Microdata User Guide

RESIDENTIAL TELEPHONE SERVICE SURVEY

December 2013

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The latest Residential Telephone Service Survey (RTSS) was conducted by Statistics Canada in December 2013.This manual has been produced to provide information about the survey and facilitate the use of the microdata file of the survey results.

How to obtain more information
Any questions about the data set or its use should be directed to:

Client Services
Special Surveys Division
Telephone: 613-951-3321 or toll-free: 1-800-461-9050
Fax: 613-951-3012
E-mail: ssd@statcan.ca

Accessing and ordering information
The Residential Telephone Service Survey (RTSS) produced one type of microdata file: the master file.

Master files
The master files contain all variables and all records from the survey collected during a collection period. These files are accessible at Statistics Canada for internal use and in Statistics Canada's Research Data Centres (RDC), and are also subject to custom tabulation requests.

Research Data Centre
The RDC Program enables researchers to use the survey data in the master files in a secure environment in several universities across Canada. Researchers must submit research proposals that, once approved, give them access to the RDC. For more information, please consult the following web page: http://www.statcan.gc.ca/rdc-cdr/index-eng.htm

Custom tabulations
Another way to access the master files is to offer all users the option of having staff in Client Services of Income Statistics Division prepare custom tabulations. This service is offered on a cost-recovery basis. It allows users who do not possess knowledge of tabulation software products to get custom results. The results are screened for confidentiality and reliability concerns before release. For more information, please contact Client Services.

2.0 Background and objectives

The Residential Telephone Service Survey (RTSS) has been conducted since the fall of 1996. The RTSS monitors residential phone penetration rates. Initially, the survey was sponsored by Bell Canada to assess reasons why certain households did not have telephone service. Later, with the increased popularity of cellular phones, the focus of the survey switched to the types of telephone service used by households. The current version of the RTSS is similar to the most recent – 2010 iteration of the survey. Survey results will be used by statisticians to calculate chances of reaching households by telephone.

3.0 Survey Methodology

The Residential Telephone Service Survey (RTSS) was administered in December 2013 to a sub-sample of the dwellings in the Labour Force Survey (LFS) sample, and therefore its sample design is closely tied to that of the LFS. The LFS design is described in the *Methodology of the Canadian Labour Force Survey* document (Catalogue no. 71-526-X).Sections 5.1 and 5.2 describe how the Residential Telephone Service Survey departed from the basic LFS design in December 2013.

3.1 Modifications to the Labour Force Survey Design for the Residential Telephone Service Survey

The Residential Telephone Service Survey used two of the six rotation groups for Ontario and Quebec and three for all other provinces in the December 2013 LFS sample. For the RTSS, the coverage of the LFS was set at the household level. However, unlike the LFS where information is collected for all eligible household members, the RTSS only collected information from one household member who reported about the household.

3.2 *Sample Size by Province for the Residential Telephone Service Survey*

The following table shows the number of households in the LFS sampled rotations that were eligible for the Residential Telephone Service Survey supplement.

Province	Sample Size
Newfoundland and Labrador	984
Prince Edward Island	702
Nova Scotia	1422
New Brunswick	1370
Quebec	3224
Ontario	5058
Manitoba	2279
Saskatchewan	1901
Alberta	2651
British Columbia	3154
Canada	**22,745**

4.0 Concepts and Definitions

The RTSS provided interviewers and respondents with the following instructions:

An active cellular phone
An “active cell phone” is one which has been activated and is used regularly to make and receive phone calls. A cell phone used only for emergency purposes is not considered an “active cell phone”.
Include cell phones used for **business purposes.** This includes cell phones for a business that is not within the household and cell phones from work that are used at home (even if the employer pays the phone bill).

Number of phone numbers in household, other than cellular
Include phone numbers that are used for computer, fax or business purposes**.**

Types of telephone service (other than cellular):

A regular landline telephone service

This is a service that has been available for decades and is provided by telephone companies such as Bell Canada, BC Tel, and Telus. Numerous other suppliers are now offering land-line service as well. There is a wire which connects to a phone jack in the wall. Landline phone works even when power and Internet are down.

