

Chapter 2 - Data Quality

Chapter 2 provides information on data quality and is composed of two parts:

Characteristics of the sample

The reader will find information on the sample design and the reliability of the estimates, covering:

- Target population
- Sample design
- Estimation
- Data Reliability

Other factors affecting data reliability

Non-sampling errors can also have an impact on data quality. The user will find information on these factors, such as the impact of rounding and adjusting high incomes and losses.

SAMPLE DESIGN AND ESTIMATION

1. Target Population and Geographical Limitations of the File

The target population for the file includes all individuals except for institutional residents and residents of incompletely enumerated Indian reserves or settlements.

In order to meet confidentiality criteria, the geographical information on the file is restricted. Aside from province, only twelve census metropolitan areas are identified.

2. Sample Design

The sample for the individual microdata file was selected in two stages. The first stage is the one-fifth (2B) sample collected during the 1986 Census. The second stage consists of a systematic sample of individuals drawn from within predetermined strata of the 2B sample.

2.1 First Stage

In the 1986 Census of Population, four out of five households were enumerated using a short questionnaire (2A). This questionnaire contained nine questions of demographic and linguistic nature. The remaining households received a more detailed questionnaire (2B) that, in addition to the nine 2A questions, contained twenty-three other questions covering a wide range of topics.

Each individual in the 2B sample represents him/herself and four other individuals, i.e. each individual carries a sampling weight of five. Statistical procedures that adjust the weight to obtain a more representative sample were also applied at this stage. Further details of the one in five sample can be obtained from the "Census Handbook," 1988, catalogue number 99-104E.

2.2 Second Stage

The second stage of sampling consisted of the following steps:

1. creation of subgroups, or strata, of individuals selected in the first stage,
2. sorting within strata, of the individuals according to specific criteria,
3. sampling within strata with special procedures for temporary residents and individuals who were sampled for other purposes.

Each one of these steps is described below.

2.2.1 Stratification

The population was initially divided into twenty geographic areas, namely:

Montreal,
Ottawa-Hull,
Toronto,
Winnipeg,
Calgary,

Edmonton,
Vancouver,
Regina and Saskatoon,
Kitchener, Hamilton, and St. Catherines,
Newfoundland,
Prince Edward Island,
Nova Scotia,
New Brunswick,
Quebec (excluding Montreal and Hull),
Ontario (excluding Ottawa, Toronto, Kitchener, Hamilton, and St. Catherines),
Manitoba (excluding Winnipeg),
Saskatchewan (excluding Regina and Saskatoon),
Alberta (excluding Calgary and Edmonton),
British Columbia (excluding Vancouver), and
Yukon and the Northwest Territories.

Six groups were then identified in each geographic region using the following classification:

1. individual is less than 15 years of age,
2. individual's age is between 15 and 44 years and in the labour force,
3. individual's age is between 15 and 44 years and not in the labour force,
4. individual's age is between 45 and 64 years and in the labour force,
5. individual's age is between 45 and 64 years and not in the labour force, and
6. individual is 65 years of age or older.

The combination of the twenty geographical areas and six classifications yields a total of 120 strata.

2.2.2 Sorting

Sorting was performed within each strata. The first sort was by ethnic origin. Four values were used:

1. British,
2. Asian,
3. European, and
4. Other (includes multiple origins).

Within each ethnic origin value, a secondary sort was applied. For temporary residents or persons residing in dwellings overseas, collective households, Prince Edward Island, or the Yukon and Northwest territories, the secondary sort was based on sex. For the remaining individuals, the secondary sort was by group. These groups consisted of all possible combinations of the these variables:

Sex;

1. female, and
2. male.

Relationship to Person 1;

1. person 1,
2. relative of person 1, and

3. not a relative of person 1.

Area of Residence;

if the geographic area was composed of cities:

1. urban core,
2. urban fringe, and
3. rural fringe.

if the geographic area was not:

4. urban area with 30,000 people or more,
5. urban area with less than 30,000 people, and
6. rural.

In order to equalize the individual effects of secondary sort variables in sampling, the groups formed were randomly ordered within ethnic origin within strata. For example, individuals in the stratum defined as persons residing in Montreal whose age is between 15 and 44 and are in the labour force would first be sorted by the four ethnic origins above. Then, within each of the four ethnic origins, individuals would be sorted into groups. The order of the groups would be random, i.e. the first group could be male relatives of person 1 in the urban core (combination 2-2-1), the next group female who are not related to person 1 in the urban fringe (1-3-2), and so on until all thirty-six combinations are exhausted.

Finally, within each of these groups, the individuals were put into a random order.

2.2.3 Sampling

Within each strata, the sample was systematically selected with a random start and probability proportional to the first stage weight. Sample sizes within stratum were determined such that each individual selected represented fifty individuals in the target population. Special procedures, described below, were required for temporary residents and overlap with other microdata files.

Temporary Residents

Temporary residents are not part of the Census one-fifth sample but are accounted for by randomly assigning each temporary resident to an individual who is part of the sample. This individual then represents him/herself, four other persons, and the temporary resident. If a temporary resident is selected for the individual microdata file, the information carried is that of the individual assigned to that temporary resident. If both an individual and his/her's associated temporary resident are selected, two identical records will appear on the microdata file.

