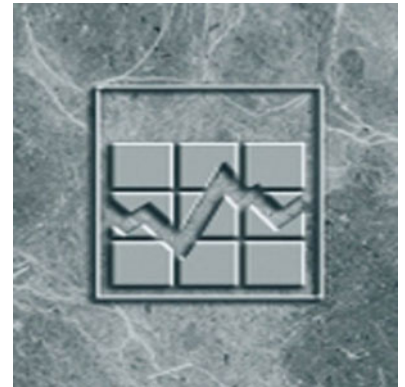**Research Paper**

Household expenditures research paper series

Survey of Household Spending 2004: Data Quality Indicators

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

Telephone: 613-951-7355

Statistics Canada   Statistique Canada

Canadä

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website at www.statcan.gc.ca, e-mail us at infostats@statcan.gc.ca, or telephone us, Monday to Friday from 8:30 a.m. to 4:30 p.m., at the following numbers:

Statistics Canada's National Contact Centre
Toll-free telephone (Canada and United States):
   Inquiries line   1-800-263-1136
   National telecommunications device for the hearing impaired   1-800-363-7629
   Fax line   1-877-287-4369

Local or international calls:
   Inquiries line   1-613-951-8116
   Fax line   1-613-951-0581

Depository Services Program
   Inquiries line   1-800-635-7943
   Fax line   1-800-565-7757

To access this product

This product, Catalogue no. 62F0026M, is available free in electronic format. To obtain a single issue, visit our website at www.statcan.gc.ca and browse by "Key resource" > "Publications."

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. 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.gc.ca under "About us" > "The agency" > "Providing services to Canadians."

Survey of Household Spending 2004: Data Quality Indicators

Note of appreciation

Symbols

The following standard symbols are used in Statistics Canada publications:

.   not available for any reference period

..   not available for a specific reference period

…   not applicable

0   true zero or a value rounded to zero

0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded

[p]   preliminary

[r]   revised

x   suppressed to meet the confidentiality requirements of the *Statistics Act*

[E]   use with caution

F   too unreliable to be published

Table of Contents

Highlights

Sampling errors

➢ The coefficients of variation (CVs) of the average estimates of total expenditure per household vary between 1.3% and 2.7% for the provinces. The CV at the national level is 0.7%.

➢ 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 1.9% at the national level and are generally lower than 5.0% at the provincial level. The results for the different categories of dwelling and household equipment characteristics are similar. Since the sample size is smaller in Prince Edward Island, the CVs tend to be higher than those of the other provinces.

Nonresponse

➢ The final response rate is 69.2%. Provincial response rates range from 58.7% in Ontario to 78.1% in Saskatchewan.

➢ The nonresponse rate is 30.8%. Nonresponse is due to refusals (18.4%), to households that could not be contacted (8.6%), and to households with data that were considered unusable (3.8%).

➢ The nonresponse rate tends to increase with the urbanization level. The nonresponse rate is 24.3% in rural areas and 33.7% in urban centres with a population of one million or more.

➢ According to an analysis of nonresponse rates in the strata consisting of the high-income geographic areas created under the sample design, the rate in high-income strata (38.9%) is higher than the rate observed in other strata (29.9%). The refusal rate for high-income strata (25.3%) is higher than the rate observed for other strata (17.6%).

Coverage errors

➢ Undercoverage of households is 8.3% at the national level. Undercoverage of households is observed for all provinces, with rates varying from 4.1% to 12.1%. British Columbia has the highest undercoverage of households.

➢ Undercoverage of persons is 9.3% at the national level. Undercoverage of persons is observed for all provinces and territories, with rates varying from 4.1% to 14.2%. British Columbia has the highest undercoverage of persons.

➢ The national slippage rate for children (aged 0 to 17) is very different from the rates for other age groups. For children, the undercoverage rate is 0.1%, whereas for adults, undercoverage is much higher at 11.9%.

Response errors

➢ Response errors include 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 data quality. Total interview time varies according to household characteristics, and it can be substantial for some households. The average interview time was one hour and forty-five minutes.

Processing errors related to imputation

i) Expenditure variables

➢ It was necessary to impute at least one expenditure variable[1] for 31.1% of households nationally. This higher rate in 2004 is due to the change made to the questionnaire regarding expenditures related to communication services in the home (telephone, cell phone and Internet access), cable television services and satellite distribution services. For these services, it is now accepted that a respondent household can provide only total expenditure for a package (bundled services), indicating which services are included. As a result, 21.5% of usable households required imputation for at least one of these five services.

➢ The overall imputation rate excluding these five services is 13.2% at the national level, which is comparable to the rates obtained in previous years. For these expenditures, most households required imputation for one or two of the 240 expenditure variables. Provincially, the imputation rates range from a low of 6.2% for Quebec to a high of 18.8% for Nova Scotia.

➢ Imputed values account for 20.4% of the estimate of total expenditure on cable television services and 32.8% of the estimate of total expenditure on Internet access services. Imputation rates for households reporting expenditures on one or the other of these two services are respectively 20.5% and 40.3%. These higher rates are mainly due to the fact that among households that reported paying for a package, a large proportion of packages included these two services.

➢ About 20% of individuals required imputation for clothing variables, but in the vast majority of such cases, the respondents provided the totals and only the components were imputed.

➢ About 4% of individuals aged 15 and over required imputation on at least one variable in the Personal Taxes, Security and Money Gifts section.

---

1. Excluding the Clothing section and Personal Taxes, Security and Money Gifts section, which are collected at the individual level and not the household level.

ii) Income variables

➢ Imputation for at least one income variable was required for 5% of individuals aged 15 and over. For 80% of them, total income was provided by the respondent but all components (income sources) were imputed.

iii) Categorical variables

➢ For 7.6% of households, it was necessary to impute at least one categorical variable. Approximately 72% of them had only one variable imputed. Provincially, imputation rates range from a low of 2.2% for New Brunswick to a high of 12.9% for Alberta. Categorical variables that required imputation can be found in the following sections of the questionnaire: Dwelling Characteristics; Facilities Associated with the Dwelling; Tenure; and 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 2004 SHS sample consists of 20,446 households[2] distributed throughout the ten provinces. 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 2004 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].

2.The initial sample is made up of 23,897 dwellings. From these dwellings, it is necessary to identify and exclude ineligible dwellings (see Section 2.1) to obtain the 20,446 households from which data on income and expenditure are collected.

---

1   Sampling errors

Sampling errors exist when inferences about the population are drawn from the survey using information collected from a sample, rather than from 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 the 2003 reference year, the jackknife resampling method was used to produce standard error estimation for the 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 median expenditure 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].