A phone service from a cable television provider
There is no phone jack in the wall, but a cable modem (an electronic box) connected to a TV cable. Include such providers as Rogers (Rogers Home Phone), Cogeco, Videotron, and Shaw.

A phone service over the Internet with a unique number (Voice over Internet Protocol – VoIP)
The phone service comes from an Internet provider. There is a modem at home and an Internet connection. Skype and Vonage are examples of VoIP service.

5.0 Data Collection

Data collection for the Labour Force Survey (LFS) is carried out each month during the week following the LFS reference week. The reference week is normally the week containing the 15th day of the month. The collection period for the 2013 RTSS ran from the 14th to the 24st of December 2013.

5.1 *Interviewing for the Labour Force Survey*

Statistics Canada interviewers are employees hired and trained to carry out the LFS and other household surveys. Each month they contact the sampled dwellings to obtain the required labour force information. Each interviewer contacts approximately 75 dwellings per month.

Dwellings new to the sample are usually contacted through a personal visit using the computer-assisted personal interview (CAPI). The interviewer first obtains socio-demographic information for each household member and then obtains labour force information for all members aged 15 and over who are not members of the regular armed forces. Provided there is a telephone in the dwelling and permission has been granted, subsequent interviews are conducted by telephone. This is done out of a centralized computer-assisted telephone interviewing (CATI) unit where cases are assigned randomly to interviewers. As a result, approximately 85% of all households are interviewed by telephone. In these subsequent monthly interviews, the interviewer confirms the socio-demographic information collected in the first month and collects the labour force information for the current month.

In each dwelling, information about all household members is usually obtained from one knowledgeable household member. Such “proxy” reporting, which accounts for approximately 65% of the information collected, is used to avoid the high cost and extended time requirements that would be involved in repeat visits or calls necessary to obtain information directly from each respondent.

If, during the course of the six months that a dwelling normally remains in the sample, an entire household moves out and is replaced by a new household, information is obtained about the new household for the remainder of the six-month period.

At the conclusion of the LFS monthly interviews, interviewers introduce the supplementary survey, if any, to be administered to some or all household members that month.

5.2 *Supervision and Quality Control*

All LFS interviewers are under the supervision of a staff of senior interviewers who are responsible for ensuring that interviewers are familiar with the concepts and procedures of the LFS and its many supplementary surveys, and also for periodically monitoring their interviewers and reviewing their completed documents. The senior interviewers are, in turn, under the

supervision of the LFS program managers, located in each of the Statistics Canada regional offices.

5.3 Non-response to the Labour Force Survey

Interviewers are instructed to make all reasonable attempts to obtain LFS interviews with members of eligible households. For individuals who at first refuse to participate in the LFS, a letter is sent from the Regional Office to the dwelling address stressing the importance of the survey and the household's cooperation. This is followed by a second call (or visit) from the interviewer. For cases in which the timing of the interviewer's call (or visit) is inconvenient, an appointment is arranged to call back at a more convenient time. For cases in which there is no one home, numerous call backs are made. Under no circumstances are sampled dwellings replaced by other dwellings for reasons of non-response.
Each month, after all attempts to obtain interviews have been made, a small number of non-responding households remain. For households non-responding to the LFS and for which LFS information was obtained in the previous month, this information is brought forward and used as the current month's LFS information. No supplementary survey information is collected for these households.

5.4 Data Collection Modifications for the Residential Telephone Service Survey

Information for the Residential Telephone Service Survey (RTSS) was obtained from a knowledgeable household member. Upon completion of the Labour Force Survey interview, the interviewer introduced the RTSS and proceeded with the interview with the respondent's permission.

The Residential Telephone Service Survey was programmed to appear on the list of surveys to be completed after the demographic component for the LFS had been completed. Any RTSS component not completed at the time the LFS was left incomplete and transmitted with the LFS.

5.5 Non-response to the Residential Telephone Service Survey

For households responding to the LFS, the next stage of data collection was to administer the Residential Telephone Service Survey. In total, 22,745 households were eligible for the supplementary survey; the RTSS interview was completed for 19,294 of these households for a response rate of 84.8%. More detailed information on response rates is presented in Chapter 7.0 (Data Quality).

