Catalogue no. 62M0004XCB

User Guide for the Public-use Microdata File

Survey of Household Spending, 2003

May 2005

Income Statistics Division
Statistics Canada, Ottawa, K1A 0T6
Telephone: 613 951-7355

Ce document est disponible en français.

Table of contents

1. Introduction

1.1 General information

1.1.1 Background

This public-use microdata file presents data from the 2003 Survey of Household Spending (SHS) conducted in January through March 2004. Information about the spending habits, dwelling characteristics and household equipment of Canadian households during 2003 was obtained by asking people in the ten provinces and the three territories to recall their expenditures for the previous calendar year (spending habits) or as of December 31 (dwelling characteristics and household equipment).

Conducted since 1997, the Survey of Household Spending integrates most of the content found in the Family Expenditure Survey and the Household Facilities and Equipment Survey. Many data from these two surveys are comparable to the Survey of Household Spending data. However, some differences related to methodology, to data quality and to definitions must be considered before comparing these data. See Section 1.1.4 “For further information.”

1.1.2 New for 2003

The following variables have been discontinued on the 2003 public-use file:

O403 – Gifts of money and other support payments to persons living inside Canada.
O404 – Gifts of money and other support payments to persons living outside Canada.
O406 – Charitable donations to religious organizations
O407 – Charitable donations to non-religious organizations

The following variables have been added to the 2003 public-use file:

O402 – Gifts of money and support payments
O405 – Charitable donations

1.1.3 Other documents

- Data dictionary (variable specifications, code sets and other information) is available in pdf format—SHS2003-PUMDF-Data dictionary.

- Record layout is available in Excel format—SHS2003-PUMDF-Record layout.

- Appendices are available in Excel format—SHS2003-PUMDF-Appendices.

    - Appendix A presents the frequency counts for non-dollar variables in the public-use microdata file. They are included to help you verify your tabulations.

- Appendix B presents expenditure data tabulated using the public-use microdata file and also using the internal survey database. They are included to help you verify your tabulations.

- Appendix C contains a table indicating the spending variables included in previous public-use microdata files of the Survey of Household Spending and the Family Expenditure Survey.

- Appendix D shows any changes in variables from the previous year.

- Appendix E presents the coefficients of variation for published data from the 2003 SHS.

1.1.4 For further information

Additional information about the SHS can now be obtained free on the Statistics Canada web site (www.statcan.ca). See especially:

- Note to former users of data from the Family Expenditure Survey (62F0026MIE2000002)

- Note to former users of data from the Household Facilities and Equipment Survey (62F0026MIE2000003)

- User Guide for the Survey of Household Spending, 2003 (62F0026MIE2004003)

- Methodology for the Survey of Household Spending (62F0026MIE2001003)

- 2002 Survey of Household Spending Data Quality Indicators (62F0026MIE2004001)

For more information about the current survey results and related products and services, or to enquire about the concepts, methods or data quality of the Survey of Household Spending, contact Client Services (613-951-7355; 1-888-297-7355; fax 613-951-3012; *income@statcan.ca*), Income Statistics Division.

1.2 Technical characteristics of the file

Content: Household spending, dwelling characteristics, and household equipment, 2003

Source: Survey of Household Spending, 2003
Income Statistics Division
Statistics Canada

Data set definition:

Data set name SHS2003.TXT

Number of records 17,265

Format
Record length 2,064

2. Technical information

2.1 Survey methodology

(For more detailed information, see the *Methodology of the Survey of Household Spending* available free on the Statistics Canada web site at www.statcan.ca).

2.1.1 The survey universe

The 2003 Survey of Household Spending was carried out in private households in Canada's ten provinces and three territories.[1]

The following groups were excluded from the survey:

- those living on Indian reserves and crown lands (with the exception of the territories);
- official representatives of foreign countries living in Canada and their families;
- members of religious and other communal colonies;
- members of the Canadian Armed Forces living in Military Camps;
- people living in residences for senior citizens; and
- people living full time in institutions: for example, inmates of penal institutions and chronic care patients living in hospitals and nursing homes.

The survey covers about 98% of the population in the ten provinces. In the Territories, coverage was restricted to 88% in the Yukon, 92% in the Northwest Territories and 89% in Nunavut.

Information was not gathered from persons temporarily living away from their families (for example, students at university), because it would be gathered from their families if selected. In this way, double counting of such individuals was avoided.

Data from part-year households were excluded from estimates of average household spending. However, these data were included in the estimates for dwelling characteristics and household equipment and in the calculation of the Survey of Household Spending response rate. Data from part-year households are also required as input into Canada's System of National Accounts. Part-year households are composed entirely of persons who were members of other households for part of the reference year. There were 570 part-year households in the sample in 2003.

2.1.2 Survey content and reference period

Detailed information was collected about expenditures for consumer goods and services, changes in assets, mortgages and other loans, and annual income. This information was collected for the calendar year 2003 (the survey reference year). Information was also collected about dwelling characteristics (e.g., type

1. In order to reduce response burden for northern households, the SHS is conducted in the north only every second year, starting in 2001

and age of heating equipment) and household equipment (e.g., appliances, communications equipment, and vehicles). This type of information was collected as of December 31st of the reference year.

Because the Survey of Household Spending is designed principally to provide detailed information on non-food expenditures, only an overall estimate of food expenditure is recorded. Detailed information on food expenditure is provided by the Food Expenditure Survey, which is conducted every four to six years. It was last conducted in 2001. In February 2003, the results were published in *Food Expenditure in Canada, 2001*, Catalogue no. 62-554-XIE.

2.1.3 The sample

The sample size for the 2003 Survey of Household Spending was 23,869 eligible households.

This year, a supplementary sample of about 2,200 dwellings in Québec was financed by the "Ministère des finances du Québec" to better suit their analytic needs. The sample was selected from a list of dwellings according to a two-stage stratified sample design. The sample frame constructed from Statistics Canada's Address Register is comprised of dwellings from selected Dissemination Areas from the 2001 Census.

The regular SHS sample was a stratified, multi-stage sample selected from the Labour Force Survey (LFS) sampling frame. Sample selection comprised two main steps: the selection of clusters (small geographic areas) from the LFS frame and the selection of dwellings within these selected clusters. The LFS sampling frame mainly uses 1991 Census geography and 1991 population counts.[2]

2.1.4 Data collection

The 2003 Survey of Household Spending was conducted from January to March 2004. Data were collected during a personal interview using a paper questionnaire. A copy of this questionnaire is available on request.

