
Administrative Social Data in Canada: Some Results and Some Implications

John M. Leyes and Judy Eisl-Culkin, Statistics Canada

Administrative records have been a source of statistics for many years. Initially, the records were on paper. In recent years, many paper records were converted to computer-readable form.

Statistics Canada, since its inception in 1918, has used administrative records to produce statistics. In 1979, Statistics Canada began to explore the feasibility of deriving social information from individual tax records. In this paper, there is a brief review of some data derived from Canada's personal income tax records. Moreover, since income tax records can be considered a biased representation of the Canadian population, some comparisons are included to illustrate the extent to which taxfiler data compare with population statistics.

From the beginning of this data development program, the work with the personal income tax (T1) file was subject to a number of expected *a priori* shortcomings. The shortcomings included the following:

- ❑ **Population Coverage Bias** -- The personal income tax system is based on individuals only. Since only 60% of Canadians were filing tax returns in the mid to late 1970's, coverage was deemed inadequate for social statistical applications.
- ❑ **Age Coverage Bias** -- The age profile of taxfilers also differed from the age profile of the population. This was judged to be an unacceptable bias.
- ❑ **Income Distribution Bias** -- The elderly and the young frequently have low incomes and do not file tax returns. Data from the tax file would, therefore, be inadequate for public policy purposes directed at these target groups.

- ❑ **Income Coverage Bias** -- Since not all income received by Canadians is taxable, the income coverage of the T1 was considered incomplete.
- ❑ **Dimensionality of Variables** -- Since any single administrative record has a specific and narrow application in program administration, the range of data variables was judged inadequate as a source of social data.
- ❑ **Concepts and Definitions** -- The concepts and definitions used in household surveys and censuses of population can only be approximated through the use of an annual tax file.

Each of the above represented a limitation or shortcoming of data derived from administrative records, in general and the T1, in particular. Since the time that the above assumptions were identified, data for taxfiler families have been developed. Beginning in the third section of this paper, the first four assumptions are assessed with empirical findings from the taxfiler family data for the period 1982-92.

■ The Development of Taxfiler Family Data

The taxfiler family concept has been designed to emulate the census family concept. A census family:

"[r]efers to a husband and a wife (with or without children who have never married, regardless of age), or a lone parent of any marital status, with one or more children who have never married, regardless of age, living in the same dwelling. For census purposes, persons living in a common-law type of arrangement are considered as now married, regardless of their legal marital

status; they accordingly appear as a husband-wife family in most census family tables." (Statistics Canada, *1991 Census Dictionary*, p. 119)

This concept is suitable for household collection methods, since respondents are asked to report on the relationships between all residents of a dwelling. With administrative records, secondary information -- such as reported marital status, value of exemptions/tax credits, ages of taxfilers, addresses, child care expenses, and so forth -- are used to form families.

It has not, therefore, been possible to emulate the exact census family concept. The major sources of difficulty arise with older children (whether they have been married or not), when they reside with their parents, and with common-law unions. In general, the census family concept works reasonably well for families with dependent children, and some success has been achieved in estimating single parent and common-law families.

Beginning with the 1982 personal income tax file, Statistics Canada began estimating families (and dependents) from the T1 in the following sequence:

- ❑ Taxfilers, reporting the Social Insurance Numbers (SIN) of their spouses, are matched to form husband-wife families.
- ❑ Other husband-wife families are formed from taxfilers who declare themselves married but do not report spousal SINS.
- ❑ Filing children [1] who reside with their parents are matched to their parents.
- ❑ There is an intermediate step to unduplicate records, to identify one-filer husband-wife family units, to assign a unique postal code to family members, and to assign a family composition type to each family unit.
- ❑ Common-law unions are identified from the pool of individuals classed as single parent families and nonfamily persons [2].

- ❑ Lastly, nonfiling family members are imputed.

With this brief introduction and description of the taxfiler family data, it is now possible to consider some data findings.

■ Some Empirical Findings

Population Coverage Comparison: 1982-92 Taxfiler Family File (T1FF) to 1983-93 Population Estimates [3]

In this section, two tables are presented. Table 1 contains comparisons at the Canada level for both population and family composition types. Table 2 contains population comparisons for the provinces, territories and Canada.

Population Coverage Bias: 1992 Taxfiler Family File (T1FF) to Statistics Canada's 1993 Population Estimates

The T1 family file (T1FF) data in Table 1 are in six classifications for comparison:

- ❑ Section one (lines 1-4) includes the number of taxfilers, the change in the number of taxfilers, the year-to-year percent change in the number of taxfilers, and the coverage (or percent of the population filing a tax return).
- ❑ Section two (lines 5-7) compares the taxfiler family (T1FF) population to the official Statistics Canada population estimates. (In creating the T1FF data, there is a record for each family member and for each nonfamily person, including the imputed population. Thus, line five of Table 1 is the T1FF population count.)
- ❑ Section three (lines 8-10) compares the number of husband-wife T1FF families to the official Statistics Canada estimate of husband-wife families. (Both sources include common-law families as husband-wife families.)