The 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) x 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 ± 2 SE).[3] 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 ± 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 expenditure, as well as the CVs of dwelling characteristics and household facilities

---

3. The confidence interval is calculated directly from the CV in similar fashion, namely Y ± 2 (CVx Y)/100.

---

and equipment, are available at the national and provincial levels in the publication *User Guide—Survey of Household Spending* (see reference [3]).

It should be noted that the estimated CVs do not take account of 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 of the true CVs. However, to assess the reliability of detailed expenditures with a high imputation rate, the CV and the imputation rate should both be considered.

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

Table 1.1
Coefficients of variation by province 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.
						%					
Total expenditure	0.7	1.4	2.7	1.7	1.3	1.3	1.5	1.5	1.4	1.3	1.6
Total current consumption	0.7	1.4	2.4	1.3	1.3	1.2	1.5	1.3	1.5	1.4	1.4
Food	0.6	1.3	2.4	1.3	1.3	1.1	1.3	1.4	1.2	1.4	1.2
Shelter	1.0	2.1	4.3	2.4	1.9	1.6	1.9	1.8	1.7	1.6	2.2
Household operation	1.0	1.7	2.8	2.2	2.3	1.8	2.0	2.4	1.9	2.1	2.3
Furnishings	2.9	4.2	8.6	3.5	3.9	4.0	5.9	4.2	5.0	3.5	3.6
Clothing	1.3	2.1	3.7	2.9	3.1	2.1	2.7	2.9	2.6	2.7	2.9
Transportation	1.6	3.8	5.0	2.8	3.4	2.8	3.1	3.2	3.4	4.1	3.2
Health care	1.5	3.1	3.9	3.3	3.1	2.8	3.6	3.4	3.6	2.5	3.2
Personal care	1.3	1.9	3.5	2.3	2.8	2.7	2.5	2.8	2.5	3.1	2.6
Recreation	1.9	4.1	8.0	3.7	3.2	4.2	3.8	4.5	3.9	3.8	3.3
Reading & printed materials	1.8	3.7	6.5	4.2	4.5	3.1	3.9	3.1	3.5	4.1	4.0
Education	3.7	7.8	9.4	8.0	8.6	6.0	6.8	7.4	8.3	7.6	7.5
Tobacco, alcoholic beverages	2.0	4.2	3.9	4.0	4.8	3.5	4.2	4.7	4.6	3.7	3.9
Games of chance (net)	5.1	7.3	18.4	8.8	7.6	15.4	7.4	8.1	7.5	9.9	6.4
Miscellaneous expenditures	3.5	7.0	17.1	6.0	5.4	5.6	7.0	4.6	5.5	9.9	6.0
Personal income tax	1.2	2.6	5.4	5.2	3.0	2.3	2.2	2.8	2.5	3.0	3.6
Personal insurance and pension contributions	1.4	2.7	6.9	2.4	2.9	1.6	2.9	4.5	1.8	3.6	4.6
Gifts and contributions	4.8	10.7	20.2	10.7	8.0	18.9	7.4	10.8	8.2	11.9	11.4
Income	0.7	1.2	2.5	1.2	1.5	1.2	1.4	1.3	1.5	2.0	1.6

The coefficients of variation of the average estimates of total expenditure per household vary between 1.3% and 2.7% for the provinces, and the national figure is 0.7%.

For summary-level expenditure categories, the CVs at the national level are less than or equal to 1.9%, except for the following categories: furnishings, education,

games of chance, miscellaneous expenditures and gifts of money and contributions. These expenditure categories represent respectively 2.9%, 1.7%, 0.4%, 1.6% and 2.6% of total expenditure. Also, with the exception of these categories, the CVs are generally less than or equal to 5% at the provincial level. Since the sample size was smaller in Prince Edward Island, the CVs tend to be higher than those of other provinces.

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

Table 1.2
Coefficients of variation (%) by province 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.
						%					
Owner	1.0	1.6	3.0	2.4	2.0	2.2	2.0	2.7	1.8	2.2	2.0
Renter	2.0	5.3	8.0	5.8	6.7	2.9	4.2	6.0	5.1	5.7	4.1
Washing machine	0.7	1.1	1.9	1.7	1.1	1.1	1.7	2.1	1.1	1.7	1.8
Clothes dryer	0.8	1.3	2.1	1.7	1.3	1.2	1.8	2.0	1.2	1.6	1.9
Dishwasher	1.3	3.3	5.1	3.6	2.9	2.4	2.7	3.2	2.7	2.3	2.8
Freezer	1.1	1.9	3.6	2.2	1.6	2.7	2.3	1.9	1.6	1.9	3.0
Microwave oven	0.4	0.6	1.0	0.7	0.6	0.7	0.7	0.6	0.5	0.8	0.9
Cellular phone	1.0	2.2	4.7	2.4	2.5	2.4	1.7	2.5	2.4	2.1	2.0
CD player	0.7	1.6	2.2	1.3	1.4	1.4	1.3	1.6	1.5	1.3	1.3
Cable TV	1.1	2.6	5.9	3.1	3.2	2.2	2.0	2.5	3.8	3.1	2.0
Satellite dish	2.7	6.8	7.8	6.3	5.5	4.3	6.0	7.0	5.4	5.8	9.1
DVD player	0.8	2.0	3.5	1.9	2.3	1.7	1.6	1.8	1.8	1.6	1.8
Home computer	0.8	2.4	2.9	1.9	2.4	1.7	1.6	1.8	1.9	1.6	1.6
Regular tel. connection to a computer (Modem)	3.3	7.2	6.8	9.3	5.9	5.6	6.4	11.8	9.9	10.7	9.4
High-Speed tel. connection to a computer	2.7	9.5	10.5	6.2	7.5	5.9	5.1	6.3	5.9	7.8	6.2
Cable connection to a computer	2.5	8.6	24.6	9.2	15.9	6.2	4.7	8.4	8.4	5.3	4.2
Use of internet (home)	1.0	3.0	3.7	2.5	3.1	2.1	1.9	2.4	2.3	1.8	1.9
Owned vehicles (one)	1.5	3.4	4.7	3.4	3.5	2.9	2.9	3.4	3.7	3.9	2.9
Owned vehicles (2 or more)	1.5	4.8	4.5	3.6	3.6	3.4	3.1	3.0	2.9	3.2	2.9

The coefficients of variation for dwelling characteristics and household equipment are generally below 4% at the provincial level, with some exceptions in the following categories: *renter, satellite dish, regular telephone connection to a computer (modem), high-speed telephone connection to a computer* and *cable connection to a computer.* Prince Edward Island is an exception. Since the sample size there is smaller, the CVs tend to be higher than those of the other provinces. Only Quebec has a CV below 4% for the renter category. It is also the province with the largest proportion of renters (43.4%). The CVs of expenditure in connection to a computer vary according to the type of connection. In Prince Edward Island, the CV (6.8%) is smaller for the regular telephone connection to a computer, with 30.0% of PEI households reporting that they have this type of connection. On the other hand, in

British Columbia the CV (4.2%) is smaller for cable connection to a computer, with 33.6% of households in that province reporting this type of connection.