2.1.5 Data processing and quality control

Data entry and automated editing for the 2003 Survey of Household Spending took place in the Statistics Canada regional offices. This allowed respondents to be contacted in the event that more information was required to resolve an inconsistency on their questionnaires.

After data entry, an automated physical edit system checked for data entry errors. Data had to pass a two-tier edit system consisting of "must-pass" edits that checked questionnaires for logic and consistency, and "warnings" that

2. A detailed description of the Labour Force Survey sampling frame can be found in *Methodology of the Canadian Labour Force Survey*, Statistics Canada, Catalogue no. 71-526-XPB.

indicated that a particular situation was unusual and could require correction. Either type of edit resulted in the intervention of a member of one of the specially trained edit resolution teams. Further editing of the data took place in head office where invalid responses were corrected.

Missing responses were imputed using the nearest neighbour method. Statistics Canada's Canadian Census Edit and Imputation System (CANCEIS) was used to insert values from donor records having similar characteristics, chosen specifically to fit the variable. For example, total household income was used for most variables; dwelling type, household size and province were also frequently used.

Tabulation for the 2003 Survey of Household Spending was completed using a PC/client server-based system. This system provides tools (database querying, searching, and viewing capabilities) for spotting systematic errors.

2.1.6 Weighting

The estimation of population characteristics from a sample survey is based on the premise that each sampled unit represents a certain number of units in the population. A basic survey weight was attached to each record in the sample to reflect this representation. For the regular sample, these basic weights were adjusted for non-response for selected metropolitan areas, additional geographical areas and for high-income strata. The additional geographical areas comprise the remaining metropolitan areas and urban and rural areas based on census definitions but do not necessarily correspond exactly. For definitions of these terms, refer to the *1996 Census Dictionary*, Catalogue no. 92-351-XPE.

For the supplementary sample in Quebec, the household nonresponse adjustment is done within groups of Dissemination Areas from the 2001 Census. These groups were created using the software Knowledge Seeker and consist of groupings of households that share similar propensity to respond to the survey. The groups were formed using Census metropolitan areas, an urbanization indicator and some population and dwelling characteristics from Dissemination Areas as of the 2001 Census, such as age, income and dwelling type. Once the nonresponse adjustment is completed, an additional adjustment is done to the household weights of the supplementary sample and the regular sample in Quebec. This adjustment is done to account for the fact that the supplementary sample in Quebec and part of the regular sample of this province cover the same sub-population.

To increase the reliability of the estimates, weights were adjusted to ensure that estimates based on relevant characteristics of the population would respect population totals from sources other than the survey. For the ten provinces, there are two sets of totals.

The first set of totals, for age/sex groups, household size and household type at the province level, is based on projections at mid-January 2004 using the 1996 Census of Population (adjusted for net undercoverage). Controls for 18 age/sex

groups are used. These are combined with totals for one-person households, two-person households and more than two-person households. There are also totals for the number of single-parent families and couples with never-married children. Finally, for the 14 selected metropolitan areas, only two age groups are used: number of persons under 18, and number of persons 18 and over.

Due to their smaller population, only two age/sex groups are used for the three territories (Yukon, Northwest Territories and Nunavut): number of persons under 18 and number of persons 18 and over, combined with the totals for one-person households, two-person households and more than two-person households for Yukon.

The second set of totals is derived from T4 information from Canada Customs and Revenue Agency (CCRA, formerly Revenue Canada) and is intended to ensure that the weighted distribution of income (based on wages and salaries) in the data set matches that of the Canadian population.

The switch from 1991 to 1996 Census-based population totals and the use of T4 information from CCRA were introduced starting with the 1999 SHS. Revised SHS estimates for earlier survey years are available and should be used for year-over-year comparisons.

2.2 Data quality

(For more detailed information, see the *Survey of Household Spending Data Quality Indicators,* soon to be available free on the Statistics Canada web site at www.statcan.ca.)

2.2.1 Sampling error

Sampling errors occur because inferences about the entire population are based on information obtained from only a sample of the population. The sample design, the variability of the data, and the sample size determine the size of the sampling error. In addition, for a given sample design, different methods of estimation will result in different sampling errors.

The design for the 2003 Survey of Household Spending was a stratified multi-stage sampling scheme. The sampling errors for multi-stage sampling are usually higher than for a simple random sample of the same size. However, the operational advantages outweigh this disadvantage, and the fact that the sample is also stratified improves the precision of estimates.

Data variability is the difference between members of the population with respect to spending on a specific item or the presence of a specific dwelling characteristic or piece of household equipment. In general, the greater these differences are, the larger the sampling error will be. In addition, the larger the sample size, the smaller the sampling error.

2.2.1.1 Standard error and coefficient of variation

A common measure of sampling error is the standard error (SE). Standard error is the degree of variation in the estimates as a result of selecting one particular sample rather than another of the same size and design. It has been shown that the "true" value of the characteristic of interest lies within a range of +/- 1 standard error of the estimate for 68% of all samples, and +/- 2 standard errors for 95% of all samples.

The coefficient of variation (CV) is the standard error expressed as a percentage of the estimate. It is used to indicate the degree of uncertainty associated with an estimate. For example, if the estimate of the number of households having a given dwelling characteristic is 10,000 households, and the corresponding CV is 5%, then the "true" value is between 9,500 and 10,500 households, 68% of the time and between 9,000 and 11,000 households, 95% of the time.

Standard errors for the 2003 Survey of Household Spending were estimated using the bootstrap method. This method is suitable for variance estimation of non-smooth statistics such as quintiles. For more information on standard errors and coefficients of variation, refer to the Statistics Canada publication, *Methodology of the Canadian Labour Force Survey,* Catalogue no. 71-526XPB.

Coefficients of variation for published data are presented in technical tables 1 and 2 in Appendix E.

2.2.1.2 Data suppression

For reliability reasons, estimates with CVs greater than 33% should be suppressed. Since CVs are not calculated for all estimates, data suppression for the Survey of Household Spending has been based on a relationship between the CV and the number of households reporting expenditure on an item. Analysis of past survey results indicates that CVs usually reach this level when the number of households reporting an item drops to about 30. Therefore, data have been suppressed for spending on items reported by fewer than 30 households.

However, data for suppressed items do contribute to summary level variables. For example, the expenditure for a particular category of clothing might be suppressed but this amount forms part of the total expenditure estimate for clothing.