Table 1. Summary of Comparisons Between T1FF (T1 Family File) and (1) Official Population Estimates, (2) Census, or (3) Survey of Consumer Finances (SCF)

T1FF DATA	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
1. Number of Taxfilers (000's) ^(a)	15,161	15,224	15,462	15,522	16,441	16,684	17,251	17,947	18,450	18,786	19,267
2. Change in # of Taxfilers	--	63	238	60	919	243	567	696	503	336	481
3. Percentage change	--	0.4	1.6	0.4	5.9	1.5	3.4	4.0	2.8	1.8	2.6
4. Coverage (1)/(20)	59.6	59.2	59.6	59.2	61.9	62.0	63.0	64.6	65.6	66.1	67.0
5. Total T1FF Pop. (000's) ^(a)	23,619	23,472	23,728	23,833	24,517	24,833	25,155	25,909	26,577	27,007	27,534
6. Percentage change	--	-0.6	1.1	0.4	2.9	1.3	1.3	3.0	2.6	1.6	2.0
7. Coverage (5)/(20)	92.8	91.3	91.5	91.0	92.3	92.3	91.9	93.2	94.5	95.0	95.8
8. # of T1FF HWFs, (000's) ^{(b)(c)}	5,506	5,509	5,555	5,515	5,610	5,738	5,866	6,103	6,212	6,284	6,491
9. Percentage change	--	0.1	0.8	-0.7	1.7	2.3	2.2	4.0	1.8	1.2	3.3
10. Coverage (8)/(22)	96.2	95.4	95.4	93.6	93.8	94.5	94.8	96.9	95.2	95.5	97.9
11. # of T1FF LPFs (000's) ^{(b)(c)}	819	819	866	932	978	942	937	1,007	1,056	1,129	1,087
12. Percentage change	--	0.0	5.7	7.6	4.9	-3.7	-0.5	7.5	4.9	6.9	-3.7
13. Coverage (11)/(24)	106.6	102.9	105.1	109.0	112.2	105.8	102.6	107.8	108.5	114.5	108.7
14. # of T1FF CLUs (000's) ^{(d)(e)(f)}	206	227	241	221	233	335	366	501	526	557	774
15. Percentage change	--	10.2	6.2	-8.3	5.4	43.8	9.3	36.9	5.0	5.9	39.0
16. Coverage (14)/(26)	--	--	--	45.4	--	--	--	--	72.5	--	--
17. # of T1FF NFPs (000's) ^{(b)(c)}	2,992	3,092	3,193	3,275	3,763	3,520	4,096	3,954	4,193	4,264	4,322
18. Percentage change	--	3.3	3.3	2.6	14.9	-6.5	16.4	-3.5	6.0	1.7	1.4
19. Coverage (17)/(27)	91.2	92.4	93.0	93.3	103.8	93.4	107.0	100.9	103.6	100.4	102.5
POPULATION ESTIMATES or SCF DATA	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
20. Official Pop. Est. (000's) ^(a)	25,456	25,702	25,942	26,204	26,550	26,895	27,379	27,791	28,118	28,436	28,753
21. Percentage change	--	1.0	0.9	1.0	1.3	1.3	1.8	1.5	1.2	1.1	1.1
22. Official est. of HWFs ^{(b)(c)}	5,722	5,773	5,824	5,894	5,979	6,071	6,190	6,298	6,525	6,579	6,632
23. Percentage change	--	0.9	0.9	1.2	1.4	1.5	2.0	1.7	3.6	0.8	0.8
24. Off. Est. of LPFs (000's) ^{(b)(c)}	768	796	824	855	872	890	913	934	973	986	1,000
25. Percentage change	--	3.6	3.5	3.8	2.0	2.1	2.6	2.3	4.2	1.3	1.4
26. # of Census CLUs (000's) ^{(d)(e)}				487					726		
27. # of NFPs, SCF (000's) ^{(b)(c)}	3,282	3,348	3,433	3,512	3,625	3,770	3,829	3,919	4,047	4,246	4,217
28. Percentage change	--	2.0	2.5	2.3	3.2	4.0	1.6	2.4	3.3	4.9	-0.7

LEGEND, Table 1:

- (a) Includes all provinces and territories.
- (b) Excludes the Yukon and Northwest Territories
- (c) Post-censal estimates are not available for the two territories, although T1FF data are available.
- (d) Counts of common law unions are available in census years only.
- (e) Estimates of nonfamily persons are identified annually through the Survey of Consumer Finances (Catalogue No. 13-208, Table 16).
- (f) The number of T1FF common law unions are included in Line 8.