Overlap with Other Files

Other microdata files are available from the 1986 Census of Population. In order to reduce the possibility of disclosure, any overlap between files has been systematically eliminated. Thus any individual selected for the Health and Activity Limitations Survey and any individual that is a member of a household or

family selected for the household or family microdata files cannot be selected for the individual microdata file. Individuals who are ineligible for the file are accounted for in the following manner:

1. Within each stratum, the sample size necessary for a one in fifty sample of the target population is determined.

E.g. A stratum contains 550,000 individuals. 11,000 individuals are required for a one in fifty sample.

2. Individuals ineligible for inclusion in the microdata file are determined and discarded.

E.g. 55,000 of the individuals in the above example are found to be ineligible. Thus the stratum now contains 495,000 individuals.

3. Sample selection occurs within the remaining individuals but retains the original sample size.

E.g. continued. The sample of 11,000 is selected from the 495,000 remaining individuals, resulting in a one in forty-five sampling rate.

4. Steps 1-3 are repeated within each stratum. In order to account for those individuals excluded, a sampling weight of fifty is used to calculate estimates.

This method assumes that within a stratum, the distribution of characteristics of interest is the same for eligible and ineligible individuals.

Caution should be applied in the analysis of estimates that might be affected by the absence of individuals who might identify themselves as health and activity limited.

3. Estimation

The individual sample is designed to allow for easy estimation. There are two types of variables on the microdata file, numeric and coded. Numeric variables, such as income, show an actual amount. Coded variables have values based on a classification, such as the variable on occupation. Some typical estimators for the two types of variable follow.

3.1 Coded Variables

Estimates of Total

Estimates of total are obtained by selecting the individuals possessing the value of interest (e.g. English as the language spoken most often at home), counting them, and multiplying the resulting count by fifty. Any estimate of total that results in less than 5000 individuals should be used with caution.

Estimates of Ratios and Percentages

Ratio estimates involve counting the number of individuals in the numerator, the number in the denominator, and dividing. For a percentage estimate, the ratio estimate is multiplied by 100. If there are less than 100 individuals in the numerator, the ratio should be used with caution.

3.2 Numeric Variables

Estimates of Total

Estimates of total are obtained by selecting the individuals having the characteristic of interest (e.g. retirement income), adding up their values, and multiplying the resulting sum by fifty. If less than 100 individuals possess the characteristic of interest, the total should be used with caution.

Estimates of Average

Average estimates involve adding up the values of the characteristic of interest and dividing the resulting sum by the number of individuals in the numerator. Do not multiply the average by fifty. If there are less than 100 individuals in the numerator, the resulting average should be used with caution.

Estimates of Ratios and Percentages

Ratios are calculated by deriving the totals for the numerator and denominator and then dividing. For a percentage estimate, the ratio estimate is multiplied by 100. Do NOT calculate the ratio first for each individual and then average the resulting ratios. If there are less than 100 individuals in the numerator, the ratio should be used with caution.

3.3 More complex analysis

Caution must be exercised in performing more complex statistical procedures to the individual microdata file. Depending on the technique, the fact that each individual on the file represents fifty individuals may have to be taken into account. In addition, procedures that use statistical tests of significance (e.g. ANOVA) may give incorrect results if the sample design is not taken into account. Users wishing further information should contact the Chief, Census Data Quality Section, Social Surveys Methods at Statistics Canada.

4. Data Reliability

The individual file is a sample of the persons enumerated in the Census, and, as a result, estimates generated from the file will not agree exactly with published estimates. This lack of agreement is due to error. In the individual file, there are two sources of error, sampling errors and non-sampling errors. Each will be discussed in turn below.

4.1 Sampling Errors

Sampling error is the error incurred because all of the individuals in the population do not contribute their values to the construction of estimates. Each

individual selected in the file represents 49 other individuals who are not exactly like the individual selected. These differences are most commonly presented as sampling variance. Sampling variance can be calculated; procedures for doing so are outlined in section 4.3 below and tabulated in Appendix B.

4.2 Non-sampling error

Sampling error is only one component of the total error of an estimate. The other components of error include such types of error as response error (e.g. an individual provides an incorrect response), coverage error (e.g. individuals are missed or counted twice), processing errors (e.g. questionnaires lost, incorrectly loaded into the computer), etc. These types of errors are difficult to estimate and their existence for the most part (q.v.) in the individual file is inherited from the original Census data. It is important to note that, unlike sampling error which decreases as sample size increases, non-sampling errors tend to be a constant proportion of the survey estimate and can result in serious biases.

Two possible sources of non-sampling error have been deliberately introduced to the individual file. Certain confidentiality criteria have been applied to the individual file that have resulted in the suppression of selected information for selected individuals. Although these criteria should not affect most analysis, the possibility does exist. A selection bias has also been introduced. Persons who identified themselves as health or activity limited or who were in households selected in the household file were not sampled. Although this selection bias is difficult to quantify, statistical tests have been performed to ensure the sample is reasonably representative of the population. Therefore selection bias can be assumed to be zero, except in cases of variables affected by the absence of health and activity limited persons. Whenever such variables are used, the resulting analysis should be subjected to extra scrutiny.

