



Catalogue no. 62F0026MIE — No. 006
ISSN: 1708-8879
ISBN: 0-662-41722-4

Research Paper

Household expenditures research paper series

Survey of Household Spending 2003: Data Quality Indicators

by Yi Li, Charles Mitchell, Sylvain Nadon and Paddison Wong

Household Survey Methods Division
R.H. Coats Building, Ottawa, K1A 0T6

Telephone: 613 951-7355



Statistics
Canada

Statistique
Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Income Statistics Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: (613) 951-7355); (888) 297-7355 income@statcan.ca.

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 62F0026MIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.



Statistics Canada
Income Statistics Division

Household expenditure research paper series

Survey of Household Spending 2003: Data Quality Indicators

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2005

All rights reserved. The content of this publication may be reproduced, in whole or in part, and by any means, without further permission from Statistics Canada, subject to the following conditions: that it is done solely for the purposes of private study, research, criticism, review, newspaper summary, and/or for non-commercial purposes; and that Statistics Canada be fully acknowledged as follows: Source (or "Adapted from", if appropriate): Statistics Canada, name of product, catalogue, volume and issue numbers, reference period and page(s). Otherwise, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopy, for any purposes, without the prior written permission of Licensing Services, Marketing Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

October 2005

Catalogue no. 62F0026MIE2005006

ISSN: 1708-8879

ISBN: 0-662-41722-4

Frequency: Occasional

Ottawa

La version française de cette publication est disponible sur demande (n° 62F0026MIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

This report was prepared by the members of the Survey of Household Spending unit of the Household Survey Methods Division (HSMD):

Yi Li
Charles Mitchell
Sylvain Nadon
Paddison Wong

The authors wish to thank the team responsible for the Survey of Household Spending in the Income Statistics Division (ISD) for their co-operation in the preparation of this report.

Table of contents

Highlights	6
Introduction	9
1. Sampling errors	10
1.1 Measures of sampling error	10
1.2 Coefficients of variation.....	10
1.3 Model for deriving an approximation of the CV	13
1.4 Suppression of unreliable data in estimation tables.....	14
2. Nonresponse	15
2.1 Response, nonresponse and vacancy rates.....	15
2.2 Nonresponse according to urbanization level	18
2.3 Nonresponse according to income strata	19
2.4 Adjustment for nonresponse	20
3. Coverage errors	20
3.1 Undercoverage and overcoverage: slippage rates	21
3.2 Adjustment at the population and household levels.....	25
4. Response errors	26
5. Processing errors	28
5.1 Proportion of households or persons requiring imputation at national, provincial and territorial levels.....	29
References	35
Appendix A Algebraic notation	36

Highlights

Sampling errors

- The coefficients of variation (CVs) of the average estimates of total expenditure per household vary between 1.2% and 2.3% for the provinces. The CV at the national level is 0.7%. The CVs are higher in the territories, namely 3.6% in the Yukon, 3.9% in the Northwest Territories and 5.9% in Nunavut.
- The coefficients of variation (CVs) of the average estimates for the different summary level expenditure categories are in most cases less than or equal to 1.7% at the national level, and generally lower than 5.0% at the provincial level. The results for the characteristics of dwelling and household equipment are similar. The sample size being smaller in the territories, the CVs tend to be higher than those of the provinces.

Nonresponse

- The final response rate is 72.3%. Provincial and territorial response rates range from 62.8% for Ontario to 91.6% in Nunavut.
- The nonresponse rate is 27.7%. It is due to refusals (17.7%), to households that could not be contacted (6.9%), and finally to households with data that were considered unusable (3.0%).
- The final nonresponse rate generally increases with the urbanization level. The nonresponse rate is 22.3% at the rural level, and 33.8% in the urban centres of one million residents or more. This tendency is also observed in the collection nonresponse rates.
- Analysis of final response rates in the strata consisting of high and low-income geographic areas created under the sample design indicates that the nonresponse rate in high-income strata (42.2%) is higher than the rates observed in regular strata (27.3%) and in low-income strata (27.3%). The refusal rate for high-income strata (30.5%) is more than twice the rate of low-income strata (13.3%) and nearly twice the rate of regular strata (17.4%).

Coverage errors

- There is undercoverage of 8.5% for households at the national level. Undercoverage of households is also observed for all provinces and territories, with rates varying from 4.9% to 14.9%. Yukon and British Columbia have the highest undercoverage of households.
- There is undercoverage of 11.3% for persons at the national level. Undercoverage of persons is also observed for all provinces and territories, with rates varying from 6.6% to 22.7%. Yukon and the Northwest Territories have the highest undercoverage of persons.

- At the national level, there is always a net undercoverage for each age group. Except for the category 65 to 69 years old, the undercoverage rates for children (0 to 6 and 7 to 17) are smaller than those of the adults. The undercoverage rate for all children combined is 5.5%, while it is 13.0% for adults.

Response errors

- Response errors include, among others, recall errors, telescopic error and errors due to proxy response. Because the Survey of Household Spending (SHS) interview is lengthy, the response burden can lead to respondent fatigue and have an impact on the data quality. Total interview time varies according to the household characteristics. For some households the interview can take more than five hours. The average length of interview was one hour and forty-six minutes.

Processing errors related to imputation

i) Expenditure variables

- For 12.8% of households, it was necessary to impute at least one expenditure variable (excluding the Clothing section and Personal Taxes, Security and Money Gifts section). For the majority of them, only one or two out of the 229 expenditure variables were imputed. Provincially and territorially, the imputation rates range from a low of 6.6% for Quebec to a high of 18.2% in Newfoundland and Labrador. The overall imputation rate is 8.3% when excluding regular mortgage payments and mortgage insurance premiums, which is similar to rates obtained in previous years. There are 15.6% of households reporting mortgages on owned and occupied dwellings that need an imputation for the mortgage insurance premiums.
- About 19% of the individuals required imputation for clothing variables. For the majority of these, the respondents provided the totals and only the components were imputed.
- About 3% of the individuals aged 15 years old and over required imputation on at least one variable in the Personal Taxes, Security and Money Gifts section.

ii) Income variables

- Less than 5% of individuals aged 15 years and over required imputation for at least one income variable. For 70% of them, total income was provided by the respondent and imputation was performed to obtain the breakdown by components.

iii) Categorical variables

- For 6.4% of households, it was necessary to impute at least one categorical variable. Nearly 75% of them had only one variable imputed. At the provincial or territorial level, the imputation rates range from a low of 2.8% in Prince

Edward Island to a high of 10.7% in Saskatchewan. Categorical variables that required imputation can be found in the following sections of the questionnaire: Dwelling Characteristics; Facilities Associated with the Dwelling; Tenure; Tobacco and Miscellaneous for variables pertaining to purchases through direct sales.

Introduction

The Survey of Household Spending (SHS) is an annual survey that collects data on household income and expenditure using personal interviews. The 2003 SHS sample is made up of 23,869 households¹ distributed throughout the 10 provinces and the 3 territories. Collection takes place in January, February and March, and income and spending figures are obtained for the period from January 1 to December 31 of the previous year. Following a redesign that took place in 1997, this survey replaces the periodic Family Expenditure Survey and the Household Facilities and Equipment Survey (with modifications to questionnaires and samples).

Like all surveys, the SHS is subject to errors, despite all the precautions taken at the different stages of the survey to control them. While there is no comprehensive measure of the quality of the data generated by a survey, some quality measures produced at the different stages of the survey can provide users with the information needed in order to interpret the data properly.

This report therefore seeks to describe the quality indicators produced for the 2003 Survey of Household Spending. It covers the usual quality indicators that generally help users interpret data, such as coefficients of variation, response and nonresponse rates, slippage rates and imputation rates.

Quality indicators have been classified according to the main types of error encountered in a survey. Section 1 deals with sampling errors—that is, errors due to the fact that the inferences about the population drawn from the survey are based on information collected from a sample of the population, rather than the entire population. The subsequent sections cover errors not due to sampling. Nonresponse and coverage errors are first discussed in sections 2 and 3. Response errors and processing errors are dealt with in sections 4 and 5 respectively.

This report focuses on data quality. For a detailed description of the methodology of the survey, see reference [1].

1. The initial sample is made up of 27,671 dwellings. From these dwellings, we have to identify and exclude the ineligible dwellings (see section 2.1) to obtain the 23,869 households from which we collect data on household income and expenditure.

1. Sampling errors

Sampling errors exist when inferences about the population are drawn from the survey using information collected from a sample, rather than the entire population. In addition to the sample design and the estimation method used in the Survey of Household Spending, the sample size and the variability of each characteristic are factors that determine sampling error. Characteristics that are rare or are distributed very unevenly in the population will have greater sampling error than characteristics that are observed more frequently or are more homogeneous in the population.