HWFs → husband-wife families
 LPFs → lone parent families
 NFPs → nonfamily persons
 CLUs → common law unions

SOURCES, Table 1:

Population Estimates (1983-1990) -- Revised Intercensal Estimates, as of July 1, 1994 (Catalogue Number 91-537).
 Population Estimates (1991) -- Updated Postcensal Estimates, as of July 1, 1994 (Catalogue Number 91-537).
 Population Estimates (1992-1993) -- Updated Postcensal Estimates, as of July 1, 1994 (Catalogue Number 91-213).
 Previously, post-censal estimates were derived from Catalogue Number 91-210.
 Husband-Wife Families: 1983-1985 are for June 1, 1986+ are for July 1
 Lone Parent Families: 1983-1985 are for June 1, 1986+ are for July 1
 Common Law Unions: Catalogue Number 93-312 (Table 4)
 Nonfamily Persons: Family Incomes, Catalogue Number 13-208, (revised in 1992 - Table 14)

- ❑ Section four (lines 11-13) compares the number of lone parent T1FF families to the official Statistics Canada estimate of lone parent families.
- ❑ Section five (lines 14-16) provides an estimate of common-law families. The only time that there is another estimate of common-law families is during census years. Thus, there are only comparisons for the 1985 T1FF and the 1990 T1FF.
- ❑ Section six (lines 17-19) includes the count of T1FF nonfamily persons for each year from 1982-1992. (The Survey of Consumer Finances (SCF) is the only other annual source for an estimate of nonfamily persons.)

There are several highlights in Table 1.

- ❑ The coverage of the T1FF population varied between 91.0% (1985) and 95.8% (1992) of the official estimates over the 1982-92 period (i.e., line 7).
- ❑ The T1FF coverage of husband-wife families varied between 93.6% (1985) to 97.9% (1992) (i.e., line 10).
- ❑ The overcoverage of lone parent families varied from a low of 102.6% (1988) to a high of 114.5% (1991) (i.e., line 13).
- ❑ Common-law unions were undercovered in the two years for which data were available (1986 Census and 1991 Census). Reviewing line 16, it can be noted that the coverage improved from about 45% (1985) to about 72.5% (1990). This improvement arose due to the record linkage procedures, in which more lone parent taxfilers and/or other lone parent taxfilers and nonfamily persons were linked to form common-law unions, especially between 1988 and 1989. In 1992, common law unions increased by 39 percent from the previous year (see footnote 2).
- ❑ The T1FF coverage of nonfamily persons, compared to the SCF, was lowest in 1982

(91.2%) and highest in 1988 (107.0%) (i.e., line 19).

- ❑ The annual percentage change in the T1FF population (taxfilers plus imputed dependents) was lower in each year than the percentage change in the number of taxfilers, except for 1985 (i.e., lines 3 and 6).

Provincial Population Coverage Comparison: 1990 Taxfiler Family File (T1FF) to 1991 Population Estimates

A comparison between the T1FF estimates and the official population estimates is incomplete without an indication of some sub-national estimates. A

comparison has, therefore, been included for the ten provinces and two territories for the 1990 T1FF versus the 1991 Census of Population, and the 1990 T1FF versus the 1991 population estimates.

Prior to the 1991 Census, the annual post-censal population estimates were based on the last census. Beginning with the 1991 Census, the post-censal population estimates are adjusted for the estimated net undercoverage of the census. This accounts for the differences in the coverage rates in Table 2 for the two different pairs of years.

It can be noted in Table 2 that the provincial coverage, compared to the 1991 Census, ranged from 101.0 percent for New Brunswick to 93.9 percent for both the Yukon and Northwest Territories.

- ❑ For Canada, the net undercoverage adjustment reduced the T1FF coverage by 2.9 percent from 97.4% to 94.5%.
- ❑ Compared to the 1991 Census, T1FF coverage exceeded 95% in all provinces, and was 93.9% in each of the territories.
- ❑ For three provinces (New Brunswick, Newfoundland and Manitoba), the enumerated T1FF population was higher than the enumerated 1991 Census population.

**Table 2. Population Comparison by Province and Territory,
1990 T1FF to 1991 Census of Population and
1990 T1FF versus 1991 Population Estimates**

Province or Territory	1990 T1FF (000's)	1991 Census (000's)	% Ratio (T1FF/Census)	1991 Population Estimates (000's)	% Ratio (T1FF/Pop. Est.)
New Brunswick	730.9	723.9	101.0	748.5	97.6
Newfoundland	573.1	568.5	100.8	579.9	98.8
Manitoba	1,095.8	1,091.9	100.4	1,113.3	98.4
Saskatchewan	982.2	988.9	99.3	1,007.0	97.5
Nova Scotia	888.6	899.9	98.7	918.1	96.8
Prince Edward Is.	128.0	129.8	98.6	130.8	97.9
Quebec	6,729.6	6,896.0	97.6	7,081.2	95.0
Canada	26,576.8	27,296.9	97.4	28,117.6	94.5
Alberta	2,471.2	2,545.6	97.1	2,600.3	95.0
Ontario	9,754.7	10,084.9	96.7	10,471.2	93.2
British Columbia	3,142.3	3,282.1	95.7	3,376.9	93.1
Northwest Terr.	54.2	57.7	93.9	61.2	88.6
Yukon	26.1	27.8	93.9	29.0	90.0

SOURCES: 1990 T1FF Unpublished T1FF tabulations.
1991 Census: Catalogue No. 93-310 (Table 4).
1991 Population Estimates: (July 1) updated postcensal estimates, 91-537 (Table 1).

