HOMMES				
(1) Provin- ce	(2) Population du Recensement	(3) Nombre de déclarants	(4) Déclarants et personnes à charge	(5) Taux de couverture (3)/(2) (4)/(2)
TN. 1PÉ. NE. NB. QUÉ. ONT. MAN. SASK. ALB. CB. YUKON T. NO.	285,690 60,940 419,575 346,005 3,172,200 4,246,790 506,510 486,075 1,143,220 1,365,155 12,175 23,965	158,476 34,690 256,023 205,514 1,945,320 2,816,990 341,816 297,457 725,338 900,999 7,741 12,604	263,713 54,403 388,812 317,151 2,860,212 4,139,461 502,569 442,055 1,044,845 1,280,808 11,163 21,558	55.5% 92.3% 56.9% 89.3% 61.0% 92.7% 59.4% 91.7% 61.3% 90.2% 66.3% 97.5% 67.5% 99.2% 61.2% 90.9% 63.4% 91.4% 66.0% 93.8% 63.6% 91.7% 52.6% 89.9%
CANADA	12,068,285	7,702,968	11,326,750	63.8% 93.8%
FEMMES				
(1) Provin- ce	(2) Population du Recensement	(3) Nombre de déclarants	(4) Déclarants et personnes à charge	(5) Taux de couverture (3)/(2) (4)/(2)
TN. 1PÉ. NE. NB. QUÉ. ONT. MAN. SASK. ALB. CB. YUKON T. NO.	281,990 61,565 427,870 350,400 3,266,205 4,378,320 519,730 482,240 1,094,505 1,379,310 10,975 21,780	135,669 31,184 222,016 182,437 1,724,469 2,553,853 319,496 245,802 601,490 789,496 6,428 10,676	299,039 60,493 440,927 360,866 3,268,276 4,621,712 569,129 483,301 1,096,422 1,397,067 11,115 22,171	48.1% 106.0% 98.0% 50.6% 98.0% 103.0% 52.1% 103.0% 52.8% 100.1% 58.3% 105.6% 61.5% 109.5% 51.0% 100.2% 54.9% 100.2% 57.2% 101.3% 58.6% 101.3% 49.0% 101.8%
CANADA	12,274,890	6,823,016	12,630,518	55.6% 102.9%