1.1 Measures of sampling error

Standard error is a commonly used measure of sampling error. Standard error is the degree of variation of the estimate considering that a particular sample was selected, rather than another, among all possible samples of the same size under the same sample design. Since the SHS uses a complex sample design and estimation method, the standard error is estimated using a resampling method known as the bootstrap technique. Prior to 2003 reference year, the jackknife resampling method was used to produce standard error estimation for SHS. Starting with the 2003 SHS, a decision was made to use the bootstrap resampling method mainly because the Income Statistics Division was going to publish the median expenditures estimates and needed the coefficient of variation of those estimates. The bootstrap resampling method is suitable for variance estimation of non-smooth statistics such as quantiles. For more details on this method, see reference [2].

Coefficient of variation (CV) is also a frequently used measure of the reliability of an estimate. It simply expresses the standard error as a percentage of the estimate. Thus, if an estimate Y is obtained for a certain characteristic and SE is the estimated standard error, then the CV will be $(SE/Y) \times 100$.

Finally, either the standard error or the coefficient of variation may be used to derive another measure of the accuracy of estimates, namely the confidence interval. This measure indicates the level of confidence that, for a characteristic observed, the true value for the population lies within the interval. An interval with a confidence level of 95% corresponds to the estimate obtained from the sample ± 2 standard errors: $(Y \pm 2 SE)$.² This means that if the sampling were repeated a large number of times, each sample would provide a different interval and 95% of the intervals would contain the true value of the characteristic. Similarly, if the sampling were repeated, the interval $Y \pm SE$ would contain the true value in 68% of cases.

1.2 Coefficients of variation

Estimates of coefficients of variation are calculated for estimates of many characteristics collected in the SHS. The CVs of detailed average household

2. The confidence interval is calculated directly from the CV in similar fashion, namely $Y \pm 2 (CV \times Y)/100$.

expenditure, as well as the CVs of dwelling characteristics and household facilities and equipment, are available at the national, provincial and territorial levels in the publication *User Guide—Survey of Household Spending* (see reference [3]).

It should be noted that the estimated CVs do not consider the fact that some of the data were imputed and thus may underestimate the true CVs. For most variables, the imputation rates are low (see Section 5) and the provided CVs represent good estimates. To assess the reliability of detailed expenditure with a high imputation rate, the CV and the imputation rate should be considered simultaneously.

Table 1.1 gives an overview of the CVs of estimates of household averages for a few of the summary-level expenditure categories and for income at the provincial, territorial and national levels.

Table 1.1
Coefficients of variation (%) by province, territory and at the national level for the estimation of average household expenditures for several summary level expenditure categories and for the estimation of average income

Summary level expenditure categories	Can.	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.	Nvt.
Total expenditure	0.7	1.6	2.3	1.6	1.2	1.7	1.4	1.4	1.5	1.3	1.5	3.6	3.9	5.9
Total current consumption	0.7	1.5	2.2	1.4	1.3	1.3	1.3	1.5	1.5	1.4	1.4	3.4	3.9	5.8
Food	0.6	1.3	2.2	1.3	1.3	0.9	1.2	1.5	1.6	1.4	1.4	3.8	3.5	2.9
Shelter	0.9	2.3	2.7	1.6	1.8	1.4	1.7	2.1	1.5	1.9	1.6	4.1	5.1	7.2
Household operation	1.0	1.6	3.5	2.6	2.0	1.6	2.1	1.9	1.9	2.3	2.1	4.7	4.1	7.1
Furnishings	1.7	3.5	5.5	4.1	3.4	3.4	3.3	3.8	4.3	3.5	4.2	10.5	9.2	12.4
Clothing	1.3	2.5	3.5	3.0	2.7	2.0	2.5	2.5	3.1	2.8	3.4	10.1	6.5	8.2
Transportation	1.5	3.7	4.7	3.6	3.0	3.3	2.8	4.6	3.7	3.7	3.3	7.2	10.4	19.7
Health care	1.3	3.0	3.5	3.7	3.2	2.0	3.3	4.0	3.8	2.5	2.8	7.8	9.9	19.7
Personal care	1.0	2.2	3.6	2.1	2.4	1.6	2.0	2.2	2.2	2.8	2.5	9.9	4.6	8.2
Recreation	1.7	3.8	6.5	3.1	4.5	3.0	3.4	3.5	3.7	3.3	3.9	7.7	7.3	9.2
Reading & printed materials	1.7	4.3	5.1	3.8	4.0	2.9	3.6	4.0	3.3	4.1	3.7	7.6	8.0	19.0
Education	3.0	7.5	11.0	8.3	8.6	4.3	5.4	7.5	6.9	7.3	6.7	12.4	19.6	29.7
Tobacco, alcoholic	1.6	3.6	6.9	4.1	4.8	2.7	3.5	4.8	4.0	4.5	4.5	8.4	8.2	8.1
Games of chance (net)	4.1	6.4	13.8	8.0	9.6	6.3	7.7	8.2	7.4	13.3	13.4	14.4	24.3	20.4
Miscellaneous expenditures	2.9	6.0	28.1	5.1	5.4	4.2	6.0	6.0	6.4	8.2	6.5	10.9	7.6	17.2
Personal income tax	1.7	3.7	4.5	2.8	2.2	3.4	3.2	2.6	2.9	2.5	2.6	6.3	5.2	9.7
Personal insurance and pension contributions	1.2	3.1	2.9	9.5	1.9	2.5	2.3	3.2	2.2	2.1	2.0	6.9	5.0	6.7
Gifts and contributions	6.0	8.3	14.9	9.3	7.9	9.3	10.8	11.5	6.3	7.6	18.6	15.1	15.1	16.5
Income	0.7	1.4	2.2	1.5	1.1	1.6	1.5	1.3	1.5	1.1	1.4	3.9	4.6	6.0

The coefficients of variation (CVs) of the average estimates of total expenditure per household vary between 1.2% and 2.3% for the provinces. The CVs are higher in the territories, namely 3.6% in the Yukon, 3.9% in the Northwest Territories and 5.9% in Nunavut. The CV at the national level is 0.7%.

For summary-level expenditure categories, the CVs at the national level are less than or equal to 1.7%, except for the following categories: education, games of chance, miscellaneous expenditures and gifts of money and contributions. These expenditure categories represent respectively 1.6%, 0.4%, 1.5% and 2.5% of total expenditure (data not shown). Also, with the exception of these categories, the CVs are generally less than or equal to 5% at the provincial level. With the exception of the same above-mentioned categories, the CVs in the territories are mostly less than 11%. Moreover, the sample size being smaller in the territories, the CVs tend to be higher than those of the provinces.

Table 1.2 gives an overview of the CVs for some dwelling characteristics and household equipment estimates at the provincial and territorial level as well as at the national level.

Table 1.2
Coefficients of variation (%) by province, territory and at the national level for some dwelling characteristics and household equipment

Categories	Can.	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.	Nvt.
Owner	0.9	1.7	3.3	2.0	1.8	1.7	1.9	1.9	1.8	1.9	1.9	6.6	10.9	15.1
Renter	1.7	5.6	8.6	5.0	5.2	2.2	4.0	4.7	5.0	4.8	3.3	10.8	16.4	8.5
Washing machine	0.6	0.9	2.1	1.5	1.3	0.8	1.5	1.6	1.2	1.1	1.3	4.7	5.3	4.9
Clothes dryer	0.6	1.1	2.3	1.5	1.3	0.9	1.6	1.6	1.1	1.2	1.5	4.4	7.0	5.5
Dishwasher	1.1	3.8	4.7	3.5	3.4	1.9	2.5	2.7	2.6	2.0	1.8	7.2	8.0	12.4
Freezer	1.0	1.4	3.0	2.3	1.8	2.0	2.1	1.7	1.5	2.3	2.2	5.6	6.1	7.9
Microwave oven	0.3	0.6	1.1	0.7	0.6	0.6	0.6	0.7	0.6	0.6	1.0	2.1	4.3	4.7
Cellular phone	1.1	2.9	5.2	2.8	3.7	2.4	2.0	2.8	2.8	2.0	2.1	12.2	8.0	24.6
CD player	0.6	1.5	2.5	1.5	1.7	1.2	1.3	1.7	1.6	1.3	1.4	4.4	3.5	6.0
Cable TV	1.0	2.4	5.0	2.7	2.9	1.9	2.0	2.7	3.0	2.6	1.6	10.4	6.2	6.8
Satellite dish	2.4	5.0	7.8	5.6	5.0	4.2	5.5	4.7	4.4	4.8	5.7	11.7	12.0	16.4
DVD player	1.1	2.9	4.7	2.4	2.9	2.1	2.2	2.3	2.6	2.1	2.6	5.4	5.0	8.2
Home computer	0.8	2.5	3.8	2.0	2.4	1.6	1.5	2.0	1.8	1.6	1.6	4.8	6.8	7.9
Regular tel. connection to a computer (Modem)	2.4	5.4	9.5	6.2	4.6	4.1	4.4	7.0	7.1	7.4	6.8	15.3	17.4	11.2
High-Speed tel. connection to a computer	2.7	9.4	12.0	5.6	8.5	5.1	5.7	6.7	5.2	5.3	5.5	9.2	12.8	40.0
Use of internet (home)	0.9	3.2	4.9	2.4	2.8	2.0	1.9	2.7	2.4	2.1	2.1	6.7	7.0	10.8
Owned vehicles (one)	1.3	2.8	6.2	3.7	3.4	2.4	2.9	3.1	3.3	3.1	2.9	9.6	5.6	14.3
Owned vehicles (2 or more)	1.4	4.4	4.9	2.9	3.5	2.9	3.1	2.8	3.2	2.4	2.7	7.1	10.2	22.3

The coefficients of variation for the estimates of dwelling characteristics and household equipment are generally below 5% at the provincial level, with some exceptions in the following categories: rent, satellite dish, regular telephone connection to a computer (modem) and high-speed telephone connection to a computer. The CVs are higher in the territories where we can find a smaller proportion of equipment. In addition, the sample size being smaller in the territories, the CVs tend to be higher than those of the provinces.