- ❑ Compared to the population estimates, T1FF coverage was equal to or exceeded 95% in eight provinces.

Age Coverage Bias, 1990 T1FF to 1991 Population Estimates

In Table 3, there are some broad age range comparisons [4]. In recent years, age and gender algorithms have been developed for the under 18 population, many of whom are identified in the taxfile as dependents.

In reviewing column 4 of Table 3 (i.e., Coverage Ratio), it can be noted that the coverage of the 1990 T1FF to the 1991 population estimates was 90% or higher for age ranges over 29.

There are several highlights in Table 3:

- ❑ The population under 18 is overcovered by 5.7%.

- ❑ For the 30 to 64 age range, the coverage was 92.9%.
- ❑ For the 65+ population, the age coverage was 92.5%.

**Table 3. Comparison by Age Group:
1990 T1FF versus
1991 Population Estimates**

Age Group	1990 T1FF (000's)	1991 Population Estimates (000's)	Coverage Ratio (T1FF/Pop. Est.)
0 - 17	7,362.9	6,967.0	105.7
18 - 29	4,606.2	5,415.7	85.1
30 - 44	6,581.0	7,081.5	92.9
45 - 64	5,057.2	5,442.5	92.9
65+	2,969.6	3,211.0	92.5
All Ages	26,576.8	28,117.6	94.5

SOURCES: 1990 T1FF: 100% T1FF tabulations (provinces + territories), unpublished.
1991 Population Estimates: (July 1) Catalogue Number 91-537

Table 4. Coverage Ratio Comparison for the 65+ Population:
T1FF versus Population Estimates,
1982-1992

Years, T1FF	T1FF Pop., 65+ (000's)	Years, Popula- tion Estimates	Population Esti- mates, 65+ (000's)	Coverage Ratios (T1FF/ Pop. Est.)	
				65+ Population	All Ages
1982	1,621.4	1983	2,502.8	64.8	92.8
1983	1,676.9	1984	2,568.9	65.3	91.3
1984	1,675.0	1985	2,654.3	63.1	91.5
1985	1,814.7	1986	2,742.3	66.2	91.0
Refundable Federal Sales Tax Credit Introduced					
1986	1,997.7	1987	2,842.6	70.3	92.3
1987	2,207.2	1988	2,929.1	75.4	92.3
1988	2,392.6	1989	3,024.6	79.1	91.9
1989	2,640.5	1990	3,117.0	84.7	93.2
1990	2,969.6	1991	3,211.0	92.5	94.5
1991	2,989.8	1992	3,299.8	90.6	95.0
1992	3,116.8	1993	3,391.1	91.9	95.8

SOURCES: T1FF (1982-1987), Unpublished tabulations from a 5% sample.
T1FF (1988-1991), 100% unpublished tabulations.
Population Estimates (1983-90), Revised Intercensal Estimates, Catalogue Number 91-537.
Population Estimates (1991), Updated Postcensal Estimates, Catalogue Number 91-537.
Population Estimates (1992), Updated Postcensal Estimates, Catalogue Number 91-213.

The taxfiling rates for the 65+ population began to increase with the introduction of the Refundable Federal Sales Tax Credit in 1986. This tax credit was directed at lower income individuals and families. Some individuals and families that had not previously filed a tax return, because they did not have a tax liability, began to file a tax return to obtain the tax credit. It can be noted in Table 4 that the coverage of the 65+ population increased from about 65% in the 1982-85 period to 90%+ in the 1990+ period.

Income Distribution Bias, T1FF to Survey of Consumer Finances

The third assumption at the beginning of this paper pertained to the expected bias that would be found in the income data derived from tax records. The general basis of this expectation can be stated as follows:

- a. In the general population, some individuals have low incomes and do not file tax returns, therefore,
- b. measures of central tendency from the tax data will be higher than similar measures from the general population.

To assess this assumption, it is possible to make T1FF data comparisons with either the Census of Population or the annual Survey of Consumer Finances. Since there would only be two years for comparison with the Census (i.e., 1986 and 1991), and since there are eleven years (1982 to 1992) for comparison with the SCF, a comparison was undertaken between the T1FF and the SCF data.

For this comparison, two different groups were compared: all families and nonfamily persons.

Median Income Comparison for Families: T1FF to SCF, 1982-92

Table 5 includes a time series comparison of median incomes between the T1FF and the Survey of Consumer Finances (SCF)[5] for the period 1982-92.

There are several highlights in Table 5:

- The T1FF medians have been lower than the SCF medians for all years.
- The T1FF medians were about 95% of the SCF medians for the period 1982-85.
- With the introduction of the Refundable Federal Sales Tax Credit in the 1986 tax year, the T1FF median fell below 95% in subsequent years except 1988 [6].

- In 1988, the T1FF/SCF median income ratio was 95.5%, the highest ratio that was found in any of the years in this time series. The ratio has been below 95% since 1988.