4.3 Estimation of Sampling Variance

It is important to consider the sampling variance in any analysis. The utility of an estimate decreases as the sampling variance increases. A frequently used measure of sampling variance is the coefficient of variation (c.v.), which expresses the square root of sampling variance as a percentage of the estimate of interest. The following guidelines have been established at Statistics Canada relating the amount of sampling variance and whether an estimate should be released.

Category	Coefficient of Variation (%)	Alphabetic code	Recommendation
Unrestricted	0 to 0.5	A	Estimates may be included in a general release with- out restriction. Use of the alphabetic code is rec- ommended. The letter A indicates that the estimate is very reliable; the letter B indicates that the estimate is reliable, but less so than category A, etc.
	0.6 to 1.0	B	
	1.1 to 2.5	C	
	2.6 to 5.0	D	
	5.1 to 10.0	E	
	10.1 to 16.5	F	

Category	Coefficient of Variation (%)	Alphabetic code	Recommendation
Restricted	16.6 to 25.0 25.1 to 33.3	G H	The estimates are sufficiently reliable for specific purposes, but must be used with great caution. Anytime they are used, it must be pointed out that their sampling variance is high.
Not to be released	33.4 and over	I	The estimates must not be released in any form or under any condition. They should be deleted from statistical tables.

Instructions for calculating approximate sampling error for both coded and numeric variables follow. Exact estimates of sampling error cannot be easily calculated from the file as information about the sampling design is not readily incorporated into the structure of the file.

4.3.1 Coded Variables

Tables of approximate sampling variability for coded variables are provided in Appendix B.

Estimates of Total

1. Locate the appropriate table in Appendix B, based on geography.
2. Look up the number closest to and greater than the estimate of interest in the "Numerator of Percentage" column.
3. The estimated c.v. is the first number to the right of the "Numerator of Percentage" column, ignoring any columns filled with asterisks.

Estimates of Percentages

1. Locate the appropriate table in Appendix B, based on geography.
2. Count the number of individuals falling in the numerator, multiply by 50, and look up the number closest to and greater than this estimate in the "Numerator of Percentage" column.
3. Select the column closest to and less than the estimated percentage.
4. The estimated c.v. is found at the intersection of the row identified in step 2 and the column selected in step 3.

Estimates of Ratios

1. If the numerator is a subset of the denominator, the instructions are identical to those used for percentages (The ratio must be converted to a percent to use the tables).
2. If the numerator is not a subset of the denominator, then the c.v.'s for the numerator and denominator must first be ascertained using the "Estimates of Total" instructions above. Each of the c.v.'s is then squared. Then the two

squared c.v.'s are added together and the square root taken of the resulting sum.

This formula will have the effect of overestimating the c.v. if there is a positive correlation between the numerator and denominator and of underestimating it if there is a negative correlation.

4.3.2 Numeric Variables

The formula given below for calculating approximate sampling variability are themselves approximations. More precise formulae can be found in any university text on sampling theory.

Estimates of Total and Average

1. Calculate the sum of squares by taking each record's value, squaring it, and then adding all the observations together.
2. Calculate the squared sum by adding up all the values and then squaring this sum.
3. Divide the sum of squares (step 1) by the squared sum (step 2).
4. Take the square root of the result of step 3 and multiply by 100.

Estimates of Ratios and Percentages

1. Perform steps 1, 2 and 3 from "Estimates of Total and Average" above for the numerator.
2. Perform steps 1, 2 and 3 from "Estimates of Total and Average" above for the denominator.
3. Take each record's value for the numerator, multiply it by the denominator, and add up the resulting products across all records of interest.
4. Add up all the values for the numerator, add up all the values for the denominator, and multiply the two sums together.
5. Multiply the result from step 3 by two and then divide by the result of step 4.
6. Add the results from steps 1 and 2 together and then subtract the result of step 5.
7. Take the square root of the result of step 6 and multiply by 100.

OTHER FACTORS AFFECTING DATA RELIABILITY

ADJUSTMENTS TO GEOSTATISTICAL AREAS

Users should be aware that census geostatistical areas are subject to change from one census to the next. Therefore, when using data from two or more censuses, the user must be aware of, and take into consideration, any changes to the geographic limits of the areas being compared. Users wishing to obtain additional information in this regard should refer to Chapter 3.

POPULATION COUNTS BASED ON USUAL RESIDENCE

The population counts shown here for a particular area represent the number of Canadians whose usual place of residence is in that area, regardless of where they happened to be on Census Day. Also included are any Canadians staying in a dwelling in that area on Census Day and having no usual place of residence elsewhere in Canada. In most areas, there is little difference between the number of usual residents and the number of people staying in the area on Census Day. For certain places, however, such as tourist or vacation areas, or those including large work camps, the number of people staying in the area at any particular time could significantly exceed the number of usual residents shown here.

