FAMILY STATISTICS FROM THE PERSONAL INCOME TAX FILE

Ann Jung and Rodger Turner, Revenue Canada - Taxation

1. INTRODUCTION

With the increased demand for socio-economic statistics and the rising cost of collecting information through censuses and surveys, the use of administrative files for the production of statistics is being widely urged. At the same time, the need for statistics on families is increasing. To these ends, research is underway into the creation of a system for family statistics utilizing existing administrative data files of Revenue Canada--Taxation (RCT).

The paper initially discusses the definition of family considered most appropriate for the purposes of this analysis. The construction technique to satisfy the chosen definition follows. The source of data is addressed briefly with mention made of alternatives considered. A brief reconciliation with external sources is provided.

The first half of the project is nearing completion. The approach taken has been to create one family type in the first year, then others are to be added in subsequent years. It is hoped to have a viable family statistical system which will produce statistical reports in the next two years.

2. FAMILY DEFINITION

The basic building block of a statistical system is the unit to which all information refers. In a fertilizer experiment it may be the plant to which treatments are applied; in a survey of political opinions it may be the individuals belonging to a particular party. For this project, the family comprised of two or more individuals is our unit of interest. Units of size one are defined as non-family persons.

There are many definitions of "family." The choice of one is based partially on the feasibility of its construction from the data available. RCT files for personal taxation consist of individual taxfilers of T1 returns. At the same time consideration is given to the appropriateness of the unit in relation to taxation statistics. Four family definitions were suggested. Each can be thought of as a subset of the prior one. First, the economic family, is a group of individuals related by blood, marriage or adoption who share a common dwelling unit. As an illustration, a husband and wife who have two children--a never-married son, of 26, attending university and a divorced daughter, of 28, who has a 4 year-old daughter from her first marriage--form one economic family. Within the economic family framework there may exist one or more census families. A census family consists, generally, of an individual with a spouse with or without never-married (single) children regardless of age or a parent with one or more never-married children who share a common dwelling unit. Our example results in two census families. The first composed of the husband, wife and son with the other being the daughter and granddaughter. A third type is a restrictive census family definition (sometimes termed nuclear family) where age limits are placed on the children. If one restricts the children's age in the definition to, say, 25, then the two nuclear families formed from the above example are the husband and wife--as one -- and the daughter and granddaughter, as the other. The son would become a non-census family person. The fourth and last considered is also a census family variant. The tax family is defined as an individual with or without a spouse but with dependent children. Under the present system of taxation, several conditions must be met to qualify as a dependent child. In general, the child's net income must be below a certain threshold. If over a certain age, the child must also be in full attendance at a school or infirm. To create a tax family or families from our example one requires additional information on the socio-economic status of the son, daughter and granddaughter. If the son were in attendance full time at school and his net income was below a certain amount, and the daughter was infirm and being fully supported financially by her parents, then it is likely that all five individuals would form a tax family.

The relative magnitudes of the economic and so-called census family definitions are worth noting. According to the 1976 Census of Canada, there were 5,769,860 economic families in Canada. Of these 5,727,895 (or 93%) were census families as well. One is able to conclude that nuclear family and tax family counts would be less than 93% of all economic families.

For reasons of comparability, usefulness of definition, and relevance to the Canadian tax framework, the choice made from the four is the census family. The tax system currently allows a number of transfers and claims within a census family structure. They include dependent exemptions, regardless of age; in some cases, transfer of unclaimed exemptions between spouses; reporting of Family Allowances; and the allowable Child Tax Credit, which is based on the sum of spouses' income.

All individuals can be partitioned into mutually exclusive census family units. The term family units refers to a generic type rather than the actual family. The three family units considered by this project are husband-wife families, one-parent families where the parent is widowed, divorced or separated and one-parent families where the parent is never married. Each is considered separately as they have different characteristics and relationships within the context of the taxfiler population.