While a definitive conclusion cannot be drawn from this time series, it does suggest that the tax data may cover the low income population better than one might expect. Of course, three programs have been introduced that have worked to increase the probability that low income individuals will file tax returns. First, the Child Tax Credit was introduced in 1978, a program directed at low income families with children. Second, the Refundable Federal Sales Tax Credit (RFSTC) in 1986 was directed at individuals and families with low incomes. The introduction of the Goods and Services Tax (GST) was accompanied with a GST tax credit that was higher than the RFSTC. This reinforced the increased taxfiling trend stimulated by the RFSTC.

Table 5. Income Comparison for Families, 1982-92:
T1FF to Survey of Consumer Finances

Year	Median Income, All Families				% Ratio (T1FF/SCF)
	T1FF	% Increase	SCF	% Increase	
1982	28,100		29,537		95.1
1983	29,000	3.2	30,419	3.0	95.3
1984	30,600	5.5	32,079	5.5	95.4
1985	32,100	4.9	33,950	5.8	94.6
Refundable Federal Sales Tax Credit Introduced					
1986	33,400	4.0	36,019	6.1	92.7
1987	35,300	5.7	38,059	5.7	92.8
1988	38,600	9.3	40,430	6.2	95.5
1989	41,100	6.5	43,547	7.7	94.4
1990	42,700	3.9	45,129	3.6	94.6
1991	42,900	0.5	45,515	0.9	94.3
1992	43,500	1.4	46,479	2.1	93.6
% Increase, 1982-1992	54.8		57.4		

SOURCES: T1FF: 1982-1992, 100% Unpublished T1FF tabulations were used. (The Territories were excluded for comparison purposes.)

SCF: 1982-1992, Catalogue Number 13-208. (Table 1) annual publication.

**Median Income Comparison for Non-Family Persons:
T1FF to SCF, 1982-92**

In Table 6, data are included from the two sources for the period 1982-1992.

There are several highlights in Table 6:

- For the years 1982-85, for nonfamily persons, the data support the assumption that data from the general (i.e., SCF) population provide medians that are lower than medians derived from tax data. For these years, the ratio of medians varied between 116.8% and 124.0%.
- With the introduction of the RFSTC in 1986, the T1FF/SCF median income ratio declined to 104.3% in 1986. In fact, in the post-RFSTC period, this ratio was below 100 for five of the seven years.

- For nonfamily persons, the results are mixed compared to the All Families data reported in Table 5. Nonetheless, since the tax credits were introduced, this ratio has been lower than in the 1982-1985 period.

Income Coverage Bias

Table 7 contains a comparison of the 1990 sources of income from the T1FF compared to the 1990 sources of income from the 1991 Census of Population. The largest component of income for both sources was employment income. Compared to the Census, T1FF covered 92.7%. For wages and salaries, the coverage was 94.0%. For self-employment income, the coverage ratio was 77.5%. For investment income it was about 126%.

In addition, two columns are included to indicate the comparison of the 1990 T1FF to the 1990 National Accounts data.

**Table 6. Income Comparison for Nonfamily Persons, 1982-92:
T1FF to Survey of Consumer Finances**

Year	Median Income, Nonfamily Persons				% Ratio (T1FF/SCF)
	T1FF	% Increase	SCF	% Increase	
1982	12,300		10,246		120.0
1983	12,100	-1.6	9,757	-4.8	124.0
1984	12,700	5.0	10,842	11.1	117.1
1985	13,300	4.7	11,383	5.0	116.8
Refundable Federal Sales Tax Credit Introduced					
1986	12,900	-3.0	12,371	8.7	104.3
1987	13,300	3.1	13,317	7.6	99.9
1988	14,600	9.8	13,880	4.2	105.2
1989	15,100	3.4	15,546	12.0	97.1
1990	15,700	4.0	16,413	5.6	95.7
1991	15,700	0.0	16,298	-0.7	96.3
1992	15,400	-1.9	16,515	1.3	93.2
% Increase, 1982-92	25.2		61.2		

SOURCES: T1FF: 1982-1992, 100% Unpublished T1FF tabulations were used. (The Territories were excluded for comparison purposes.)

SCF: 1982-1992, Catalogue Number 13-208. (Revised in 1992 - Table 14).

On the one hand, the National Accounts data provide some new insights, while, on the other hand, they can be misleading.

- ❑ The T1FF coverage of employment income, wages & salaries and self-employment income is marginally higher in the National Accounts comparison than for the Census comparison.
- ❑ The T1FF coverage of investment income is higher than the Census (126%) and lower for the National Accounts comparison (63%). Of course, the National Accounts include all investment income, and much of the income is paid to recipients other than persons in Canada (e.g., foreigners, corporations, etc.).

In the government transfers section of Table 7, some additional observations emerge:

- ❑ The OAS/GIS/SPA (i.e., Old Age Security, Guaranteed Income Supplement and Spouse's Allowance) program is directed at the elderly. Both the GIS and SPA parts of this program are directed at the low income elderly and are not subject to taxation. Thus, the tax system

coverage can be expected to be well under 100%, and this is the case.