The coefficients of variation for dwelling characteristics and household equipment at the national level are below or equal to 1.7% with the exception of the following categories: satellite dish, regular telephone connection to a computer (modem) and high-speed telephone connection to a computer. There is a smaller proportion of households with such equipment for these three categories. At the national level, this proportion represents 22.6%, 21.2% and 17.3% respectively (data not shown).

1.3 Model for deriving an approximation of the CV

Estimates for different domains of interest (for example, by income quintile) for the summary level expenditure categories are available in the publication *Spending Patterns in Canada* (see reference [4]). Estimates for different domains of interest for detailed expenditures are available upon request from the Income Statistics Division. (For more details on tables available upon request from the Income Statistics Division, see references [3] or [4].) For operational reasons, it is not possible to produce CVs for all the characteristics collected by the survey at all the different levels of aggregation that may interest users.

1.3.1 Model for deriving an approximation of the CV for domain estimates

It is, however, possible to calculate an approximation of the CV by using a relationship between the number of households in the sample that reported expenditures for a given category and the CV at an aggregated level. This relationship, based on the CV's tendency to increase in proportion to a decrease in the square root of the number of households reporting an expenditure, is illustrated below.

Formula for approximating the CV for a domain (subgroup of the population)

If $CV(Y)$ represents the CV for the estimate of the average per household of a certain characteristic for the entire population, then an approximation of the CV of the estimate of that characteristic can be calculated for a domain (which may be considered as a subgroup of the population, such as household type, income quintile, urbanization level) according to the following equation:

$$CV(Y_d) = CV(Y) \times \sqrt{\frac{nP}{n_d P_d}}$$

where

- n: number of households in the sample
P: estimate of the proportion of households reporting a value > 0 for this characteristic in the population
 n_d : number of households in the sample in domain d
 P_d : estimate of the proportion of households reporting a value > 0 for this characteristic in domain d

Generally, approximations for the different domains are calculated using the CV, size n and proportion P at the national level. If an approximation of a CV is desired for a domain that is entirely contained within a single province (for example, a metropolitan area), then it is preferable to use the values at the provincial level, since provincial CVs are published for the 2003 SHS (reference [3]). It should be noted that a CV obtained using this approach is only an approximation of the real value.

1.3.2 Method of computation of an approximate CV from the microdata file

The microdata file users can obtain an approximation of the CV of the estimates using another method which will generally provide better results than the method described in the previous section for the CVs of detailed expenditure estimates. This approach is described in detail in the documentation provided with the 2003 microdata file. This method of approximation can be used only with the microdata file since it is necessary to have data and weights for each household.

The document on data quality for the 1997 SHS contains the results from the performance evaluation of these two CV approximation methods.

1.4 Suppression of unreliable data in estimation tables

Since the coefficient of variation is an indicator of the reliability of data, we would like to use it to determine whether or not the estimates should be published. Estimates for which the CV is more than 33% are not considered sufficiently reliable to be published. However, CV estimates are not calculated for many of the published estimates. The suppression rule for expenditure estimates is therefore based on the number of households reporting a value greater than zero.³

It can be shown that CVs are usually below 33% when the number of households reporting an expenditure is greater than 30. Since this is an approximate rule, some estimates may be published even though the CV is greater than 33%, and some estimates will not be published even though the CV is less than 33%. The document on data quality for the 1997 SHS gives the results from the evaluation of the risk of error in the use of the suppression rule.

3. In practice, we use the estimate of the proportion of households reporting an expenditure, which is multiplied by the sample size.

2. Nonresponse

Errors due to nonresponse result from the fact that some potential respondents do not provide the necessary information or the provided information proves to be unusable. When the respondent has failed to respond to only some questions, this is referred to as partial nonresponse. In such a case, the missing data are imputed. Errors associated with imputation are described in Section 5, which deals with processing errors. In the present section, nonresponse includes collection nonresponse, which is mainly due to the inability to contact the household or to the refusal of the members of the household to participate partially or completely in the survey, as well as data collected from households that prove to be unusable.

The main impact of nonresponse on data quality is that it can introduce a bias in the estimates if the characteristics of respondents and nonrespondents differ and the difference has an impact on the characteristics studied. Nonresponse rates may easily be calculated, but they have only an indicative value with regard to data quality, since they do not allow estimation of the bias associated with the estimates. The scope of nonresponse may be considered as an indicator of the risks of bias in the estimates.

2.1 Response, nonresponse and vacancy rates

Since the units selected in the SHS are dwellings, interviewers must first identify ineligible dwellings, that is, dwellings occupied by persons who are not part of the target population, as well as dwellings that no longer exist (demolished, mobile home moved or dwelling converted to business) and vacant dwellings (unoccupied, seasonal or under construction).

Among eligible dwellings, the proportion of households that did not respond to the survey is evaluated next. This is called the collection nonresponse rate. Included are households that refused to participate in the survey and households where no contact could be made with the respondents, either because they were absent or because of special circumstances (language problem, illness, death).

Again among eligible dwellings, the rate of unusable data is determined. Unusable data refers to the number of households whose questionnaires were at least partially completed but which were rejected during data processing. There are two main causes for rejection. First, when many questions on income or expenditures have been left unanswered, the questionnaire is classified as incomplete and is not used. The other source of rejection consists of questionnaires in which the difference between receipts (income and other sources of money received by the household) and disbursements (expenditures and net change in assets and liabilities) is greater than 20%. These questionnaires are also excluded from the estimation and are considered as nonresponse.

Note that all rates provided in this section are unweighted. For the 2003 Survey of Household Spending, the final response rate is 72.3%. Table 2.1a shows the final response rates as well as the sample size (eligible households) broken

down by refusals, no-contacts, usable and unusable data. These rates are provided at the national level as well as at the provincial and territorial level.

This year, a supplementary sample of about 2,000 households in Quebec was financed by the “Ministère des finances du Québec” to better suit their analytic needs. The sample was selected from a list frame of dwellings according to a two-stage stratified sample design. The sampling frame constructed from Statistics Canada’s Address Register is comprised of dwellings from selected Dissemination Areas of the 2001 Census. For a detailed description of the sample design of the supplementary sample, see reference [6].

Table 2.1a
Sample size and response rate (%) by province, territory and at the national level

Province or territory	Eligible households	Non-contacts	Refusals	Unusables	Usables	Final response rate (at estimation stage) ¹ %
Canada	23,869	1,647	4,234	723	17,265	72.3
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	2,58	205	500	45	1,83	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

1. Usable/eligible x 100

Table 2.1b shows the final nonresponse rates; the collection nonresponse rates, broken down by refusals and no-contacts; and the rate of unusable data broken down into incomplete and out-of-balance questionnaires. The vacancy rate is also included. These rates are provided at the national level as well as at the provincial and territorial level.

Note that the vacancy rates shown in tables of Section 2 include vacant dwellings (unoccupied, seasonal or under construction) as well as dwellings that no longer exist (demolished, mobile home moved or dwelling converted to business).