2.2.2 Non-sampling error

Non-sampling errors occur because certain factors make it difficult to obtain accurate responses or responses that retain their accuracy throughout processing. Unlike sampling error, non-sampling error is not readily quantified. Four sources of non-sampling error can be identified: coverage error, response error, non-response error, and processing error.

2.2.2.1 Coverage error

Coverage error results from inadequate representation of the intended population. This error may occur during sample design or selection, or during data collection and processing.

2.2.2.2 Response error

Response error may be due to many factors, including faulty design of the questionnaire, interviewers' or respondents' misinterpretation of questions, or respondents' faulty reporting. In the Survey of Household Spending, the difference between receipts and disbursements is calculated as a check on respondents' recall. This important quality control tool involves the balancing of receipts (income and other money received by the household) and disbursements (total expenditure plus the variable *Money flows—assets, loans, and other debts*) for each questionnaire. If the difference is greater than 10% of the larger of receipts or disbursements, respondents are contacted again for additional information. This ensures that expenditures, at least at the aggregate level, match household income and other sources of funds.

Several features of the survey help respondents recall their expenditures as accurately as possible. First, the survey period is the calendar year because it is probably more clearly defined in people's minds than any other period of similar length. Second, expenditure on food (about 11% of the average budget in 2003) can be estimated as either weekly or monthly expenses depending on the respondent's purchasing habits. Third, expenses on smaller items purchased at regular intervals are usually estimated on the basis of amount and frequency of purchase. Purchases of large items (automobiles, for example) are recalled fairly easily, as are expenditures on rent, property taxes, and monthly payments on mortgages. However, even with these items, the accuracy of data depends on the respondent's ability to remember and willingness to consult records.

2.2.2.3 Non-response error

Non-response error occurs in sample surveys because not all potential respondents cooperate fully. The extent of non-response varies from partial non-response to total non-response.

Total non-response occurs when the interviewer is unable to contact the respondent, no member of the household is able to provide information, or the respondent refuses to participate in the survey. Total non-response is handled by adjusting the basic survey weight for responding households to compensate for non-responding households. For the 2003 Survey of Household Spending, the overall response rate was 72.3%. See Figure 1 for provincial and territorial response rates.

In most cases, partial non-response occurs when the respondent does not understand or misinterprets a question, refuses to answer a question, or is unable to recall the requested information. Imputing missing values compensates for this partial non-response.

The importance of the non-response error is unknown but in general this error is significant when a group of people with particular characteristics in common refuse to cooperate and where those characteristics are important determinants of survey results.

Figure 1
Response rates, Canada, provinces and territories, 2003

	Eligible households (1)	Non-contacts	Refusals	Unusables (2)	Usables	Response rate (3)
Newfoundland and Labrador	1,772	134	198	69	1,371	77.4%
Prince Edward Island	781	52	121	27	581	74.4%
Nova Scotia	2,048	150	347	132	1,419	69.3%
New Brunswick	1,845	98	303	118	1,326	71.9%
Quebec	4,817	360	927	68	3,462	71.9%
Ontario	3,149	282	787	103	1,977	62.8%
Manitoba	1,887	104	312	49	1,422	75.4%
Saskatchewan	1,837	72	280	77	1,408	76.6%
Alberta	2,087	128	336	18	1,605	76.9%
British Columbia	2,580	205	500	45	1,830	70.9%
Yukon	418	28	74	9	307	73.4%
Northwest Territories	410	21	42	8	339	82.7%
Nunavut	238	13	7	-	218	91.6%
Canada	**23,869**	**1,647**	**4,234**	**723**	**17,265**	**72.3%**

(1) Part-year households are included in the calculation of response rates. There were 570 part-year households in 2003.
(2) Rejected at the editing stage.
(3) Usable/eligible*100

2.2.2.4 Processing error

Processing errors may occur in any of the data processing stages, for example, during data entry, editing, weighting, and tabulation. See Data Processing and Quality Control (above) for a description of the steps taken to reduce processing error.

2.2.3 The effect of large values

For any sample, estimates can be affected by the presence or absence of extreme values from the population. These extreme values are most likely to arise from positively skewed populations. The nature of the subject matter of the SHS lends itself to such extreme values. Estimates of totals, averages and standard errors may be greatly influenced by the presence or absence of these extremes.

2.2.4 Comparability over time

Conducted since 1997, the Survey of Household Spending integrates most of the content found in the Family Expenditure Survey and the Household Facilities and Equipment Survey. Many variables from these two surveys are comparable to those in the Survey of Household Spending. However, some differences related to the methodology, to data quality and to definitions must be considered before making comparisons.

For more information, refer to Note to Former Users of Data from the Family Expenditure Survey, Catalogue no. 62F0026MIE2000002 and Note to Former Users of Data from the Household Facilities and Equipment Survey, Catalogue no. 62F0026MIE2000003. Both documents are available free of charge on the Statistics Canada web site (www.statcan.ca).

Historical data from the 1997 and 1998 surveys of household spending, the 1996 Family Expenditure Survey and the 1996 Household Facilities and Equipment Survey have been re-weighted using the weighting methodology described in the section "Weighting". Historical comparisons between data from those surveys and data from recent years of the Survey of Household Spending should generally be made with re-weighted data, although the differences between survey estimates from the old and new methodologies appear to be minimal at a summary level. Certain populations or variables, however, may be more strongly affected.

Starting with the 1997 Survey of Household Spending, "Tenants' maintenance, repair and alterations" and "Insurance premiums" were reduced by the proportion of rent charged to business. This may affect comparisons with data from previous years.

For the 2001 reference year, extra questions were included for use in the weighting of the Consumer Price Index. This change may affect some historical comparisons. For example, in 2001, questions were added under "personal care" to collect extra information about hair care products, makeup, fragrances, deodorants and oral hygiene products. As a result of these extra questions, respondents may have given more precise information and the increase in the estimated expenditures for "personal care" in 2001 may have been caused by an improvement in respondent recall. The effect of additional questions on estimates is difficult to quantify. However, in 2002, when extra questions were removed, the estimate for "personal care" spending decreased again.

The mortgage payment calculation has been revised for the years from 1997 to 2001. This has resulted in downward revisions for those years in the following expenditure categories: "Regular mortgage payments", "Owned living quarters", "Principal accommodation", "Shelter", "Total current consumption", and "Total expenditure". The effect of the revision has been an average decrease of about 0.5% in "Total expenditure" for those years.