6.0 *Data Processing*

Processing transforms survey responses obtained during collection into a form that is suitable for tabulation and data analysis. It includes all data handling activities – automated and manual – after collection and prior to estimation.The main output of the Residential Telephone Service Survey (RTSS) is a "clean" microdata file. This chapter presents a brief summary of the processing steps involved in producing this file.

6.1 *Data Capture*

For computer-assisted interviews, responses to survey questions are captured directly by the interviewer at the time of the interview using a computerized questionnaire. The computerized questionnaire reduces processing time and costs associated with data entry, transcription errors, and data transmission. The response data are encrypted to ensure confidentiality and sent via modem to the appropriate Statistics Canada Regional Office. From there they are transmitted over a secure line to Ottawa for further processing. In total 22,745 documents were captured and transmitted for the survey.

Some editing is done directly at the time of the interview. Where the information entered is out of range (too large or small) of expected values, or inconsistent with the previous entries, the interviewer is prompted, through message screens on the computer, to modify the information. However, for some questions interviewers have the option of bypassing the edits, and of skipping questions if the respondent does not know the answer or refuses to answer. Therefore, the response data are subjected to further edit and imputation processes once they arrive in head office.

6.2 *Editing*

The first stage of survey processing undertaken at head office was the replacement of any “out-of-range” values on the data file with blanks. This process was designed to make further editing easier.

The first type of error treated was errors in questionnaire flow, where questions which did not apply to the respondent (and should therefore not have been answered) were found to contain answers. In this case a computer edit automatically eliminated superfluous data by following the flow of the questionnaire implied by answers to previous, and in some cases, subsequent questions.

The second type of error treated involved a lack of information in questions which should have been answered. For this type of error, a non-response or "not-stated" code was assigned to the item.

6.3 Coding of Open-ended Questions

No data items on the questionnaire were recorded by interviewers in an open-ended format. One partially open-ended question was included in the survey. This related to responses “other – specify” to the RTS_Q08 (Excluding cell phones, what other telephone service do you have in your household?). A total of 111 responses were entered in this category. About 40 of these responses indicated that there was no other telephone service in the household besides cellular phone(s). The remaining responses were potentially suitable for recoding to the existing categories (e.g. Magic Jack to “VOIP “or a cable company name to “a phone from a cable television provider”), were truly “Other” (e.g. Satellite phone) or were not clear. Given the negligible impact of potential recoding it was not done.

6.4 Weighting

The principle behind estimation in a probability sample such as the LFS is that each person in the sample "represents", besides himself or herself, several other persons not in the sample. For example, in a simple random 2% sample of the population, each person in the sample represents 50 persons in the population. The same principle also applies to households.

The weighting phase is a step which calculates, for each record, what this number is. This weight appears on the microdata file, and **must** be used to derive meaningful estimates from the survey. For example, if the number of households with one or more telephone numbers for their residence is to be estimated, it is done by selecting the records referring to those households in the sample with that characteristic and summing the weights entered on those records.

Details of the method used to calculate these weights are presented in Chapter 8.0.

6.5 Disclosure Control

Statistics Canada is prohibited by law from releasing any data that would divulge information obtained under the *Statistics Act* that relates to any identifiable person, business or organization without the prior knowledge or the consent in writing of that person, business or organization. Various confidentiality rules are applied to all data that are released or published to prevent the publication or disclosure of any information deemed confidential. If necessary, data are suppressed to prevent direct or residual disclosure of identifiable data.

In the final RTSS data file, all names and other personal identifiers were removed to ensure confidentiality of the data.

7.0 Data Quality

Survey errors come from a variety of different sources. They can be classified into two main categories: non-sampling errors and sampling errors.

7.1 Non-sampling errors

Non-sampling errors can be defined as errors arising during the course of virtually all survey activities, apart from sampling. They are present in both sample surveys and censuses (unlike sampling error, which is only present in sample surveys). Non-sampling errors arise primarily from the following sources: non-response, coverage, measurement and processing.

7.1.1 Non-response

Non-response errors result from a failure to collect complete information on all units in the selected sample.