The husband-wife family unit mentioned earlier, uses all married taxfilers who form the parents for the unit. The grouping of husbands and wives is called marital units for the purposes of the project. A married taxfiler either has a spouse who did file a return or there is a spouse present who did not file a return. In each case, they can be brought together (in one case by proxy) [1] to form marital units. According to the 1976 Census, there were 5,168,560 marital units with or without never-married children representing 90% of all census families (see Table 1). The individuals who are members of marital units plus their single children when present make up 81% of the population. An earlier version of this project for the 1976 tax year provides figures from RCT files indicating there were 5,126,528 marital units formed from the file of individual taxfilers. The 99% correspondence between Census and RCT counts is worthy of note in light of conceptual differences, such as timing (December 31, 1976 for RCT and June 1, 1976 for Census); exclusion from Census counts of data on collective households; treatment of common-law relationships (RCT does not presently recognize them); Canadian non-residents (Census tries to place them in their usual province of residence); and the lack of RCT units where both spouses are not required to file tax returns.

By considering the remainder of the filer universe, a partition of individuals who are single with those who are divorced, widowed or separated leads to the start of one-parent family definitions. The latter group provides the basis of the second census family unit, namely, non-single one-parent families. The 1976 Census determined there were 519,740 of these family units (9% of all census families). There were no equivalent figures produced from RCT files for 1976.

The third census family unit is the least numerous but, potentially, the most complex. It is the one-parent family with children where the parent is never-married. The 1976 Census reported 39,585 (1% of census families) never-married one-parent family units.

As well, within the single population there exist those persons who reside with their parents. They are designated census family children. Most are children in the colloquial sense, but there will be some who are normally considered adults as no restrictions are placed on the age of children. Figures from the 1976 Census indicate that there were 8,886,750 individuals (40% of population) in this category. From the RCT viewpoint, a substantial portion of this 40% will not be taxfilers. This is because more than 32% were under 18 years of age and may not have received income (or received less than the taxfiling threshold) and, hence, would not be required to file a tax return. An estimate from the statistical sample used to produce "Taxation Statistics" reveals less than 342,000 individuals (2% of the population) under the age of 18 filed a tax return for 1976.

The remaining population of individuals are those termed non-census family persons. They include persons with marital status of divorced, widowed, separated or single persons not living in a parentally dependent relationship.

A numerical summary of the family unit types with their respective population composition is given by Table 1.

3. FAMILY CONSTRUCTION

One inherent weakness of any computerized data collection system is the necessity to do that which it is designed to do at the expense of that which it would be nice to do. The system in place at RCT is set up to administer the Income Tax Act as it applies to individual taxfilers. This includes the assessment of T1 tax returns as well as the collection and administration of the Canada Pension Plan and Unemployment Insurance program. It was not designed to create family units for the purposes of statistical research and probable data dissemination. Therefore, the family statistics project is at the mercies of the items collected for administrative purposes.

The initial assumption made for family statistics is that the variable, marital status at year-end, is correct. Whether the assumption is good or bad becomes academic. There is no direct way to verify this item in the context of the RCT mandate.

Construction of family units initially creates the parent or parents of a family. Once all are designated as such the remainder become either children in families or non-census family persons. However, no unique family or household identifier exists on the tax returns which could facilitate linkage of all family members. In particular, there is no administrative requirement which could facilitate the linkage of child taxfiler records to their parents. This, however, does not presume matching is impossible, only that research and testing is necessary before results can be presented with confidence.

If a taxfiler is married, then there exists a spouse to whom the filer, theoretically, can be linked. The bringing together of these spouses forms the marital unit. As is often the case, there may be no requirement for one or both spouses to file a tax return. The latter situation is beyond our grasp. The former can be managed by creating a proxy return. A test linkage of married taxfilers for 1976 was carried out. The result was 195,898 wife-only-filer marital units (4% of marital units), 2,499,072 (49%) husband-only-filer units and 2,431,558 (47%) husband-wife-filer units. It should be noted that tax legislation in the last few years will increase this last percentage with an associated reduction in the second.