2.3 Guidelines for tabulation, analysis and dissemination

This section describes the guidelines that users should follow when totalling, analysing, publishing or releasing data taken from the public-use microdata file.

2.3.1 Important note to users about full and part-year households

In 1997, the Survey of Family Expenditure (FAMEX) and the Household Facilities and Equipment Survey (HFE) were replaced by the Survey of Household Spending (SHS). FAMEX microdata files included full-year households[3] only, as only such households could give a clear picture of income and expenditures over an entire year. HFE microdata, on the other hand, included all households, since data were collected as of December 31. To meet user needs, all households are listed on the SHS file, along with a variable indicating each household's status (full-year, part-year).

To create statistics for average annual expenditures, users should use records for full-year households. To tabulate dwelling characteristics, household equipment or create other types of expenditure statistics such as totals (aggregates) or market share, users should use records for full-year and part-year households.

2.3.2 Guidelines for rounding

To ensure that estimates from this microdata file intended for publication or any other type of release correspond to estimates that would be obtained by Statistics Canada, we strongly recommend that users comply with the following guidelines for rounding estimates.

a) Estimates in the body of a statistical table must be rounded to the nearest hundredth using the traditional rounding technique, i.e., if the first or only number to be eliminated is between 0 and 4, the preceding number does not change. If the first or only number to be eliminated is between 5 and 9, the value of the last number to be retained increases by 1. For example, when using the traditional technique of rounding to the nearest hundredth, if the last two numbers are between 00 and 49, they are replaced by 00 and the preceding number (denoting hundredths) stays as is. If the last two numbers are between 50 and 99, they are replaced with 00 and the preceding number increased by 1.

b) Total partial sub-totals and total sub-totals in statistical tables must be calculated using their unrounded corresponding components, then rounded in turn to the closest hundredth using the traditional rounding technique.

3. A *full-year household* has at least one member present throughout the year. A *part-year household* consists entirely of members present only part of the year. A member present for part of the year is a member of a household who has been present less than 52 weeks. Income and expenditure data for members present just part of the year are collected for only that part of the year they were included in the household.

c) Means, ratios, rates and percentages must be calculated using unrounded components (i.e., numerators and/or denominators), and then rounded to a decimal using the traditional rounding technique.

d) Totals and differences in aggregates (or ratios) must be calculated using their corresponding unrounded components, then rounded to the nearest hundredth (or decimal place) using the traditional rounding technique.

e) If, due to technical or other limitations, a technique other than traditional rounding is used, with the result that the estimates to be published or released differ in any form from the corresponding estimates that would be obtained by Statistics Canada using this microdata file, we strongly advise users to indicate the reasons for the differences in the documents to be published or released.

f) Unrounded estimates cannot under any circumstances be published or released in any way whatsoever by users. Unrounded estimates give the impression that they are much more precise than they actually are.

2.3.3 Guidelines for the weighting of the sample for totalling purposes

The sample design used for the SHS is not self-weighted, meaning that the households in the sample do not all have the same sampling weight. To produce simple estimates, including standard statistical tables, users must use the appropriate sampling weight. Otherwise, the estimates calculated using the microdata files cannot be considered as representative of the observed population and will not correspond to those that would be obtained by Statistics Canada using this microdata file. See Section 2.1.6 “Weighting.”

Users should also note that depending on the method they use to process the weight field, some software packages may not produce estimates that correspond exactly to those of Statistics Canada using this microdata file.

2.3.4 Types of estimates: categorical versus quantitative

Before discussing how SHS data can be totalled and analysed, it is useful to describe the two main types of estimations that may be produced from the microdata file for the Survey of Household Spending.

2.3.4.1 Categorical estimates

Categorical estimates are estimates of the number or percentage of households in the survey’s target population that have certain characteristics or belong to a defined category. The number of households reporting a particular expenditure is an example of this type of estimate. The expression ‘aggregate estimate’ can also be used to refer to an estimate of the number of individuals with a given characteristic.

Examples of categorical questions:
Did you have a cellular phone for personal use? _yes _no

When was this dwelling originally built?
_ 1920 or earlier
_ 1921-1945
_ 1946-1960
_ 1961-1970
_ 1971-1980
_ 1981-1990
_ 1991-2000
_ 2001-2002
_ 2003

On December 31, 2003, was your dwelling:
_ Owned without a mortgage by your household?
_ Owned with (a) mortgage(s) by your household?
_ Rented by your household?
_ Occupied rent-free by your household?

Totalling of categorical estimates
Estimates of the number of persons with a given characteristic can be obtained from the microdata file by adding the final weights of all records containing the desired characteristic or characteristics. Percentages and ratios in the X/Y form are obtained as follows:

a) by adding the final weights of records containing the desired characteristic for the numerator X;
b) by adding the final weights of records containing the desired characteristic for the denominator Y;
c) by dividing the estimate for the numerator by the estimate for the denominator.

2.3.4.2 Quantitative estimates

Quantitative estimates are estimates of totals or means, medians or other central tendency measurements of quantities based on all members of the observed population or based on some of them. They also explicitly include estimates in the form X/Y where X is an estimate of the total quantity for the observed population and Y is an estimate of the number of individuals in the observed population who contribute to that total quantity.

An example of a quantitative estimate is mean annual expenditure for personal and health care per household in the target population. The numerator corresponds to an estimate of total annual expenditure for personal and health care, and the denominator corresponds to an estimate of the number of households in the population.

Example of quantitative question:
In 2003, how much did your household spend for telephone service? _____

Totalling of quantitative estimates
Quantitative estimates can be obtained from the microdata file by multiplying the value of the desired variable by the final weight of each record, and then adding this quantity for all records of interest. For example, to obtain an estimate of total expenditure by households that were owners on December 31 for electricity, the value reported for the question "In 2003, how much did your household spend on electricity?" is multiplied by the final weight of the record, and then that result is summed over all records with a positive response to the question "On December 31, 2003, was your house: 'Owned mortgage-free by your household' or 'Owned with one or more mortgages by your household'."

To obtain a weighted mean expressed by the formula X/Y, the numerator X is calculated as a quantitative estimate and the denominator Y as a categorical estimate. For example, to estimate mean household expenditures for electricity by owners, you must:

a) estimate the total expenditure for electricity for households where the residence is owned, using the method described above;
b) estimate the number of owned households by adding the final weights for all records with a positive response to the question "As at December 31, 2003, was your house: 'Owned mortgage-free by your household' or 'Owned with one or more mortgages by your household"; and then,
c) divide the estimate obtained in a) by the one calculated in b).