The CVs for dwelling characteristics and household equipment at the national level are less than or equal to 2.0% with the exception of the following categories: *satellite dish, regular telephone connection, high-speed telephone connection to a computer* and *cable connection to a computer*. There is a smaller proportion of households with such equipment for these four categories. At the national level, the proportions are respectively 22.0%, 15.7%, 21.3% and 22.0%.

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 reference [3] or [4].) For operational reasons, it is not possible to produce CVs for all the different levels of aggregation that may interest users.

1.3.1   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 who 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 a household type, an income quintile, an 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 these values at the provincial level, since provincial CVs are published for the 2004 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 Approximation of the CV from the microdata file

Microdata file users can obtain an approximation of the CV of the estimates using another method that will generally provide better results than the method described in the previous section for the CVs of detailed expenditure categories. This approach is described in detail in the documentation provided with the 2004 microdata file. This method of approximation can be used only with the microdata file, since it requires having 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.[4]

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.

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

---

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

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 estimate and are considered as nonresponse.

Note that all rates provided in this section are unweighted. For the 2004 Survey of Household Spending, the final response rate is 69.2%. Table 2.1-1 shows the final response rate as well as the sample size (eligible households) broken down by refusals, no-contacts, unusable data and usable data. This rate is provided at both the national and the provincial level.

Table 2.1-1
Sample size and response rate by province and at the national level

Province	Eligible households	Non-contacts	Refusals	Unusables	Usables	Final response rate (at estimation stage)[1]
	number					%
Canada	20,446	1,751	3,759	782	14,154	69.2
Newfoundland and Labrador	1, 829	170	206	92	1,361	74.4
Prince Edward Island	778	67	98	24	589	75.7
Nova Scotia	2,080	204	401	151	1,324	63.7
New Brunswick	1, 861	182	316	114	1,249	67.1
Quebec	2, 645	172	526	32	1,915	72.4
Ontario	3, 033	315	758	181	1,779	58.7
Manitoba	1, 922	174	295	38	1,415	73.6
Saskatchewan	1, 807	105	264	27	1,411	78.1
Alberta	2, 038	166	385	30	1,457	71.5
British Columbia	2, 453	196	510	93	1,654	67.4

1. Usable/eligible x 100

Table 2.1-2 shows the final nonresponse rate; the collection nonresponse rate, broken down by refusals and no-contacts; and the rate households with unusable data, broken down into incomplete and out-of-balance questionnaires. The vacancy rate is also included. These rates are provided at the national and provincial 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 2.1-2
Nonresponse and vacancy rates by province and at the national level

Province	Vacancy rate	Collection nonresponse rate			Unusable data rate			Final nonresponse rate (at estimation stage)
		Total	No contact	Refusal	Total	Incomplete	Out-of-balance	
	%							
Canada	11.9	26.9	8.6	18.4	3.8	1.5	2.3	30.8
N.L.	14.4	20.6	9.3	11.3	5.0	0.9	4.1	25.6
P.E.I.	19.2	21.2	8.6	12.6	3.1	1.0	2.1	24.3
N.S.	12.9	29.1	9.8	19.3	7.3	1.0	6.3	36.3
N.B.	14.1	26.8	9.8	17.0	6.1	1.2	4.9	32.9
Que.	9.4	26.4	6.5	19.9	1.2	1.2	0.0	27.6
On.	9.2	35.4	10.4	25.0	6.0	2.4	3.6	41.3
Man.	8.4	24.4	9.1	15.3	2.0	1.1	0.8	26.4
Sask.	14.0	20.4	5.8	14.6	1.5	1.2	0.3	21.9
Alta.	9.9	27.0	8.1	18.9	1.5	1.3	0.1	28.5
B.C.	12.9	28.8	8.0	20.8	3.8	2.7	1.1	32.6

16

The final nonresponse rate in Canada is 30.8%. It is due to refusals (18.4%), to households that could not be contacted (8.6%), and finally to households for which the data were unusable (3.8%). For each province, 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 final nonresponse rate varies from one province to another. Saskatchewan has the lowest nonresponse rate at 21.9%. This is also the province with the lowest no-contact rate (5.8%). As may be seen, in Quebec there were no out-of-balance questionnaires. The nonresponse rates in Saskatchewan and Prince Edward Island are less than 25%, while the rates in Nova Scotia and Ontario are greater than 35%. The higher final nonresponse rate in Nova Scotia is partly attributable to a higher rate of out-of-balance questionnaires (6.3%). Ontario has an especially high nonresponse rate at 41.3%. The highest rates of no contact (10.4%) and refusal (25.0%) are also observed in Ontario.

Vacancy rates are shown in Table 2.1-2, 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 2004 SHS is 11.9%.

2.2   Nonresponse according to urbanization level

Nonresponse varies according to urbanization level. The various rates at the national scale are shown by urbanization level in Table 2.2.[5]

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	8.3	30.4	8.0	22.3	3.3	1.9	1.5	33.7
500,000 to 999,999	7.2	28.2	8.8	19.4	1.5	1.0	0.5	29.8
250,000 to 499,999	8.6	36.2	11.4	24.8	7.2	2.2	4.9	43.3
100,000 to 249,999	10.2	27.8	9.3	18.4	5.8	1.8	3.9	33.5
30,000 to 99,999	10.1	28.1	10.1	17.9	3.5	1.2	2.3	31.6
Less than 30,000	11.7	22.3	7.3	15.0	3.0	1.3	1.7	25.3
Rural	22.2	20.4	7.1	13.3	3.9	1.2	2.6	24.3
Total	11.9	26.9	8.6	18.4	3.8	1.5	2.3	30.8

The final nonresponse rate generally increases with urbanization level. The urbanization category "250,000 to 499,999" has the highest rates of units with no contact (11.4%), refusal (24.8%) and unusable data (7.2%). The final nonresponse

---

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

rate of 43.3% for this urbanization category is due both to the nonresponse obtained during collection (36.2%), which is mostly attributable to the rate observed in Ontario (42.1%, data not shown), and to the rate of unusable data (7.2%), which is mostly attributable to the rate observed in Nova Scotia (8.5%, data not shown).

