



87M0006GPE

Canadian Travel Survey Microdata User's Guide

aussi disponible en français (87M0006GPF)



Statistics Canada
Statistique Canada

Canada

Canadian Travel Survey

CTS

Table of contents

	Page
OVERVIEW	1
1. INTRODUCTION TO CTS	3
Background	3
Objectives	5
CTS products	6
Uses of the CTS data	7
2. CONCEPTS AND DEFINITIONS	9
LFS concepts and definitions	9
CTS concepts and definitions	13
3. SAMPLE	22
LFS population coverage	22
LFS sample design	23
LFS sample size	29
LFS sample rotation	29
CTS as a supplement	30
4. DATA COLLECTION	32
LFS data collection	32
LFS interviewers	33
Non-response to the LFS	34
CTS data collection	35
Non-response to the CTS	36

<i>Table of contents - continued</i>	Page
5. DATA PROCESSING	38
Data capture	38
Preliminary editing	39
Linking to the LFS	39
Geographic coding	39
Secondary editing	41
Weighting	41
Consistency edits	41
Expenditure imputation	41
Reallocation of expenditures	43
Data review	50
6. WEIGHTING	51
Weighting procedures for the LFS	51
Weighting procedures for the CTS	54
7. USING THE MICRODATA	58
File description	58
Person dataset	58
Trip dataset (available only for microdata files users)	62
Person-trip dataset	63
Visit dataset	67
Using the CTS weights	73
Person weight	74
Person-trip weight	75
Household-trip weight	75
Person-visit weight (for Beyond 20/20 users only)	76
Reallocated expenditures weight (for Beyond 20/20 users only)	77
Using the data dictionary	77
8. GUIDELINES FOR ANALYSIS AND PRESENTATION	79
Rounding guidelines	79
Types of estimates	80
Tabulation of categorical estimates	82
Tabulation of quantitative estimates	84

<i>Table of contents - continued</i>	Page
9. GUIDELINES FOR RELEASE (DATA QUALITY)	86
Introduction	86
Survey errors	87
Hypothesis tests provided by statistical software packages	89
Minimum sizes of estimates for release	91
Using the coefficient of variation (CV)	91

OVERVIEW

The Canadian Travel Survey (CTS) is a biennial survey whose purpose is to gather information on domestic trips and travellers. The CTS is a supplement using the Labour Force Survey (LFS) sampling frame. The survey collects more than 30 characteristics which include socio-demographic information on travellers, trips and expenditures. In 1996, a monthly sample of approximately 16,000 persons was interviewed. That same year, additional information was also collected. For the first time, the CTS is measuring the number of visits and providing reallocated expenditures at the national, provincial and sub-provincial level.

The Canadian Travel Survey was conducted by Statistics Canada in 1996 with the cooperation and support of the Canadian Tourism Commission and the ten provincial governments.

This document has been produced to facilitate the manipulation of the Beyond 20/20 files as well as the microdata files of the survey results.

Any questions about the data sets or their use should be directed to:

Sylvie Bonhomme
Tourism Statistics Program
Culture, Tourism and the Center for Education
Statistics
Statistics Canada
3-K, R.H. Coats Building
Tunney's Pasture
Ottawa, Ontario
K1A 0T6

Tel: 951-1672
Fax: 951-2909

Questions concerning the collection and processing
of these data should be directed to:

Jack Beauregard
Special Surveys Division
Statistics Canada
5D-7 Jean Talon Building
Tunney's Pasture
Ottawa, Ontario
K1A 0T6

Tel: 951-3328
Fax: 951-0562

1

INTRODUCTION TO CTS

Background

The tourism and travel industry is an important driving force in the Canadian economy. In 1996, tourism spending amounted to \$41.8 billion, up from \$39.9 billion in 1995 - an increase of 4.7%. The industry employed 491,900 people in 1996, a 2% increase over 1995; over the same period, there was little or no growth in employment in the overall business sector of the economy (Statistics Canada, Catalogue No. 13-220-XPB and Catalogue No. 13-009-XPB).

The Canadian Travel Survey, along with other surveys such as the International Travel Survey, are important sources of data to measure the impact and state of the tourism and travel industry in Canada.

The Canadian Travel Survey was developed to measure the volume, characteristics and economic impact of domestic travel by Canadians. In official publications, only those trips having a Canadian destination of 80 kilometres or more from home are included. However, the survey collects information for all overnight trips, regardless of distance, and for all same-day trips having a one-way distance from home of 80 kilometres or more (40 km in Ontario) regardless of the destination (Canada or other countries). The survey was first conducted in 1979, and, beginning in 1980, has been conducted every two years.

Changes to the sampling frame have been introduced in the last few years. From 1980 to the first quarter of 1992, the survey had been conducted on a quarterly basis, that is the respondents were contacted at the end of the three-month period and had to recall all trips taken in that period. Because it was hard to remember all the trips taken in such a long period, starting at the second quarter of 1992, the same respondent was contacted every month in that three-month period. This led to trips being under reported by the second and third months of collection. To improve the data quality, in 1994, the collection method was changed again to have a new sample of respondents every month.

However, the improvement in the collection method represents a break in the series. Even though the 1992 data has been adjusted to be comparable with the historical series since 1980, the 1994 data have not been and therefore are not comparable with those of previous years. The 1996 data, however, is comparable to the 1994 data.

Other changes concerning the content of the survey questionnaire have also been introduced. For example, for the first time in 1996, the survey collected information on visits at the subprovincial level. Previously, in terms of geographic information, only origin, destination and number of nights spent in each province were available. With the introduction of the visit concept, it is possible to keep track of every location visited on a particular trip. This means that the expenditure data can be reallocated at different geographic levels. Before

1996, the money reported as having been spent during the trip could only be redistributed at the provincial level. Because we now have information on each geographic area visited, the expenditures can be reallocated at the sub-provincial level.

These changes came about after the Tourism Research Working Group (TRWG) made recommendations. This group has been formed so that the Canadian Tourism Commission (CTC) as well as the provinces' could participate in recommending changes to the survey.

The Canadian Travel Survey is a federal-provincial project. Until 1990, Tourism Canada (now the Canadian Tourism Commission) assumed the cost of the basic survey and the provinces shared the cost of increasing the sample size. Since 1992, the cost of the basic survey has been shared by the CTC and the provincial governments.

Objectives

The Canadian Travel Survey's general objective is to provide measures of the volumes and characteristics of Canadian travellers and the domestic trips they take. Its purpose is to gather information on these journeys and those who take them.

Specific survey objectives are to:

- provide estimates of the volume of travel by Canadians to Canada by origin and destination, mode of transportation used and trip purpose;

- provide estimates of the volume of visits by Canadians to Canada by origin and location visited, length of stay and types of accommodation used;
- provide a picture of the socio-demographic characteristics of travellers and non-travellers;
- provide estimates of the expenditures on travel by category of expenditure, e.g., transportation, accommodation, food, clothing, etc.
- provide estimates of the contribution of domestic demand for touristic goods and services on the national and provincial economies.

CTS products

The results of the survey are published in *Touriscope - Domestic Travel* (Catalogue 87-504) which is prepared by the Tourism Statistics Program of Statistics Canada. The publication contains a selection of statistics on travel by Canadians within Canada. It is a valuable reference document for various players in the tourism industry, because it

- contains the most current statistics on domestic tourism;
- indicates provincial differences in terms of trips and expenditures by tourists;
- identifies market segments;
- gives an idea of potential uses of the CTS database.

Statistical profiles and customized products may also be obtained from the subdivision. Microdata files are also available for users who wish to produce their own outputs. The Beyond 20/20 software is also available with the microdata files for easier manipulation of data in tables and charts.

Other sources of tourism data include CANSIM, Statistics Canada's machine-readable database and retrieval system. As well, analytical articles regarding the Canadian Travel Survey and other tourism data can be found in *Travel-log* (Catalogue 87-003), a quarterly publication produced by the Tourism Statistics Program.

Uses of the CTS data

The CTS data is used in different ways. The main users of the data are the Canadian Tourism Commission, the provinces and tourism boards. Other users include the media, businesses, consultants as well as researchers. They refer to the data to better understand the habits of Canadian travellers as well as their socio-demographic characteristics. The survey allows them to get information on their destinations, the purpose of their trips, the length of their stay, the places they visited, the activities they participated in as well as the expenditures involved in the trip.

The survey also provides input into the Tourism Satellite Account (TSA). This is a structured information system that allows the statistics describing all the measurable aspects of tourism to be collected, classified and linked. It provides a measurement of the economic demand for, and

production of, tourism commodities. It makes it possible to evaluate the direct contribution of tourism expenses to Canada's Gross Domestic Product (GDP). For further information on the TSA, refer to the winter 1997 issue of *Travel-log* (Catalogue 87-003) which is prepared by the Tourism Statistics Program of Statistics Canada. One can also refer to the list of other tourism data sources included on the CD-ROM to get more specific information on the TSA and on other data on tourism.

2

CONCEPTS AND DEFINITIONS

This chapter outlines concepts and definitions of interest to the users. The concepts and definitions used in the Labour Force Survey (LFS) are described in the first part while those specific to the Canadian Travel Survey are given in the next part. Users are referred to the CTS questionnaire (included with the CTS documentation on CD-ROM) for an electronic copy of the computer-assisted interviewing questionnaire.

LFS concepts and definitions

- **Target population.** Civilian, non-institutionalised population 15 years of age and over in Canada's ten provinces. Specifically excluded from LFS coverage are residents of the Yukon and Northwest Territories, persons living on Indian Reserves, full-time members of the Canadian Armed Forces and persons living in institutions (e.g., inmates of penal institutions and patients in hospitals or nursing homes who have resided in the institution for more than six months); these groups together represent an exclusion of approximately 2% of the Canadian population aged 15 and over.
- **LFS reference week.** Entire calendar week (Sunday to Saturday) covered by the LFS each month. It is usually the week containing the 15th day of the month.

- **Labour Force Status.** Status of the respondent in the labour market, that is **employed**, **unemployed** or **not in the labour force**.

Employed—did any work at all at a job or business during reference week (excluding unpaid housework and volunteer work) or absent from a job for a reason other than layoff.

Unemployed—not employed (but available for work) during reference week and looked for work at some time during the month OR absent from a job due to temporary layoff OR had a new job to start within four weeks of reference week.

Not in the labour force—unwilling or unable to offer or supply labour services, that is not employed or unemployed during reference week.