Non-response produces errors in the survey estimates in two ways. First, is that non-respondents often have different characteristics from respondents, which can result in biased survey estimates if non-response is not corrected properly. Secondly, it reduces the effective size of the sample, since fewer units than expected answered the survey. As a result, the sampling variance increases and the precision of the estimate decreases.
The following table summarizes the response rates to the Labour Force Survey (LFS) and to the Residential Telephone Service Survey (RTSS) in December 2013.

Province	Household Response Rate for Full LFS* December 2013	Household Response Rate for LFS Rotations 1, 2 in ON,QC and 1,2,4 in other provinces *	Household Response Rate for RTSS**	RTSS Responding Households
	%			
Newfoundland and Labrador	89.9	88.6	88.0	866
Prince Edward Island	92.4	91.4	90.9	638
Nova Scotia	90.3	90.3	89.1	1,267
New Brunswick	91.3	91.6	89.4	1,225
Québec	91.2	91.7	88.6	2,855
Ontario	88.5	88.4	83.3	4,214
Manitoba	88.1	87.4	85.0	1,938
Saskatchewan	87.8	87.5	85.4	1,624
Alberta	87.7	87.5	82.0	2,173
British Columbia	86.7	87.1	79.1	2,494
Canada	**89.0**	**88.8**	**84.8**	**19,294**

* The LFS response rate is the number of responding households as a percentage of the number of eligible households.

** The RTSS response rate is the number of households responding to the RTSS as a percentage of the number of households responding to the LFS, in the rotations sampled.

7.1.2 Coverage Errors

Coverage errors consist of omissions, erroneous inclusions, duplications and misclassifications of units in the survey frame. Since they affect every estimate produced by the survey, they are one of the most important types of error; in the case of a census they may be the main source of error. Coverage errors may cause a bias in the estimates and the effect can vary for different sub-groups of the population.

Since the RTSS was a supplement to the LFS, the frame used was the LFS frame. Any non-response to the LFS had an impact on the RTSS frame. The quality of the sampling variables in the frame was very high. The RTSS sample consisted of households in a subset of LFS rotation groups. The criteria used for the RTSS selection (such as a rotation group) were not missing for any LFS records.

Note that the LFS frame excludes less than 3% of all households in the 10 provinces of Canada. Therefore, the RTSS frame also excludes a similar proportion of households. It is unlikely that this exclusion introduces any significant bias into the survey data.

7.1.3 Measurement Errors

Measurement errors (or sometime referred to as response errors) occur when the response provided differs from the real value; such errors may be attributable to the respondent, the interviewer, the questionnaire, the collection method or the respondent's record-keeping system. Such errors may be random or they may result in a systematic bias if they are not random.

It is very costly to accurately measure the level of response error and very few surveys conduct a post-survey evaluation. However, interviewer feedback and observation reports usually provide clues as to which questions may be problematic (poorly worded question, inadequate interviewer training, poor translation, technical jargon, no help text available, etc.).

Several measures are taken to reduce the level of response error. These measures include questionnaire review and testing using cognitive methods, the use of highly skilled interviewers, extensive training of interviewers with respect to the survey procedures and content, observation and monitoring of interviewers to detect problems of questionnaire design or misunderstanding of instructions.

Interviewer training consisted of reading the RTSS Procedures Manual, practicing with the RTSS training cases on the computer, and discussing any questions with senior interviewers before the start of the survey. A description of the background and objectives of the survey was provided, as well as a glossary of terms and a set of questions and answers.
Item non-response is usually very low for the RTSS. For questions RTS_Q01 to RTS_Q08 item non-response ranged from a low of 0.047% for RTS_Q01 to a high of 0.687% for RTS_Q08.

7.1.4 Processing Errors

Processing error is the error associated with activities conducted once survey responses have been received. It includes all data handling activities after collection and prior to estimation. Like all other errors, they can be random in nature, and inflate the variance of the survey’s estimates, or systematic, and introduce bias. It is difficult to obtain direct measures of processing errors and their impact on data quality especially since they are mixed in with other types of errors (non-response, measurement and coverage).