Table 2.1b
Nonresponse and vacancy rates (%) by province, territory and at the national level

Province or territory	Vacancy rate	Collection nonresponse rate			Unusable data rate			Final nonresponse rate (at estimation stage)
		Total	No contact	Refusal	Total	Incomplete	Out-of-balance	
Canada	11.0	24.6	6.9	17.7	3.0	1.3	1.7	27.7
Newfoundland and Labrador	15.4	18.7	7.6	11.2	3.9	1.8	2.1	22.6
Prince Edward Island	18.1	22.2	6.7	15.5	3.5	1.7	1.8	25.6
Nova Scotia	13.8	24.3	7.3	16.9	6.4	1.2	5.2	30.7
New Brunswick	13.7	21.7	5.3	16.4	6.4	1.9	4.5	28.1
Quebec	7.3	26.7	7.5	19.2	1.4	1.2	0.2	28.1
Ontario	8.2	33.9	9.0	25.0	3.3	1.9	1.4	37.2
Manitoba	9.6	22.0	5.5	16.5	2.6	0.7	1.9	24.6
Saskatchewan	15.7	19.2	3.9	15.2	4.2	1.3	2.9	23.4
Alberta	9.0	22.2	6.1	16.1	0.9	0.7	0.1	23.1
British Columbia	8.9	27.3	7.9	19.4	1.7	1.4	0.4	29.1
Yukon	15.3	24.4	6.7	17.7	2.2	0.2	1.9	26.6
Northwest Territories	12.2	15.4	5.1	10.2	2.0	0.2	1.7	17.3
Nunavut	14.8	8.4	5.5	2.9	0.0	0.0	0.0	8.4

The final nonresponse rate at the national level is 27.7%. It is due to refusals (17.7%), to households that could not be contacted (6.9%), and finally to households for which the data were unusable (3.0%). For most provinces, refusals are the main cause of nonresponse, followed by the households that could not be contacted, and by the households for which the data were unusable. The only exceptions are Nunavut for whom the refusal rate is smaller than the no-contact rate, as well as New Brunswick and Saskatchewan where the proportion of unusable data is higher than the proportion of households that we were unable to contact.

The final nonresponse rate varies from one province to another. Nunavut has the lowest nonresponse rate at 8.4%. The lowest refusal rate (2.9%) is also observed for Nunavut. In addition, in Nunavut, all data collected are usable which means that the final nonresponse is due only to the collection nonresponse rate. Nonresponse rates in Nunavut and Northwest Territories are smaller than 20% while rates over 30% are observed in Nova Scotia and Ontario. The higher final nonresponse rate in Nova Scotia is due in part to a higher rate of out-of-balance questionnaires (5.2%). Ontario has the highest nonresponse rate at 37.2%. The highest rates of no contact (9.0%) and refusal (25.0%) are also observed in Ontario.

The vacancy rates are shown in Table 2.1b, but it should be kept in mind that vacant dwellings do not contribute to the bias of the sample if they are correctly identified. By analysing vacancy rates, we can detect dwelling identification problems associated with the collection process. The national vacancy rate for the 2003 SHS is 11.0%.

2.2 Nonresponse according to urbanization level

Nonresponse varies according to urbanization level. The various rates at the national level are shown by urbanization level in Table 2.2.⁴

Table 2.2
Nonresponse and vacancy rates (%) by urbanization level

Urbanization category	Vacancy rate	Collection nonresponse rate			Unusable data rate			Total nonresponse rate (at estimation stage)
		Total	No contact	Refusal	Total	Incomplete	Out-of-balance	
Urban								
1,000,000 or more	4.8	31.8	9.4	22.4	2.0	1.5	0.5	33.8
500,000 to 999,999	5.7	25.3	6.3	19.0	1.6	0.8	0.7	26.9
250,000 to 499,999	5.4	30.1	8.2	21.9	6.2	1.2	5.0	36.3
100,000 to 249,999	7.6	24.0	6.5	17.5	4.5	1.3	3.2	28.5
30,000 to 99,999	6.7	26.4	7.1	19.4	3.4	1.7	1.7	29.8
Less than 30,000	10.6	20.0	5.8	14.2	2.7	1.2	1.5	22.7
Rural	23.1	18.9	5.4	13.5	3.4	1.3	2.0	22.3
Total	11.0	24.6	6.9	17.7	3.0	1.3	1.7	27.7

The final nonresponse rate generally increases with urbanization level. According to Table 2.2, the “500,000 to 999,999” and “100,000 to 249,999” groups and, to a lesser extent, the “1,000,000 or more” group, go against this rule. For the “500,000 to 999,999” group, this can be explained in part by the fact that this group has the lowest rate of unusable data (1.6%). For the “100,000 to 249,999” group, this is due to a lower refusal rate (17.5%) than those of urbanization regions of 30,000 inhabitants and more. The final nonresponse rate of 36.3% for the urbanization category “250,000 to 499,999” can be attributed to the collection nonresponse rate (30.1%), which itself is mostly due to the rate observed in Ontario (38.1%, data not shown), as well as the rate of unusable data (6.2%), which itself is mostly due to the rate observed in Nova Scotia (9.8%, data not shown).

The collection nonresponse rate also increases with urbanization level. There is a difference of nearly 12% between the urbanization categories “less than 30,000” and “1,000,000 or more”. Refusals account for more than 60% of the

4. Tables on nonresponse rates by urbanization level and province are available on request from the Household Survey Methods Division.

total nonresponse at each level of urbanization. Although the urbanization category “500,000 to 999,999” has the lowest total nonresponse rate among urban regions of 30,000 inhabitants or more, this category has the highest proportion of nonresponse (70.6%) due to refusals. The urbanization category “1,000,000 or more” has the highest rates of no contact (9.4%) and refusal (22.4%).

From an examination of the vacancy rate by urbanization level, it appears that the vacancy rate at the rural level (23.1%) is twice that for low-population urban areas (10.6%). The latter also shows a rate that is higher than for the high-population urban areas. This phenomenon is also observed in the LFS and may be explained by a greater number of seasonal dwellings in rural areas. The same factor also explains the higher vacancy rates in the Atlantic provinces, as illustrated in Table 2.1b, and especially in Prince Edward Island, which has a high proportion of rural dwellings.

2.3 Nonresponse according to income strata

Since income information is not available for nonrespondents, it is not possible to compare nonresponse rates according to income. However, the LFS sample design, used for the SHS, was designed in such a way that in nine large cities there are strata consisting of geographic areas where the average household income exceeds \$100,000, and in seven large cities there are strata consisting of apartments inhabited by households with an average income of less than \$20,000. Even though the number of such strata is small and accounts for only a small number of dwellings in the SHS sample (approximately 580 for high income and 160 for low income strata, or less than 3% of the sample), the comparison of nonresponse rates for these two groups in relation to the other strata is revealing. Table 2.3 shows these results. Note that results from this table exclude the supplementary sample in Quebec. Since the supplementary sample was not selected from the LFS sampling frame, such strata classification (high-income, regular, low-income) is not available for that particular sample.

Table 2.3
Comparison of nonresponse and vacancy rates (%) in high-income and low-income strata in relation to other strata

Stratum type based on income ¹	Vacancy rate	Collection nonresponse rate			Unusable data rate			Final nonresponse rate (at estimation stage)
		Total	No contact	Refusal	Total	Incomplete	Out-of-balance	
High-income	4.1	39.8	9.4	30.5	2.4	1.5	0.9	42.2
Regular	11.7	24.1	6.6	17.4	3.2	1.3	1.9	27.3
Low-income	3.7	25.2	11.9	13.3	2.1	1.4	0.7	27.3
Total	11.4	24.5	6.7	17.7	3.2	1.3	1.9	27.6

1. Excluding the supplementary sample in Quebec.

The final nonresponse rate (42.2%) in high-income strata is much higher than in regular or low-income strata. The refusal rate for high-income strata is over 30%, which is about twice the rate of other types of strata.

Households in low-income strata have the same final nonresponse rate and a collection nonresponse rate similar to those of the regular strata. However, the causes of the nonresponse are different for the two types of strata. For the regular strata, collection nonresponse rate is mainly due to refusals, while for the low-income strata it is approximately equally attributable to refusals and no-contact.

As for the 1997 to 2002 SHS, the vacancy rate is higher for regular strata than for each of the other two strata. In high-income and low-income strata, the vacancy rates are similar. The vacancy rate in low-income strata is at 3.7%, up from a rate of 1.8% in 2002.

2.4 Adjustment for nonresponse

To compensate for nonresponse, the weights in the SHS are inflated by the inverse of the weighted response rate within certain groups defined on the basis of the different urbanization levels in each province. The weighted rates differ from the rates presented in this section since the former takes into account the sampling weight of each household. An algebraic description of the nonresponse adjustment is provided in Appendix A.

The weights adjustment for nonresponse takes into account the differences in nonresponse by urbanization level as described in Section 2.2. It will reduce the bias to the extent that the characteristics of respondents and nonrespondents are similar for a given urbanization level.

3. Coverage errors

In the design of the survey, the target population was defined. It is useful to go over this definition, since a good understanding of the target population is necessary in order to properly interpret the survey data. One should note that SHS, with the exception of Quebec's supplementary sample this year, uses the LFS sampling frame.