Persons are regarded as available for work if they reported that they could have worked in the reference week if a suitable job had been offered (or they had been recalled if on temporary layoff), or if the reason they could not take a job was of a temporary nature such as own illness or disability, personal or family responsibilities, vacation or they already had a job to start in the near future. Full-time students currently attending school and looking for full-time work are not considered to be available for work during the reference week since they are

assumed to be looking for a job to start sometime in the future.

- **Class of worker.** Category of employer-employee relationship or self-employment for the respondent's job or business. If the respondent did not have a job during the reference week, the information is collected for the last job held, providing the respondent worked within the previous twelve months.

Paid worker - private sector—employee of a private firm or business, or self-employed with an incorporated business.

Paid worker - public sector—employee of a local, provincial or federal government, a government service or agency, a Crown corporation, or a government owned public establishment such as a school or a hospital.

Self-employed - unincorporated with employees—self-employed with an unincorporated business (i.e., a business, farm or professional practice that is not incorporated) or without a business (e.g., baby-sitters, newspaper carriers), with employees.

Self-employed - unincorporated without employees—self-employed with an unincorporated business (i.e., a business, farm or professional practice that is not incorporated) or without a business (e.g., baby-sitters, newspaper carriers), without employees.

Unpaid family worker—person working without pay on a farm or in a business or professional practice owned and operated by a related member of the same household.

- **Industry.** General nature of the business carried out by the employer for whom the respondent works. If the respondent did not have a job during the reference week, the information is collected for the last job held, providing the respondent worked within the previous twelve months. This variable is not available on the Beyond 20/20 files or on the microdata files but can be used in special requests.
- **Occupation.** Kind of work the respondent was doing during the LFS reference week, as determined by the kind of work reported and the description of the respondent's most important duties. If the respondent did not have a job during the reference week, the information is collected for the last job held, providing the respondent worked within the previous twelve months. This variable is not available on the Beyond 20/20 files or on the microdata files but can be used in special requests.
- **Household income.** Total household income, before taxes and deductions, including income from wages, salaries, tips, commissions, pensions, interest, rents, etc. for all household members, for the year preceding the reference year.

CTS concepts and definitions

- **Educational attainment.** Highest level of schooling completed by the respondent.
- **Relationship to the family reference person.** Relationship of the respondent to the person who has been identified as the family reference person (who is normally an adult with responsibility for the care or support of the family).
- **CTS reference month.** Month that ended before the collection period. For example, if the collection period is in February, we are collecting information on trips that ended in January. The collection period for the CTS includes the two weeks following the LFS reference week.
- **Trip.** For the purposes of the CTS, a trip is defined as travel by the respondent accompanied or not by one or more household members for any reason (except as noted below) to a Canadian destination of at least 80 km one-way from home. The following types of travel are excluded: travel to and from work or school (i.e., commuting); one way travel involving a change of residence; travel of operating crew members of buses, airplanes, boats, etc.; travel in an ambulance to a hospital or clinic; trips that did not originate in Canada; trips longer than a year. Only the microdata files include all trips regardless of destination or distance in the

case of overnight trips and same-day trips from Ontario of at least 40 km.

- **Household-trip.** Trip involving one or more persons from the same household leaving and returning together. If four persons from the same household go on a trip together, that is they leave and return together, it counts as one household-trip. If the same household takes two trips, it counts as two household-trips.
- **Person-trip.** Trip taken by one person. If this person took more than one trip and/or travelled with other members of the same household, we will count as many person-trips as there are trips and persons who took these trips. If four persons from the same household go on a trip together, it counts as four person-trips. If the same person takes two trips, it counts as two person-trips.
- **Person-night.** Night spent away from home by a person taking a trip. If two persons take a trip involving three nights away from home, there is a count of six person-nights.
- **Same-day trip.** Trip which does not include a night away from home.
- **Overnight trip.** Trip that includes at least one night away from home.
- **Visit.** Place of overnight stay of the traveller or for a same-day trip, the location of destination

of the trip. If the respondent travelled twice to the same location during the same trip, only one visit is recorded at that location.

- **Person-visit.** Visit taken by a traveller either single or travelling as a group. If four persons go on a visit together, it counts as four person-visits.
- **Origin.** Starting point of a trip is the respondent's place of residence.
- **Destination.** Place reported by the respondent as the destination of a trip. If a respondent visited more than one place during a trip, the destination is the place furthest from their home.
- **Trip distance.** Refers to the one way distance between the trip origin and its destination expressed in kilometres.
- **Traveller.** Any person who completes a trip. Any person who does not take a trip may be described as a *non-traveller*.
- **Intraprovincial travel.** Travel within a province by its residents, e.g., Ontario residents travelling in Ontario.
- **Interprovincial travel.** Travel by the resident of one province to and in another province, e.g., Ontario residents travelling elsewhere in Canada.

- **Non-resident travel.** Travel by residents from another province to a specific province, e.g., travel of Canadians except Ontario residents to Ontario.
- **Census Metropolitan Area (CMA).** Main labour market area of an urbanized core having 100,000 or more population. For more details, refer to the Standard Geographical Classification SGC 1991.
- **Census Division (CD).** Areas established by provincial laws which are intermediate geographic areas between the municipality (Census Subdivision) and the province level. Census Divisions represent counties, regional districts, regional municipalities and other types of provincially legislated areas. For more details, refer to the Standard Geographical Classification SGC 1991.
- **Tourism region.** From a touristic point of view, homogeneous regions defined by each province. See map included with the CTS documentation.
- **Mode of transportation.** Mode of transportation used to travel the greatest distance during a trip; if two modes of transportation were used to travel *equal* distances, the mode used to reach the destination is recorded.
- **Purpose.** Main reason the respondent to the survey went on a trip, regardless of the reason

anyone else from the household had for taking the same trip. Categories included in this variable are the following: visiting friends and relatives, pleasure, personal, business and non-business convention.

- **Package deal.** Package deals cover the expenses of two or more components of a trip but in most cases the costs are not identified separately. These items may be a combination of transportation and accommodation or accommodation with food and beverages, etc. A package deal is usually purchased from a travel agency or social organization.

- **Expenditures.** Trip expenditures made by all household members who went on the trip. Expenditures may have been paid for by individuals, by government or by a private sector business. In the CTS, the expenditures are broken down in the following categories: vehicle rental, vehicle operation, local transportation, intercity transportation, food or beverages purchased at restaurants or bars, food or beverages purchased at stores during the trip, accommodation, recreation and entertainment, clothing and other expenditures. The following items are excluded:
 - ▶ food purchased before the trip for use while on the trip;
 - ▶ items purchased to be resold or used in business (including items used on farm);
 - ▶ vehicles such as cars, caravans, boats;
 - ▶ capital investments such as real estate, works of arts, rare articles and stocks;

- ▶ cash given to relatives or friends during a holiday trip which does not represent payment of tourism goods or services, as well as donations made to institutions.
- **Reported expenditures.** Reported expenditures are the expenditures made by the respondent and other household members who went on the trip. Expenditures are reported in Canadian dollars and include all taxes and tips. As described above, these expenditures are reported in different categories. They do not represent where money was spent on a specific trip, but rather how much money was spent, and in which category of expenditures. It is associated with a trip as a whole.
- **Reallocated expenditures.** Refers to the process by which the respondent's reported expenditures are *redistributed* to specific geographic regions where money was spent. The CTS uses an expenditure reallocation model by which money included in each expenditure category is redistributed to a geographic region according to specific rules. Generally, money is allocated at the destination location in the case of same-day trips and redistributed according to the number of nights spent in each geographic region in the case of overnight trips. However, this rule may vary according to the expenditure item (for more details, refer to [Chapter 5](#)). Once redistributed, the expenditures are referred to as reallocated expenditures.

The level of geography for which the expenditures are redistributed in the CTS are the following: province, tourism region, Census Metropolitan Area (CMA) and Census Division (CD).

Here is an example of the reallocation process for accommodation expenditures declared by a respondent going from Montreal to Vancouver using his own car, and sleeping in motels in Sault-Ste-Marie, Winnipeg, Calgary and Vancouver. The expenditures will be redistributed according to the number of nights spent in each location, that is, in Ontario, Manitoba, Alberta and British Columbia at the provincial level.

- **Vehicle rental expenditures.** Includes all operation, rental or users' fees encountered for the use of any vehicle such as an automobile, a truck, motorcycle, bicycle, boat, motor home, snowmobile, etc. (including insurance).
- **Vehicle operation expenditures.** It includes expenditures related to the operation of a private or rental vehicle, namely gasoline, repairs and parking costs while on the trip.
- **Local transportation expenditures.** Local transportation means within a city or metropolitan area, that is, intracity transportation. It includes the cost of taxis, city bus fares, subway fares and such things as bus

tours in the place of visit and/or destination while on the trip.

- **Intercity transportation expenditures.** Includes expenses between cities or metropolitan areas to get to or from the destination. These transportation expenditures include ticket fares for intercity aircrafts, boats, hovercrafts, trains, buses as well as intercity ferries.
- **Food or beverages purchased at restaurants or bars.** Includes the total amount spent on meals and drinks purchased from restaurants, bars, cafeterias, fast food take out counters, and minibars located in some hotel/motel rooms, regardless where they were consumed.
- **Food or beverages purchased at stores during the trip.** Includes the total amount spent at local stores while on the trip, regardless of where they were consumed. For example, groceries purchased to bring home or to eat during the trip are included in this category.
- **Accommodation expenditures.** Total expenditures for nights spent in hotels, motels, resorts, cabins, cottages, campgrounds, etc. This primarily includes rental fees. However, in the case of trips to private cottages or the home of friends or relatives, it could also include any money given to owners for the use of their accommodation.

- **Recreation and entertainment expenditures.** Includes the cost of admission to theatres, sports events, art galleries, nightclubs, etc. It also includes gambling expenses and sporting activity costs such as daily ski lift passes, green fees, rental of sports equipment, entrance fees to a cultural or leisure activity, boat and bus tours but excludes recreation and entertainment purchased for the season (e.g., season ski passes).

- **Clothing expenditures.** Includes any clothing purchases whether they are gifts or for personal use.

- **Other expenditures.** Includes items such as souvenirs, shoes, belts, jewellery, handbags and household items, registration fees for courses or conferences, customs duties, postcards and stamps, insurance fees, medication, books, craft supplies, film and developing, costs for child care, telephone charges, etc.

The Canadian Travel Survey is administered to one or more sub-samples of households in the Labour Force Survey (LFS) sample, and therefore its sample design is closely tied to that of the LFS. Except for differences in sample size, the design features of the LFS sample are maintained in the CTS sample. Non-response to the LFS also carries over to CTS except in those cases where the LFS interviews could not be conducted in the current reference month because of temporary circumstances and when in an earlier month, the LFS data was collected. The LFS has response rates of 95% or higher. The LFS design is briefly described in this chapter. The last part of the chapter describes how the Canadian Travel Survey departed from the basic LFS design in 1996.