During processing of the data, 340 RTSS records did not match to corresponding records in the LFS. Thus they were coded as out-of-scope and were dropped from further processing. When supplementary survey records do not match to host survey records they must be dropped since a weight cannot be derived for them.

Conversely, 171 records in the LFS were found that should have matched to an RTSS record but did not. These records were coded as in-scope, since they were eligible records from the frame which, for one reason or another, did not have corresponding RTSS records. These records were considered to be non-responding records, and were used in the weighting process to adjust for non-response.

Data processing of the RTSS was straightforward since there were only eight questions on the CAI application. Any record that contained a "Refused" or "Don't know" response in question RTS_Q01 (has a cellular phone) and in RTS_Q05 (has other type of phone service) was coded as non-response. Note that 3,451 records were treated this way. Since the data was collected using a CAI instrument, data quality before processing was very high. Very few changes were made to the data during editing.

7.2 Sampling errors

Sampling error is defined as the error that results from estimating a population characteristic by measuring a portion of the population rather than the entire population. For probability sample surveys, methods exist to calculate sampling error. These methods derive directly from the sample design and method of estimation used by the survey.

The most commonly used measure to quantify sampling error is sampling variance. Sampling variance measures the extent to which the estimate of a characteristic from different possible samples of the same size and the same design differ from one another. For sample designs that use probability sampling, the magnitude of an estimate's sampling variance can be estimated. The key issue is the magnitude of an estimate's estimated sampling variance relative to the size of the survey estimate: if the variance is relatively large, then the estimate has poor precision and is unreliable.

Factors affecting the magnitude of the sampling variance include:

1. The variability of the characteristic of interest in the population: the more variable the characteristic in the population, the larger the sampling variance.
2. The size of the population: in general, the size of the population only has an impact on the sampling variance for small to moderate sized populations.
3. The response rate: the sampling variance increases as the sample size decreases. Since non-respondents effectively decrease the size of the sample, non-response increases the sampling variance.
4. The sample design and method of estimation: some sample designs are more efficient than others in the sense that, for the same sample size and method of estimation, one design can lead to smaller sampling variance than another.

The standard error of an estimator is the square root of its sampling variance. This measure is easier to interpret since it provides an indication of sampling error using the same scale as the estimate whereas the variance is based on squared differences.

However, even standard error might be difficult to interpret in terms of "How big a standard error is acceptable?" What is large depends on the magnitude of the estimate. For example, a standard error of 100 would be considered large for measuring the average weight of people but would not be considered large for estimating average annual income.

It is more useful in many situations to assess the size of the standard error relative to the estimate of the characteristic being measured. The coefficient of variation (CV) provides such a measure. It is the ratio of the standard error of the survey estimate to the average value of the estimate itself, across all possible

samples. The coefficient of variation is usually computed as the estimate of the standard error of the survey estimate to the estimate itself. This relative measure of sampling error is usually expressed as a percentage (10% instead of 0.1). It is very useful in comparing the precision of sample estimates, where their sizes or scale differ from one another.

8.0 Weighting

The principle behind estimation in a probability sample is that each unit in the sample "represents", besides itself, several other units not in the sample. For example, in a simple random 2% sample of the population, each unit in the sample represents 50 units in the population.

The weighting phase is a step which calculates, for each record, what this number is. This weight appears on the microdata file, and must be used to derive meaningful estimates from the survey.

Since the Residential Telephone Service Survey used a sub-sample of the Labour Force Survey (LFS) sample, the derivation of weights for the survey records is clearly tied to the weighting procedure used for the LFS. The LFS weighting procedure is briefly described below in sub-section 8.1. In sub-section 8.2, details of the method used to calculate sampling weights for the CSEW are provided.

8.1 Weighting procedures for the Labour Force Survey

In the LFS, the final weight attached to each record is the product of the following factors: the basic weight, the cluster sub-weight, the stabilization weight, the balancing factor for non-response, and the province-age-sex and sub-provincial area ratio adjustment factor. Each is described below.

Basic weight
In a probability sample, the sample design itself determines weights which must be used to produce unbiased estimates of population. Each record must be weighted by the inverse of the probability of selecting the person to whom the record refers. In the example of a 2% simple random sample, this probability would be 0.02 for each person and the records must be weighted by 1 / 0.02 = 50. Due to the complex LFS design, dwellings in different regions will have different basic weights. Because all eligible individuals in a dwelling are interviewed (directly or by proxy), this probability is essentially the same as the probability with which the dwelling is selected.