To do the linkage of spousal records, the initial step considers the Social Insurance Number (SIN) of a taxfiler and that of the filer's spouse. Every taxfiler is supposed to report his/her own SIN along with the spouse's on the filed return. Should a SIN match occur then sex and surname are checked and a code is assigned depending on the combination of the latter two variables. Where only one spouse filed, a proxy response is the result. Results for the 1979 tax year show 254,570 wife-only-filer units (5% of marital units), 1,234,469 (23%) husband-only-filer units and 3,795,588 (72%) husband-wife units given a total of 5,284,627 (100%) marital units. One can observe an approximate 25% switch from husband-only-filer units to husband-wife units between 1976 and 1979.

Consider now the group of taxfilers whose marital status is widowed, divorced or separated. In theory, the latter two can be linked to their spouses. By definition of family, however, linkage is not done. The group forms the pool for selection of the majority of one-parent family parents. To isolate them, characteristics of income, deduction, exemption and tax credit on their records are the conditions used. To be a parent a taxfiler must have a child to support. This support can take many forms. Given the information at our disposal, they should report or claim at least one or more of:

- Family Allowances;
- Child Care Expenses;
- Child Tax Credit;
- Exemption for Wholly Dependent Children;
- Education Deduction;
- Equivalent to Married Exemption.

The mechanism employed to ensure their parental status will be to assign a code based on the "majority of" principle. While we do not plan to go into all the details of the heirarchy, one example is a taxfiler reporting Family Allowances and Child Tax Credit. This person will be assigned a high code in the range for it is likely that the filer fully supports one or more children as defined within the T1 tax return system. Estimates for the 1979 tax year indicate that 1.75 million taxfilers had a marital status of widow(er), divorced or separated of which less than 20% of them were parents in one-parent families. The more than 80% remaining are designated as non-census family persons. They have a marital status which is neither married nor never-married. Hence, they cannot be considered components of a marital unit or children in families under the census family definition.

The last group of taxfilers are individuals with a marital status of single (never-married). There are three distinct classes into one of which each person can be placed.

The first, in operational sequence, is the single one-parent family parents. The criteria chosen for them essentially duplicates the mechanism for individuals who are widowed, divorced or separated. The assignment of a parental status code is based on the recorded values of income, deduction, exemption and tax credit taken from their tax return.

The next class are individuals who form the body of individuals known as children in families. The census family definition allows no discrimination on the basis of age as mentioned earlier. The requirement is the sharing of a common dwelling unit. Thus, a youth of 18 years with a summer job, who files a return to retrieve the tax deducted at source, is a child under this definition, along with the never-married business executive, who cares for an infirm parent in a common dwelling unit. The linkage of these individuals back to their parents is more difficult than that described for marital unit matching. There is no reporting of dependant's SIN either by the parent or child. In some instances it is available on the tax return but it is not captured in a machine-readable fashion. The software of the family project attempts to link principally by address. Secondary consideration is given to surname. A final prerequisite will be common characteristics. This takes the form of one or more items on a child's record having a relationship to items on the parental return. As an example, parents can claim the education deduction for a child should the child not need to claim the deduction. Another is the deduction for children. The parent accrues an exemption amount which is directly related to a child's net income. If the two amounts should match following address and surname agreement, then a linkage code will be assigned. In such a fashion, all children can be linked to their parents.

The non-census family persons who are single are those taxfilers remaining after stripping off all single one-parent family parents and census family children.

To summarize the construction, the first step is to split off all married tax filers who will form marital units. The second creates one-parent family parents based on their reported characteristics from the portion of tax filers who are widowed, divorced or separated. The remainder become the bulk of non-census family persons. Thirdly, all single tax filers are partitioned into single one-parent family parents, children in census families or non-census family persons. Thus all tax filers belong either to their particular census family or are designated as non-census family persons.