LFS population coverage

The LFS is a monthly household survey whose sample of individuals is representative of the civilian, non-institutionalised population 15 years of age or older in Canada's ten provinces. Specifically excluded from the survey's coverage are residents of the Yukon and Northwest Territories, persons living on Indian Reserves, full-time members of the Canadian Armed Forces and inmates of institutions. These groups together

represent an exclusion of approximately 2% of the population aged 15 or over.

LFS sample design

The LFS has undergone an extensive redesign culminating in the introduction of the new design at the end of 1994. The LFS sample is drawn from an area frame and is based on a stratified, multi-stage design that uses probability sampling. The successive stratification and sampling stages are outlined below. A detailed description of the previous LFS design can be obtained in *Methodology of the Canadian Labour Force Survey*, Statistics Canada Catalogue No. 71-526. A diagram summarizing the design stages appears at the end of this section.

The main advantage of a stratified sample design is that when sampling is applied separately to many diverse strata, each stratum having a relatively homogeneous population, the efficiency of the sample is higher than that obtained from simple random sampling without stratification.

Primary stratification

First, each province is divided into economic regions (set by federal-provincial agreements) and employment insurance regions. Economic regions (ERs) are geographic areas of more or less homogeneous economic structure and are relatively stable over time. Employment insurance economic regions (EIERS) are also geographic areas, and are roughly the same size and number as ERs, but they

do not share the same definitions. Labour force estimates are produced for the EIER regions for the use of Human Resources Development Canada.

The intersections of the two types of regions form the first level of stratification for the LFS. These ER/EIER intersections are treated as primary strata and further stratification is carried out within them. Note that a third set of regions, Census Metropolitan Areas (CMAs) is also respected by stratification in the current LFS design, since each CMA is also an EIER.

Types of areas

Each ER/EIER intersection is in turn divided into rural, urban and remote areas. Urban and rural areas are loosely based on the Census definitions, with some exceptions to allow for the formation of strata in some areas. Urban areas include the largest CMAs down to the smallest villages categorized by the 1991 Census as urban (1 000 people or more), while rural areas are made up of areas not designated as urban or remote.

All urban areas are then further subdivided into two types, that is those using an apartment list frame as well as an area frame, and those using only an area frame.

Approximately 1% of the LFS population is found in remote areas of provinces which are less accessible to LFS interviewers than other areas. For administrative purposes, this portion of the

population is sampled separately through the remote area frame. Some populations, not congregated in places of 25 or more people, are excluded from the sampling frame.

Secondary stratification

In urban areas with sufficiently large numbers of apartment buildings, the strata are subdivided into apartment frames and area frames. The apartment list frame is a register which is based upon information supplied by Canadian Mortgage and Housing Corporation and is maintained in the 18 largest cities across Canada. The purpose of this is to ensure better representation of apartment dwellers in the sample, as well as to minimize the effect of growth in clusters due to construction of new apartment buildings. In the major cities, the apartment strata are further stratified into low income strata and regular strata.

Where it is possible and/or necessary, the urban area frame is further stratified into regular strata, high income strata, and low population density strata. Most urban areas fall into the regular urban strata, which, in fact, cover the majority of Canada's population. High income strata are found in major urban areas, while low density urban strata consist of small towns that are geographically scattered.

In rural areas, the population density can vary greatly from relatively high population density areas to low population density areas, resulting in

the formation of strata that reflect these variations. The different stratification strategies for rural areas were based not only on concentration of population, but also on cost-efficiency and interviewer constraints.

In each province, remote settlements are sampled proportional to the number of dwellings in the settlement, with no further stratification taking place. Dwellings are selected using systematic sampling in each of the places sampled.

Cluster delineation and selection

Households in final strata are not selected directly. Instead, each stratum is divided into clusters, and then a sample of clusters is selected within the stratum. Dwellings are then sampled from selected clusters. Different methods are used to define the clusters, depending on the type of stratum.

Within each urban stratum in the urban area frame, a number of geographically contiguous groups of dwellings, or clusters, are formed based upon 1991 Census counts. These clusters are generally a set of one or more city blocks or block faces. The selection of a sample of clusters (always 6 or a multiple of 6 clusters) from each of these secondary strata represents the first stage of sampling in most urban areas. In some other urban areas, Census Enumeration Areas (EAs) are used as clusters. In the low density urban strata, a three-stage design is followed. Under this design, two towns within a stratum are sampled, and then six or 24 clusters within each town are sampled.

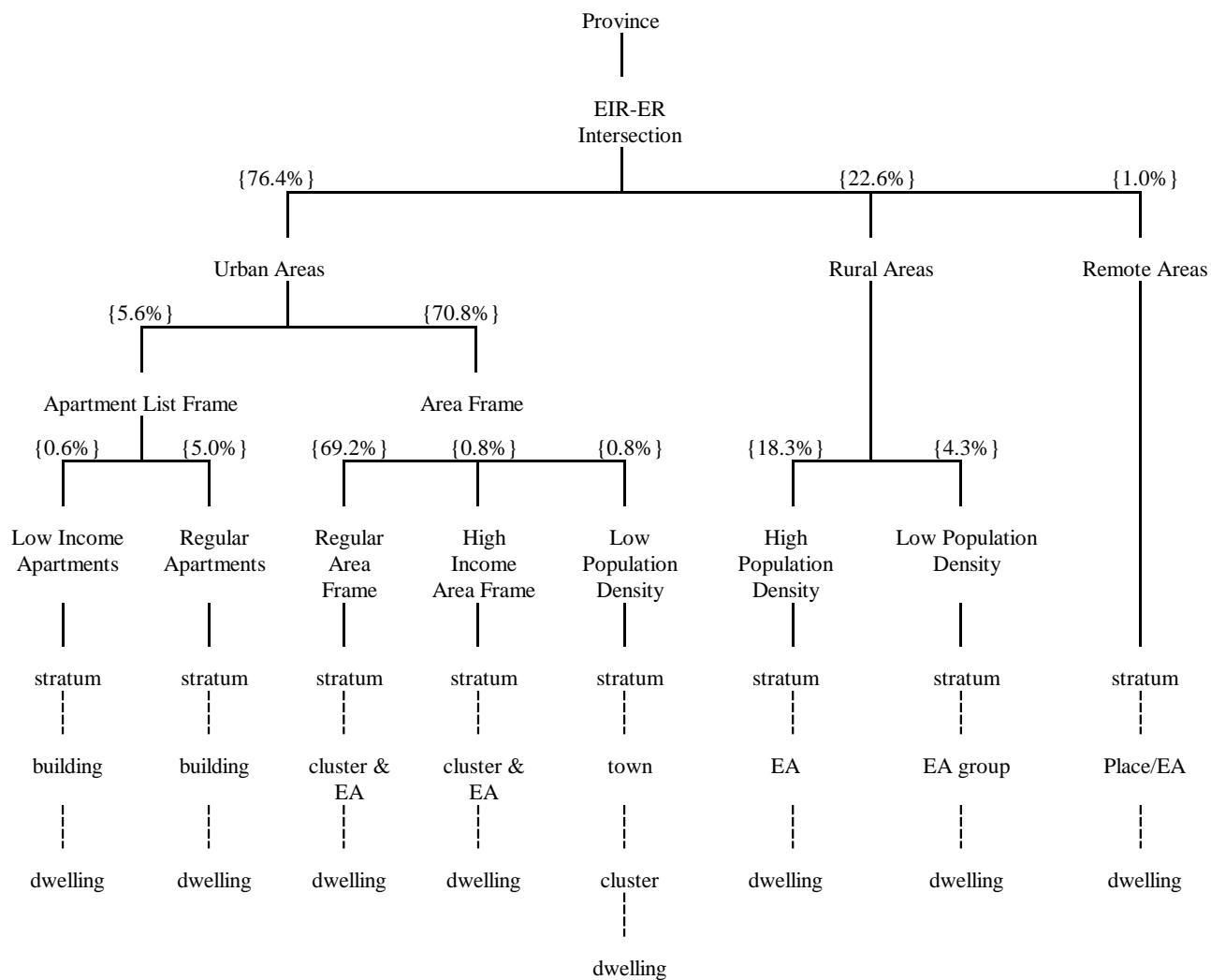
For urban apartment strata, instead of defining clusters, the apartment building is the primary sampling unit. Apartment buildings are sampled from the list frame with probability proportional to the number of units in each building.

Within each of the secondary strata in rural areas, where necessary, further stratification is carried out in order to reflect the differences among a number of socioeconomic characteristics within each stratum. Within each rural stratum, six EAs or two or three groups of EAs are sampled as clusters.

Dwelling selection

In all three types of areas (urban, rural and remote areas), selected clusters are visited by enumerators in the field to prepare a listing of all private dwellings in the cluster. From the listing, a sample of dwellings is then selected. The sample yield depends on the type of stratum. For example, in the urban area frame, sample yields are either 6 or 8 dwellings, depending on the size of the city. In the urban apartment frame, each cluster yields 5 dwellings, while in the rural areas and EA parts of cities, each cluster yields 10 dwellings. In all clusters, dwellings are sampled systematically. This represents the final stage of sampling.

Labour Force Survey Sample Design - 1995+



| = level of stratification

EIR - Employment Insurance Region

EA - Census Enumeration Area

ER - Economic Region

cluster - set of blockfaces

{%} - percentage of total sample

| = stage of sampling

Person selection

Demographic information is obtained for all persons for whom the selected dwelling is the usual place of residence. LFS information is obtained for all civilian household members 15 years of age or older. Response burden is minimized for the elderly (70 years of age or older) by carrying forward their responses for the initial interview to the subsequent five months in the survey.

LFS sample size

The sample size of eligible persons in the LFS is determined so as to meet the statistical precision requirements for various labour force characteristics at the provincial and subprovincial level, to meet the requirements of federal, provincial and municipal governments as well as a host of other data users.

The monthly LFS sample consists of approximately 59,000 dwellings. After excluding vacant dwellings, dwellings demolished or converted to non-residential uses, dwellings containing only ineligible persons, dwellings under construction, and seasonal dwellings, about 52,350 dwellings remain which are occupied by eligible persons. From these dwellings, LFS information is obtained for about 102,000 civilians aged 15 or over.