Target population

The target population consists of individuals living in private households. It therefore excludes residents of institutions such as prisons, chronic care hospitals or senior citizens' homes, as well as members of religious orders and other groups living communally, members of the Armed Forces living in military compounds, and individuals residing permanently in hotels or rooming houses. Also excluded are foreign countries' official representatives residing in Canada and their families as well as individuals residing on Indian reserves or public lands (with

exception for the Territories). With these exclusions, the survey covers nearly 98% of the population in the 10 provinces. In the Yukon, the coverage of the population is 88 %. The coverage of the Northwest Territories represents 92 % of the population and the coverage of Nunavut represents 89 % of the population.⁵

We did not collect data from persons temporarily living away from their families (for example, students at university) because the data would be gathered from their families if selected in the sample.

Coverage errors result from inadequate representation of the target population based on the units of the sampling frame. Some units of the target population may be omitted from the sampling frame, in which case there is undercoverage. Other units that are not in the target population may be included by error, or some units may be included more than once. These units are responsible for overcoverage.

3.1 Undercoverage and overcoverage: slippage rates

In the SHS, the sample is selected using a list of dwellings in each selected cluster. Factors contributing to undercoverage are: the omission of dwellings in the creation of the list, new dwellings that are added between the creation of the list and the interviewer's visit (mainly in developing areas), and the erroneous classification of vacant dwellings. The inclusion of dwellings that are not within the boundaries of the cluster is a source of overcoverage. Similarly, errors can take place during data collection, due to improper identification of persons as members of the selected household. These errors also contribute to undercoverage or overcoverage.

A good representation of the target population is essential to the production of realistic expenditure estimates. The number of people per household is also an important characteristic in the estimation of average household expenditures. Therefore, it is necessary that the sample not only adequately represent the individuals in the target population, but also the distribution of households according to their size.

In 1999, a weighting strategy that uses new controls was introduced. This method results in a better correction of the representation of the target population by using a more detailed age-sex grouping than was used previously and for which the coverage varies from one group to the other.

There is generally a net undercoverage of the number of persons in the SHS. This undercoverage is corrected by an adjustment of weights using auxiliary data based on post-censal demographic estimates. The slippage rate (see Appendix A) is a measure of the percentage of difference between the auxiliary data and the survey estimates calculated using weights not adjusted with these data.⁶ Tables 3.1 and 3.2 show slippage rates by age-sex group at the national level

5. In terms of households, the coverage of the Yukon, the Northwest Territories and Nunavut represents 88%, 93% and 90% respectively of households.

6. The subweight which is the survey weight adjusted for nonresponse is used (see Appendix A).

and at the provincial and territorial level respectively, while Table 3.3 presents these rates for the household size categories used for the weight adjustment. A positive rate indicates overcoverage of the number of persons or households in the survey.

Table 3.1
National slippage rates by age-sex group

National slippage rates by age-sex group				
	Age	Sex		Total
		Male	Female	
Canada ¹	0-6 years	-1.0	-6.5	-3.7
	7-17 years	-6.0	-6.8	-6.4
	18-24 years	-16.9	-14.7	-15.8
	25-34 years	-24.2	-15.7	-20.0
	35-54 years	-15.0	-10.1	-12.5
	55-59 years	-12.2	-8.0	-10.1
	60-64 years	-10.1	-6.8	-8.4
	65-69 years	-4.5	-5.2	-4.9
	+	-8.9	-7.5	-8.1
	Total	-12.8	-9.8	-11.3

1. Excluding the territories.

For the 2003 SHS, the national undercoverage rate was 11.3%. If we analyze Table 3.1, we can see that at the national level, there is always a net undercoverage for each age group. Except for the category 65 to 69 years old, the undercoverage rates for children (0 to 6 and 7 to 17) are smaller than those of the adults. The undercoverage rate for all children combined is 5.5%, while it is 13.0% for adults (data not shown). The slight undercoverage for the 0 to 6 year-old males is mostly due to the overcoverage in Quebec (1.7%) and Ontario (3.6%) for that age-sex group (see Table 3.2). Similarly, the lower undercoverage rate for the category 65 to 69 years is due to the overcoverage in Quebec (11.9%) among women of that age group, as well as the overcoverage in Ontario (11.2%) among men of that age group (see Table 3.2).

The highest national rates occurred among 25 to 34 year-old men, 18 to 24 year-old men and 25 to 34 year-old women. Except for the category 65 to 69 years, the undercoverage rate for women is always smaller than the undercoverage rate for men.

As mentioned previously, the SHS uses the LFS sampling frame. Over the same period, the national LFS undercoverage rate was 10.7% (reference [5]). This is slightly lower than the 12.6% SHS rate for those 15 years or older (data not shown).

Table 3.2
Slippage rates for the provinces and territories by age-sex group

Slippage rates by age-sex group														
Sex	Age	Newfoundland and Labrador	Prince Edward Island	Nova Scotia	New Brunswick	Quebec	Ontario	Manitoba	Saskatchewan	Alberta	British Columbia	Yukon	Northwest Territories	Nunavut
Male	0-6	-10.7	9.9	-4.5	2.5	1.7	3.6	-3.5	-17.3	-2.2	-14.3	-0.9	-11.4	-12.7
	7-17	-6.0	-4.6	-6.5	-7.2	4.0	-10.7	-0.9	-12.3	-4.2	-9.6			
	18-24	-35.5	-21.0	-27.4	-21.5	-2.9	-21.2	-30.4	-33.3	-11.5	-20.4			
	25-34	-29.6	-18.7	-27.2	-18.6	-18.1	-30.0	-20.2	-19.0	-16.4	-26.5			
	35-54	-11.4	-12.5	-3.6	-15.7	-9.5	-16.3	-14.6	-22.5	-23.2	-16.4			
	55-59	-20.2	-24.6	-29.4	-11.1	-13.2	-3.9	-26.1	-14.9	-20.0	-18.4	-28.2	-22.0	-16.0
	60-64	-20.7	-9.4	-9.2	-12.9	-22.0	2.0	-30.6	-5.9	-11.6	-14.2			
	65-69	-0.3	-3.6	-3.8	1.9	-16.3	11.2	-32.5	-0.2	-11.5	-19.6			
	70 +	-2.7	3.8	-19.5	-10.6	-2.4	-13.6	7.8	5.9	-10.1	-12.8			
	Total	-15.5	-10.2	-12.7	-12.6	-7.9	-13.6	-14.0	-16.6	-14.4	-16.9	-21.7	-18.8	-14.6
Female	0-6	-10.3	14.3	0.5	9.9	6.3	-13.3	-7.7	-5.6	-10.6	-8.6	-20.4	-25.3	-16.8
	7-17	-9.8	-3.4	-4.0	-10.7	-5.4	-6.0	7.4	-19.8	-3.5	-15.1			
	18-24	-17.1	-20.5	-10.9	-11.5	-6.8	-17.6	-9.4	-19.3	-8.0	-26.6			
	25-34	-18.3	-12.5	-16.3	-17.9	-7.2	-16.9	-22.1	-24.2	-9.7	-27.0			
	35-54	-8.5	-5.0	-0.3	-6.6	-3.9	-12.0	-3.0	-15.4	-17.1	-14.4			
	55-59	-6.9	-12.4	-22.0	-10.3	-17.4	6.5	-22.0	-3.4	-17.9	-18.2	-24.6	-15.4	-8.6
	60-64	-8.1	0.7	-12.1	-5.4	-14.2	-1.0	-2.3	-1.7	-5.1	-10.6			
	65-69	-5.4	-21.1	-1.5	-8.7	11.9	-11.2	-18.8	-3.9	-25.5	-3.8			
	70 +	-13.0	-7.1	-1.0	-0.5	-8.3	-9.0	12.0	9.1	-11.4	-11.9			
	Total	-11.1	-7.0	-6.0	-7.6	-5.3	-10.5	-4.9	-12.2	-11.9	-16.4	-23.7	-18.6	-12.1
Total		-13.3	-8.6	-9.3	-10.1	-6.6	-12.1	-9.4	-14.4	-13.1	-16.6	-22.7	-18.7	-13.4

We observe a net undercoverage for all provinces and territories, with the rates varying from 6.6% to 22.7%. The magnitude of the undercoverage is greater for the Northwest Territories and Nunavut. Quebec has the lowest undercoverage rate at 6.6%. Nevertheless, this lower rate observed in Quebec is to some extent attributed to the supplementary sample selected in that province in 2003. When excluding Quebec's supplementary sample, the net undercoverage for that province went from 6.6% to 8.8 % (data not shown). However, a low overall rate of undercoverage is not a guarantee of a better coverage. For example, Prince Edward Island's overall slippage rate (-8.6%) concealed the worst case of overcoverage for a provincial age-sex group (14.3% among 0 to 6 year-old females). On the other hand, despite an overall slippage rate of -16.6%, British

Columbia has only the twelfth worst case of undercoverage (27.0% among 25 to 34 year-old women).

Among the provinces, the highest undercoverage rate occurred among 18 to 24 year-old men in Newfoundland and Labrador (35.5%). The magnitude of the undercoverage is higher among 25 to 34 year-old men. Indeed, for all provinces we observe an undercoverage rate greater than 16.0%, which explains the overall undercoverage rate of 24.2% at the national level (see Table 3.1). We can also see that the pattern of slippage rates differs substantially for age-sex groups from one province to the next.