IMMIGRANT POPULATION AND POPULATION BORN OUTSIDE CANADA

All persons born outside Canada are not necessarily immigrants to Canada. Individuals who have reported their place of birth outside Canada, but who are Canadian citizens by birth, are not considered immigrants to Canada. Consequently, they do not have a period of immigration or age at immigration when they take up permanent residence in Canada. These individuals will be included in the non-immigrant population. This approach was used in the 1981 Census. By contrast, in the 1971 Census, all persons born outside Canada were categorized as immigrants and required to respond to the question on period of immigration.

MOBILITY STATUS

The geographic areas reflect boundaries as of January 1, 1986, the geographic reference date for the 1986 Census of Canada.

The counts for total "migrants" (a migrant is anyone who, five years earlier, did not have his/her usual place of residence within the census subdivision (CSD) where he/she was enumerated) are additive across any geographic level - e.g., the migrant count at the Canada level is the sum of the migrants at the provincial level.

At the CSD level, users are advised to exercise caution in the use of data on migrants, particularly for suburban municipalities within large metropolitan areas. Counts for total migrants, including in- and out-migrants, could be distorted due to suspected types of mis-response such as: (a) respondents in metropolitan areas reporting the main city rather than the municipality they actually lived in five years earlier (e.g., reported Toronto instead of Scarborough); (b) respondents failing to indicate a move from a different CSD if they perceived that they were still in the same main city (e.g., moved from Toronto to Scarborough but indicated that they still lived in the same municipality); and (c) respondents reporting moves according to out-of-date boundaries.

The concept of "migrant" is defined at the CSD level. For geographic levels below the CSD, such as enumeration areas (EAs) and census tracts (CTs), please note that the distinction between the migrant and non-migrant population refers to the corresponding CSD of the EA or CT. For example, migrants of a CT are those persons who moved from a different CSD, while non-migrants are those who moved within the same CSD - they moved either between different CTs or within the same CT.

Names and boundaries of particular census subdivisions may undergo trivial or, in some cases, substantial modifications during the five-year intercensal period; therefore, comparisons of data for a specific subprovincial area between any two censuses will not be valid unless these changes, if any, are accounted for.

Details of intercensal boundary changes can be found in the **Standard Geographical Classification** (Cat. No. 12-573).

Boundaries and CSD components of CMAs and CAs will often undergo modifications during the intercensal period; therefore, comparisons of data for specific areas between any two censuses will not be valid unless these changes are accounted for. A publication is available which provides comparisons of 1986 CMAs and CAs, and their 1981 versions. **Census Metropolitan Areas and Census Agglomerations: A 1986 and 1981**

Comparison (Cat. No. 99-105E or F) lists census subdivisions that make up the 1986 version of each CMA and CA, and shows corresponding delineations for 1981.

NUMBER OF WEEKS WORKED

The data for the 40-48 and 49-52 weeks worked categories for 1985 must be interpreted with caution because some respondents tend to exclude their paid leave of absence due to vacation or for other reasons from their work weeks, when in fact such leave of absence should be included. As a result, the 49-52 week category may be understated.

LABOUR FORCE ACTIVITY

The census labour force activity concepts have not changed between 1981 and 1986. However, the processing of the data was modified causing some differences. In the 1986 Census, contrary to previous censuses, a question on school attendance was not asked. This question was used to edit the labour force activity variable, specifically unemployment. Consequently, the processing differences affect the unemployed population and are mostly concentrated among the 15-19-year age group. The table on the following page indicates the magnitude of the effect upon the data, at the Canada level.

COMPARABILITY AND QUALITY OF LANGUAGE DATA

Comparison between 1981 and 1986

Mother tongue and home language. The language questions were the same in the last two censuses, but the instructions to respondents were modified for mother tongue and home language. In 1981, respondents were asked to indicate only one mother tongue and only one home language; nevertheless, 597,980 persons (2.5% of the population) reported more than one mother tongue and 535,735 persons (2.2% of the population) reported more than one home language.

To better reflect the linguistic reality in Canada, these instructions were dropped from the 1986 Census. Under the new guidelines, individuals could report more than one mother tongue if they had learned them at the same time and had spoken one as frequently as the other when they were children. Similarly, respondents could indicate

more than one home language if they were now speaking them equally often at home.

The number of multiple responses given in the 1986 Census was significantly higher than in the 1981 Census. In 1986, 954,940 persons or 3.8% of the population reported a multiple response to the mother tongue question, while 1,159,675 or 4.6% of the population indicated more than one home language.

This increase was the result either of the changes made in the questionnaire, of changes in the way in which the population answers language questions or of an increase in the number of persons who had more than one mother tongue or spoke more than one language at home. A combination of these factors may also explain the increase in multiple responses.

When the 1981 data were processed, only one language was retained for publication, even in cases where the respondent reported more than one. In 1986, responses indicating more than one language were accepted.