4. SOURCE OF FAMILY DATA

The reader will recall that at the start of the previous section a limitation on family construction was the purposive design for the computer system. A limitation which is just as imposing is the source of data. The T1 tax return is geared toward the individual not the family. It may be possible to go out and gather family data such as that which can be done by survey. A good example is the Survey of Consumer Finances carried out by Statistics Canada. However, an imposed restriction on the project was to generate family information from RCT administrative files. A question was raised as to whether a sample would meet our needs. Probably it would, but there are also drawbacks to it. Examination of the sampling alternative will be addressed later in this section. The decision was made to generate family statistics from 100% to the taxfilers.

The Taxpayer Master (TAPMA) file is large (138 reels of standard magnetic tape for the 1980 tax year). It is comprised of records for individuals who have filed one or more tax returns within five years of the taxation year. It has selected fields captured from the filed tax returns. In deciding to do a census of census families in which one or both parents filed a return, we chose to create family records which have a minimum length but contain a maximum amount of information. The information on such records should allow us to reconcile these data to other sources of comparable information and to provide a selected quantity of taxation statistics. Should the first year's set of information prove viable, extension through additional fields would be considered. This does not mean that every year more fields will be added indiscriminately, but rather evaluation of each field and assorted inter-relationships will be done. Thus a field may be substituted for a new one, should the first not meet our requirements. Additional fields will be added but it should be obvious that every new field of 5 bytes adds approximately 40 million bytes to this family file. Due to its size, there is presently scheduled to be only limited amounts of manipulation and tabulation on the computer file.

The sampling alternative was considered at length. One can form a sample design to select representative returns in many ways. Anything from simple random selection to a complex multi-stage multi-type of selection sample design is possible. Each has its own particular merits and demerits. The most popular (within the group discussing the problem) was a stratified random sample with post-stratification adjustment of selection probabilities. Simply stated, a stratification of the universe would produce mutually exclusive strata from which a sample of taxfiler records could be selected. Once selected, the software determines the family status of the return and would go back to retrieve the remaining family members still present but unselected. The inverse of the selection probability would then be adjusted for this modification to the sampling frame and would eliminate the possibility of multiple family member selection.

Another alternative considered to a lesser extent, was to create a frame from the previous year's universe and select all family members as they pass through the current year's program. At the same time, slight adjustments and additional sampling would be necessary to represent those in families filing this year who were not in families in the previous year. Changes in marital status could present complexities as it is estimated that as many as 25% of taxfilers change status from year to year. Changes in filing patterns, as well, affect this approach. For the 1979 tax year 1.7 million persons who filed a return in the previous year (1978) did not file a return for the current year (1979).

Perhaps one of the most difficult problems to resolve in sampling of families is one phase in data reconciliation. The phase relates to comparison with another sample chosen from the same universe. The sample referred to here is that used to produce "Taxation Statistics", a report on the administration and assessment of T1 tax returns by Revenue Canada Taxation. In theory, estimated totals of counts or amounts like net income from one sample should agree statistically with another sample. Some differences can be attributed to sampling. But clients look at you quizically when trying to explain such differences, especially when the client's program is critically dependent on your numbers which is the case of "Taxation Statistics".

An even greater difficulty arises when one tries to explain to one's supervisors when one of them is a "wild sample." (That term is left to the reader to derive his own definition.)

In light of sampling intricacies, reconciliation difficulties, the low computer cost (done on RCT computers) and person cost (a project leader and program analyst charged 0.8 person years between the start of the project in the fall of 1981 to April 1, 1983), the decision was made to use all taxfilers in the creation of family statistics.