Table 3.3
Slippage rates for the provinces and the Yukon⁷ by household size

Province or territory	Slippage rate			
	Households	One-person households	Two-person households	Three-person and more households
Canada ¹	-8.5	-2.2	-7.5	-13.2
Newfoundland and Labrador	-8.1	11.0	-5.1	-17.2
Prince Edward Island	-8.3	-9.8	-3.3	-11.3
Nova Scotia	-6.7	0.6	-8.2	-9.8
New Brunswick	-4.9	6.2	3.1	-17.6
Quebec	-5.0	-3.8	-3.0	-8.0
Ontario	-9.6	-2.9	-8.6	-13.9
Manitoba	-5.7	5.5	-9.5	-10.5
Saskatchewan	-9.4	2.8	-8.0	-19.4
Alberta	-10.6	-5.7	-14.2	-10.5
British Columbia	-11.8	-0.3	-9.6	-21.5
Yukon	-14.9	-2.0	-12.7	-26.7

1. Excluding the Northwest Territories and Nunavut.

Nationally, the number of households was underestimated by 8.5%. This slippage rate is slightly lower than the -11.3% slippage rate for individuals.

Among the provinces and the Yukon, there is a substantial variation in the slippage rate for one-person households. This rate ranges from -9.8% for Prince Edward Island to 11.0% for Newfoundland and Labrador. The same phenomenon occurs with respect to two-person households but with less variation. For two-person households, the slippage rate ranges from -14.2% in Alberta to 3.1% in New Brunswick. The slippage rate for households of three or more persons ranges from -26.7% for the Yukon to -8.0% in Quebec.

Except for New Brunswick, there is always undercoverage for two-person households and households of three or more persons. Note that for households

7. For 2003 reference year, the sampling weights were not adjusted to reflect post-censal estimates of the number of households in the Northwest Territories and in Nunavut (see Section 3.2).

of three or more persons, the undercoverage is generally greater than 10%. Again here, the lower rate in Quebec (8.0%) for that latter category is to some extent attributed to the supplementary sample selected in that province in 2003. When excluding Quebec's supplementary sample, the net undercoverage for that province went from 8.0% to 10.3 % (data not shown). Except for Alberta, the greatest undercoverage is observed for households of three or more persons.

3.2 Adjustment at the population and household levels

To correct the problem of the sample's representativeness shown in Tables 3.1 and 3.2 and to reduce the resulting bias, the survey data are adjusted during weighting using demographic estimates for the age groups defined in these tables, for each province and territory. For more information on the adjustment's methodology, see reference [1]. This adjustment reduces the bias but does not completely eliminate it entirely if the characteristics of the individuals omitted from the survey differ from those of individuals included for a given age group in a province.

It should also be noted that the effectiveness of the adjustment based on demographic estimates depends largely on the quality of those estimates and their accuracy in representing the target population of the survey. The demographic estimates are not error-free. They are post-censal estimates based on the population counts from the 1996 Census adjusted for net undercoverage, and they take into account recent statistics on migration, births, deaths, etc. These demographic estimates are adjusted to account for certain exclusions specific to household surveys, such as persons living in institutions. Conceptually, they differ slightly from the SHS target population in that they include persons living in non-institutional collective dwellings, such as members of groups living communally and individuals permanently residing in hotels or rooming houses. However, this difference is considered negligible, since such individuals represent less than 0.4% of the Canadian population.

To remedy the issue of the representativeness of the sample in terms of the number of households by size as illustrated in Table 3.3, the survey data are adjusted using supplementary data. By adjusting the SHS weights to reflect post-censal estimates of the number of households by size, the goal is to compensate for the bias resulting from inadequate representation of households. However, the bias will not necessarily be eliminated if characteristics of households not interviewed (i.e., omitted or non-respondent households) differ from those of responding households for a given household size. As in the case of demographic estimates of population, the effectiveness of the adjustment will depend on the quality of the supplemental data on the number of households. As such, controls on the number of households computed for SHS showed a break in the series of controls that have been produced up till now for Northwest Territories and Nunavut. While controls up to January 2001 were similar to those of the Census data, a gap was observed from those of January 2002. In an effort to reduce the impact of such a gap on the series of estimates from SHS, some changes were made to the methodology used in the production of the supplemental data on the number of households for those two territories for the 2001 SHS weighting. On the other hand, for all provinces and in the Yukon, no such problem is observed for the controls based on the 1996 Census. For the

2003 SHS, the gap in the series of controls on the number of households was still present for those two territories. Note that the new series of controls on the number of households based on the 2001 Census is available and this last series does not reveal any problems for those two territories. Before introducing the new series based on the 2001 Census, some evaluations must be done in order to measure the impact of the new controls on the survey estimates. For the above mentioned reasons and since the changes made for the 2001 SHS were meant to be temporary in nature, it has been decided to not use the controls on the number of households for the Northwest Territories and Nunavut for the 2003 SHS weighting. Although the sampling weights were not adjusted to reflect post-censal estimates of the number of households (based on 1996 Census) in the Northwest Territories and in Nunavut, the estimates of the total number of households obtained from the final weights are similar to the controls on the number of households based on the 2001 Census for those two territories. That is, we obtain a net overcoverage of 1.3 % for the Northwest Territories and a net undercoverage of 4.0 % for Nunavut.

In addition to demographic estimates of age-sex groups by province, three other groups of supplementary data are used during weighting to adjust survey data and thereby improve their representativity. The first set of data is used to control for the number of children and adults in certain major cities. The second is designed to control for the number of lone-parent households and couples with children by province. Finally, counts for major categories of income from wages and salaries are used when adjusting weights to ensure a certain degree of consistency between the income distributions from the SHS and from outside sources.

4. Response errors

Response errors represent a lack of accuracy in responses to questions. They can be attributed to different factors, including a questionnaire that requires improvements, misinterpretation of questions by interviewers or respondents, and errors in respondents' statements.

In the SHS, there can be various reasons for errors in respondents' statements. First, there are recall errors that occur when a respondent forgets expenditures made during the period covered by the survey (which corresponds to the calendar year), or when a respondent provides an erroneous value because of the time interval that has elapsed between the time of purchase and the date of the interview. Recall errors are probably the survey's largest source of response error, since the reference period is long (12 months) and a great variety of information is requested.

One of the main measures taken to minimize recall error in the SHS is to calculate the difference between receipts (income and other amounts received by the household) and disbursements (expenditures plus net change in assets and liabilities) for each household. When the difference exceeds 10% of receipts or disbursements, with the higher amount being retained, respondents are contacted again in order to obtain additional information and to try to identify

errors or omissions. The respondent is also encouraged to consult various documents (invoices, bank statements, etc.) in order to provide more accurate data. To determine expenditures for small items purchased at regular intervals, interviewers generally suggest that respondents estimate the frequency of the purchases and the price generally paid in order to derive expenditures for a 12-month period.

A second source of error in respondents' reporting is telescopic error, which consists of including in the reference period events that occurred before or after it. In the SHS, the use of the calendar year is considered to provide a good marker for the start of the reference period. Furthermore, since the reference period is a long one, telescopic error has less impact.

Responses by proxy can also contribute to response error. The household member who made an expenditure is generally best able to report it accurately. This is definitely the case with, say, personal purchases. Expenditures reported by an intermediary are more likely to be tainted by response error, and this type of error tends to have a greater effect on certain types of expenditures.

Among other sources of response error, the extent of the respondent's cooperation should not be overlooked. For personal reasons, the respondent may decide not to mention particular expenditures or decide to twist the facts.

In the SHS, another factor is the response burden, owing to the length of the interview and the great variety of items to be reported, as well as the pace of the interview. This can lead to respondent fatigue and affect the quality of the responses obtained. The interview time varies greatly from one household to another, depending on household size, income and various other characteristics. For some households, the interview can take more than five hours.

It should be stressed that questions were added to the 2001 SHS. Among other things, for 2001 only, extra questions were included in the survey so that data from the SHS could be used in the weighting of the Consumer Price Index. This change may affect historical comparisons for a few variables. For example, questions were added under "Personal care preparations" to collect specific 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 estimate for "Personal care preparations" may have been at least partly caused by an improvement in respondent recall. The effect of additional questions on estimates is difficult to quantify. However, in 2002 when the extra questions were removed, the estimate for personal care spending decreased again.

While response errors are a major source of error in a historical interview, they are the aspects of data quality that are the hardest to measure. Generally, it is necessary to conduct quite costly special studies in an attempt to measure them. Efforts are made to combat response errors by using survey techniques designed to reduce them.