Note: Because average expenditures are being estimated, "part-year" households must first be excluded from calculations. (For further details, see Section 2.3.1 "Important note to users about full and part-year households.")

2.3.5 Guidelines for statistical analysis

The Survey of Household Spending is based on a complex survey design that includes stratification and multiple stages of selection, as well as uneven respondent selection probabilities. The use of data from such complex surveys poses problems for analysts, because the survey design and the selection probabilities influence the estimation and variance calculation methods to be used.

Although numerous analytical methods in statistical software packages allow for the use of weights, the meaning or definition of weights differs from that suitable for a sample survey. As a result, although the estimates done using those packages are in many cases accurate, **the variances calculated have almost no significance.**

For numerous analytical techniques (for example, linear regression, logistic regression, variance analysis), there is a way to make the application of standard packages more significant. If the weights of the records contained in the file are converted so that the mean weight is (1), the results produced by standard packages will be more reasonable and will take into account uneven selection

probabilities, although they still cannot take into account the stratification and the cluster distribution of the sample. The conversion can be done using in the analysis a weight equal to the original weight divided by the mean of original weights for sampling units (households) that contribute to the estimator in question. However, because this method still does not take into account sample design stratification and clusters, the estimates of the variance calculated in this way will very likely be underestimates of true values.

2.3.6 Guidelines for release

Before releasing and/or publishing estimates taken from the microdata file, users must first determine the level of reliability of the estimates. The quality of the data is affected by the sampling error and the non-sampling error as described above. However, the level of reliability of estimates is determined solely on the basis of sampling error, as evaluated using the coefficient of variation (CV) as shown in the table below. In addition to calculating CVs, users should also read the section of this document regarding the characteristics of data quality.

Whatever CV is obtained for an estimate from this microdata file, users should determine the number of sampled respondents who contribute to the calculation of the estimate. If this number is less than 30, the weighted estimate should not be released regardless of the value of the CV for this estimate. For weighted estimates based on sample sizes of 30 or more, users should determine the CV of the rounded estimate following the guidelines below.

Figure 2
Sampling variability guidelines

Type of Estimate	CV (in %)	Guidelines
1. Acceptable	0.0 – 16.5	Estimates can be considered for general unrestricted release. Requires no special notation.
2. Marginal	16.6 – 33.3	Estimates can be considered for general unrestricted release but should be accompanied by a warning cautioning subsequent users of the high sampling variability associated with the estimates. Such estimates should be identified by the letter M (or in some other similar fashion).
3. Unacceptable	Greater than 33.3	Statistics Canada does not recommend the release of estimates of unacceptable quality. However, if the user chooses to do so then estimates should be flagged with the letter U (or in some other similar fashion) and the following warning should accompany the estimates: "The user is advised that . . . (specify the data) . . . do not meet Statistics Canada's quality standards for this statistical program. Conclusions based on these data will be unreliable and most likely invalid."

2.3.6.1 Computation of approximate CVs

In order to provide a way of assessing the quality of estimates, Statistics Canada has produced a coefficient of variation table (CV table) which is applicable to estimates of averages, ratios and totals obtained from this public use microdata file for the major variables of the SHS by province and at the Canada level (see Appendix E). The CV of an estimate is defined to be the square root of the variance of the estimate divided by the estimate itself and expressed as a percentage. The numerator of the CV is a measure of the sampling error of the estimate, called the standard error, and is calculated at Statistics Canada with the bootstrap method. This method requires, among other things, information about the strata and the clusters, which can't be given on the public use microdata file for reasons of confidentiality. So that users may estimate CVs for variables not included in the CV tables, Statistics Canada has produced a set of rules to obtain approximate CVs for a wide variety of estimates. It should be noted that these rules provide approximate and, therefore, unofficial CVs. The quality of the approximation, however, is quite satisfactory, especially for the most reliable estimates. Note that accuracy of this approximation is reduced when the domains become smaller. Therefore, the CV approximation method must be used prudently when the domains are small. The document on data

quality for the 1997 SHS contains the results of the evaluation of the performance of the CV approximation method.

How to obtain approximate CVs

The following rules should enable the user to determine the approximate coefficients of variation for estimates of totals, means or proportions, ratios and differences between such estimates for sub-populations (domains) for which the Jackknife CV is not provided in the CV tables.

Important: If the number of observations on which an estimate is based is less than 30, the weighted estimate should not be released regardless of the value of the CV for this estimate.

Rule 1: Approximating CVs for estimates of totals (aggregates)

All the steps below must be followed to obtain an approximate CV (ACV) for an estimate of a total (either a number of households possessing a certain characteristic (categorical estimate) or a total of some expense for all households (quantitative estimate)) for a sub-population (domain) of interest:

1) Create a binary variable for each household, say I, equalling 1 if the household is part of the domain of interest, i.e. possesses the desired characteristic and 0 otherwise;
2) To estimate a quantitative variable, create a variable Y representing the product of the binary variable I and the variable of interest. To estimate a categorical variable, create a variable Z equal to 1 if the categorical variable is equal to the value of interest, and equal to 0 otherwise. Define variable Y as the product of I and Z;
3) Do step (4) to step (9) for each province separately;
4) Calculate the sum over all the households of the product of the final weight (section Weighting), and Y (this sum represents the estimate of the total for the domain of interest in the province under consideration);
5) Calculate the sum over all the households of the product of the final weight and the household size;
6) Divide the result obtained in step (4) by the result obtained in step (5);
7) For each household, multiply the result obtained in step (6) by the household size;
8) For each household, define a variable, say E, by the subtraction of the result obtained in step (7) from Y;
9) Calculate the sum over all the households of the product of the final weight minus 1, the final weight and E squared; (this sum represents the estimated variance of the total estimated at step 4);
10) Add up the result obtained in step (9) for each province;
11) The ACV is defined to be 100 times the square root of the result obtained in step (10), divided by the estimate. The estimate is the sum over all the provinces of the result obtained in step (4).