5. PROJECT STATUS AND DATA RECONCILIATION

To date, the project has constructed all marital units for 1979 tax year. This linkage of husband-wife (or proxy) taxfilers is the extent of the project for 1979 tax year. Current activity for the 1980 tax year centres around the creation of one-parent family parents who are widowed, divorced or separated and the single one-parent family parents. Depending on time and results of work on this file, some investigation into children-in-family linkages will also take place. It is expected that the broad focus of the project will finish with 1981 tax year files. At that time, only minor refinements should be necessary to produce family statistics annually.

With the information available today, reconciliation of the 1979 tax year information considers two sources of external statistics; namely, the Survey of Consumer Finances for 1979 and the Census of Canada for 1981. There are, of course, differences in reference periods, the concept of family and of income which pose problems in doing direct comparisons. Nevertheless, an approximate level of agreement can be observed from the comparisons.

According to the 1981 Census, there were 5,610,970 marital units with or without never-married children. RCT file for the 1979 tax year indicates that there were 5,284,627 marital units formed from the file of individual taxfilers. This is 94% of the 1981 Census count. The undercoverage can be mainly explained by two factors: the increase in population from 1979 to 1981, and the marital units not in the RCT file because both spouses have not filed a tax return.

Table 2 shows the number of marital units derived from the project, and the number of husband-wife families from the censes of 1976 and 1981 by province. For the project, the province of residence for a taxfiler is at the time of filing the 1979 tax return. For the 1981 Census, the reference period is June 3, 1981. The highest coverage for a province is 99% for Manitoba, the lowest is 85% for Yukon territories. It seems reasonable to assume that the representation would be higher if the comparison could be made against the actual 1979 count.

The number of marital units and the percentage distribution by number of children are given in Table 3. Presented are the results from the project covering the 1979 tax year and the Censuses of Canada for 1976 and 1981. Direct comparison of the data is not quite appropriate due to the differences in the time period and the definition of children. However, the comparison does give an indication of the coverage of husband-wife families at a finer level.

Compared against the 1981 Census, the marital units produced through matching covers over 90% of the husband-wife families with two or less children. The coverage of over 100% for the husband-wife families with no children can be explained by the families with children who do not qualify as dependants. In some families with children, one or more children are not qualified as dependants. These families would be classified as having less numbers of children than they actually do. This would, in part, account for the lower coverage of the husband-wife families with many children.

Table 4 provides comparison between the result of the project and the statistics produced by the Survey of Consumer Finances (SCF) relative to the 1979 calendar year. Surveys of Consumer Finances are conducted by Statistics Canada on an annual basis. They are household surveys of income of families and individuals. SCF uses a sample that represents the civilian, non-institutional population of families and individuals in the ten provinces of Canada. The census family income of SCF includes the incomes of children 15 years of age and over as well as the incomes of parents.

The concept of income for the RCT project includes only those income items subject to assessment and excludes such items as: Veteran's pensions and allowances, workmen's compensation benefits, a variety of municipal and provincial welfare payments as well as the federal Guaranteed Income Supplement to the Old Age Security pension and the Spouse's Allowance. Also, the windfall gains such as receipts from lotteries, gifts, gambling and similar activities are excluded. On the other hand, the aforementioned items are included in the income defined by the Survey of Consumer Finances. At the same time, there are some items which are included in the RCT income concept but are excluded from the SCF income. These include capital gains or losses, receipts from the sale of property or personal belongings, death benefits, retiring allowance and termination payment.

Despite the differences in the concept of income, the comparison of the result from the project and the SCF data shows a general degree of comparability. Table 4 presents the average income of marital units from RCT data and the average income of husband-wife families from SCF for provinces. The RCT income data was adjusted to exclude the net taxable capital gains or losses, and to include only the actual amount, not taxable amount, of dividends received. The adjustment brings the RCT data closer to the SCF concept. Overall, the RCT average income is 93% of the SCF average. The closest agreement occurs for Alberta with 99%, and the lowest for Quebec with 91%. The difference could be attributed to the non-assessed income items that are excluded from the RCT income concept but included in the SCF concept.