LFS sample rotation

The LFS employs a panel design whereby the entire monthly sample of dwellings consists of six panels, or rotation groups, of approximately equal size. Each of these panels can be considered by itself to

be representative of the entire LFS population. All dwellings in a rotation group remain in the LFS sample for six consecutive months after which time they are replaced (rotated out of the sample) by a new panel of dwellings selected from the same or similar clusters.

This rotation pattern was adopted to minimize any problems of non-response or respondent burden that would occur if households were to remain in the sample for longer than six months. It also has the statistical advantage of providing a common sample base for short-term month-to-month comparisons of LFS characteristics.

Because of the rotation group feature, it is possible to readily conduct supplementary surveys using the LFS design but employing less than the full size sample.

CTS as a supplement

The Canadian Travel Survey used two of the six rotation groups in the 1996 LFS sample in all provinces except in Newfoundland where the sample was augmented to three rotation groups. For the Canadian Travel Survey, the coverage of the LFS remains the same, that is, it includes all members of the household aged 15 and over. However, unlike the LFS where information is collected for all eligible household members, the Canadian Travel Survey only collects information from one household member and proxy responses are not permitted.

For further information on the CTS sample size by month and by province, refer to the Data Quality Report ([Chapter 10](#)).

4

DATA COLLECTION

LFS data collection

Data collection for the LFS is carried out each month during the week following the LFS reference week, usually the third week of the month.

Statistics Canada interviewers, who are part-time employees hired and trained specifically to carry out the Labour Force Survey, contact each of the sampled dwellings to obtain the required labour force information. Each interviewer contacts approximately 70 dwellings per month.

Dwellings new to the sample are contacted through a personal visit. The interviewer first obtains socio-demographic information for each household member and then obtains labour force information for all eligible members. Provided there is a telephone in the dwelling and permission has been granted, subsequent interviews are conducted by telephone. As a result, approximately 85% of all dwellings are interviewed by telephone. In these subsequent monthly interviews, as they are called, the interviewer confirms the socio-demographic information collected in the first month and collects the labour force information for the current month.

Collection by computer-assisted interviewing (refer to [Chapter 4](#), CTS data collection) is decentralised: interviewers generally work at home. They carry out the interview by telephone or at the

respondents' home and transmit their data by modem to the nearest Statistics Canada regional office. For security reasons, the data are encrypted during transmission.

In all dwellings, information about all household members is obtained from a knowledgeable household member - usually the person at home when the interviewer calls. Such 'proxy' reporting, which accounts for approximately 55% 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.

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.

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.

LFS interviewers

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 eight Statistics Canada regional offices.

Non-response to the LFS

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. For a supplementary survey such as the Canadian Travel Survey, if the LFS demographics section is not completed then, the CTS cannot be generated.

CTS data collection

The Canadian Travel Survey uses computer-assisted interviewing (CAI), i.e., the questionnaire is designed for a laptop computer and the interviewer keys in reported information during the interview. Some of the benefits of CAI come from on-line editing of information as it is received (refer to [Chapter 5](#) on data processing). This collection method also helps in avoiding errors associated with the electronic transcription of data written on the paper questionnaire.

The survey is administered to one randomly selected individual per household. Only old respondents to the LFS who were already interviewed at least one month before, can be selected as respondents to the CTS. The CTS program automatically selects the respondent from the LFS demographics component. It prefills the respondents first and last names. Because of the possibility of duplicate names within the household and because some names are gender-neutral, it also indicates the age and sex of the respondent.

Upon completion of the Labour Force Survey interview, the interviewer asks to speak to the selected person for the Canadian Travel Survey. If the selected person is not available, the interviewer arranges for a convenient time to phone back. Proxy response is not allowed, hence the collection period is extended by one week to allow the interviewers time to contact the selected individuals.

In some cases the interviewers will not interview the selected CTS individuals with the following characteristics:

- ▶ incapacity (for example, deafness, mental retardation, etc.);
- ▶ language difficulties (a person speaks a language other than English or French and no one in the household is available to translate);
- ▶ respondents away at school or on business (outside the local dialing area);
- ▶ respondents who are in a hospital;
- ▶ respondents who move out of the household between the time the LFS demographic component was completed and the CTS is completed.

Non-response to the CTS

For all households responding to the LFS, the next stage of data collection is to administer the Canadian Travel Survey. Overall, the response rate for the CTS is around 90%. Response rates for the CTS reflect the proportion of respondents eligible for the Canadian Travel Survey who has reported information. These response rates are not cumulative, that is, they don't take into account those people who would have been eligible for CTS but have been non-respondents to LFS. This is because those individuals who don't respond to the LFS are not even asked if they would like to answer

the Canadian Travel Survey. Thus, they cannot be considered as non-respondents to the CTS. More detailed information on response rates is presented in the Data Quality Report ([Chapter 10](#)).

5

DATA PROCESSING

The main output of the Canadian Travel Survey is a set of "clean" microdata files. This chapter presents a brief summary of the processing steps involved in producing these files.

Data capture

Capture of survey data is accomplished as the interview is done. Computer-Assisted Interviewing (CAI) allows the interviewers to key in the data as the respondent reports it. After the interview is completed, the data are electronically transmitted to regional offices for further processing.

As the data are captured, on-line editing is done. CAI controls the sequence or flow of questions resulting from previous question responses, so that problems caused by incorrect flows are minimized.

Other checks are also done during the interview to reduce the number of errors resulting from typing mistakes or misunderstandings. For example, if the number of nights spent in different types of accommodation does not add up to the total number of nights spent away from home, an edit message appears. The interviewer can then correct the mistake and less editing has to be performed at head office. Although some checks are done during the interview, there are almost no cases where the interviewer must correct the information; at worst,

the application will assign an incomplete code for that individual.

Preliminary editing

The data files transmitted from the regional offices are downloaded to the Special Surveys Division computers for processing.

The data are edited in a series of iterations to detect errors in the data and to prepare the data files for subsequent weighting and expenditure imputation. At this step of preliminary editing, duplicate records are detected, and some minor edits are performed.

Linking to the LFS

The CTS files received from the regional offices contain an identifier which is used to link the records to their respective Labour Force Survey information. Once the data have been unduplicated in the preceding step, records are linked to the LFS files to retrieve information required for weighting, to geo-code the respondents place of residence, to pick up demographic variables (age, education, industry, occupation, LFS status, etc.) and to derive the response status of each record.

Geographic coding

This step involves the geographic coding of the trip and location information to derive the province, Census Metropolitan Area, tourism region and Census Division of each origin, destination and location on the file.

The trip origin is the respondent's place of residence unless this has changed since the previous month (recall that the month of interview is the month following the CTS reference month). Information from the LFS is used to determine the place of residence.

The trip destination and locations (and the origin, in cases where the respondent has moved) consist of place names which were entered in the computer by the interviewer. These place names are matched to files containing all of the place names registered in the Standard Geographic Classification (SGC) and supplementary files containing the names of destinations outside Canada. Each place name on these files has a standard geographic code and can be matched to a tourism region code. In addition, the longitude and latitude can be determined for each geo-code - this is used to compute the distance between the origin and destination.

All records requiring destination and location coding have key information copied to coding files. Trip place names which match to the SGC files are automatically coded on these files. Place names which do not find a match are reviewed manually to determine an appropriate code. When all geocoding is complete, geocodes are linked back to the CTS data files. Once the files have been fully geo-coded, they proceed to the secondary editing process described next.

Secondary editing

Once linked, more editing is performed on the files. Gross errors, such as the numbers of nights away from home not matching in different questions, as well as errors in question flows, are reviewed and corrected.

Weighting

The edited files are then processed through the weighting system which derives appropriate response rates for each record and computes the person, household-trip and person-trip weights. Refer to [Chapter 6](#) for more information on weighting procedures and [Chapter 7](#) for more information on using the different CTS weights.

Consistency edits

This series of edits consists of more thorough validity checks to detect inconsistencies and outliers in the data. This step occurs after the origin and destination place names have been geo-coded. Once again, errors are reviewed and corrected.

Expenditure imputation

The expenditures are the only data in the CTS which are imputed. All other values which are found to be missing or in error are either corrected or converted to a “not stated” code. The expenditures, however, must be complete so that aggregate estimates of expenditures can be produced.

Missing expenditure information is imputed based on the average expenditures of respondents whose expenditure information was reported. These are

the “donor” respondents whose expenditures will be used to compute average expenditures used to impute the missing expenditure values.

Donor expenditure averages are calculated by type of trip since expenditures vary considerably depending on the characteristics of the trip. Averages are computed for trips which have the following characteristics in common: destination (Canada, U.S., other country); duration (overnight or same-day); number of persons on the trip; reason for the trip (business or non-business); type(s) of accommodation and mode of transportation used.

There must be at least three donor records for each imputation category; if not, the trip characteristics are collapsed, i.e., less specific, and a set of averages are calculated for this next level of trip types. If there are insufficient donors for this level, the trip characteristics are collapsed further and another set of averages is computed. This process is repeated until all levels of trip collapsing have donor averages computed.

For example, the first level of imputation may be for all trips to a Canadian destination with one or more nights duration, a party of two adults, hotel accommodation, mode of travel by air and business as the reason for the trip. If insufficient numbers of donors are available at this level, then the trip characteristics will be collapsed to include trips with any type of commercial accommodation; if sufficient donors are still not available, the

characteristics will be collapsed to include trips with any reason for the trip, and so on.

Once a set of donor averages has been computed for all levels of trips characteristics, the trips requiring imputation are then matched to the averages for trips with the same characteristics and the missing expenditures are calculated.

An additional step in the imputation process is the distribution of package deal expenditures to specific expenditure categories. This is accomplished in the same fashion as expenditure imputation: donor averages are used to impute the expected value of the expenditure items included in the package deal; these imputed amounts are then ratio-adjusted to arrive at the total amount reported for the package deal.

Imputed expenditures are then reedited to ensure that no outlier values have been created by the expenditure imputation process.

Reallocation of expenditures

Reallocation of expenditures is the process by which the respondent's reported expenditures are *redistributed* to specific geographic regions where money was spent. The CTS uses an expenditure reallocation model by which money included in each expenditure category is redistributed to a geographic region according to specific rules. Generally, money is allocated at the destination location in the case of same-day trips and redistributed according to the number of nights

spent in each geographic region in the case of overnight trips. However, this rule may vary according to the expenditure item (for more details, refer to tables 1 and 2 on [pages 48 and 49](#)). Once redistributed, the expenditures are referred to as *reallocated expenditures*.

In 1996, data providing more detailed information have been collected and allowed the redistribution of the reported expenditures at a sub-provincial level. By using the additional information available at the *locational level*, namely number of nights spent at a location, types of accommodation used and number of nights spent by type of accommodation, it was possible to refine the expenditure reallocation model for the following categories of expenditures: accommodation, food or beverages purchased at restaurants or bars and food or beverages purchased in stores during the trip.