5. Processing errors

Processing errors can arise in all types of data handling. The main stages of data processing are coding, data entry, editing, imputation of partial nonresponse and weighting. In the SHS different procedures are applied at each stage in order to minimize processing errors and the survey estimates are compared with other data sources prior to release. Errors related to the adjustments made at the weighting stage have been described in sections 2 and 3. The other types of processing errors are covered in this section.

Coding is necessary for only a few questions. This is done by the interviewer and subsequently verified by a senior interviewer. Before 2001, data entry was done with the help of an automated verification system that grouped the questionnaires into batches and chose some questionnaires from each batch to be entered a second time. Any errors found were to be corrected. If the number of errors in a batch was greater than a certain threshold, then the entire batch was submitted for re-entry. Due to the introduction of a new data capture system (BLAISE), there was no questionnaire batch verification procedure in 2003. However, some edits were implemented in the new data capture system to ensure consistency of data captured. The results of a preliminary data capture study seem to show that data capture error rates of the new system are similar to the ones of the old system.

The first stage of automated verification is done after each questionnaire has been verified manually by both the interviewer and the senior interviewer. It is ensured that the respondent's answers follow some essential consistency rules. Unusual situations that may justify corrections are also identified. This stage of verification is done in Statistics Canada's regional offices in case it is necessary to recontact respondents if some supplementary information is required to resolve inconsistencies in the answers provided. Specially-trained members of the verification teams solve any problems identified. Thereafter, other verification checks are done at head office and invalid responses are corrected.

The processing of SHS data also involves imputation for partial nonresponse. Partial nonresponse occurs when the respondent refuses to answer or does not know the answer to certain questions. The imputation approach differs depending on whether the data is categorical or continuous. Categorical data takes on only specific values (as in yes/no questions or type of dwelling questions), while continuous data can take any numerical value (as for income and expenditure data).

Income and expenditure data are imputed using the nearest neighbour technique. The imputation is done on one group of variables at a time, with the groups chosen by taking the relationships among the variables into account. A group generally corresponds to a section of the questionnaire. For every group, the missing values of a recipient (a household that has some missing data for at least one of these variables) are imputed from data of the most similar record among all donors (households that have no missing values for these variables). For each recipient the closest donor is chosen as the one that minimizes a particular distance function. This function is based on matching variables that

are chosen because they are correlated with the variables to be imputed. For example, the total income of a household is chosen as a matching variable for all sections pertaining to expenditures. It must also be ensured that, after receiving the donor values, the recipient household satisfies some consistency rules. In general, the imputation is done at the household level, but in some groups (e.g., income and clothing expenditures), the imputation is done at the person level since the original data is collected at that level.

Note that since 2001, the imputation of all expenditure and income data is done using the Canadian Census Edit and Imputation System (CANCEIS) of Statistics Canada. This new system is based on slightly different methodology from the one used by the previous system. The new system allows a better use of categorical variables as matching fields when selecting a donor. The new system was tested prior to its implementation and the results it gave were similar to those with the old system. Moreover, this system allows for the imputation of both continuous and categorical data. Starting with 2003, categorical data, which are obtained mainly in the facilities and equipment section of the questionnaire, are imputed with the CANCEIS system. The categorical data were previously imputed with the help of a "hot deck" imputation technique that randomly chooses a donor from a group of answering households with similar characteristics.

The bias caused by imputation of partial nonresponse is difficult to evaluate. It depends on the differences between respondents and nonrespondents as well as the ability of the imputation method to produce unbiased estimates. However, the imputation rates indicate the importance of partial nonresponse. They are presented in the following section.

5.1 Proportion of households or persons requiring imputation at national, provincial and territorial levels

A preliminary indication of the magnitude of partial nonresponse is the proportion of households requiring imputation and the number of variables imputed by household. The questionnaire can be divided into two major groups of variables: those collected at the household level and those collected at the individual level (such as income and clothing expenditure). For the latter, the respondent may provide only the total income or total clothing expenditures if he/she is unable to provide the breakdowns by source of income or type of expenditure. The level of imputation for the components of income and clothing expenditure is then larger, but this does not affect the total income, total clothing expenditure or total expenditure.

The percentage of households requiring imputation for household expenditure (excluding clothing expenditures and expenditures in the section on Personal Taxes, Security and Money Gifts) is presented in the next sub-section. The subsequent sub-section presents the percentage of persons requiring imputation for a clothing expenditure variable, the percentage of persons requiring imputation for an income variable and the percentage of persons requiring imputation for a variable in the section on Personal Taxes, Security and Money Gifts. Finally, the last sub-section presents the percentage of households requiring imputation for at least one of the categorical variables. After data imputation by the system, some corrections might have been needed on both

imputed and non-imputed variables aiming to ensure data consistency. The actual percentage of records requiring this kind of modification is about 1% at the household level. The results are provided at the national, provincial and territorial levels. This gives an indication of which provinces are more affected by imputation.

5.1.1 Household expenditure imputation by province and territory

The percentage of usable households that required imputation for an expenditure variable (excluding clothing expenditures and expenditures in the section on Personal Taxes, Security and Money Gifts) is presented in Table 5.1. Usable households correspond to all households living in eligible dwellings, excluding households who could not be contacted, who refused to participate in the survey, or who provided unusable data (see Section 2.1). The table is broken down by the number of imputed variables (out of 229) for a household.

Note that regular mortgage payments and mortgage insurance premiums are included under the shelter costs and thus under the total expenditure. Starting with 2002, these two variables were added in the calculation of imputation rates shown in Table 5.1. The impact of this modification is a higher overall imputation rate.

Table 5.1
Households requiring expenditure imputation by province and territory

Province or territory	Households (%) requiring imputation for expenditure variables ¹ (excluding clothing expenditures and expenditures in the section on Personal Taxes, Security and Money Gifts)			
	Number of variables imputed (out of 229)			Total
	1	2	3 or more	
Canada	9.9	1.8	1.1	12.8
Newfoundland and Labrador	15.5	2.0	0.7	18.2
Prince Edward Island	16.0	1.5	0.0	17.6
Nova Scotia	13.2	2.5	1.8	17.5
New Brunswick	8.4	1.2	0.7	10.3
Quebec	5.8	0.5	0.2	6.6
Ontario	11.3	3.2	3.0	17.5
Manitoba	8.9	1.6	0.4	10.9
Saskatchewan	8.7	1.5	0.7	10.9
Alberta	9.3	1.3	0.7	11.3
British Columbia	9.9	3.2	2.5	15.7
Yukon	13.0	1.3	0.7	15.0
Northwest Territories	12.1	3.2	0.3	15.6
Nunavut	7.3	1.4	0.5	9.2

1. Includes regular mortgage payments and mortgage insurance premiums

Table 5.1 indicates that 12.8% of households required some expenditure imputation (excluding the Clothing section and the Personal Taxes, Security and Money Gifts section) at the national level, but nearly 77% of them had only one variable imputed. The overall imputation rate is 8.3% (data not shown) when excluding regular mortgage payments and mortgage insurance premiums, which is similar to rates obtained in previous years. There are 5.3% of households (or 15.6% of households when selecting only households that reported mortgages on dwellings that they owned and occupied) that need an imputation for the mortgage insurance premiums.

There are very few households that had more than one variable imputed (2.9%). At the provincial or territorial level, Quebec (6.6%) and Nunavut (9.2%) have the lowest imputation rates. In fact, this province and that territory are the only ones with an imputation rate smaller than 10%. The highest rates are observed in Newfoundland and Labrador (18.2%), Prince Edward Island (17.6%), Nova Scotia (17.5%) and Ontario (17.5%). Ontario and British Columbia have the highest rates among households that required imputation for more than one variable. In those two provinces, about one-third of the households that required imputation had two or more variables imputed. The low percentage of households (particularly when excluding regular mortgage payments and mortgage insurance premiums) for which some variables had to be imputed, combined with the low number of variables that had to be imputed when imputation was necessary, suggests that the imputed values should not have a strong impact on the estimates.

5.1.2 Person expenditure and income imputation by province and territory

Since some respondents provide only totals for clothing expenditure and income variables, a two-step procedure is used to impute these variables (at the individual level). Individuals who require imputation of only certain components are imputed first. Then, they are followed by those for which totals are available but imputation on all components is required (see reference [1] for a more detailed description of this process).

The percentage of usable individuals (persons who are members of usable households) requiring imputation for an income variable are presented by province and territory in Table 5.2. The percentage of persons who had exactly one variable imputed, those who had two or more variables (but not all) imputed and the percentage of persons for which only total income was available (and hence required having all their components imputed) are shown. The total percentage of persons requiring some form of income imputation is also provided. The second to last column of Table 5.2 indicates the total percentage of persons requiring some form of imputation, except for clothing expenditure variables. The last column of Table 5.2 indicates the total percentage of persons requiring some form of imputation for the Personal Taxes, Security and Money Gifts section of the questionnaire.