In order to facilitate the determination of the trends between the two censuses, the 1986 Census results have been adjusted. In cases where more than one language was reported, the multiple responses were distributed among the component languages in the same proportions as in the 1981 Census. The results have been published in a special document entitled "**Adjusted Language Data**", April (1988). Also, data from the 1981 Census have been adjusted to show the multiple responses reported at that time. The data are presented in Table 4 of publications 93-102 (mother tongue) and 93-103 (home language). These adjustments to the mother tongue and home language figures make it easier to relate the 1986 data to the 1981 data, but do not make the results of the two censuses entirely comparable. Consequently, considerable care must be exercised in the interpretation of changes between 1981 and 1986.

The 1986 Classification of languages differs from that used in 1981, especially with regard to aboriginal languages. Appendix B of the **1986 Census Dictionary** (Catalogue No. 99-101E) provides a description of the changes.

Official language - Some respondents report speaking English or French or both at home, while on the other hand they indicate in the official language question, that they cannot carry on a conversation in these languages.

Labour Force Activity, 1981 Census of Canada

Canada	1981 Census (as published in 1981)	1981 Census (using 1986 processing)	% change
Labour force 15 years and over	12,054,150	12,081,280	0.23
Employed	11,167,915	11,167,915	no change
Unemployed	886,235	913,365	3.06
Not in the labour force	6,555,135	6,528,005	-0.41
Labour force 15-19 years	1,073,945	1,098,390	2.28
Employed	906,705	906,705	no change
Unemployed	167,240	191,680	14.61
Not in the labour force	1,229,630	1,205,190	-1.99
Labour force 20 years and over	10,980,205	10,982,890	0.02
Employed	10,261,210	10,261,210	no change
Unemployed	718,995	721,685	0.37
Not in the labour force	5,325,505	5,322,815	-0.05

In such cases, in the 1981 Census, the answer to the official language question was considered erroneous. Consequently, during data processing, this answer was changed to show that the person could speak the official language(s) they had reported to the home language question.

In the 1986 Census, not all of these responses were considered erroneous. If the respondent indicated being able to speak only one official language - either English or French - and this language matched the person's mother tongue, no correction was made during processing. Consequently, these response patterns appear as such in the 1986 tabulations.

For further information on language data, contact the Housing, Family and Social Statistics Division, Statistics Canada, Ottawa, Canada K1A 0T6.

COMPARABILITY OF DATA ON ETHNIC ORIGIN

Comparison between 1981 and 1986. The 1981 and 1986 ethnic origin data are not directly comparable.

The 1981 ethnic origin question: To which ethnic or cultural group did you or your ancestors belong on first coming to this continent?, was modified for the 1986 Census. The phrase "on first coming to this continent" was removed from the 1986 version as it was viewed as being inappropriate for persons of aboriginal origin. The 1986 question was: To which ethnic or cultural group(s) do you or did your ancestors belong?

In 1986, respondents were instructed to mark or specify as many groups as apply. This instruction

along with the addition of two more write-in spaces contributed significantly to an increase in multiple ethnic origin responses.

As well, the mark boxes in the question were ordered on the basis of 1981 incidence reporting of single ethnic origins. This changed the relative position of the mark boxes Chinese and Polish.

In light of the recommendations of a Parliamentary Commission on Visible Minorities in Canadian Society in the report Equality Now and the Abella Commission on Equality in Employment, the mark box Black was added to the 1986 ethnic origin question.

The mark boxes for aboriginal peoples were also changed. In 1986, status and non-status Indian categories which had been part of the 1981 ethnic origin question were replaced by North American Indian. It should be noted that persons of non-aboriginal cultural origin but status Indian under the Indian Act of Canada, for example, persons who obtained Indian status at marriage, could have been included in 1981 data for aboriginal peoples. These persons may not have identified their ethnic origin to be North American Indian in 1986 and thus would not be included in the 1986 count of aboriginal peoples. Also, in 1986, an undetermined number of persons of Métis origin could have indicated their ethnic origin as being the multiple response North American Indian and some other ethnic or cultural origin(s).

Single and Multiple Response

A **Single Response** occurs when the respondent provides only one origin. For example, for Canada, 709,585 gave Italian as their only ethnic origin.

A **Multiple Response** occurs when the respondent provides more than one origin. Some 297,325 Canadians gave a response which included Italian and one or more ethnic or cultural origin(s). For example, 31,495 provided the multiple response combination: Italian and French.

In the ethnic origin legend for this profile, the single origins are shown as unique groups. The multiple origins are shown as one group: multiple origins. In the case of the 31,495 Italian and French multiple response combination, it would be included in the multiple origins count (6,986,345 for Canada).

For further information regarding the data on ethnic origin, please contact the Housing, Family and Social Statistics Division, Statistics Canada, Ottawa, Ontario K1A 0T6, telephone (613) 951-2574.

HOUSEHOLD MAINTAINER

Users of data on household maintainers, such as sex of maintainer or mother tongue of maintainer, should be aware of certain limitations which can potentially have a large impact on the use and analysis of these data.

The household maintainer variable is a derived variable, a combination and manipulation of the responses that users have provided to the question on "person responsible for payments" and the question on "relationship to Person 1". The purpose of the household maintainer variable is to classify families within a household as primary (i.e. families of which the maintainer is a member) or secondary (i.e. families of which the maintainer is not a member). The variable is neither designed nor recommended for use as the equivalent of the previous "Household Head" variable for analytical purposes.