- ❑ For the C/QPP (i.e., Canada and Quebec Pension Plans), the T1FF coverage is higher compared to the Census (111.1%) and lower compared to the National Accounts (97.6%).
- ❑ For FA/CTC (i.e., Family Allowances and Child Tax Credit), the T1FF coverage is 92.2% compared to the Census and 88.7% compared to the Accounts.
- ❑ For UI (i.e., Unemployment Insurance), T1FF coverage is 107.0% compared to the Census and 93.1% compared to the Accounts.
- ❑ For Other Government Transfers, the T1FF coverage exceeds the Census coverage, 129.8%, and falls short of the Accounts, 60.5%.

■ Major Directions for 1994+

Two new data development initiatives are in progress. The first uses the T1FF database, while the second is designed to reduce the deterioration in

Table 7. Sources of Income Comparison:
1990 T1FF versus (i) 1991 Census and
(ii) 1990 National Accounts

Sources of Income	1990 T1FF (\$Billions)	1991 Census (\$Billions)	Coverage Ratio (T1FF/Census)	National Accounts (\$Billions)	Coverage Ratio (T1FF/Prog. Exp.)
Employment Income	336.3	362.6	92.7	361.1	93.1
Wages & Salaries	315.6	335.9	94.0	334.6	94.3
Self-Employ. Inc.	20.7	26.7	77.5	26.5	78.1
Investment Inc.	38.0	30.1	126.2	60.4	62.9
Gov't Transfers					
OAS/GIS/SPA	9.9	15.5	63.9	15.5	63.9
C/QPP	12.0	10.8	111.1	12.3	97.6
FA + CTC	4.7	5.1	92.2	5.3	88.7
Unemp. Insurance	12.2	11.4	107.0	13.1	93.1
Other Gov't Transfers	13.5	10.4	129.8	22.3	60.5
TOTAL	426.6	446.0	95.7	490.1	87.0

Source: A. Rashid. *Comparison of 1990 Income Estimates from T1FF and Census*. Unpublished paper. Statistics Canada, April 1994 (Table 4).

the quality of T1FF that has been created by a change in the tax law.

- A Longitudinal Administrative Database (LAD) is under development. This database will enable research studies of low income/welfare/income dynamics in Canada for the period 1982+. The LAD is a 1% sample of all taxfilers.
- In 1993, the Canadian government transformed the tax credits directed at children. The Family Allowance program (directed at all qualifying children) and the Child Tax Credit (directed at children in families with low and middle incomes) were eliminated. In addition, the non-refundable tax credit for all families with children, irrespective of income, was eliminated. These three were replaced with the Child Tax Benefit, that is directed at low to medium income families with dependent children. The maximum amount in 1993 was \$1,733 for a husband-wife family with one dependent child (includes a \$500 Working Income Supplement). Furthermore, the amount of the Child Tax Benefit is inversely related to family income and directly related to the number of dependent children. Thus, for a husband-wife family with one dependent child, a Child Tax Benefit would be received for a qualifying family up to a family income of \$66,720. The benefit would be zero for this family with a family income of \$75,241 and over.

The changes affecting children have led to a deterioration in the data for dependent children. Independent studies are currently underway to create an "historical imputation file" to impute dependent children for taxfiler families that have higher income -- incomes that preclude them from obtaining Child Tax Benefits.

■ Privacy and the Statistical Uses of Administrative Records [7]

Administrative Records and Privacy

In this paper, administrative data have been used to indicate the overall coverage of the population. Not only is the coverage comparatively high, but the cost of deriving statistics from administrative records is comparatively low, if it is assumed that the program administration costs will be incurred whether or not the records are used for statistical applications. Before concluding this paper, we would like to touch on an issue that is of concern when accessing and using, for statistical purposes, administrative records pertaining to individuals who, in all likelihood, are not aware that their information is being used in this manner -- i.e., the issue of privacy. We would like to explain how legislation in Canada has recognized the legitimacy of such access and use.

Confidentiality and Statistical Reporting

In Statistics Canada, statisticians collect, compile and report numerical descriptors of the population, the economy, the industrial organization, the institutions, and so on for various geographical areas (including the nation and small areas). One of the guiding principles has been a legal requirement that:

"... No person who has been sworn under [this Act] shall disclose or knowingly cause to be disclosed, by any means, any information obtained under this Act in such a manner that it is possible from the disclosure to relate the particulars obtained from any individual return to any identifiable individual person, business or organization." (*Statistics Act*, §17.1(b))

Thus, there is a formal legal requirement to release data in forms that preserve the confidentiality

of those persons, businesses and organizations to whom the data relate, whether the data have been derived from surveys, censuses or administrative records.

Privacy and the Statistical Uses of Administrative Records

Some observers of the statistical scene believe that if data are collected directly and voluntarily from respondents, an invasion of privacy is visible. When statistics, however, are based on the use of administrative records, it is sometimes argued that this is unacceptable without the prior approval of the individuals concerned. This requirement would be prohibitively expensive -- perhaps impossible -- to achieve on an individual basis. With millions of administrative records, the requirement of prior approval would be unrealistic.