Note that questions related to personal income, personal taxes, security and money gifts are asked to each household member aged 15 years old or over on December 31 of the reference year. Thus, this year, the percentage of persons requiring some form of imputation for income variables as well

as for the Personal Taxes, Security and Money Gifts section was calculated using only the persons aged 15 years old or over and was not based on all persons as done in previous years. The impact of this modification resulted in an imputation rate slightly higher for those variables. As done in previous years, the percentage of persons requiring imputation for clothing expenditure variables is based on all persons since those expenditure questions are asked for each household member.

Table 5.2
Persons requiring income imputation, persons requiring clothing expenditure imputation and persons requiring imputation for variables in personal taxes, security and money gifts section by province and territory

Province or territory	Percentage of persons requiring imputation for income variables				Percentage of persons requiring imputation for at least one of the 11 clothing expenditure variables	Percentage of persons requiring imputation for at least one of the 15 variables in the section on personal taxes, security and money gifts
	1 income variable imputed	2 or more income variables imputed (not all)	All income variables imputed (total income known)	Total (any form of income imputation)		
Canada	1.0	0.3	3.3	4.7	18.6	3.2
Newfoundland and Labrador	0.5	0.1	2.9	3.7	7.4	1.9
Prince Edward Island	0.1	0.2	8.1	8.6	12.5	1.0
Nova Scotia	0.6	0.1	3.9	4.7	15.2	5.1
New Brunswick	0.5	0.2	6.3	7.1	14.8	2.4
Quebec	0.6	0.1	1.9	2.6	19.9	1.2
Ontario	2.3	0.7	3.2	6.2	16.3	6.8
Manitoba	1.6	0.4	2.1	4.2	19.1	2.4
Saskatchewan	0.4	0.6	3.0	4.2	24.5	3.0
Alberta	0.3	0.3	3.1	3.8	20.9	2.3
British Columbia	2.4	0.4	4.1	7.0	20.0	5.6
Yukon	0.5	0.4	4.4	5.3	40.9	4.2
Northwest Territories	0.7	0.0	1.3	2.0	34.1	1.3
Nunavut	0.0	0.0	0.8	0.8	26.3	0.8

These results show that nearly 5% of persons from usable households had some imputation performed on at least one income variable. For 70% of them, the respondent gave the total income but all their components had to be imputed. For many of the remaining persons requiring imputation, only one component of

income (one variable) had to be imputed. Note that this year, the imputation rates for income variables are based on persons aged 15 years old and over. The overall imputation rate is 3.8% (data not shown) when computed over all persons, which is comparable to the 4.1% rate obtained in 2002. Provincially, the percentages of persons requiring some imputation on at least one income variable range from a low of 0.8% for Nunavut to a high of 8.6% for Prince Edward Island. The highest imputation rates are observed in Prince Edward Island (8.6%), New Brunswick (7.1%), British Columbia (7.0%) and Ontario (6.2%). In Prince Edward Island and New Brunswick, almost all income imputation is attributed to cases where totals are available but imputation on all components is required. Conversely, in Ontario and British Columbia, the income imputation is mainly attributable to cases where only one component had to be imputed and cases where totals are available but imputation on all components is required.

From the second to last column of the table, it can be seen that about 19% of persons required imputation for at least one of the clothing expenditure variables. The rates at the provincial and territorial levels range from 7.4% for Newfoundland and Labrador to 40.9% in the Yukon. Almost all these people provided their total expenditure on clothing but required imputation of the components. Newfoundland and Labrador is the only province with imputation rate smaller than 10%. The imputation rates are lower in the Atlantic provinces and conversely, higher in the Territories. The higher level of imputation required on clothing expenditure components implies that the estimates for these components could be greatly affected by imputation, while the effect on the estimates for total clothing expenditures will be negligible.

From the last column of the table, results show that about 3% of persons had some imputation performed on at least one variable in the Personal Taxes, Security and Money Gifts section. Note that this year, the imputation rate is based on persons aged 15 years old and over. The overall imputation rate is 2.6% (data not shown) when computed over all persons, a rate identical to the one obtained in 2002. Provincially, these percentages are also low, ranging from a low of 1.2% for Prince Edward Island to a high of 6.8% for Ontario. The provinces of Nova Scotia, Ontario and British Columbia have an imputation rate greater than 5%. In those three provinces, the higher imputation rate is due to the variable corresponding to the income tax paid on the 2003 income that required in proportion more imputation than for other provinces or territories.

5.1.3 Imputation of categorical variables by province and territory

The percentage of usable households that required imputation for at least one of the categorical variables is presented in Table 5.3. The table is broken down by the number of imputed variables (out of 56) for a household. Categorical variables that required imputation can be found in the following sections of the questionnaire: Dwelling Characteristics (with the exception of the dwelling type variable); Facilities Associated with the Dwelling; Tenure (with the exception for variables related to a tenure change during the reference year); Tobacco and Miscellaneous for variables pertaining to purchases through direct sales (yes/no questions). Note that other categorical variables from the questionnaire, such as the household composition variables or questionnaire skips, are edited and

validated by subject matter personnel from the Income Statistics Division. Therefore, those latter variables are not imputed using the nearest neighbour technique.

Table 5.3
Households requiring imputation of categorical variables by province or territory

Province or territory	Households (%) requiring imputation of categorical variables			
	Number of variables imputed (out of 56)			Total
	1	2	3 or more	
Canada	4.7	0.6	1.2	6.4
Newfoundland and Labrador	3.1	0.1	0.9	4.2
Prince Edward Island	2.1	0.3	0.3	2.8
Nova Scotia	3.0	0.2	0.7	3.9
New Brunswick	2.5	0.2	0.8	3.5
Quebec	3.5	0.5	0.6	4.6
Ontario	6.5	1.0	2.4	9.9
Manitoba	4.7	0.6	1.5	6.8
Saskatchewan	7.5	1.3	1.9	10.7
Alberta	5.8	0.2	1.0	7.0
British Columbia	6.7	1.0	2.1	9.8
Yukon	6.2	0.3	1.6	8.1
Northwest Territories	3.5	0.3	0.3	4.1
Nunavut	2.8	0.9	0.5	4.1

Table 5.3 indicates that 6.4% of households required some categorical imputation for dwelling characteristics, facilities associated with the dwelling, tenure and purchases through direct sales at the national level, but nearly 75% of them had only one variable imputed. At the provincial or territorial level, Prince Edward Island (2.8%), New Brunswick (3.5%) and Nova Scotia (3.9%) have the lowest imputation rates of categorical variables. The highest rates are observed in Saskatchewan (10.7%), British Columbia (9.9%) and in Ontario (9.8%).

References

- [1] Tremblay, J. and Arsenault, S. 2001. *Methodology of the Survey of Household Spending*. Catalogue no. 62F0026MIE2001003. Ottawa. Household Survey Methods Division, Statistics Canada.
- [2] Rao, J.N.K., Wu, C.F.J., Yue, K. 1992. Some recent work on resampling methods for complex surveys. *Survey Methodology* (Statistics Canada. Catalogue no. 12-001); 18(2):209-217.
- [3] Statistics Canada, Income Statistics Division. 2003. *User Guide—Survey of Household Spending*. Catalogue no. 62F0026MIE2004003. Ottawa.
- [4] Statistics Canada, Income Statistics Division. 2003. *Spending Patterns in Canada*. Catalogue no. 62-202. Ottawa.
- [5] Statistics Canada, Household Survey Methods Division. 2003. *Labour Force Survey, Operations Report*. Survey 200312. Ottawa.
- [6] Lapierre, B., Nadeau, C., Tremblay, J. et Gaudet, J. 2004. Improving the quality of estimates for a low-income population: use of a dual frame in the Survey of Household Spending, *Proceedings of Statistics Canada Symposium 2004* (to appear).

Appendix A

Algebraic notation

1. Nonresponse adjustment

The subweight (i.e. the design weight adjusted for nonresponse) for a household k , denoted as w_k^{NR} , is

$$w_k^{NR} = \pi_k^{-1} * \frac{1}{rate_g} \quad \text{with} \quad rate_g = \frac{\sum_{k \in s_{g,r}} \pi_k^{-1}}{\sum_{k \in s_{g,r}} \pi_k^{-1} + \sum_{k \in s_{g,nr}} \pi_k^{-1}}$$

where

$s_{g,r}$ is the set of respondents in nonresponse group g ,

$s_{g,nr}$ is the set of nonrespondents (refusals, no contacts, unusable data) in nonresponse group g , and

π_k^{-1} is the design weight assigned to household k .

2. Calculation of the slippage rate

The slippage rate for a control group c , denoted as $rate_c$, is

$$rate_c = 100 * \frac{\left(\sum_{k \in s_{c,r}} w_k^{NR} \right) - t_c}{t_c}$$

where

$s_{c,r}$ is the set of respondents in control group c ,

w_k^{NR} is the subweight of household k , and

t_c is the total of the auxiliary data for the control group c .