Cluster sub-weight
The cluster delineation is such that the number of dwellings in the sample increases very slightly with moderate growth in the housing stock. Substantial growth can be tolerated in an isolated cluster before the additional sample represents a field collection problem. However, if growth takes place in more than one cluster in an interviewer assignment, the cumulative effect of all increases may create a workload problem. In clusters where substantial growth has taken place, sub-sampling is used as a means of keeping interviewer assignments manageable. The cluster sub-weight represents the inverse of this sub-sampling ratio in clusters where sub-sampling has occurred.

Stabilization weight
Sample stabilization is also used to address problems with sample size growth. Cluster sub-sampling addressed isolated growth in relatively small areas whereas sample stabilization accommodates the slow sample growth over time that is the result of a fixed sampling rate along with a general increase in the size of the population. Sample stabilization is the random dropping of dwellings from the sample in order to maintain the sample size at its desired level. The basic weight is adjusted by the ratio of the sample size, based on the fixed sampling rate, to the desired sample size. This adjustment factor is known as the

stabilization weight. The adjustment is done within stabilization areas defined as dwellings belonging to the same employment insurance economic region and the same rotation group.

Non-response
For certain types of non-response (i.e. household temporarily absent, refusal), data from a previous month's interview with the household if any, is brought forward and used as the current month's data for the household.

In other cases, non-response is compensated for by proportionally increasing the weights of responding households. The weight of each responding record is increased by the ratio of the number of households that should have been interviewed, divided by the number that were actually interviewed. This adjustment is done separately for non-response areas (commonly named response homogeneity group), which are defined by employment insurance economic region, type of area, and rotation group. It is based on the assumption that the households that have been interviewed represent the characteristics of those that should have been interviewed within a non-response area.

Labour Force Survey sub-weight
The product of the previously described weighting factors is called the LFS sub-weight. All members of the same sampled dwelling have the same sub-weight.

Sub-provincial and province-age-sex adjustments
The sub-weight can be used to derive a valid estimate of any characteristic for which information is collected by the LFS. However, these estimates will be based on a frame that contains some information that may be several years out of date and therefore not representative of the current population. Through the use of more up-to-date auxiliary information about the target population, the sample weights are adjusted to improve both the precision of the estimates and the sample's representation of the current population.

Independent estimates are available monthly for various age and sex groups by province. These are population projections based on the most recent census data, records of births and deaths, and estimates of migration. In the final step, this auxiliary information is used to transform the sub-weight into the final weight. This is done using a calibration method. This method ensures that the final weights it produces sum to the census projections for the auxiliary variables, namely totals for various age-sex groups, economic regions, census metropolitan areas, rotation groups, household and economic family size. Weights are also adjusted so that estimates of the previous month's industry and labour status estimates derived from the present month's sample, sum up to the corresponding estimates from the previous month's sample. This is called composite estimation. The entire adjustment is applied using the generalized regression technique.

8.2 Weighting procedures for the Residential Telephone Service Survey

The principles behind the calculation of the weights for the RTSS are identical to those for the LFS. However, further adjustments are made to the LFS sub-weights in order to derive a final weight for the individual records on the RTSS microdata file.

1) An adjustment to account for the use of a two-sixths sub-sample for Ontario and Quebec and three-sixths sub-sample for all other provinces, instead of the full LFS sample.

2) An adjustment to account for the additional non-response to the supplementary survey i.e., non-response to the RTSS households that did respond to the LFS. The procedure is similar to the LFS non-response weight adjustment, but the response homogeneity groups are based on different variables.

At this stage the weight is comprised of two components: the inverse of the sampling rate and the non-response adjustment. Independent estimates are available monthly for various age and sex groups by

province. These are population projections based on the most recent census data, records of births and deaths, and estimates of migration. Using a linear regression model, auxiliary information is used to arrive at the final weight. The regression is set up to ensure that the final weights it produces sum to the census projections for the auxiliary variables, namely population counts by various age-sex groups, household counts by size of households. This improves the reliability of estimates that can be produced by the RTSS.