More formally, steps 1 to 10 above can be obtained with the following formula:

$$\sum_{p=1}^{12} \sum_{k \in S_p} (w_k - 1) w_k \left(Y_k - m_k \sum_{k \in S_p} w_k Y_k \Big/ \sum_{k \in S_p} w_k m_k \right)^2$$

where the index *p* corresponds to provinces, S_p is the sample of respondents for the province *p*, the index *k* corresponds to households, w_k is the final weight for the k^{th} household, m_k is the household size for the k^{th} household and Y_k is the value of the variable Y, defined in step (2) above, for the k^{th} household. As you can see, index *p*, the province indicator, takes values ranging from 1 to 12. Twelve distinct province codes appear on the microdata file: one for each of the ten provinces, one for the territories and a "00" province code assigned to a set of records for reasons of confidentiality. (See Section 2.4 "Confidentiality of the public-use microdata.")

Note: Two household size variables appear in the microdata file. To calculate approximate CVs, the variable used to define household size is "Household size at December 31," rather than "Household size (number of persons a member sometime in reference year)."

Important: When estimating variance for a given domain, do not limit yourself to units belonging to the domain. The entire sample should always be used to estimate variance. Units that do not belong to the domain of interest are not considered when computing the point estimate of the total, but do contribute when estimating the variance.

Rule 2: Approximating CV for estimates of averages or proportions

An estimated mean or proportion is obtained by the ratio of two estimated totals. For a proportion, the numerator is an estimate that is a sub-set of the denominator, for example the proportion of expenditures for households in Manitoba compared to all Canadian households. The CV of an estimated mean or proportion tends generally to be slightly lower than the corresponding CV of the numerator. The CV of an estimated mean or proportion can thus be approximated with the CV of the numerator and the technique described in rule (1) can be used.

Rule 3: Approximating CV for estimates of ratios

Ratio refers to the relationship between any two estimates of totals for which rule (2) does not apply. Approximate CVs for any other types of ratio, may be calculated using the following formula:

$$\mathrm{ACV_R} = \sqrt{\mathrm{ACV_N^2} + \mathrm{ACV_D^2}}$$

where $\mathrm{ACV_R}$ is the approximate CV of the ratio, $\mathrm{ACV_N}$ is the approximate CV of the numerator of the ratio and ACVD is the approximate CV of the denominator of the ratio. The formula will tend to overestimate the CV if the two estimates forming the ratio are positively correlated and underestimate the CV if these two estimates are negatively correlated.

Rule 4: Approximating CVs for estimates of differences

The approximate CV of a difference between any two estimates ($EST_{DIFF} = EST_1 - EST_2$) is given by:

$$ACV_{DIFF} = \frac{\sqrt{(EST_1 ACV_1)^2 + (EST_2 ACV_2)^2}}{| EST_{DIFF} |}$$

where ACV_1 is the approximate CV associated with EST_1 and ACV_2 is the approximate CV associated with EST_2. The formula will tend to overestimate the CV if the two estimates forming the difference are positively correlated and underestimate the CV if these two estimates are negatively correlated.

Examples

Detailed calculations of approximate CVs used for estimating totals are initially presented using fictional cases. Then actual cases of estimating totals, averages (or proportions) ratios and differences, based on microdata file data, will be presented so users can check results and ensure that the method used was valid.

Part 1: Fictional case: details of calculating an approximated CV for estimating a total

A) Quantitative variable

Let us assume we wanted to estimate the total for a (quantitative) expenditure variable X, for households containing at least one person less than 18 years of age. To illustrate this procedure, we will use a fictional sample (see Figure 3) on which we will present calculation details (see Figure 4) for each of the eleven steps described above. As this procedure is applied independently within each province, we shall merely describe calculations for one province.

Let us use the following sample for Ontario:

Figure 3
Fictional example

Initial Data						
Identifier	Province	Entire year	Weight	Household size	Number of children aged 0-17	Variable of Interest X
00001	Ontario	Yes	5	3	2	30
00002	Ontario	Yes	20	5	3	0
00003	Ontario	Yes	25	2	1	20
00004	Ontario	No	5	4	2	50
00005	Ontario	Yes	15	3	0	20
00006	Ontario	Yes	10	1	0	10
00007	Ontario	Yes	15	4	0	15

In step 1, we define the domain of interest by creating a binary variable equal to 1 for all units belonging to the domain. In the present case, these are households with at least one child between the ages of 0 and 17 years. We then proceed to

steps 2 through 9 to estimate variance, which will lead to calculation of the CV. We thus obtain the following results:

Figure 4
Calculation details for approximating the CV of a total (steps 1 to 9)

	Step 1	Step 2	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9
Ident.	Binary variable I	Quantitative variable Y	Weigted Y	Variable K		Step 6 * size	(Y - step 7)	(Weight -1) * Weight * (Step 8)^2
		(X * I)	(Weight * Y)	(Weight * size)				
00001	1	30 * 1 = 30	5 * 30 = 150	5 * 3 = 15		3 * 3 = 9	30 - 9 = 21	(4 * 5 * 21 * 21) = 8,820
00002	1	0 * 1 = 0	20 * 0 = 0	20 * 5 = 100		3 * 5 = 15	0 - 15 = -15	(19 * 20 * (-15) * (-15)) = 85,500
00003	1	20 * 1 = 20	25 * 20 = 500	25 * 2 = 50		3 * 2 = 6	20 - 6 = 14	(24 * 25 * 14 * 14) = 117,600
00004	1	50 * 1 = 50	5 * 50 = 250	5 * 4 = 20		3 * 4 = 12	50 - 12 = 38	(4 * 5 * 38 * 38) = 28,880
00005	0	20 * 0 = 0	15 * 0 = 0	15 * 3 = 45		3 * 3 = 9	0 - 9 = -9	(14 * 15 * (-9) * (-9)) = 17,010
00006	0	10 * 0 = 0	10 * 0 = 0	10 * 1 = 10		3 * 1 = 3	0 - 3 = -3	(9 * 10 * (-3) * (-3)) = 810
00007	0	15 * 0 = 0	15 * 0 = 0	15 * 4 = 60		3 * 4 = 12	0 - 12 = -12	(14 * 15 * (-12) * (-12)) = 30,240
			Total: 900	**Total: 300**	900 / 300 = 3			**Total = 288,860**

If we wanted to know the CV for Ontario, we would perform the following calculation:

$$CV_{ONT} = 100 * \frac{\sqrt{Variance_{ONT}}}{Estimation_{ONT}} = 100 * \frac{\sqrt{Step\ 9_{ONT}}}{Step\ 4_{ONT}} = 100 * \frac{\sqrt{288860}}{900} = 59.7$$