Accommodation

In cases where the traveller used one or multiple accommodation types in one location, reported expenditures are allocated at the location, the same as in 1994.

In cases where the traveller used one accommodation type in multiple locations, reported expenditures are redistributed according to the number of nights spent in each location, the same as in 1994.

In cases where the traveller went to different locations and used different types of accommodation, the reallocation of expenditures model is more refined than it was in 1994. Knowing that the amount of money spent on accommodation depends on the type of accommodation (see table 3, [page 50](#)), it is possible to redistribute expenditures according to the number of nights spent in each location and the type of accommodation used.

Example of expenditure reallocation for accommodation

A traveller reports having spent \$350 in accommodation. He spent two nights in a hotel in Ottawa and three nights camping in Mont-Tremblant. Considering that on average, travellers spend \$72 / night when staying in a hotel and \$13 / night when camping, the money spent in accommodation can be redistributed as:

Average spending per night in hotel: $\$72 \times 2 \text{ nights} = \144
Average spending per night camping: $\$13 \times 3 \text{ nights} = \39
Total average amount spent: $\$183$

Actual amount spent: $\$350$
Amount spent in hotel: $(\$144 / \$183) * \$350 = \275.41
Amount spent camping: $(\$39 / \$183) * \$350 = \74.59

Using this approach, the Ottawa region would have \$275.41 allocated in accommodation and the Mont-Tremblant region would have \$74.59. Using the 1994 approach, the Ottawa region would have had \$140 whereas the Mont-Tremblant region would have had \$210.

***Food or beverages bought in restaurants or bars
and food or beverages bought in stores during the
trip***

Spending in these categories depends on the type of accommodation used. As a basic rule, travellers staying in commercial accommodation tend to spend more on food bought in restaurants or bars than travellers staying in cottages, camping or staying with friends or relatives (table 3).

The reallocation process for these two categories is the same as the one used for accommodation expenditures.

***Example of expenditure reallocation for food or
beverages bought in restaurants or bars***

A traveller reports having spent \$200 for food or beverages bought in restaurants or bars. He spent two nights in a hotel in Ottawa and three nights camping in Mont-Tremblant. Considering that on average, travellers spend \$34 / night on food or beverages in restaurants or bars when staying in a hotel and \$8.50 / night when camping, the money spent in accommodation can be redistributed as:

Average spending per day in hotel: $\$34 \times 2 \text{ nights} = \68
Average spending per day camping: $\$8.50 \times 3 \text{ nights} = \25.50
Total average amount spent: $\$93.50$

Actual amount spent: $\$200$
Amount spent in hotel: $(\$68 / \$93.50) * \$200 = \145.45
Amount spent camping: $(\$25.50 / \$93.50) * \$200 = \54.55

Using this approach, the Ottawa region would have \$145.45 allocated for food or beverages bought in restaurants or bars and the Mont-Tremblant region would have \$54.55. Using the 1994 approach, the Ottawa region would have had \$80 whereas the Mont-Tremblant region would have had \$120.

TABLE 1. *Allocation of expenditures at a sub-provincial level, same-day trips, 1996*

SAME-DAY TRIPS	Location of origin	Location of destination
Vehicle rental	X	
Vehicle operation	X	
Local transportation		X
Intercity transportation	X	
Food or beverages bought in restaurants or bars		X
Food or beverages bought in stores during the trip		X
Accommodation		X
Recreation and entertainment		X
Clothing		X
Other expenses		X

TABLE 2. *Allocation of expenditures at a sub-provincial level, overnight trips, 1996*

OVERNIGHT TRIPS	Loc. of origin	Loc. of desti- nation	Loc. of over- night stay
Vehicle rental - primary mode = automobile - primary mode = commercial	X (1)	X (1)	
Vehicle operation			X
Local transportation			X
Intercity transportation	X		
Food or beverages bought in restaurants or bars			X (2)
Food or beverages bought in stores during the trip			X (2)
Accommodation			X (2)
Recreation and entertainment			X
Clothing			X
Other expenses			X

- (1) *If main mode of transportation is airplane, train, buses or boat then vehicle rental expenditures are allocated to the province of destination. If main mode of transportation is a car or other vehicle, rental fees are allocated to the province of origin. This is the same as was done in 1994.*
- (2) *In cases of one accommodation type and multi-locations, money is allocated according to the number of nights spent in each location. In cases of mixed accommodation types and multi-locations, the amount of money redistributed would vary depending on the type of accommodation and the number of nights spent in each location.*

TABLE 3: *Average expenditures per type of accommodation, travellers who used only one type of accommodation, overnight travel, 1996*

	Accommodation	Food or bev. bought in restaurants or bars	Food or bev. bought in stores during the trip
	\$ per night	\$ per day	\$ per day
Hotel	72.02	34.20	9.67
Resort and Lodge	71.52	29.33	11.02
Commercial Cottage	57.81	15.69	9.54
Motel	48.81	26.08	6.79
Bed and Breakfast	45.52	22.15	9.33
Private Cottage	42.29	11.50	10.03
Camping and Trailer Park	13.03	8.50	8.21
Home of friends or relatives	11.47	12.84	4.50
Other	35.87	26.51	5.56
Not Stated	24.36	17.16	31.94

Source: Canadian Travel Survey, 1994.

Data review

The final weighted and imputed files are then subject to an analytical review to detect problems at the aggregate level. Any anomalies are corrected and the final files are then made available to the end users.

6

WEIGHTING

Since the Canadian Travel Survey used a sub-sample of the LFS sample, the derivation of weights for the survey records is clearly tied to the weighting procedure used for the LFS. The LFS as well as the CTS weighting procedures are briefly described below.

Weighting procedures for the LFS

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 balancing factor for non-response and the province-age-sex 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 the 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 .02 for each person and the records must be weighted by $1/.02=50$. 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.

Non-response

Notwithstanding the strict controls of the LFS, some non-response is inevitable, despite all the attempts made by the interviewers. The LFS non-response rate is approximately 5%. For certain types of non-response (e.g., 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, which are defined by employment insurance 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. To the extent that this assumption is not true, the estimates will be somewhat biased.

LFS 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.

Subprovincial 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. In particular, estimates are produced of the total number of persons 15+ in provincial economic regions and the 24 large metropolitan areas as well as of designated age-sex groups in each of the ten provinces.

Independent estimates are available monthly for various age and sex groups by province. These are population projections based on the 1991 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 various age-sex groups, economic regions, Census Metropolitan Areas, and rotation groups.

This weighting procedure ensures consistency with external Census counts and that each rotation group is representative of the population, and also ensures that every member of the economic family is assigned the same weight.

Weighting procedures for the CTS

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

- 1- An adjustment to account for the use of a two-sixths or three-sixths sub-sample, instead of the full LFS sample. Note that this adjustment may differ depending on the sample month and the province.
- 2- An adjustment to account for the additional non-response to the supplementary survey, i.e., non-response to the Canadian Travel Survey for individuals who did respond to the LFS or for which previous month's LFS data was brought forward.
- 3- An adjustment to account for all persons eligible for CTS within a household.

- 4- A readjustment to account for independent province-age-sex projections, after the above adjustments are made.
- 5- A readjustment to account for independent economic region-Census Metropolitan Area projections, after the above adjustments are made.

The first two adjustments are taken into account by multiplying the LFS sub-weight for each responding Canadian Travel Survey record by:

$$\frac{\text{sum of LFS subweights from each household responding to LFS}}{\text{sum of LFS subweights from each household responding to CTS}}$$

to obtain a non-response adjusted Canadian Travel Survey sub-weight.

The CTS sub-weight is further adjusted by multiplying it by the number of eligible persons for the CTS within a household (WEIGHT1).

Adjustment 4 is calculated by multiplying WEIGHT1 for each Canadian Travel Survey respondent by:

$$\frac{\text{population total for province-sex-age group } i}{\text{sum of WEIGHT1 for survey respondents in province-sex-agegroup } i}$$

The resulting weight (WEIGHT2) is then used in adjustment 5 by multiplying by another factor

$$\frac{\text{population total for ER-CMA } j}{\text{sum of WEIGHT2 for survey respondents in ER-CMA } j}$$

Steps 4 and 5 are repeated iteratively until the adjustment factors converge to 1.

The resulting person weight (called PERSONWT or _PERSWT) is the final weight which appears on the Canadian Travel Survey microdata files. This weight is then used to calculate the person-trip weight (PTRIPWT or _PTRIPWT) and the household-trip weight (called HHTRIPWT or _HHTRWT) and .

Person-Trip and Household-Trip Weighting

Once the person weight has been created, the person-trip weight and household-trip weight are calculated. For each trip, the person weight of the respondent who took the trip is multiplied by the sum of the number of identical trips taken + 1, resulting in the intermediate weight. This accounts for the trip that was described by the respondent, as well as other trips the respondent reported taking which were identical to the described trip. To calculate the person-trip weight, the intermediate weight is adjusted by the following ratio: total number of trips that the respondent reported, divided by the number of trips that the respondent described. The adjustment of trips depends on the type of trip: same-day or overnight. For example, the weight for a same-day trip is adjusted by the ratio of the total number of same-day trips that the respondent reported, divided by the number of same-day trips that the respondent described. This adjustment is necessary since not all respondents provide descriptions of all of the trips that they

reported taking. For instance, a respondent may report having taken five trips, yet after providing details for three trips, may decline to provide details for the remaining two trips. Finally, in order to calculate the household-trip weight, the person-trip weight is simply divided by the number of adults (including the respondent) who went on the trip.

The person-trip weight is calculated in the following manner:

$$\frac{\text{PERSON WEIGHT} * (\text{number of identical trips} + 1) * (\text{number of reported trips})}{\text{number of described trips}}$$

The household-trip weight is calculated by adjusting the person-trip weight in the following way:

$$\frac{\text{PERSON-TRIP WEIGHT}}{\text{number of adults from same household who went on the trips}}$$

The weights are then used to produce the appropriate estimates. A description of the different weights and their uses is provided in [Chapter 7](#) (Using the Microdata).

7

USING THE MICRODATA

File description

The Beyond 20/20 microdata files contain three different *annual* datasets with one being provided at four different geographic levels, that is, the visit dataset provided at the province, Census Metropolitan Area, tourism region and Census Division level. It also contains a person and person-trip dataset. The information provided on those files is for trips of 80 km or more with a Canadian destination.