The collection nonresponse rate also tends to increase with urbanization level. There is a difference of nearly 8% between the urbanization categories "Less than 30,000" and "1,000,000 or more." Refusals account for more than 50% of total nonresponse at each level of urbanization.

From an examination of the vacancy rate by urbanization level, it emerges that the vacancy rate in rural areas (22.3%) is nearly twice that for low-population urban areas (11.7%). These low-population urban areas also have a higher vacancy rate than higher-population urban areas. This phenomenon is also observed in the Labour Force Survey (LFS) and is probably attributable to a greater number of seasonal dwellings in rural areas. This explains, among other things, why the vacancy rate is higher in the Atlantic provinces, as illustrated in Table 2.1-2, and especially in Prince Edward Island, since that province has a higher proportion of rural dwellings in the sample.

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 constructed in such a way that in census metropolitan areas, there are strata consisting of geographic areas with a high concentration of high-income households. While the number of high-income strata remains relatively small (51 out of a total of 1,060 strata), the comparison of response rates in this group in relation to the other strata provides relevant information on the potential effect of nonresponse.

The 2004 SHS sample was selected using the new sample design for the LFS. Owing to changes that were made to stratification, especially with respect to income-based strata, and to the fact that strata were redefined according to data from the 2001 Census, the response rates by income strata in the 2004 SHS are not directly comparable with those of previous surveys.

Under the new design, the number of high-income strata has increased. These strata consist of geographic areas with a high concentration of households with an income exceeding $125,000. They account for approximately 5% of all households in Canada. For the 2004 SHS sample, 9.9% of eligible households came from these strata, because of oversampling to obtain better representation of high-income households in the sample.

Also, under the new LFS sample design, there is no longer an apartment frame, and therefore there is no longer a frame of low-income apartment buildings such as was previously used to form low-income strata. Apartment buildings are now incorporated into the regular LFS design. Thus, there are no longer any separate low-income strata, as was the case for previous years.

Table 2.3 shows the nonresponse and vacancy rates for high-income strata in relation to other strata. Note that in addition to regular strata, the "Other" strata

category includes four types of strata that were added to the new LFS sample design. These are: strata with a high vacancy rate, high-cost strata, strata with a concentration of immigrants, and strata with a concentration of Aboriginals. Since the portion of the SHS sample allocated to the latter four strata was smaller, the results for them are not broken out in Table 2.3.

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

Stratum type based on income	Vacancy rate	Collection nonresponse rate			Unusable data rate			Total nonresponse rate (at estimation stage)
		Total	No contact	Refusal	Total	Incomplete	Out-Of-balance	
		%						
High-income	5.8	34.8	9.5	25.3	4.1	1.8	2.3	38.9
Others	12.5	26.1	8.5	17.6	3.8	1.5	2.3	29.9
Total	11.9	26.9	8.6	18.4	3.8	1.5	2.3	30.8

In high-income strata, the final nonresponse rate (38.9%) is approximately 30% higher than in the other strata. The refusal rate for high-income strata is 25.3%, which is higher than for the other strata. However, high-income strata and the other strata have similar rates of unusable data.

As may be seen, the vacancy rate is lower for high-income strata than for the other strata. This phenomenon was also observed for previous surveys.

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 predefined groups. Following the overhaul of the LFS sample design, the nonresponse adjustment groups were redefined. As was the case for previous years, these groups are defined on the basis of the different urbanization levels in each province and of sub-provincial geographic areas for Quebec, Ontario and British Columbia. Also, specific nonresponse adjustment groups were created for high-income strata. As was seen in the previous section, the number of high-income strata was increased with the introduction of the new LFS sample design. Thus, it is now possible to form nonresponse adjustment groups for high-income strata in all provinces except Prince Edward Island. Such a group cannot be formed in Prince Edward Island because that province has no high-income strata.

The weighted rates differ from the rates presented in this section, since the former take the sampling weight of each household into account. An algebraic description of the adjustment for nonresponse is provided in Appendix A.

The adjustment of weights for nonresponse serves to take account of differences in nonresponse by urbanization level (as illustrated in Section 2.2) and geographic area or by groups of high-income strata. It will serve to reduce the bias insofar as the

characteristics of respondents and non-respondents are similar for a given urbanization level and geographic area or for a given group of high-income strata.

It should be noted that a nonresponse adjustment group can be combined with another group if the number of households is too small or the adjustment factor is too high.

3 Coverage errors

The target population was defined in the design of the survey. 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. It is important to note that the SHS uses the sampling frame of the Labour Force Survey (LFS).

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 ten provinces. The Territories are excluded from the target population for the 2004 SHS, since the survey covers this region only every second year.

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

Coverage errors result from inadequate representation of the target population based on the units in 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 occur during data collection, due to improper identification of persons as members of the selected household. These errors also contribute to undercoverage or overcoverage.

Also, as described in Section 2.4, reweighting methods are implemented to take account of nonresponse. However, when these adjustments are made, it is impossible to correct the survey weights to ensure that all subgroups within the population are well represented.

A good representation of the target population is essential to the production of realistic expenditure estimates. The sample must adequately represent the individuals in the target population as well as the distribution of households according to their size.

There is generally a net undercoverage of the number of persons and the number of households in the SHS. This undercoverage is corrected by an adjustment of weights using auxiliary or reference data based on post-censal demographic estimates. The slippage rate (see Appendix A) is a measure of the percentage of difference between the estimates from these auxiliary data and the survey estimates calculated using weights not adjusted with these data.[6] Slippage therefore represents the combined effect of undercoverage and unbalance in certain subgroups of the population created by survey nonresponse that could not be corrected at the reweighting stage.

For the 2004 survey, several changes were made to the approach with a view to adjusting the weights by using auxiliary data. First, post-censal demographic estimates now come from the 2001 Census rather than the 1996 Census; this makes the 2004 slippage rates hard to compare with those from previous surveys. Also, changes were made to the weights adjustment strategy [6]. For example, in previous surveys, demographic adjustments were made in cross-tabulations of nine age groups by sex, whereas the new strategy is limited to eight age groups. Analyses showed that sex does not appear to be associated with household expenditures, since households are generally made up of persons of both sexes. Therefore, adjustments according to sex do not serve to improve the quality of the estimates. However, it is clear that sex has an effect on expenditures for one-person households. But since there are no annual demographic statistics on the number of one-person households, broken down by sex, it is not possible to make adjustments for this particular case.

Slippage rates by age group and sex at the national level

Slippage rates by age group and sex at the national level are shown in Table 3.1-1. A positive rate indicates overcoverage of the number of persons in the survey.