Nonetheless, many Canadian residents require detailed data on Canada — the population, the economy, and on the structure and functioning of the industrial and organizational structure. In Canada, we have a body of statutes that recognizes the importance of using administrative records for statistical purposes, while, at the same time demanding the protection of personal information and the preservation of confidentiality. Canada's *Privacy Act*, in particular, reflects the importance of preserving the principles of confidentiality and privacy, while acknowledging that this information can be used for reasons authorized by other federal statutes. In reading the *Privacy Act*, we find that:

"Personal information under the control of a government institution shall not, without the consent of the individual to whom it relates, be disclosed by the institution except in accordance with this section." (*Privacy Act*, §8(1))

Then, in §8(2), we find that:

"Subject to any other Act of Parliament, personal information under the control of a government institution may be disclosed
 "... (b) for any purpose in accordance with any

Act of Parliament or any regulation made thereunder that authorizes its disclosure;

"... (j) to any person or body for research or statistical purposes if the head of the government institution

"(i) is satisfied that 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

"(ii) obtains from the person or body a written undertaking 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*, §8(2))

Thus, the *Privacy Act* enables federal departments and agencies to disclose personal information in accordance with any Act of Parliament (§8(2b)), in particular, and, in general, for research and statistical purposes (§8(2j)).

Additionally, in the *Statistics Act*, we find the following general statement regarding access to administrative records:

"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 of that information, shall grant access thereto for those purposes to a person authorized by the Chief Statistician to obtain that information or aid in the completion or correction of that information." (*Statistics Act*, §13)

Then, and with regard to access to income tax records, we find the following statement in the *Statistics Act*:

"For the purposes of this Act ...

"(a) The Chief Statistician or any person authorized by the Chief Statistician to do so may inspect and have access to any returns, certificates, statements, documents or other records

obtained on behalf of the Minister of National Revenue for the purposes of the *Income Tax Act or Part IX of the Excise Tax Act*, and "(b) the Minister of National Revenue shall cause the returns, certificates, statements, documents, or other records to be made available to the Chief Statistician or person authorized by the Chief Statistician to inspect such records,

"in such manner and at such times as the Governor in Council may prescribe on the recommendation of the Minister and the Minister of National Revenue." (*Statistics Act*, §24)

In reviewing both the *Privacy Act* and the *Statistics Act*, we find that the former allows access if another Act of Parliament authorizes it (§8(2b)). The *Statistics Act*, in turn, specifically includes a provision for access to administrative records in general (§13) and specifically to taxation records (§24). The conclusion is that these two Acts of Parliament recognize the need to strike a reasonable balance between the privacy rights of individuals and the need for aggregated, non-confidential information to support the public interest.

Administrative Records Obtained and Used by Statistics Canada

In the Federal Government's annual index of holdings of personal information, *Info Sources*, Statistics Canada contributes a listing of the administrative record files obtained and used for statistical and research purposes. Also, for the personal income tax records, Revenue Canada registers its personal income tax database and clearly notes that the "[u]se of selected information from this bank is made by Statistics Canada in light of entitlement prescribed in the *Statistics Act*." (*Info Sources*, p. 745) In some cases, "[t]his has allowed Statistics Canada to exempt many respondents, particularly small businesses, from the filing of regular questionnaires." (*Info Sources*, p. 840)

Statistics Canada has recognized the importance of enriching administrative records through the linkage of records. Statistics Canada is, nonetheless, sensitive to the privacy implications and has adopted an internal

policy governing all of the linkages it performs. Before any linkage can be carried out, it must be reviewed and approved by an internal expert committee and the Chief Statistician. The Chief Statistician then refers each internally approved record linkage request to the Minister responsible for Statistics Canada.

"It is the Minister who (as proxy for the general public) applies the ultimate judgement regarding the trade off between the expected public benefit and the degree of privacy invasion which may be involved. Dissemination of information resulting from record linkages, like all other statistical information, is done only at levels of aggregation which protect the confidentiality of the individual elements included in the published totals." (*Statistics Canada, 1992-93 Annual Report on Access to Information and Privacy*, p. 6)

Only after this multi-layered approval process has been completed can record linkages be undertaken. Finally, Statistics Canada not only follows this record linkage review and approval process, but reports record linkages involving personal information in its Annual Report to Parliament on the administration of the *Privacy Act* (1992-93 *Annual Report*, p. 6). The individual files resulting from the record linkages are, in turn, described in *Info Sources*.

Administrative Records, Privacy and the Cost of Data Collection

Governments everywhere are facing a dual problem -- a desire for more data to manage and plan programs and the need to do so with dwindling resources. Statistical agencies are caught in this vice. On the one hand, there is a demand for more data, while statistical budgets are being squeezed.

Administrative records often cover large portions of the population and are a partial solution. Since the costs of maintaining them are attributed to program administration, the cost of deriving data from them is much less than direct data collection. In other words, administrative records represent one means of providing data at a lower cost [8].

Privacy and Administrative Records -- A Concluding Comment

In this last section of the paper, we have reviewed the extant legislation regarding Statistics Canada's access to and use of the administrative records. The conclusion is that both the *Privacy Act* and the *Statistics Act* enable Statistics Canada to access and use the personal income tax records for statistical applications. This right of access is conditioned on the requirement to report data from these records in groups and classes so that no person can be identified.