The Master file has been created at the household level and, consequently, the weights on the file are "household weights".

9.0 Guidelines for Tabulation, Analysis and Release

This chapter of the documentation outlines the guidelines to be adhered to by users tabulating, analysing, publishing or otherwise releasing any data derived from the survey microdata files. With the aid of these guidelines, users of microdata should be able to produce the same figures as those produced by Statistics Canada and, at the same time, will be able to develop currently unpublished figures in a manner consistent with these established guidelines.

9.1 Rounding Guidelines

In order that estimates for publication or other release derived from these microdata files correspond to those produced by Statistics Canada, users are urged to adhere to the following guidelines regarding the rounding of such estimates:

a) Estimates in the main body of a statistical table are to be rounded to the nearest hundred units using the normal rounding technique. In normal rounding, if the first or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is raised by one. For example, in normal rounding to the nearest 100, if the last two digits are between 00 and 49, they are changed to 00 and the preceding digit (the hundreds digit) is left unchanged. If the last two digits are between 50 and 99 they are changed to 00 and the preceding digit is incremented by 1.

b) Marginal sub-totals and totals in statistical tables are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units using normal rounding.

c) Averages, proportions, rates and percentages are to be computed from unrounded components (i.e. numerators and/or denominators) and then are to be rounded themselves to one decimal using normal rounding. In normal rounding to a single digit, if the final or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is increased by 1.

d) Sums and differences of aggregates (or ratios) are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units (or the nearest one decimal) using normal rounding.

e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released which differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s).

f) Under no circumstances are unrounded estimates to be published or otherwise released by users. Unrounded estimates imply greater precision than actually exists.

9.2 Sample Weighting Guidelines for Tabulation

The sample design used for the Residential Telephone Service Survey (RTSS) was not self-weighting. When producing simple estimates, including the production of ordinary statistical tables, users must apply the proper sampling weight.

If proper weights are not used, the estimates derived from the microdata files cannot be considered to be representative of the survey population, and will not correspond to those produced by Statistics Canada.

Users should also note that some software packages may not allow the generation of estimates that exactly match those available from Statistics Canada, because of their treatment of the survey weight field.

9.3 Guidelines for statistical analysis

The Residential Telephone Service Survey is based upon a complex sample design, with stratification, multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used. In order for survey estimates and analyses to be free from bias, the survey weights must be used.

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures differ from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are poor. Approximate variances for simple estimates such as totals, proportions and ratios (for qualitative variables) can be derived using the accompanying Approximate Sampling Variability Tables.

For other analysis techniques (for example linear regression, logistic regression and analysis of variance), a method exists which can make the variances calculated by the standard packages more meaningful, by incorporating the unequal probabilities of selection. The method rescales the weights so that there is an average weight of 1.

For example, suppose that analysis of all Quebec households is required. The steps to rescale the weights are as follows:

1) select all households from the file who reported PROVINCE = 24, Quebec;

2) calculate the AVERAGE weight for these records by summing the original household weights from the microdata file for these records and then dividing by the number of households who reported PROVINCE = 24;

3) for each of the these records, calculate a RESCALED weight equal to the original household weight divided by the AVERAGE weight;

4) perform the analysis for these households using the RESCALED weight.

However, because the stratification and clustering of the sample's design are still not taken into account, the variance estimates calculated in this way are likely to be under-estimates.

The calculation of more precise variance estimates requires detailed knowledge of the design of the survey. Such detail cannot be given in this microdata file because of confidentiality. Variances that take the complete sample design into account can be calculated for many statistics by Statistics Canada on a cost-recovery basis.

9.4 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimate from the Residential Telephone Service Survey, users should first determine the quality level of the estimate. The quality levels are *acceptable*, *marginal* and *unacceptable*. Data quality is affected by both sampling and non-sampling errors as discussed in Chapter 7.0. However for this purpose, the quality level of an estimate will be determined only on the basis of sampling error as reflected by the coefficient of variation as shown in the table below. Nonetheless users should be sure to read Chapter 7.0 to be more fully aware of the quality characteristics of these data.