The CTS microdata files contain four different *monthly* datasets with one being provided at four different geographic levels, that is, the person-visit dataset provided at the province, Census Metropolitan Area, tourism region and Census Division level. It also contains a person, trip and person-trip dataset. The information provided on those files is for same-day trips of 80 km or more (40 km in Ontario) and all overnight trips regardless of distance. Following is a description of each dataset.

Person dataset

The person dataset provides information on travellers and non-travellers, a traveller being a person who takes at least one trip during the reference month (or the reference year) and a non-traveller being someone who does not take a trip.

Each respondent to the CTS has one record on the person file whether or not that person went on a trip during the reference month or the reference year. The sum of all travellers and non-travellers adds up to the non-institutional Canadian population aged 15 and over on each monthly file and on the annual file (December supplementary section, see below).

The person microdata file includes socio-demographic information on both travellers and non-travellers. It is useful to produce socio-demographic profiles and to calculate travel incidences. For example, what is the age, sex or educational attainment of travellers versus non-travellers? What percentage of the population 55 years of age and over travelled?

The person microdata file *does not* provide information on the volume of trips or person-trips taken but rather on the volume of travellers and non-travellers. If a person travelled more than once during the reference period, that person will be counted only once as a traveller.

When using the person file and compiling travellers' information (alone or linked with the TRIP file), *only* the person weight can be used.

From the annual dataset provided with the Beyond 20/20 software and on the microdata file (supplementary information included on the December person dataset), it is possible to get annual socio-demographic information on travellers

and non-travellers from this file. That information is derived from an additional question included in the December questionnaire. The question is as follows:

DURING THE ELEVEN MONTH PERIOD FROM JANUARY 1 TO NOVEMBER 30, (survey year), DID YOU TAKE ANY NON-BUSINESS TRIPS OF ONE NIGHT OR MORE TO A DESTINATION...

- 1 WITHIN THE PROVINCE?
- 2 TO SOME OTHER PROVINCE?
- 3 TO THE UNITED STATES?
- 4 TO A FOREIGN COUNTRY OTHER THAN THE UNITED STATES?

This question for which results appear in the variables YRSAMEPR (the person took at least one intraprovincial trip), YRCANADA (interprovincial trip), YRUSA (trip to the United States) and YROTHER (trip to a foreign country other than the United States) gives users the information determining if a respondent has travelled to specific destinations within the reference year. Microdata users have to combine the results from this question with the information collected for the month of December (to know about trips taken in December) to determine if a respondent travelled during the complete reference year. Results for the complete reference year are already combined under these same variables for Beyond 20/20 users.

Except for the month of December, the monthly person datasets are not included on the Beyond 20/20 microdata files.

For Beyond 20/20 users

To define if a person is a domestic traveller or a non-traveller one should use the variables YRSAMEPR and YRCANADA. If these variables are NOs, the person is a non-traveller within Canada, otherwise, that person is considered as a domestic traveller.

For microdata files users

To define if a person is a traveller or a non-traveller one should use the variable TOTTRIP (total number of trips). If TOTTRIP is 0, the person is a non-traveller, otherwise, that person is considered as a traveller, regardless of the number of trips taken. This variable includes all overnight trips regardless of distance and same-day trips of 80 km or more from home (40 km in Ontario) to any destination.

Besides the annual travellers' information provided on the December dataset, microdata file users also have access to the monthly travellers' information. Available are twelve monthly person datasets providing socio-demographic information on travellers and non-travellers for specific reference months.

Each person dataset can be used on a *monthly* basis *only* because the question asked in the CTS relates to trips taken within the last month *only*. *Contrary to the information included in the other datasets, it is not additive with other time periods.*

Trip dataset
(available only for
microdata files users)

The trip dataset provides information on the trip characteristics taken by travellers. It contains one record for each trip taken. It can be used only when linked to the person dataset and provides information on the travel incidence based on selected trip characteristics. It is useful to answer questions such as: How many persons flew at least once during January? How many persons took at least one business trip in July?

To get this information and additional trip information on travellers, one has to link the selected monthly person dataset with the appropriate monthly trip dataset using the unique person identifier (UNIQID). The person file should be sorted by the UNIQID and the trip file should be sorted by the UNIQID as well as the trip number identifier (TRIPNUM). These sorting sequences are the ones required to properly link those two files together.

After producing the merged file, duplicate records will appear on the file for travellers who reported more than one trip. To avoid double-counting, duplicate records of the same traveller have to be eliminated based on selected trip characteristics. For example, if one wants to know how many travellers flew at least once in January, one has to eliminate any record with the same UNIQID (i.e., the same traveller) having more than one record where the mode of transport is air. Counts on the number of travellers who flew at least once in January can then be produced.

When using the trip dataset merged with the person dataset, *only* the person weight can be used. The trip file cannot be used alone; no weight is included on this file.

Remember, according to Statistics Canada's definition of a trip, a person must have travelled for 80 km or more from home and must have travelled to a Canadian destination. Do not forget to make the proper selection on the variables DISTANC (distance) and DEPROV (province of destination) when defining the eligible population.

Contrary to the monthly person dataset, it is not possible to link the additional information included on the December person dataset to the monthly December trip dataset to gather information on trip characteristics of those people who travelled at least once during the year. The information on the trip file *only* provides information on characteristics of trips taken in December. For example, one cannot find out the main purpose or the mode of transport of those travellers who went to the United States at least once during the year from the December trip file.

The monthly trip datasets are not included on the Beyond 20/20 microdata files.

Person-trip dataset

The person-trip dataset contains socio-demographic and trip characteristics. On that file, there are adult records and child records. There is *only one* record

for adults from the same household despite the number of adults who went on the trip (the person-trip weight for that record is adjusted accordingly). If there were children from the same household accompanying the adult(s), there is *only one* record for children despite the number of children who went on the trip (the person-trip weight for that record is adjusted accordingly). From the person-trip file, one can obtain person-trip or household-trip information.

A person-trip is defined as a trip taken by one person. If this person takes more than one trip and/or travels with other members of the same household, we will count as many person-trips as there are trips and persons who took these trips. If four persons from the same household go on a trip together, it counts as four person-trips. If the same person takes two trips, it counts as two person-trips. The person-trip dataset takes into account the number of persons from a same household who accompanied the respondent on a specific trip and the frequency of travel.

The person-trip concept is used to measure the total number of trips combined with the total number of people who went on the trip. To make this concept more understandable, let us take a family of two adults with two children. If they all go to a zoo, using the person-trip concept, they will each have to go through the turnstile. We would count them as four person-trips. This concept can be associated

with both the socio-demographic variables and the trip characteristics variables.

If a user wants to know how many person-trips were made by plane in April (how many adults and children took a trip in April and used the plane as their mode of transportation) or how many person-trips were made to the Calgary CMA as a final destination in November (how many adults and children took a trip to the Calgary CMA in November), one has to use the person-trip concept from the person-trip dataset.

When compiling person-trip information, one has to use the person-trip weight available on the files (called `_PTRIPWT` on the Beyond 20/20 files and `PTRIPWT` on the microdata files).

A household-trip is defined as a trip involving one or more persons from the same household leaving and returning together. If four persons from the same household go on a trip together, that is, they leave and return together, it counts as one household-trip. If the same household takes two trips, it counts as two household-trips.

The household-trip concept is used to measure the total number of trips. In our same example of a family of two adults and two children going to the zoo, using the household-trip concept, they would all go through the turnstile as a group and would be counted as one household-trip.

A user would apply the household-trip concept to know how many business trips were taken within the last year (how many groups of people went on business trips within the year) or how many trips were taken by plane in July (how many groups of people went on trips by air in July). Notice that the socio-demographic characteristics CANNOT be associated with the household-trip weight.

When compiling household-trip information, one has to use the household-trip weight available on the files (called `_HHTRWT` on the Beyond 20/20 files and `HHTRIPWT` on the microdata files).

Both the person-trip and household-trip information provided on each monthly person-trip file is additive. It is possible to obtain monthly, quarterly or annual information for each variable on the person-trip dataset. Beyond 20/20 users only, have to make the proper preselection on the variables `MONTH` (reference month) or `QUARTER` (quarter) and then produce the required table. Microdata file users have to first combine the monthly person-trip files to create the file for the required time period. Then by making the proper selection when defining their population on the variables `MONTH` or `QUARTER`, they can obtain the required information.

Reported expenditures information

Reported expenditures are only available on the person-trip file. When compiling reported

expenditures (refer to [Chapter 2](#) for difference between reported and reallocated expenditures) information from the person-trip file, *only* the household-trip weight can be used because these represent the household expenditures and not just the traveller's expenses.

For microdata file users only

Microdata file users will notice that some records included on the person-trip file contain partial information only. Those are child records, that is, records for children from the same household as the respondent who accompanied them on the trip. Those child records have blank fields or 0s for the socio-demographic variables, the household-trip weight and the expenditures information. They must be included when compiling person-trip information. Because these records have a household-trip weight equal to 0, by default, they will not appear in household-trip counts.

Remember, according to Statistics Canada's definition of a trip, a person must have travelled for 80 km or more from home and must have travelled to a Canadian destination. Do not forget to make the proper selection on the variables DISTANC (distance) and DEPROV (province of destination) when defining the eligible population.

Visit dataset

The introduction of the visit concept in 1996 gives data users a more thorough picture of Canadians travelling to a specific region. Besides providing

trip information (main destination, length of trip, etc.), it also provides information on place(s) visited by those travellers, whether it was their main destination or an overnight stop on their journey. In addition to providing visit information, the visit dataset also includes reallocated expenditures information.

A visit represents a trip as a whole or part of a trip. A visit is defined as the place of overnight stay of the traveller or for a same-day trip, the location of destination of the trip. If the respondent travelled twice to the same location during the same trip, only one visit is recorded to that location. For instance, for a person going on a same-day trip to Toronto from Kingston, the count of visits and person-trips would be the same, that is, one. However, for a person leaving Montreal in destination of Winnipeg who spent a night in Ottawa and then spent a night in Winnipeg, we would have one record of a person-trip to Winnipeg (as a destination), but two records of person-visits: one in Ottawa-Hull and one in Winnipeg.

The visit concept is used to measure the number of person-visits who went to a specific location. The visit files have at least one record for every visit made for every person-trip found on the person-trip dataset.