NOTES

For further statistical results, the authors may be contacted at:

Statistical Services Division
Revenue Canada Taxation
875 Heron Road
Ottawa, Ontario, Canada
K1A 0L8

[1] A spouse who did not file a tax return is assumed to have had no income other than what is reflected in the Married Exemption of his or her spouse.

[2] TAPMA file (1976-80) contained records of 19,911,911 individual taxfilers.

TABLE 1 - FAMILY UNIT TYPES AND COMPOSITION (1976 CENSUS)

FAMILY UNIT TYPE AND COMPOSITION	FAMILY UNITS		POPULATION	
	NUMBER	%	NUMBER	%
MARITAL UNITS	5,168,560	90	18,176,250	81
A) WITHOUT CHILDREN*	3,446,010	60	6,892,020	31
B) WITH CHILDREN*	1,729,205	30	3,458,410	15
NUMBER OF CHILDREN	-		7,825,820	35
ONE-PARENT FAMILY UNITS	559,335	10	1,610,260	7
A) DIVORCED, SEPARATED, WIDOWED	519,740	9	519,740	2
B) SINGLE	39,585	1	39,585	-
NUMBER OF CHILDREN	-		1,060,930	5
NON-CENSUS FAMILY PERSON	-		2,628,800	12
TOTAL	5,727,895	100	22,415,310	100

* BASED ON 1976 CENSUS SAMPLE DATA. - LESS THAN 1%.

TABLE 2 - NUMBER OF HUSBAND-WIFE FAMILIES

PROVINCE	(A) CENSUS 1976	(B) RCT 1979	(C) CENSUS 1981	(B) AS % OF (C)
CANADA	5,168,560	5,284,627	5,610,970	94
NEWFOUNDLAND	113,855	113,608	121,665	93
PRINCE EDWARD ISLAND	24,685	24,120	26,560	91
NOVA SCOTIA	179,010	175,176	190,045	92
NEW BRUNSWICK	145,875	144,773	155,090	93
QUEBEC	1,381,505	1,319,291	1,463,100	90
ONTARIO	1,902,090	1,995,436	2,028,690	98
MANITOBA	227,240	230,885	232,920	99
SASKATCHEWAN	206,585	215,728	222,030	97
ALBERTA	407,570	446,877	508,720	88
BRITISH COLUMBIA	568,250	579,104	648,965	89
YUKON	4,430	4,230	4,970	85
NORTHWEST TERRITORIES	7,465	7,396	8,215	90

Table 3 - Number of Husband-Wife Families

No. of Children	1976 Census	% of Total	(a) 1979 RCT	% of Total	(b) 1981 Census	% of Total	(a) as % of (b)
0	1,729,205	33	2,081,444	39	2,012,560	36	103
1	1,072,085	21	1,110,986	21	1,192,140	21	93
2	1,219,235	23	1,298,006	25	1,441,595	26	90
3	655,015	13	548,712	11	651,140	11	84
4	298,910	6	172,949	3	216,830	4	80
5 or more	200,760	4	72,530	1	96,700	2	75
TOTAL	5,175,210	100	56,284,627	100	5,610,965	100	94

Table 4 - Average Income of Husband-Wife Families (1979)

Province	(a) RCT (M.U.)	(b) SCF (H-W)	(a) as % of (b)
NEWFOUNDLAND	17,484	18,947	92
PRINCE EDWARD ISLAND	17,602	18,792	94
NOVA SCOTIA	19,412	19,976	97
NEW BRUNSWICK	18,462	19,559	94
QUEBEC	21,186	23,400	91
ONTARIO	23,698	25,298	94
MANITOBA	20,051	21,916	91
SASKATCHEWAN	20,621	22,874	90
ALBERTA	25,518	25,884	99
BRITISH COLUMBIA	24,915	26,644	94
ALL	22,606	24,245	93