---

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

Table 3.1-1
National slippage rates by age-sex group, Canada

Age	Sex		Total
	Male	Female	
	%		
0-6 years	-2.9	2.8	-0.1
7-17 years	-1.4	1.3	-0.1
18-24 years	-14.8	-10.7	-12.8
25-34 years	-17.1	-13.1	-15.1
35-44 years	-13.9	-7.3	-10.6
45-54 years	-10.3	-12.8	-11.6
55-64 years	-15.1	-12.3	-13.7
65 years and +	-9.0	-7.6	-8.2
Total	-10.8	-7.9	-9.3

For the 2004 SHS, the national undercoverage rate was 9.3%. The slippage rates for children (aged 0 to 6 and 7 to 17) are quite different from those for other age groups. The undercoverage rate for all children combined is 0.1%, while it is 11.9% for adults (data not shown). Also, for girls aged 0 to 17, a slight overcoverage is observed. For girls aged 0 to 6, this is due to the overcoverage obtained in Ontario (4.3%), Alberta (13.0%) and British Columbia (14.9%) for this age-sex group (see Table 3.2-1). Similarly, the overcoverage for girls aged 7 to 17 is due to the overcoverage obtained in Quebec (10.9%) and Alberta (16.7%) for this age group.

The highest national rates occurred among men aged 18 to 24, 25 to 34 and 55 to 64. Note that the undercoverage rate for women is consistently lower than the corresponding rate for men.

As mentioned previously, the SHS uses the LFS sampling frame. Over the same period, the national LFS undercoverage rate was 9.8% (reference [5]). This is slightly lower than the 11.6% SHS rate for those aged 15 and over (data not shown).

Since sampling weights are no longer adjusted according to the sex of the persons in respondent households, net undercoverage or overcoverage by age-sex group persists. This net under- or overcoverage is measured by the residual slippage rate. The residual slippage rate is a measure of the percentage of difference between estimates from the auxiliary data and the survey estimates, which this time are calculated using the final weights. Residual slippage rates by age-sex group at the national level are shown in Table 3.1-2.

Table 3.1-2
National residual slippage rates by age-sex group, Canada

Age	Sexe		Total
	Male	Female	
	%		
0-6 years	-1.8	1.9	0.0
7-17 years	-1.5	1.5	0.0
18-24 years	-0.8	0.9	0.0
25-34 years	0.0	0.0	0.0
35-44 years	-1.1	1.1	0.0
45-54 years	1.0	-1.0	0.0
55-64 years	-1.2	1.1	0.0
65 years and +	-0.1	0.1	0.0
Total	-0.6	0.6	0.0

Residual slippage rates for males and females are necessarily of opposite signs, since the weights were adjusted to correspond to the benchmark demographic estimates for each age group. Although sex is no longer taken into account in adjusting the weights, the fact remains that controlling for age greatly reduces the size of the undercoverage by age-sex group observed in Table 3.1-1. After adjusting the weights by age group, the residual slippage rates are less than 1.2% for adults and less than 1.9% for children at the national level.

Provincial slippage rates by age group and sex

Slippage rates by age group and sex at the provincial level are shown in Table 3.2-1.

Table 3.2-1
Slippage rates for provinces by age-sex group

	Slippage rates by age-sex group									
Age	Newfoundland and Labrador	Prince Edward Island	Nova Scotia	New Brunswick	Quebec	Ontario	Manitoba	Saskatchewan	Alberta	British Columbia
male										
0-6	-8.2	12.0	2.7	-13.1	-1.2	-7.9	-12.2	-20.9	18.2	-0.5
7-17	-9.0	11.0	-7.9	3.5	8.0	-5.0	-8.0	-14.7	8.4	-8.9
18-24	-34.4	-29.6	-26.3	-22.4	0.0	-11.7	-25.0	-22.9	-23.3	-30.0
25-34	-20.2	-22.7	-13.9	-20.2	-16.8	-17.9	-23.3	-25.3	-4.4	-22.9
35-44	-19.4	-10.9	-19.2	-12.5	-3.8	-18.9	-13.1	-24.5	-2.8	-22.2
45-54	-21.2	-4.5	-17.0	-4.5	-8.8	-6.7	-13.6	-14.6	-14.3	-16.8
55-64	-11.5	-11.4	-7.4	-11.0	-19.6	-14.0	-6.7	-11.0	-23.4	-10.5
65 +	-2.2	-21.6	0.1	-13.1	-3.7	-10.7	-3.2	-9.8	-18.1	-11.1
Total	-16.1	-9.3	-11.9	-10.9	-6.1	-11.9	-13.1	-17.8	-6.8	-16.1
female										
0-6	-16.0	0.7	-1.3	4.0	-7.1	4.3	-6.8	-5.8	13.0	14.9
7-17	-3.5	-8.5	1.6	-4.2	10.9	-2.2	-5.9	-5.5	16.7	-12.2
18-24	-22.0	-7.1	-18.9	-26.5	1.5	-12.0	-9.9	-15.1	-13.9	-18.5
25-34	-8.6	-3.4	-8.1	-15.0	-8.5	-21.2	-17.4	-21.3	4.7	-9.6
35-44	-10.1	-0.2	-9.1	-12.2	-1.3	-9.2	-14.5	-11.9	4.7	-16.5
45-54	-22.2	3.6	-15.4	-5.2	-10.9	-13.6	-16.2	-12.3	-7.1	-17.2
55-64	-19.3	-12.3	11.1	-0.7	-19.2	-9.5	-6.3	-0.8	-21.0	-12.2
65 +	7.1	-20.1	-2.3	-19.6	-9.0	-4.4	0.8	-1.1	-16.5	-12.9
Total	-11.5	-6.1	-5.5	-10.4	-5.6	-9.2	-9.7	-9.2	-1.4	-12.2
Total	-13.8	-7.7	-8.6	-10.6	-5.8	-10.5	-11.3	-13.5	-4.1	-14.2

We observe net undercoverage for all provinces, with the rates varying from 4.1% to 14.2%. Alberta has the lowest undercoverage rate at 4.1%. However, a low overall rate of undercoverage is not a guarantee of better coverage. For example, the overall slippage rate observed in Alberta (4.1%) conceals the worst case of overcoverage at the provincial level among age groups (18.2% for boys) and the worst cases of undercoverage for the 55-64 age group.

Among the provinces, the highest undercoverage rate occurred among 18 to 24 year-old men in Newfoundland and Labrador (34.4%). Also, 18 to 24 year-old men had the highest undercoverage rates in the other Atlantic provinces and in British Columbia. Although this age group was found to have the highest undercoverage rates, the highest rate at the national level instead occurred among men aged 25 to 34 (see Table 3.2-1). This is because in Quebec, among men aged 18 to 24, there was no coverage error (to be more precise, there was a slippage rate of -0.014%)

prior to adjustment of the weights according to age groups. Prince Edward Island had the highest slippage rates for the 65 and over age group. Another point worth noting is that the pattern of slippage rate variation differs substantially for age-sex groups from one province to the next.