Each monthly CTS visit file is provided at four different geographic levels, that is, the province, Census Metropolitan Area, tourism region and

Census Division level. Depending on which level of geography one wants to obtain visit information, one has to use the corresponding file. The visits are recorded at the level of geography represented on the file. For example, a person who takes a trip from Vancouver to Edmonton and spends one night in Calgary and one in Edmonton would show one visit to Alberta on the provincial visit file. This traveller would also show one visit to Calgary and one visit to Edmonton on the CMA visit file. The records on the CMA visit file would show:

- 1- ORCMA (CMA of origin) = Vancouver
DECMA (CMA of destination) = Edmonton
REALTYPE (CMA visited code) = Edmonton

- 2- ORCMA (CMA of origin) = Vancouver
DECMA (CMA of destination) = Edmonton
REALTYPE (CMA visited code) = Calgary

Each monthly CTS visit dataset also includes reallocated expenditures information (refer to [Chapter 2](#) for definition). Again, the reallocated expenditures included on each file would represent money spent in each location. Depending on which level of geography one wants to get reallocated expenditures information, one has to use the corresponding file. The reallocated expenditures are recorded at the level of geography represented on the file. For example, that same person who takes a trip to Edmonton and spends \$200 in Calgary and \$100 in Edmonton would show reallocated expenditures of \$200 (total reallocated visit expenditures shown in RREPCOST) in

Calgary and \$100 in Edmonton on the CMA-visit file. On the province-visit file, this person would show \$300 (total reallocated visit expenditures shown in RREPCOST) in Alberta.

The reallocation of expenditures is performed based on expenditures reported by respondents using the model of expenditures reallocation developed with the CTS Working Group (refer to [Chapter 5](#) on Data Processing). Historically, the reallocation of expenditures was done at the provincial level, however, in 1996, because of the new visit information available, the process of expenditures reallocation was expanded at the sub-provincial level. It is now possible to know how much money travellers spent in the Regina CMA or in the tourism region of Charlevoix.

As for the person-trip and household-trip information, the person-visit information provided on each monthly visit file is additive. It is possible to obtain monthly, quarterly or annual information from each variable from these files. For Beyond 20/20 users only, make the proper preselection on the variables MONTH or QUARTER and then produce the required table. Microdata file users first have to combine the monthly person-visit files to create the file for the required time period. Then by making the proper selection when defining their population on the variables MONTH or QUARTER they can obtain the required information.

For Beyond 20/20 users only

This visit dataset contains both the visit and reallocated expenditures information combined with selected traveller and trip characteristics. One dataset is provided for each level of geography. On each visit file, some traveller and trip characteristics are included *only to be used* as a selection criteria to compile visit information and not to compile trip information; the trip characteristics cannot be associated with the weights provided on the visit files (see table 4 to find out which weights can be used with different variables). To get information on trips, one should use the person-trip file.

When compiling visit information, one has to use the person-visit weight (_PVISWT) included on the datasets. One must also make the proper selection of visit records by using only the ones where the visit flag (VISFLAG) is equal to 1.

When compiling reallocated expenditures information, one has to use the reallocated expenditures weight (_REXPWT) included on the datasets.

For microdata file users only

The visit and reallocated expenditures datasets include province, Census Metropolitan Area, tourism region or Census Division visit and reallocated expenditures information only, that is, visit location, visit length, money spent by travellers in that location, etc.

To use the information provided on the visit and reallocated expenditures files, one has to merge the file presenting the proper geographical level of information with the appropriate person-trip file (same time period). This will give the user the socio-demographic characteristics of the travellers who visited a specific location as well as the trip characteristics.

When the person-trip and visit files are merged, one has to be careful about which variables can be used in association with the different weights. If one wants to compile visit information, one can use traveller and trip characteristics to select a population subgroup and should use the PTRIPWT variable on that file; the trip characteristics cannot be associated with the weights provided on the visit files (see table 4 to find out which weights can be used with different variables). To get information on trips, one should use the person-trip file.

Trip destination and trip distance are two criterions that are essential to have when compiling visit and reallocated expenditures information. Remember, according to Statistics Canada's definition of a trip, a person must have travelled for 80 km or more from home and must have travelled to a Canadian destination. Do not forget to make the proper selection on the variables DISTANC (distance) and DEPROV (province of destination) when defining the eligible population.

Having this additional information will allow the user to compile information such as the socio-demographic profile of persons who visited the Avalon tourism region; places visited by travellers whose destination was Vancouver, etc.

To link the person-trip and the visit files, one has to sort the person-trip file by MONTH, UNIQID, TRIPNUM and ADCH. The visit files should be sorted by MONTH, UNIQID, TRIPNUM, ADCH, the province where money was spent (REXPROV) and the location of visit (REALTYPE) depending of the file one is using. These sorting sequences are the ones required to properly link those files together. The linking between those two files can be done using the UNIQID and TRIPNUM variables.

When compiling visit information, one has to use the person-trip weight (PTRIPWT) provided on the datasets. One must also make the proper selection of visit records by using only the ones where the visit flag (VISFLAG) is equal to 1.

When compiling reallocated expenditures information, one has to use the household-trip weight (HHTRIPWT) provided on the datasets.

Using the CTS weights

The sample design used for the Canadian Travel Survey is not self-weighting. When producing simple estimates, including the production of ordinary statistical tables, users must apply the

proper sampling weight. This part describes the use of the different weights associated with the survey.

If proper weights are not used, the estimates derived from the microdata files cannot be considered 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 weight field.

When using the microdata datasets, one should try replicating some tables that are already included in Acrobat under the Quality Indicators on the CD-ROM and use the figures as a benchmark to verify the validity of the tables generated.

Person weight

The person weight can be used to measure the number of travellers and non-travellers. It can be associated with all the socio-demographic variables describing the respondent. Used with the person records, the user can get a socio-demographic profile of travellers and non-travellers or calculate travel incidences. The person weight can also be used with trip information to calculate incidences based on certain trip characteristics. For example, how many persons flew at least once during the month of January? However, proper selection has to be made when combining the person weight with

trip information (refer to [Chapter 7](#), person and trip dataset). The person weight *cannot* be used to measure the volume of person-trips because even used with a trip record, it does not take into account the identical trips nor the number of persons who went on the trip.

Person-trip weight

This weight can only be associated with those people who have taken a trip, that is the travellers. It can be used to measure the volume of person-trips (refer to [Chapter 7](#), Person-trip dataset). It can be associated with the socio-demographic characteristics of the travellers as well as the trip characteristics such as origin, destination or purpose. It *cannot* be associated with expenditures of trips because these represent the household expenditures and not just the traveller's expenses.

For microdata files users only

The person-trip weight is also used to calculate the number of person-visits when using the person-visit and reallocated expenditures file.

Household-trip weight

The household-trip weight can be associated with the trip characteristics as well as the expenditures information. It *cannot* be associated with socio-demographic characteristics of the respondents. The household-trip weight is used to measure the number of household-trips taken (refer to [Chapter 7](#), Person-trip dataset).

For Beyond 20/20 users only

The household-trip weight is to be used when compiling information on the number of household-trips taken and reported expenditures (refer to [Chapter 7](#), Person-trip dataset).

For microdata file users only

The household-trip weight is to be used when compiling information on the number of household-trips taken and both reported expenditures (provided on the person-trip file) and reallocated expenditures (provided on the person-visit and reallocated expenditures files) (refer to [Chapter 7](#), Person-trip dataset).

Person-visit weight

(for Beyond 20/20 users only)

This weight is to be used when compiling person-visit information from the visit files. It is used to measure the number of person-visits. It can be associated with all the socio-demographic and visit variables (refer to table 4). For example, one would apply the person-visit weight to get the number of person-visits who went to the tourism region of Regina or to get a socio-demographic profile of those people who visited the Toronto CMA.

***Reallocated
expenditures
weight***

*(for Beyond 20/20
users only)*

This weight should be used to compile reallocated expenditures information using the visit files. It can be associated only with reallocated expenditures related variables (refer to table 4). For example, one would apply the reallocated expenditures weight to know how much money was spent on recreation and entertainment in the Niagara Regional Municipality (Census Division located in Ontario).

***Using the data
dictionary***

The electronic data dictionary provided with this package presents the complete information about each variable on the Beyond 20/20 files and on the microdata files. For each variable is shown: the variable name, the description or definition, code lists with descriptions or alternatively the range of values that the variable can take on.

TABLE 4: Use of weights in association with different variables

	Person	Person-trip	Household-trip	Person-visit	Reallocated expenditures
<i>Socio-demographic characteristics</i>					
Age group	✓	✓		✓	
Sex	✓	✓		✓	
Marital status	✓	✓		✓	
Relation to family reference person	✓	✓		✓	
Class of worker	✓	✓		✓	
Educational attainment	✓	✓		✓	
Labour force status	✓	✓		✓	
Household income	✓	✓		✓	
<i>Trip characteristics</i>					
Province of origin		✓	✓		
CD of origin		✓	✓		
CMA of origin		✓	✓		
Tourism region of origin		✓	✓		
Province of destination		✓	✓		
CD of destination		✓	✓		
CMA of destination		✓	✓		
Tourism region of destination		✓	✓		
Distance (one-way)		✓	✓		
Total number of nights		✓	✓		
Nights spent in Newfoundland ...		✓	✓		
Nights spent in hotel ...		✓	✓		
Mode of transportation		✓	✓		
Main purpose of trip		✓	✓		
Activities		✓	✓		
<i>Visit characteristics</i>					
Number of visit-nights				✓	
Number of visit-nights - hotel ...				✓	
<i>Expenditures variables</i>					
Total cost of packages			✓		
Reported expenditures variables			✓		
Reallocated expenditures variables			✓*		✓

* For microdata files users only.

8

GUIDELINES FOR ANALYSIS AND PRESENTATION

This chapter outlines the guidelines to be adhered to by users tabulating and analysing any data derived from the survey Beyond 20/20 files and 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.

Rounding guidelines

Once it has been determined whether the results obtained are reliable, the level of rounding indicates the level of *precision* that the data can actually support. The following guidelines for rounding should be used:

- Estimates of population subgroups should be rounded to the nearest thousand units.
- Rates and percentages should be rounded to one decimal point.

Note that all calculations are to be derived from their unrounded components, and then rounded 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 1000, the estimate 49,448 would be rounded down to 49,000 and an estimate of 49,552 would be rounded up to 50,000. The figure 1.78% would be rounded to 1.8%.

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 released document(s).

Types of estimates

Before discussing how the Canadian Travel Survey data can be tabulated, it is useful to describe the two main types of point estimates of population characteristics which can be generated from the files for the Canadian Travel Survey, that is categorical or quantitative estimates.

Categorical Estimates

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of people 15

years of age or more who took at least one same-day trip in the reference month or the proportion of trips taken which were business-related are examples of such estimates. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of Categorical Questions:

Q: Did you take any trips of one night or more which ended in reference month?