If we wanted to know the CV for Canada, we would proceed in similar manner, by totalling the results for each province. In other words,

$$CV_{CAN} = 100 * \frac{\sqrt{Variance_{CAN}}}{Estimation_{CAN}}$$

$$= 100 * \frac{\sqrt{Variance_{NF} + ... + Variance_{BC} + Variance_{TERR} + Variance_{PROV\,00}}}{Estimation_{NF} + ... + Estimation_{BC} + Estimation_{TERR} + Estimation_{PROV\,00}}$$

Comment: In this example, we wanted to estimate the total for expenditure variable X. If, for example, we had wanted to use this total to calculate average expenditure X per household, the approach would have been slightly different. As unit 00004 was a "part-year" household, it would not have been considered part of the domain of interest. Binary variable I would thus have taken the value of 0 and the final result would have been different. (For further details, see Section 2.3.1 "Important note to users about full and part-year households.")

B) Qualitative variable (categorical)

In the event a categorical variable is estimated, the steps in calculating the approximate CV will be the same as in the quantitative variable example presented. Instead of a quantitative value for variable of interest X, we would create a dichotomous variable that would be equal to 1 if the household has the features we want to estimate. If not, it would be equal to 0.

To estimate categorical variables, various approaches may be used for defining the domain and the variable of interest, both of which will produce the same results.

Let us assume we want to estimate the number of households consisting of more than one person living in a single-family dwelling. We could proceed in different ways:

1) Binary variable I is equal to 1 for all households and variable X is equal to 1 for households consisting of more than one person living in a single-family dwelling.
2) Binary variable I is equal to 1 for all households consisting of at least one person and variable X is equal to 1 for all households the members of which live in a single-family dwelling.
3) Binary variable I is equal to 1 for all households the members of which live in a single-family dwelling and variable X is equal to 1 for all households made up of more than one person.
4) Binary variable I is equal to 1 for all households made up of more than one person living in a single-family dwelling and X is equal to 1 for all households.

Whatever approach is used, the resulting Y variable (step 2) will be equal to 1 if the household possesses all the necessary features (more than one person and living in a single-family dwelling). If not, it will be equal to 0. Results in terms of point estimates and estimates of variance (CV) will thus be the same.

Part 2: Actual cases based on the microdata file

Example 1a: Approximation of CV for estimates of totals (quantitative variable)

Let us assume that we have estimated that household furnishings and equipment expenditures for one-person households in Manitoba total $85,777,298. We have to estimate the approximate CV for this estimate. Users must therefore follow steps (1) to (11) of rule 1.

1) Create a binary variable I whose value is 1 if the household is a one-person household and resides in Manitoba, otherwise I equals 0.
2) Y is defined for each household as the product of the binary variable I and the ‘total household furnishing and equipment expenditures’ variable.

Note that the estimate of spending on household furnishings and equipment is obtained by adding the product of variable Y defined in 2) and the final weight of the household.

Figure 5 shows the results of some of the steps in the approximate CV calculation.

Figure 5
Calculation of ACV

Step	Total spending on household furnishings and equipment for one-person households in Manitoba
4	85,777,298
5	1,066,289
6	80.44
9	8.4180×10^{13}
10	8.4180×10^{13}
11	10.70

Example 1b: Approximation of CV for estimates of totals (qualitative variable)
Let us assume we now want to estimate the total number of Canadian one-person households, as well as the total number of Canadian households made up of one person living in different types of accommodations.

In this case, variable I is defined as having the value 1 if the household is one-person. If not, it is 0. We must create five Z variables: Z1 with a value of 1 if the type of residence occupied is a "single-family dwelling," and 0 if not; Z2 equals 1 if the type of residence is semi-detached, and 0 if it is not. Z3 equals 1 if the type of residence is a townhouse, and 0 if it is not. Z4 equals 1 if the type of residence is a row house, and 0 if it is not. Finally, Z5 equals 1 if the type of house is "other," and 0 if it is not. Y1 is defined as the product of I and Z1, Y2 as the product of I and Z2, etc.

The estimates obtained are 3,164,930 for the set of one-person households, 1,028,603 for single-family dwellings[4], 114,802 for semi-detached houses[5], 169,425 for town houses[6] and 1,852,100 for "other[7]" We want to calculate the approximate CVs for these estimates.

Figure 6 shows the results for some steps in the calculation of the approximate CV. The results presented for steps 4 to 9 are the results for Manitoba (presented as an example, for a province, they will be used for comparison in the next example), while those presented for steps 10 and 11 are Canada-wide.

4. Single family = single detached
5. Semi-detached = double
6. Town houses = row or terrace
7. Other = duplex, apartment, hotel, mobile home, other

Figure 6
Calculation of ACV

Step	Number of one-person households	Number of one-person households living in a single-family dwelling	Number of one-person households living in a semi-detached dwelling	Number of one-person households living in a townhouse	Number of one-person households living in other housing
4	117,285	58,050	747	3,194	55,294
5	1,066,289	1,066,289	1,066,289	1,066,289	1,066,289
6	0.11	0.05	0.00	0.00	0.05
9	48,310,931	22,087,229	342,917	938,852	20,415,020
10	5,675,505,959	1,694,346,875	208,459,405	280,243,854	3,070,767,506
11	2.38	4.00	12.58	9.88	2.99

Example 1c: Approximation of CV for estimates of totals used in the calculation of average expenditure

Let us assume we want to estimate average expenditure on furnishings and household equipment for one-person households in Manitoba. To do so, we would have to estimate the number of one-person households in Manitoba, as well as the total of their expenditure on furnishings and household equipment.

Because we are interested here in calculating average expenditures, “part-year” households are outside the domain of interest. (For further details, see Section 2.3.1 “Important note to users about full and part-year households.”) This is why estimates of totals are slightly different than those obtained in the prior two examples.

Figure 7
Calculation of ACV

Step	Number of one-person households in Manitoba	Total expenditure on furnishings and household equipment for households consisting of one person in Manitoba
4	110,577	79,951,359
5	1,066,289	1,066,289
6	0.10	74.98
9	44,588,794	7.6904×10^{13}
10	44,588,794	7.6904×10^{13}
11	6.04	10.97

The estimate of the mean would be \$79,951,359/110,577 = \$723. How do we determine the CV of this estimate?