There are cogent arguments on both sides of this debate over privacy. One of the hallmarks of a modern nation is its ability to produce clear and consistent statistics on:

- the performance of its economy;
- the location and demographic structure of its population;
- the structure and location of business activity; etc.

Statistics Canada has a responsibility to provide statistical information, and to use cost-effective and efficient means of meeting this responsibility. In an era of dwindling budgets, the income tax records represent an opportunity to provide reliable and valid data each year on a large percentage of the population for a comparatively low cost. The current use of the income tax records strikes a balance between the preservation of personal privacy, while enabling the reporting of data on the population of Canada.

■ **Footnotes**

- [1] Unmarried persons who (a) declare themselves to be single, (b) reside with their parents, and (c) file a tax return are defined to be "filing children."
- [2] Beginning with the 1992 tax year, taxfilers were asked to report if they were "living common

law." As a result, there was a change in the time series data for common-law couples between 1991 and 1992. Moreover, since some lone parents were living common law and not identifiable, there was an increase in common-law unions and a reduction in the number of lone parent families between 1991 and 1992.

- [3] There is not an exact reference period relationship between T1FF and the population estimates. The closest is to use T1FF for one year and the population estimates for the next year. For example, the 1992 T1FF population is compared to the 1993 population estimates.
- [4] Taxfilers, in completing their T1, are asked to report information on their dependent children — relationship to taxfiler and birthdate. This information is not, however, captured. Until 1992, Revenue Canada linked the Family Allowance file to the tax file and added the ages of children (about 92% of children for the 1991 tax year). With the introduction of the Child Tax Benefits program in 1993, this information is no longer available.
- [5] The SCF is an annual supplement to the Canadian Labour Force Survey. The SCF is similar to the March supplement to the Current Population Survey (CPS) in the United States.
- [6] The Canadian government introduced tax reform in the 1986 tax year. Part of this tax reform included a tax credit for the Federal Sales Tax, a tax credit directed at the low income population, including low income families. In 1986, the credit was \$25 per person and available to everyone whose income fell below the prescribed maximum. Further, the credit was available to taxfilers who had no tax liability (i.e., a negative income tax?). By 1989, the credit had increased to \$70 per person who qualified (including non-taxfiling children). In 1990, the Federal Sales Tax was replaced by the Goods and Services Tax (GST). With the GST, the tax credit was increased to a maximum of \$190 per qualifying adult. (The maximum amount for de-

pendent children was lower.) By 1992, the GST Tax Credit had increased to \$199. For both of these tax credits, and in the case of families, the credits were based on family income, including non-taxable income from social assistance, Guaranteed Income Supplements, Spouse's Allowances and Workers' Compensation Benefits. In general, the two tax credit programs provided an incentive for individuals and families with low incomes to file tax returns. One impact of these two tax credit programs was increased taxfiling, a fact that is evident in reviewing the population coverage for the 65+ population. Furthermore, it was expected that the tax medians could decline relative to the SCF, and they did.

[7] Mrs. L. Desramaux, Director, Data Access and Control Services, provided invaluable assistance in preparing this section.

[8] For the period 1982 through 1990, the cost per record to enumerate the taxfiler plus dependents population has varied between a low of 6.8 cents per record in the 1987 tax year processing to a high of 10 cents per record for the 1984 tax year processing. (Statistics Canada, *Annual Program Report: Small Area and Administrative Data Division*. 1992-93, p. 6)

■ References

Canada. *Privacy Act*.

Canada. *Statistics Act*.

Ekos Research Associates (1993). *Privacy Revealed: The Canadian Privacy Survey*. Ottawa, Canada.

Rashid, A. (1994). *Comparison of 1990 Income Estimates from TIFF and Census*. Statistics Canada, Ottawa, Canada. Unpublished.

Régie des Rentes du Québec. *Les Allocations d'aide aux Familles, Statistiques 1992*. Québec, Canada.

Statistics Canada. *1991 Census Dictionary*. Catalogue Number 92-301E.

Statistics Canada. *Age, Sex and Marital Status: The Nation*. 1991 Census. Ottawa, Canada. Catalogue Number 93-310.

Statistics Canada. *Annual Demographic Statistics, 1993*. Demography Division, Ottawa, Canada. Catalogue Number 91-213.

Statistics Canada (1993). *Annual Program Report: Small Area and Administrative Data Division, 1992-93*. Ottawa, Canada. (Unpublished annual report).

Statistics Canada. *Annual Report: Access to Information and Privacy, 1992-93*.

Statistics Canada. *Family Incomes: Census Families. Survey of Consumer Finances*, Ottawa, Canada. Catalogue Number 13-208.

Statistics Canada. *Revised Intercensal Population Estimates, 1971-1991*. Demography Division, Ottawa, Canada. Catalogue Number 91-537.

Statistics Canada. *Unemployment Insurance Statistics, 1991*. Labour Division, Ottawa, Canada. Catalogue Number 73-202S.

Treasury Board Secretariat. *Info Source: Sources of Federal Government Information, 1993-1994*. ■