Residual slippage rates by age-sex group at the provincial level are shown in Table 3.2-2.

Table 3.2-2
Residual slippage rates for provinces by age-sex group

Age	Newfoundland and Labrador	Prince Edward Island	Nova Scotia	New Brunswick	Quebec	Ontario	Manitoba	Saskatchewan	Alberta	British Columbia
					male					
0-6	0.9	5.8	1.9	-9.3	5.4	-5.3	-3.0	-10.3	4.1	-6.0
7-17	-4.8	10.2	-5.5	3.4	-2.0	-0.5	-1.3	-6.8	-5.3	1.0
18-24	-6.9	-11.6	-3.8	3.1	-0.8	2.5	-4.5	0.7	0.6	-10.1
25-34	-4.1	-6.5	-0.5	3.5	-3.8	4.3	-1.6	1.3	-2.3	-3.8
35-44	-4.9	-2.3	-4.2	2.7	2.7	-2.8	2.6	-2.4	-2.6	-1.4
45-54	1.2	-6.5	1.0	1.5	-0.7	4.3	1.2	-1.5	-5.5	0.7
55-64	5.8	-0.4	-8.3	-6.0	0.2	-2.2	0.4	-2.9	-0.7	0.5
65 +	-7.1	-0.5	1.3	4.1	5.3	-3.0	-1.4	-4.3	-1.9	1.1
Total	-2.5	-1.3	-2.5	1.0	0.4	0.0	-0.7	-3.1	-2.2	-1.7
					female					
0-6	-0.9	-6.0	-2.0	9.8	-5.6	5.5	3.1	10.7	-4.3	6.3
7-17	5.1	-10.5	5.8	-3.6	2.1	0.5	1.4	7.3	5.6	-1.1
18-24	7.0	11.7	3.9	-3.2	0.8	-2.6	4.7	-0.8	-0.6	10.5
25-34	3.9	5.9	0.5	-3.3	3.9	-4.3	1.6	-1.3	2.4	3.7
35-44	4.7	2.2	4.0	-2.7	-2.8	2.8	-2.6	2.4	2.7	1.3
45-54	-1.2	6.1	-0.9	-1.5	0.7	-4.2	-1.2	1.5	5.7	-0.7
55-64	-5.7	0.4	8.0	6.0	-0.2	2.1	-0.4	2.9	0.7	-0.5
65 +	5.8	0.4	-1.0	-3.2	-4.0	2.4	1.1	3.4	1.6	-0.9
Total	2.4	1.3	2.3	-0.9	-0.4	0.0	0.7	3.1	2.3	1.7
Total	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Just as for the national results, the adjustment of weights by age group tends to moderate the size of the under- or overcoverage by age group seen in Table 3.2-1. The highest residual slippage rates are on the order of 10% to 12%. These rates occurred among children aged 0 to 6, children aged 7 to 17 and adults aged 18 to 24. As noted above, analyses have shown that sex does not appear to be a factor in household expenditures except perhaps in the case of one-person households, and almost none of the persons in the latter three age groups constitute one-person households. Thus, while the residual slippage rate is higher for these age groups,

there would likely be no bias in the estimates owing to poorer representativeness in these age groups.

It should also be noted that the overall slippage rates for females and males in each province are relatively low.

Slippage rates by household size

Table 3.3 shows the slippage rates by household size that were used in adjusting the weights. A negative rate corresponds to undercoverage of the number of households in the survey.

Table 3.3
Slippage rates for provinces by household size

Province	Slippage rate			
	Households	One-person households	Two-person households	Three-person and more households
	%			
Canada	-8.3	-7.9	-9.0	-8.0
Newfoundland and Labrador	-11.7	-16.7	-5.8	-14.2
Prince Edward Island	-6.2	-4.0	-4.6	-8.9
Nova Scotia	-6.3	-2.2	-4.6	-10.9
New Brunswick	-10.6	-15.9	-11.0	-6.8
Quebec	-4.9	-3.9	-4.4	-6.1
Ontario	-10.1	-8.7	-13.7	-8.2
Manitoba	-9.1	-5.0	-7.4	-13.9
Saskatchewan	-9.5	-9.3	-2.3	-16.5
Alberta	-4.1	-6.9	-4.5	-2.2
British Columbia	-12.1	-14.4	-11.6	-10.7

Nationally, the number of households was underestimated by 8.3%. This underestimation is slightly lower than the 9.3% underestimation observed for the number of individuals. Both nationally and provincially, undercoverage is observed for all sizes of household. Nationally, there are few differences in undercoverage according to household size, with rates ranging from 7.9% to 9.0%.

Provincially, there is greater variation in the slippage rate. However, the gap observed is nearly the same between household sizes. For one-person households, the rate varies from -2.2% for Nova Scotia to -16.7% for Newfoundland and Labrador For two-person households, the slippage rate ranges from -2.3% in Saskatchewan to -13.7% in Ontario. For households of three or more persons, the slippage rates vary from -2.2% for Alberta to -16.5% for Saskatchewan.

Except for Ontario, undercoverage was highest for one-person households and households of three or more persons, at nearly equal levels.

3.2 Adjustment at the population and household levels

To correct the problem of the sample's representativeness, shown in Tables 3.1-1 and 3.2-1, 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. For more details on the adjustment methodology, see references [1] and [6]. This adjustment reduces the bias but does not 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 2001 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 problem 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 auxiliary 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-responding 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 auxiliary data on the number of households.

In addition to demographic estimates of age groups by province, two other sets of auxiliary data are used during weighting to adjust survey data and thereby improve their representativeness. The first set of data is used to control for the number of children and adults in certain major cities. For the second set, counts for major categories of income from wages and salaries are used when adjusting weights to ensure a degree of consistency between the income distributions from the SHS and those 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 15% 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. Also, respondents are encouraged to consult various documents (bills, 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 co-operation 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.

While response errors are a major source of error in a recall interview, they are the aspect of data quality that is 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

Errors can arise in all types of data handling. The main stages of data processing are response 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 edit 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), no questionnaire batch edit procedure has been used since 2001, unlike in previous years. 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 in the new system are similar to those in the old system.

An initial automated edit is carried out after each questionnaire has been verified manually by both the interviewer and the senior interviewer. This ensures that the respondent's answers follow some essential consistency rules. Unusual situations that may justify corrections are also identified. This automated edit stage takes place in Statistics Canada's regional offices in case it is necessary to recontact respondents if some supplementary information is needed to resolve inconsistencies in the answers provided. Specially-trained members of the edit teams solve any problems identified. Thereafter, other edit 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 are categorical or continuous. Categorical data take 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 each group, the missing values for a recipient (a household that has some missing data for at least one of these variables) are imputed from data on 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 certain 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 are collected at that level for these variables.