The variable itself was not treated, during processing, as a variable to be used in analysis. For example, if a respondent listed more than one name under the "person responsible for payments" question, only the first name inscribed was captured; the others were discarded. In addition, if a respondent indicated that no person in the household made shelter payments, the household was left without a primary family, but Person 1 was arbitrarily assigned to be the household maintainer. The basis for these processing decisions was the priority of categorizing families as primary or secondary, not providing a reference person for the household.

Users are cautioned, therefore, to refrain from making unjustified inferences based solely on direct comparisons of characteristics of household maintainers. For example, one should be careful when comparing female maintainers with male maintainers because an unknown number of each may have been entered as a second entry in the "person responsible for payments" question, and subsequently discarded. Similarly, a number of cases may have occurred in which a person outside the household has been replaced by "Person 1" in the derivation of the household maintainer, resulting in a person of a different sex ending up as the household maintainer.

Misinterpretation of results can also occur when using other maintainer characteristics, such as mother tongue or ethnic origin, to classify a household because these characteristics can be different for the other members of the household. It is suggested that analyses using these variables also take into account the characteristics of the maintainer's spouse.

STRUCTURAL TYPE

Users of structural type of dwelling data are cautioned about certain limitations of the data. Initial investigation of these data reveals the following limitations which may affect the quality of the data:

- (1) In the 1986 Census, there was a higher rate of non-response to the structural type of dwelling question than in 1981 (2.3% compared with 0.5%). The impact of this higher non-response on overall data quality should be small, except in a limited number of geographic areas where non-responses may have been concentrated. It should also be noted that the information on structural type was reported by the Census Representative in 1986, whereas, in 1981, it was reported by the household respondent.
- (2) Sharp declines between the 1981 and 1986 Censuses were found in every province for mobile homes and other movable dwellings. This is thought to be due to the misclassification of a number of mobile homes as other structural types, primarily single-detached dwellings. For larger geographic areas, this error is not expected to have a significant impact upon other dwelling categories because of the relatively small number of mobiles and movables.

- (3) Apartments in buildings of less than five storeys present some differences with 1981 Census counts, especially in Quebec and particularly in Montréal. Also, high overcounts in 1981 of duplexes, double houses and row houses resulted in sharp declines for these types in 1986 in certain provinces. An initial historical analysis indicated the 1986 counts were quite realistic.

INCOME DATA

The 1986 Census collected income information from all individuals, 15 years and over, in private households and non-institutional residents of collective households. The total income concept included, for the first time, federal child tax credits. Income statistics for families and households are for those in private households only.

Census income statistics are subject to sampling variability. Although such sampling variability may be quite small for large population groups, its effects cannot be ignored in the case of very small subgroups of population in an area or in a particular category. This is because, all other things being equal, the larger the sample size, the smaller is the error. For this reason, published income data for areas below the provincial level, where the non-institutional population was less than 250, have been suppressed. The users of this microdata file are strongly advised to exercise caution in the interpretation of statistics based on relatively small totals.

Rounding and Adjustment of High Incomes and Losses

In planning this microdata file, it was deemed essential to utilize procedures to guard against the possibility of associating a particular income with an identifiable individual, family or household. To accomplish this, the income of individuals selected for this microdata file were subjected to the following rounding and adjustment procedure.

The incomes of individuals on this file were subjected to two separate operations. Initially, the amounts in wages, self-employment income (farm plus non-farm), investment income, retirement

pensions, other money income and total income were rounded to the limits specified in Table 1. This rounding procedure created certain inconsistencies between the sum of sources of income and total income. These inconsistencies were rectified by applying the adjustment procedure specified in Table 2. The income variables on this microdata file were derived after the individual records had been rounded and adjusted.

The number of records affected by this procedure and its impact on individual income are summarized in Tables 3 to 9.

Table 3 provides a distribution of individuals who had one or more sources of income and/or total income outside the limits imposed by confidentiality consideration.

Table 4 provides a summary of the changes, at the sample level, in the aggregate and average individual income, by source, as a result of the rounding/adjustment procedure.

Tables 5 and 6 provide distributions of the weighted aggregate income in 1985, by source and province, respectively, from the Census and the Public Use Microdata File.

Table 7 provides a comparison of the aggregate income in 1985, by source, from the Census, the Public Use Microdata File and personal income estimates from the System of National Accounts.

Table 8 provides a comparison of the average income of individuals in 1985, by province, from the Census and the Public Use Microdata File.

Table 9 provides comparable 1985 individual income size distributions, for Canada, from the Census and the Public Use Microdata File.