Rule (2) should be applied in this case. Thus, the CV of this mean may be approximated with the CV of the numerator, the total expenditure on furnishings

and household equipment in Manitoba for one-person households. This CV is 10.97%.

Example 2: Approximation of CV for estimating ratios
Let us assume we want to estimate the ratio between the average expenditures on furnishings and household equipment for one-person households in urban Manitoba and rural Manitoba.

Figure 8
Calculation of ACV

Step	Total expenditure on furnishings and household equipment for households consisting of one person in Manitoba (urban)	Total expenditure on furnishings and household equipment for households consisting of one person in Manitoba (rural)
4	71,607,232	14,170,066
5	1,066,289	1,066,289
6	67.16	13.29
9	7.4117×10^{13}	8.7789×10^{12}
10	7.4117×10^{13}	8.7789×10^{12}
11	12.02	20.91

The estimate of the ratio would be equal to $71,607,232/$14,170,066 = 5.1 (one-person households in urban Manitoba spend 5 times more on furnishing than those in rural Manitoba). How does the user determine the CV of this estimate?

We have already calculated CVs for each of the two estimates involved in estimating the ratio. We would thus apply rule (3) to obtain the desired CV:

$$\mathrm{CVA}_{\mathrm{R}} = \sqrt{\mathrm{CVA}_{\mathrm{N}}^2 + \mathrm{CVA}_{\mathrm{D}}^2} \quad = \sqrt{12.02^2 + 20.91^2} = 24.12$$

This CV should be identified as “Marginal” (see Section 2.3.6 “Guidelines for release”) as it is quite high, being between 16.6% and 33.3%.

Example 3: Approximation of CV for estimating differences
Let us assume we wanted to estimate the difference between total expenditures on furnishings and household equipment in Alberta and in Manitoba, as well as the CV for this difference.

We would estimate total expenditures on furnishings and household equipment, along with their respective CVs for Manitoba (total = 639,539,796; CV = 4.04) and for Alberta (total = 2,130,282,394; CV = 3.84).

Estimation of the difference would thus be 2,130,282,394 – 639,539,796 = 1,490,742,598. Rule (4) can be applied to obtain the desired CV.

$$CVA_{DIFF} = \frac{\sqrt{(EST_1 CVA_1)^2 + (EST_2 CVA_2)^2}}{|EST_{DIFF}|}$$

$$= \frac{\sqrt{(2{,}130{,}282{,}394 * 3.84)^2 + (639{,}539{,}796 * 4.04)^2}}{|1{,}490{,}742{,}598|} = 5.75$$

2.3.6.2 How to obtain confidence limits

Although coefficients of variation are widely used, a more intuitively meaningful measure of sampling error is the confidence interval of an estimate. A confidence interval constitutes a statement on the level of confidence that the true value for the population lies within a specified range of values. For example a 95% confidence interval can be described as follows.

If sampling of a population is repeated many times, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.

Using the CV of an estimate, its confidence intervals may be obtained assuming that, under repeated sampling of the population, the various estimates obtained for a characteristic are normally distributed around the true population value. Using this assumption, the chances are about 68 out of 100 that the difference between a sample estimate and the true population value would be less than one standard error, about 95 out of 100 that the difference would be less than two standard errors, and about 99 out 100 that the differences would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

Confidence intervals for an estimate, EST, are generally expressed as two numbers, one below the estimate and one above the estimate, as (EST - k, EST + k) where k is determined depending on the level of confidence desired and the sampling error of the estimate.

Confidence intervals for an estimate can be calculated by first determining the ACV of the estimate and then using the following formula to convert to a confidence interval CI:

$$(EST - z \times EST \times ACV / 100, EST + z \times EST \times ACV / 100)$$

where

z = 1 if a 68% confidence interval is desired,
z = 1.6 if a 90% confidence interval is desired,
z = 2 if a 95% confidence interval is desired,
z = 3 if a 99% confidence interval is desired.

Note: Release guidelines, which apply to the estimate, also apply to the confidence interval. For example, if the estimate is not releasable, then the confidence interval is not releasable either.

Example 4
A 95% confidence interval for the estimated mean of spending on household furnishings and equipment for one-person households in Manitoba would be calculated as follows:

EST = 723
z = 2
ACV = 10.97
CI = (723 – 2 x 723 x 10.97/100; 723 + 2 x 723 x 10.97/100) = (564.4; 881.6)

2.3.6.3 How to do a Z-test

Coefficients of variation may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The sample estimates can be totals, averages, ratios, etc. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

Let EST_1 and EST_2 be sample estimates for 2 characteristics of interest. Let the approximate CV of the difference $EST_1 – EST_2$ be ACV_{DIFF}.

If $z = 100 / ACV_{DIFF}$ is less than 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is larger than 2, the observed difference is significant at the 5% level.

Example 5
Let us suppose we wish to test, at the 5% level of significance, the hypothesis that there is no difference between the total of spending on furnishings and equipment in Alberta and the same total in Manitoba. From example 3, the approximate CV of the difference between these two estimates was found to be 5.75 and z = 17.4. Since this value is greater than 2, it must be concluded that there is significant difference between the two estimates at the 0.05 level of significance.

2.4 Confidentiality of the public-use microdata

Microdata files for public use differ in many ways from the master file of the survey held by Statistics Canada. These variations are due to measures taken to preserve the anonymity of respondents to the survey.

The confidentiality of this file is ensured mainly by reducing information, i.e., deleting variables or suppressing or collapsing some of their detail.

To protect confidentiality

- All explicitly identifying information, such as identification numbers, was removed from the file. (Names and addresses are not data captured).

- 112 records had their *province codes* set to 0 due to special characteristics (e.g., exceedingly high or low expenditure values). These records were reweighted.

- Other records were also reweighted for confidentiality reasons.

- The three territories are grouped under the same code.

- There was *top-coding* and *collapsing* of code sets for non-spending variables.

- Income values at the household, reference person and spouse of reference person levels were *rounded* in the following manner:

    - For income values between $1 and $9,999: round to the nearest $100
    - For income values between $10,000 and $99,999: round to the nearest $1,000
    - For income values between $100,000 and $999,999: round to the nearest $10,000
    - For income values between $1,000,000 and $9,999,999: round to the nearest $100,000
    - For income values between $10,000,000 and $99,999,999: round to the nearest $1,000,000 (there are no such values on the 2003 file).

The variables “Purchase price of dwelling” and “Selling price of dwelling” were also rounded.