Note that since 2001, the imputation of all expenditure and income data has been done using the Canadian Census Edit and Imputation System (CANCEIS) of Statistics Canada. This new system is based on methodology that is slightly different from that in the system used previously. The new system allows a better use of categorical variables as matching fields when selecting a donor. Moreover, this system lends itself to the imputation of both continuous and categorical data. The new system was tested prior to its implementation and the results it gave were similar to those with the old system. Starting with 2003, categorical data, which are found mainly in the dwelling characteristics and facilities sections 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 respondent 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 give an indication of the importance of partial nonresponse. They are presented in the following section.

5.1   Proportion of households or individuals requiring imputation, at the national and provincial levels

A first indication of the magnitude of partial nonresponse is the proportion of households requiring imputation and the number of variables imputed per 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, it is important to note that 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 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 results for 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, in order to ensure data consistency. In reality, these changes constitute only a very small percentage. The results are provided at the national and provincial levels. This gives an indication of which provinces are most affected by imputation.

5.1.1   Household expenditure imputation by province

Table 5.1-1 shows the percentage of usable households requiring imputation of at least one expenditure variable. Usable households are all households living in eligible dwellings, excluding households who could not be contacted, who refused to

participate in the survey, or who provided incomplete data or who were out of balance (see definitions in Section 2.1). The table is broken down by the number of imputed variables (out of 240) for a household.

Note that regular mortgage payments and mortgage insurance premiums are included under shelter costs and thus under total expenditure. Starting with 2002, these two variables were added to the calculation of imputation rates shown in Table 5.1-1. The impact of this change is a higher overall imputation rate.

Starting in 2004, a change was made to the questionnaire regarding expenditures on communication services in the home (telephone, cell phone and Internet access), cable television services and satellite distribution services. Because of the growing use of packages (bundled services), a household may be billed for combined services, with the result that it is more difficult for it to provide expenditures for individual services. In such a case, the respondent household may provide only the total expenditure for these services while indicating which services are included in the package. Expenditures for individual services are then imputed in two stages. First, we impute households for which only a few services are missing, followed by households for which only the total expenditure for the package is available. For the latter households, the imputed expenditures for services (those included in the package) are adjusted proportionally so that their sum corresponds to the total expenditure on the package as provided by the respondent household. Since this change has had a major impact on the overall imputation rate for expenditures, the imputation rates in Table 5.1-1 are shown separately with and without the costs of communications services in the home, rental of cable television services and rental of satellite distribution services. Also, since this change has had an impact on the level of imputation of expenditures for these five services, Table 5.1-2 is provided, showing the imputation rate and a measure of the impact of imputation for each of these services.

Table 5.1-1
Households requiring expenditure imputation by province

Province	Households requiring imputation for **expenditure variables**[1] (excluding clothing expenditures and expenditures in the section on Personal Taxes, Security and Money gifts)							
	Excluding expenditures related to communications services in the home and rental services for cable television and satellite distribution				Including expenditures related to communications services in the home and rental services for cable television and satellite distribution			
	Number of variables imputed (out of 240)			**Total**	Number of variables imputed (out of 245)			**Total**
	1	**2**	3 or more		**1**	**2**	3 or more	
	%							
Canada	10.0	1.9	1.3	13.2	7.8	15.1	8.2	31.1
N.L.	15.2	1.9	0.9	18.0	12.0	10.7	6.7	29.3
P.E.I.	10.9	2.2	0.8	13.9	6.1	15.1	21.1	42.3
N.S.	13.7	2.7	2.4	18.8	8.8	15.4	24.2	48.4
N.B.	9.8	1.5	0.6	11.9	8.2	16.9	5.9	31.1
Que.	5.5	0.5	0.2	6.2	5.3	7.1	3.1	15.5
On.	8.9	3.2	2.8	15.0	8.2	14.3	8.3	30.7
Man.	5.1	1.1	0.4	6.6	4.5	13.6	2.5	20.6
Sask.	13.5	1.8	0.9	16.2	10.6	15.0	7.4	33.0
Alta.	15.0	2.3	1.4	18.6	11.3	18.1	6.4	35.8
B.C.	5.6	2.1	2.0	9.7	4.1	25.8	6.4	36.3

1. Includes regular mortgage payments and mortgage insurance premiums.

Table 5.1-1 shows that it was necessary to impute expenditures for 31.1% of households nationally. This higher rate in 2004 is due to the change made to the questionnaire regarding expenditures related to communications services in the home (telephone, cell phone and Internet access), cable television services and satellite distribution services. Thus, 21.5% (data not shown) of usable households required imputation of at least one of these five services. In almost all of these cases, the household had reported paying for a package (bundled services) and the expenditures associated with the services included in the package were imputed. The higher imputation rates when these five variables are taken into account, such as shown in the column "2 variables imputed" and the column "3 or more variables imputed," are due to the fact that a package usually includes two or more services. Excluding expenditures related to communications services in the home, cable television services and satellite distribution services, the overall imputation rate is 13.2% at the national level, which is comparable to the rates obtained in previous years. Just for the variable representing mortgage insurance premiums, imputation is required for 5.4% of usable households (or 14.4% of households when selecting only households that reported mortgages on dwellings that they owned and occupied).

When expenditures related to telecommunications services in the home (telephone, cell phone and Internet access), cable television services and satellite distribution services are excluded, it may be seen that nearly 75% of usable households required imputation of a single variable. Also, very few households had more than one variable imputed (3.2%). At the provincial level, Quebec (6.2%) and Manitoba

(6.6%) have the lowest proportions of households requiring imputation of at least one expenditure variable. The highest rates are in Nova Scotia (18.8%), Alberta (18.6%) and Newfoundland and Labrador (18.0%). Ontario and British Columbia have the highest percentages of households that required imputation for more than one expenditure variable. In those two provinces, more than 40% of the households that required imputation had two or more expenditure variables imputed.

If we exclude regular mortgage payments, mortgage insurance premiums, expenditures related to communications services in the home, cable television services and satellite distribution services, then the low percentage of households for which variables had to be imputed, combined with a generally low number of variables to be imputed when imputation is required, suggests that the impact of imputed values on the estimates should not be too high.