R: Yes / No

Q: What was your main reason for taking this trip?

R: Visiting friends or relatives/ Pleasure / Personal / Business/ Non-business convention

Quantitative Estimates

Quantitative estimates are estimates of totals or of means, medians and other measures of central tendency of quantities based upon some or all of the members of the surveyed population. They also specifically involve estimates of the form \bar{X}/\hat{Y} where \bar{X} is an estimate of surveyed population quantity total and \hat{Y} is an estimate of the number of persons in the surveyed population contributing to that total quantity.

An example of a quantitative estimate is the average number of trips taken per person who travelled during the reference month. The

numerator is an estimate of the total number of trips taken (same-day and overnight) by persons who took at least one trip during the reference month, and its denominator is the number of persons who took at least one trip during the reference month.

Examples of Quantitative Questions:

Q: How many nights were you away from home on this trip?

R: |_|_|_| nights

Q: Including yourself, how many persons now living in this household went on this trip?

R: |_|_| people

Tabulation of categorical estimates

Categorical estimates of the number of people with a certain characteristic can be obtained from the Beyond 20/20 files or from the microdata files by applying the weight to the records possessing the characteristic(s) of interest. Proportions and ratios of the form X/Y are obtained by:

- ▶ applying the weights of records having the characteristic of interest for the numerator (X),
- ▶ applying the weights of records having the characteristic of interest for the denominator (Y), then
- ▶ dividing the numerator estimate by the denominator estimate.

Example using the Beyond 20/20 person file

From the Beyond 20/20 person file, to obtain an estimate of the number of people who took at least one non-business, overnight, interprovincial trip in 1996,

- ▶ drag the YRSAMEPR tile in the column dimension and the AGEGRP tile in the row dimension (although it is not necessary to look at age groups to get the estimated number, Beyond 20/20 does not allow you to use only one variable).

- ▶ Apply the _PERSWT and load the data. (Because only one weight is included on the person file, data are already weighted. For the other extracts, one will have to apply a weight factor and then load the data.) In the top left corner cell, you will have the denominator (total population = 23,537,562) and in that same row, where the column title yes, you will have the numerator (number of people who took at least one non-business, overnight, interprovincial trip in 1996 = 11,375,951).

- ▶ To get a proportion, divide the numerator by the denominator ($11,375,951/23,537,562 = 0.48$).

Tabulation of quantitative estimates

Quantitative estimates can be obtained by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest.

Example using the Beyond 20/20 person-trip file

To obtain an estimate of the total reported expenditures for the first quarter of 1996 from the Beyond 20/20 person-trip file,

- ▶ drag the QUARTER tile into the column dimension of the table;
- ▶ apply the _HHTRWT;
- ▶ drag, drop and sum the REPCOST tile into the cell area of the table view and load the data. The total reported expenditures for all persons who travelled in the first quarter of 1996 can be found under the first quarter column (3,731,742,037).

To obtain a weighted average of the form X/Y, the numerator (X) is calculated as for a quantitative estimate and the denominator (Y) is calculated as for a categorical estimate. For example, to estimate the average expenditures per person who travelled in the first quarter of 1996,

- ▶ estimate the total expenditures as described above (3,731,742,037),

- ▶ estimate the number of person-trips by using the Beyond 20/20 person-trip file and applying the _PTRIPWT weight. To do this, drag and drop the QUARTER tile into the column dimension of the table, drag and drop the ORPROV tile (all trips have a province of origin) into the row dimension of the table, apply the _PTRIPWT and load the data. The total number of person-trips will be found at the first quarter row (31,193,087),
- ▶ divide the first estimate by the second estimate ($3,731,742,037/31,193,087 = \119.63).

9

GUIDELINES FOR RELEASE (DATA QUALITY)

Microdata users should apply the rules for assessing data quality, below, to all estimates they produce, and retain only those that satisfy the release criteria. Estimates that do not satisfy the release criteria are not reliable.

Introduction

The guidelines for release and publication make use of the concept of *sampling variability* to determine whether estimates obtained from the microdata files are reliable. Sampling variability is the error in the estimates caused by the fact that we survey a sample rather than the entire population. The concept of *standard error* and the related concept of *coefficient of variation* and *confidence interval* provide an indication of the magnitude of the sampling variability.

The standard error and coefficient of variation do not measure any systematic biases in the survey data which might affect the estimate. Rather, they are based on the assumption that the sampling errors follow a normal probability distribution.

Subject to this assumption, it is possible to estimate the extent to which different samples that have the same design and the same number of observations

would give different results. This indicates the margin of error that is likely to be included in the estimates derived from our single sample.

For a detailed description of the measures of sampling variability, see A. Satin and W. Shastry, *Survey Sampling: A Non-Mathematical Guide*, Statistics Canada, Catalogue Number 12-602E.

Survey errors

The survey produces estimates based on information collected from and about a sample of individuals. Somewhat different estimates might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc., as those actually used in the survey. The difference between the estimates obtained from the sample and those resulting from a complete count taken under similar conditions is called the sampling error of the estimate.

Errors which are not related to sampling may occur at almost every phase of a survey operation. Interviewers may misunderstand instructions, respondents may make errors in answering questions, the answers may be incorrectly entered on the computer and errors may be introduced in the processing and tabulation of the data. These are all examples of non-sampling errors.

Over a large number of observations, randomly occurring errors will have little effect on estimates derived from the survey. However, errors occurring

systematically will contribute to biases in the survey estimates. Considerable time and effort are made to reduce non-sampling errors in the survey. Quality assurance measures are implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures include the use of highly skilled interviewers, extensive training of interviewers with respect to the survey procedures and questionnaire, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions, procedures to ensure that data capture errors are minimized and coding and edit quality checks to verify the processing logic.

A major source of non-sampling errors in surveys is the effect of non-response on the survey results. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. Total non-response occurs because the interviewer is either unable to contact the respondent, no member of the household is able to provide the information, or the respondent refuses to participate in the survey. Total non-response is handled by adjusting the weight of households who respond to the survey to compensate for those who do not respond (refer to [Chapter 6](#)).

In most cases, partial non-response to the survey occurs when the respondent does not understand or misinterprets a question, refuses to answer a

question, cannot recall the requested information, or cannot provide proxy information.

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error. This chapter of the documentation outlines the measures of sampling error which Statistics Canada commonly uses and which it urges users producing estimates from these microdata files to use also.

The basis for measuring the potential size of sampling errors is the standard error of the estimates derived from survey results.

***Hypothesis
tests provided
by statistical
software
packages***

The Canadian Travel Survey is based upon a complex design, with stratification and 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.

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 almost meaningless.

For many analysis techniques (for example linear regression, logistic regression, analysis of variance), a method exists which can make the application of standard packages more meaningful. If the weights on the records are rescaled so that the average weight is one (1), then the results produced by the standard packages will be more reasonable; they still will not take into account the stratification and clustering of the sample's design, but they will take into account the unequal probabilities of selection. The rescaling can be accomplished by dividing each weight by the overall average weight before the analysis is conducted.

In order to provide a means of assessing the quality of tabulated estimates, Statistics Canada has produced a set of Approximate Sampling Variability Tables (commonly referred to as "C.V. Tables") for the Canadian Travel Survey. These tables can be used to obtain approximate coefficients of variation for categorical-type estimates and proportions. Refer to [Chapter 10](#) for more details.

Minimum sizes of estimates for release

Before releasing and/or publishing any estimate from these microdata files, users should first determine the number of respondents who contribute to the calculation of the estimate. If this number is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the rounded estimate and follow the guidelines given in Table 5.

When the unweighted estimate is satisfactory, the user should look at the weighted estimate to see if it satisfies the release criteria. The release cutoffs for the weighted estimates of the CTS can be found in the Data Quality Report ([Chapter 10](#)).

Using the coefficient of variation (CV)

The standard error of an estimate is frequently expressed as a percentage of the estimate itself, in which case it is called the coefficient of variation. Whereas the standard error is measured in the same units as the estimate, the coefficient of variation is simply a ratio. This makes it easier to use as a criterion for the reliability of any estimate.

For example, suppose that, based upon the survey results, one estimates that 25.9% of Canadians aged 15 or more took at least one trip in March 1996, and this estimate is found to have standard error of 0.009. Then the coefficient of variation of the estimate is calculated as:

$$\left(\frac{.009}{.259} \right) \times 100\% = 3.47\%$$

The coefficients of variation (C.V.) are derived using the variance formula for simple random sampling and incorporating a factor which reflects the multi-stage, clustered nature of the sample design. This factor, known as the design effect, was determined by first calculating design effects for a wide range of characteristics and then choosing from among these a conservative value to be used in the look-up tables which would then apply to the entire set of characteristics.

To have more information on the design effects, sample sizes and population counts by province that were used to produce the Approximate Sampling Variability Tables for person-weights, refer to the Data Quality Report ([Chapter 10](#)).

Note that Approximate Sampling Variability Tables are also available for person-trip weights, household-trip weights, person-night weights and expenditures weights.

In the Data Quality Report provided with this Guide, a set of Approximate Sampling Variability Tables has been provided to give microdata users some approximate coefficients of variation for groups of estimates at a time, such as all estimates pertaining to a particular province. In most cases, these will be adequate to determine whether an

TABLE 5. Acceptable levels of the coefficient of variation

Approximate coefficient of variation (%)	Restriction on use
0.0 - 16.5	ACCEPTABLE. Estimates can be considered for general unrestricted release.
16.6 - 25.0	MARGINAL. Estimates can be considered for general unrestricted release but should be accompanied by a warning cautioning subsequent users of the high sampling variability associated with the estimates.
25.1 - 33.3	CONFIDENTIAL. Estimates can be considered for general unrestricted release only when exact coefficients of variation have been calculated on a cost-recovery basis and are found to be acceptable. Otherwise, such estimates should not be used or released.
33.4 or greater	UNACCEPTABLE. Estimates should not be used or released under any circumstances.

estimate can be released. The Data Quality Report ([Chapter 10](#)) explains how to obtain the approximate CV from the tables, depending on whether the estimate is a simple population count or a percentage, difference or ratio of population subgroups. In the case of numeric totals or means, the CV is generally larger than the CV of the population count on which it is based.

All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Estimates of actual variance for specific variables may be obtained from Statistics Canada on a cost-recovery basis. The use of actual variance estimates would allow users to release otherwise unreleaseable estimates, i.e., estimates with coefficients of variation in the 'confidential' range.

Remember: if the number of observations on which an estimate is based is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. This is because the formulas used for estimating the variance do not hold true for small sample sizes.