Table 1: High and Low Income Limits

(1)	The following income sources were subject to lower and upper limits for all individuals 15 years and over in the sample on the Individual Microdata File:		
	(a)	Wages and salaries	
	(b)	Income from self-employment	
	(c)	Investment income	
	(d)	Retirement pensions	
	(e)	Other money income	
(2)	The limits were as follows:		
		<u>Negative</u>	<u>Positive</u>
	Limit 1 (L1)	-\$30,000	\$100,000
	Limit 2 (L2)	-\$50,000	\$140,000
(3)	Amounts above or below the limits in (2) above were rounded to the appropriate limits as indicated in (4) below.		
(4)	Individual Income	<u>Atlantic Region</u>	<u>Other Regions</u>
	Males	L1	L2
	Females	L1	L1
	Family/Household Income		
	1 income recipient - female	L1	L1
	1 income recipient - male	L1	L2
	All other families/households	L2	L2
(5)	In cases where only total income was beyond the limit, as a first step, it was rounded to the applicable limit.		
(6)	To ensure consistency between the sum of sources and total income, individual records were then subjected to the adjustment procedure described in Table 2.		

Table 1: High and Low Income Limits

(1) The following income sources were subject to lower and upper limits for all individuals 15 years and over in the sample on the Individual Microdata File:		
(a) Wages and salaries		
(b) Income from self-employment		
(c) Investment income		
(d) Retirement pensions		
(e) Other money income		
(2) - The limits were as follows:		
	<u>Negative</u>	<u>Positive</u>
Limit 1 (L1)	-\$30,000	\$100,000
Limit 2 (L2)	-\$50,000	\$140,000
(3) Amounts above or below the limits in (2) above were rounded to the appropriate limits as indicated in (4) below.		
(4) Individual Income	<u>Atlantic Region</u>	<u>Other Regions</u>
Males	L1	L2
Females	L1	L1
Family/Household Income		
1 income recipient - female	L1	L1
1 income recipient - male	L1	L2
All other families/households	L2	L2
(5) In cases where only total income was beyond the limit, as a first step, it was rounded to the applicable limit.		
(6) To ensure consistency between the sum of sources and total income, individual records were then subjected to the adjustment procedure described in Table 2.		

Table 2: Adjustments Made to Remove Inconsistencies Introduced by Rounding

After rounding of applicable sources and/or total income as outlined in Table 1, individual sources and total income were subjected to the following adjustment routine in order to ensure consistency between the sum of sources and total income:

I. Adjustment of Sources

- (1) If $A > 0$ and $B > 0$ and $A < B$ then $S_f = (S_i)(C/D)$
- (2) If $A < 0$ and $B < 0$ and $A > B$ then $S_{ef} = S_e + A - B$
- (3) No adjustment in all other cases

II. Adjustment of Total Income

$Y = \text{Sum of Sources (after adjustments above and including transfer payments)}$

- A = Total income after rounding
- B = Sum of sources after rounding
- C = A less transfer payments
- D = B less transfer payments
- S_i = Rounded wages, self-employment, investment, retirement and other money income
- S_f = Final wages, self-employment, investment, retirement and other money income on PUMF
- S_e = Rounded self-employment income
- S_{ef} = Final self-employment income on PUMF
- Y = Final total income on PUMF

TABLE 3. Distribution of individuals (unweighted sample) with Incomes Outside Positive and Negative Limits(1) in 1985, PUMF (individual), 1986 Census

Source outside limits	Sample count	Percent
One source	751	73.0
Wages and salaries	335	32.6
Self-employment income	304	29.5
Investment income	108	10.5
Retirement Income	2	0.2
Other money income	2	0.2
Two sources	27	2.6
Wages and self-employment	2	0.2
Wages and investment	15	1.5
Wages and Retirement Income	1	0.1
Self-employment and investment	6	0.6
Self-employment and Other Money Income	1	0.1
Investment and Retirement Income	1	0.1
Investment and Other Money Income	1	0.1
Three sources	1	0.1
Wages, self-employment and investment	1	0.1
Total income only	250	24.3
TOTAL	1,029	100.0

(1)See Table 1 for limits

TABLE 4. Number of Individuals, (unweighted sample) their Original and Changed Aggregate and Average Incomes, by Source, and Composition of Income in 1985, PUMF (Individual), 1986 Census

Number, aggregate income, average income and composition of income	Wages and salaries	Self-employment income	Investment income	Retirement pensions	Other money income	Government transfer payments	Total income
1. Number of records							
a) Total	500,434	500,434	500,434	500,434	500,434	500,434	500,434
b) With income	246,084	24,201	107,438	24,186	12,820	176,584	341,769
c) Changed	597	453	670	133	53	-	1,024
d) (c/b)							
Percent changed	0.2	1.9	0.6	0.5	0.4	-	0.3
2. Aggregate income - dollars ('000)							
a) Original	4,553,521	354,778	413,424	176,282	51,318	686,341	6,235,665
b) Change	-4007	-14817	-15677	-1493	-1184	-	-73279
c) Final	4,513,414	339,961	397,747	174,789	50,134	686,341	6,162,386
d) (b/a) -							
Percent changed	-0.9	-4.4	-3.9	-0.9	-2.4	-	-1.2
3. Average - dollars							
a) Original per recipient	18,504	14,660	3,848	7,289	4,003	3,887	18,245
b) Change	-163	-612	-146	-62	-92	-	-214
c) Final	18,341	14,047	3,702	7,227	3,911	3,887	18,031
4. Composition of income - percent							
a) Original	73.02	5.69	6.63	2.83	0.82	11.01	100.00
b) Final	73.24	5.52	6.45	2.84	0.81	11.14	100.00