Since there is a higher level of imputation for expenditures related to communications services in the home, cable television services and satellite distribution services, it is important to measure the effect of imputation on the estimates of totals for these five variables. This measure, along with the imputation rate, can be used to see how the amount of imputation done for these variables changes over time. Owing to the growing popularity of packages (bundled services) within the population, the imputation level should increase over time. To measure the impact of imputation, the weighted total of the imputed data is divided by the total estimate (sum of weighted values). This measure represents the proportion of the total value of the estimate that is obtained from imputed data.

Table 5.1-2

Impact of imputation of communications services in the home, cable television services and satellite distribution services at the national level

Characteristics	Impact of Imputation	Households requiring imputation of communications services in the home, cable television services or satellite distribution services
	%	
Telephone services	7.2	12.2
Cell phone, pager and handheld text messaging services	3.5	6.1
Rental of cable television services	20.4	20.5
Satellite distribution services	6.8	3.7
Internet access services	32.8	40.3

According to Table 5.1-2, the imputation rate and the impact of imputation are greater for expenditures related to Internet access services and the rental of cable television services. This is mainly due to the fact that among households that reported paying for a package, a large proportion of packages included these two services. The high level of imputation performed on the components in Table 5.1-2

suggests that the estimates of these components might be greatly affected by imputation, while the effect on the estimate of the total of these five services combined will be negligible, since households must provide the total expenditure associated with the package. While the imputation rate and the impact are high for expenditures on Internet access services, the increase that occurred in 2004 for average Internet access expenditures was consistent with the trends observed from other independent sources of information. Internet access services accounted for 16% of all household expenditures on communications. Total expenditures on the five services in Table 5.1-2 combined represent only 2.6% of total household expenditure.

5.1.2 Person expenditure and income imputation by province

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, followed by those for whom only 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 is presented by province in Table 5.2. The table shows the percentage of persons who had exactly one variable imputed, the percentage of those who had two or more variables (but not all) imputed and the percentage of persons for whom only total income was available (and hence required having all their components imputed). 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 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 for each household member aged 15 or over on December 31 of the reference year. Thus, since the 2003 reference 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 persons aged 15 or over and was not based on all persons as done in previous years. This modification resulted in an imputation rate slightly higher for those variables. As was 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

Province	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	**0.7**	**0.2**	**4.0**	**5.0**	**20.2**	**3.8**
N.L.	0.3	0.2	3.5	4.2	13.6	2.6
P.E.I.	0.1	0.4	4.8	5.6	15.1	3.7
N.S.	1.2	0.3	6.9	8.5	18.8	6.1
N.B.	0.6	0.2	3.9	4.7	15.1	3.1
Que.	0.2	0.1	3.3	3.6	28.1	1.2
Ont.	1.8	0.8	2.8	5.4	16.6	7.0
Man.	0.2	0.1	2.1	2.4	22.5	2.9
Sask.	0.5	0.1	3.2	3.8	17.4	3.7
Alta.	0.8	0.1	3.7	4.7	23.7	3.7
B.C.	0.8	0.2	6.5	7.6	25.3	3.9

These results show that 5% of persons from usable households had imputation performed on at least one income variable. For 80% of them, the respondent gave the total income but all their components had to be imputed. For a very large proportion of the remaining persons requiring imputation, only one component of income (one variable) had to be imputed. Provincially, the percentages of persons requiring imputation on at least one income variable range from a low of 2.4% for Manitoba to a high of 8.5% for Nova Scotia. A higher imputation rate is observed for Nova Scotia (8.5%) and British Columbia (7.6%). In Ontario, 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 20% of persons required imputation for at least one of the clothing expenditure variables. Provincial rates range from 13.6% for Newfoundland and Labrador to 28.1% for Quebec. Almost all these people provided their total expenditure on clothing but required imputation of the components. Newfoundland and Labrador is the only province with an imputation rate smaller than 15%. As may also be seen, with the exception of Nova Scotia, these rates are somewhat lower in the Atlantic provinces than in the other provinces. While Quebec has the highest imputation rate for clothing expenditure, it has among the lowest imputation rates for income or for variables in the Personal Taxes, Security and Money Gifts section. The higher level of imputation required on clothing expenditure components suggests that the estimates for these components could be greatly affected by imputation, while the effect on the estimates for total clothing expenditure will be negligible.

From the last column of the table, results show that approximately 4% of persons had some imputation performed on at least one variable in the Personal Taxes, Security and Money Gifts section. Provincially, these percentages are also low, ranging from a low of 1.2% for Quebec to a high of 7.0% for Ontario. Only Nova Scotia and Ontario have a rate exceeding 4%. In both those provinces, the higher imputation rate is due to the variable corresponding to personal income tax paid on 2004 income, which required proportionally more imputation than for the other provinces.

5.1.3   Imputation of categorical variables by province and territory

Table 5.3 shows the percentage of usable households that required imputation for at least one categorical variable. The table is broken down by the number of imputed variables (out of 58) 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 of 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 experts from the Income Statistics Division. Therefore, the latter variables are not imputed using the nearest neighbour technique.

Table 5.3
Households requiring imputation of categorical variables by province

Province	Households requiring imputation of categorical variables			
	Number of variables imputed (out of 58)			
	1	2	3 or more	Total
	%			
Canada	5.5	1.1	0.9	7.6
Newfoundland and Labrador	4.0	0.4	0.2	4.6
Prince Edward Island	5.1	1.0	0.7	6.8
Nova Scotia	3.2	0.5	0.0	3.6
New Brunswick	1.8	0.3	0.1	2.2
Quebec	6.5	0.7	0.7	8.0
Ontario	5.4	1.3	1.3	8.0
Manitoba	5.7	1.3	1.1	8.1
Saskatchewan	7.2	1.1	1.6	9.9
Alberta	8.5	2.3	2.1	12.9
British Columbia	6.4	2.1	1.2	9.7

Table 5.3 indicates that at the national level, 7.6% of households required some categorical imputation for dwelling characteristics, facilities associated with the dwelling, tenure and purchases through direct sales, but approximately 72% of those households had only one variable imputed. The lowest proportions of households requiring imputation for at least one categorical variable are observed for the Atlantic Provinces. By contrast, the Western provinces have the highest proportions.

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. 2004. *User Guide for the Survey of Household Spending.* Catalogue no. 62F0026MIE2005007. Ottawa.

[4] Statistics Canada, Income Statistics Division. 2004. *Spending Patterns in Canada.* Catalogue no. 62-202. Ottawa.

[5] Statistics Canada, Household Survey Methods Division. 2004. *Labour Force Survey, Operations Report.* Survey 200412. Ottawa.

[6] Lessard, S. 2005. *Revision of the Calibration Strategy for the Survey of Household Spending.* Ottawa. Household Survey Methods Division, Statistics Canada.

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

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

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 attributed 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 control group c.