First, the number of household records that contribute to the calculation of the estimate should be determined. If this number is less than 30, the weighted estimate should be considered to be of unacceptable quality.

For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the estimate and follow the guidelines below. These quality level guidelines should be applied to weighted rounded estimates.

All estimates can be considered releasable. However, those of marginal or unacceptable quality level must be accompanied by a warning to caution subsequent users.

Quality level guidelines

Category 1 - Acceptable
The estimates have low coefficients of variation in the range of 0.0% to 16.5%. No release restrictions: data are of sufficient accuracy that no special warnings to users or other restrictions are required.

Category 2 - Marginal
The estimates have high coefficients of variation in the range of 16.6% to 33.3%. Release with caveats: data are potentially useful for some purposes but should be accompanied by a warning to users regarding their accuracy.

Estimates should be flagged with the letter E (or some similar identifier).

Category 3 - Unacceptable
The estimates have very high coefficients of variation in excess of 33.3%. Not recommended for release: data contain a level of error that makes them so potentially misleading that they should not be released in most circumstances. If users insist on inclusion of Category 3 data in a non-standard product, even after being advised of their accuracy, the data should be accompanied by a disclaimer. The user should acknowledge the warnings given and undertake not to disseminate, present or report the data, directly or indirectly, without this disclaimer.

Estimates should be flagged with the letter F (or some similar identifier) and the following warning should accompany the estimates:

“Please be warned that these estimates [flagged with the letter F] do not meet Statistics Canada’s quality standards. Conclusions based on these data will be unreliable, and most likely invalid.”

Appendix A – Variance estimation

In order to determine the quality of the estimate and to calculate the CV, the standard deviation must be calculated. Confidence intervals also require the standard deviation of the estimate. The RTSS uses a multi-stage survey design and calibration, which means that there is no simple formula that can be used to calculate variance estimates. Therefore, an approximate method was needed, the bootstrap method. With the use of the bootstrap weights and the BOOTVAR program, discussed in the next section, CV's and other variance estimates can be derived with accuracy.

Bootstrap method for variance estimation

Independently, in each stratum, a simple random sample of $(n-1)$ of the n units in the sample is selected with replacement. Note that since the selection is with replacement, a unit may be chosen more than once. This step is repeated R times to form R bootstrap samples. An average initial bootstrap weight based on the R samples is calculated for each sample unit in the stratum. The entire process (selecting simple random samples, recalculating weights for each stratum) is repeated B times, where B is large, yielding B different initial bootstrap weights.

These weights are then adjusted according to the same weighting process as the regular weights: non-response adjustment, calibration and so on. The end result is B final mean bootstrap weights for each unit in the sample. The variation among the B possible estimates based on the B bootstrap weights are related to the variance of the estimator based on the regular weights and can be used to estimate it.

Statistical packages for variance estimation

Bootvar

Users should note that bootstrap weights are provided and should be used for variance estimation. BOOTVAR is a macro program that can be used to do the variance calculation using the bootstrap weights. The Bootvar program is available in SAS or SPSS format. It is made up of macros that compute variances for totals, ratios, differences between ratios and for linear and logistic regression.

Bootvar may be downloaded from Statistics Canada's Research Data Centre (RDC) website. Users must accept the Bootvar Click-Wrap Licence before they can read the files. There is a document on the site explaining how to adapt the system to meet users' needs.

SAS: http://www.statcan.gc.ca/rdc-cdr/bootvar_sas-eng.htm
SPSS: http://www.statcan.gc.ca/rdc-cdr/bootvar_spss-eng.htm

Other packages

Other than Bootvar, there are different commercial software packages that can carry out some design-based analysis for variance estimation; Stata 9 or 10, SUDAAN and WesVar.

These methods can be adapted for the RTSS from a paper by Owen Phillips "*Using bootstrap weights with Wes Var and SUDAAN*" (Catalogue no. 12-002-X20040027032) in *The Research Data Centres Information and Technical Bulletin, Chronological index*, Fall 2004, vol.1 no. 2 Statistics Canada, Catalogue no. 12-002-XIE.