TABLE 5. Comparison Between Census and PUMF Income Estimates (Weighted), by Source of Income, 1985

Income source	Census income estimates ⁽¹⁾	PUMF income estimates	Difference PUMF/CENSUS
	\$'000,000		percent
Wages and salaries	227,074.3	225,670.7	-0.6
Self-employment income	17,851.3	16,998.1	-4.8
Investment income	20,496.9	19,887.4	-3.0
Retirement Pension	8,727.4	8,739.5	0.1
Other money income	2,700.3	2,506.7	-7.2
Government transfer payments	34,540.3	34,317.0	-0.6
TOTAL	311,390.5	308,119.3	-1.1

(1)1986 Cenus of Canada, Population, Total Income, Catalogue no. 93-114.

TABLE 6. Distribution of Aggregate Income of Individuals (weighted) in 1985, by Province - Census and PUMF (Individual) Estimates

Province	Dollars		Percent		Difference PUMF/Census
	Census ⁽¹⁾	PUMF	Census	PUMF	
	\$'000,000		percent		
Newfoundland	4,680.0	4,693.2	1.5	1.5	0.3
PEI	1,176.1	1,191.0	0.4	0.4	1.3
Nova Scotia	9,200.7	9,010.1	3.0	2.9	-2.1
New Brunswick	6,762.0	6,718.4	2.2	2.2	-0.6
Quebec	73,830.5	73,596.5	23.7	23.9	-0.3
Ontario	123,709.4	121,968.9	39.7	39.6	-1.4
Manitoba	12,101.8	11,954.0	3.9	3.9	-1.2
Saskatchewan	11,149.5	11,017.0	3.6	3.6	-1.2
Alberta	31,091.9	30,770.1	10.0	10.0	-1.0
British Columbia	36,798.3	36,373.0	11.8	11.8	-1.2
Yukon - N.W.T.	890.1	827.1	0.3	0.3	-7.1
CANADA	311,390.5	308,119.3	100.0	100.0	-1.1

(1)1986 Census of Canada, Population, Total Income, Catalogue no. 93-114

TABLE 7. Comparison Between Census, PUMF and Adjusted (1) Personal Income Estimates, by Source of Income, 1985

Income source	Census income estimates	PUMF income estimates	Adjusted personal income estimates	Census/Personal income	PUMF/Personal income
	\$'000,000			percent	
Wages and salaries	227,074.3	225,670.7	230,776.3	-1.6	-2.2
Self-employment income	17,851.2	16,998.1	19,212.6	-7.1	-11.5
Investment Income	20,496.9	19,887.4	36,929.0	-44.5	-46.1
Government Transfer payments ⁽²⁾	33,128.2	32,919.4	44,567.1	-25.7	-26.1
Total ⁽³⁾	298,550.7	295,475.6	331,485.0	-9.9	-10.9

(1) Adjustments to the Personal Income Estimates in the National Accounts were to compensate for differences of concepts and coverage.

(2) Total of comparable sources only; excludes child tax credits.

(3) Total of comparable sources only; excludes child tax credits, retirement pensions and other money income.

TABLE 8. Average Income of Individuals in 1985, by province, Census and PUMF (Individual)

	Census ⁽¹⁾	PUMF	Difference
	Dollars		Percent
Newfoundland	14,156	14,245	0.6
PEI	13,739	13,962	1.6
Nova Scotia	16,030	15,784	-1.5
New Brunswick	14,870	14,828	-0.3
Quebec	17,057	17,077	0.1
Ontario	19,462	19,184	-1.4
Manitoba	16,796	16,643	-0.9
Saskatchewan	16,828	16,703	-0.7
Alberta	19,661	19,470	-1.0
British Columbia	18,571	18,365	-1.1
Yukon - N.W.T.	19,838	19,101	-3.7
Canada	18,188	18,031	-0.9

(1) 1986 Census of Canada, Population, Total Income, Catalogue no. 93-114

TABLE 9. Percentage Distribution of Individuals 15 Years and Over, With Income, by 1985 Income Size Groups, Canada - Census and PUMF (Individual)

Income size group	Census ⁽¹⁾	PUMF
	percent	
Under 2,000 ⁽²⁾	8.2	8.7
2,000 - 4,999	10.3	10.2
5,000 - 6,999	7.6	7.6
7,000 - 9,999	12.5	12.5
10,000 - 14,999	13.8	13.9
15,000 - 19,999	11.3	11.3
20,000 - 24,999	9.4	9.4
25,000 - 29,999	7.5	7.5
30,000 - 34,999	6.1	6.1
35,000 - 39,999	4.1	4.1
40,000 and over	8.7	8.7
TOTAL	100.0	100.0
Average income	\$18,188	\$18,031
Median income ⁽³⁾	\$13,911	\$13,937

(1) 1986 Census of Canada, Population, Total Income, Catalogue no. 93-114

(2) Includes Loss

(3) Median Income calculated from the distribution in this table