Microdata User Guide

National Travel Survey

2020

Statistics Canada Statistique Canada

Canada

How to obtain more information
Specific inquiries about this product and related statistics or services should be directed to:

Statistics Canada
150 Tunney's Pasture Driveway, Ottawa, Ontario K1A 0T6
Telephone: 1-800-263-1136
Email : STATCAN.infostats-infostats.STATCAN@canada.ca

Accessing and ordering information
The National Travel Survey (NTS) produces two types of microdata files - master files, and public use microdata files (PUMF).

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

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

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

Public use microdata files (PUMF)
The public use microdata files are developed from the master files using a technique that balances the need to ensure respondent confidentiality with the need to produce the most useful data possible. The PUMF must meet stringent security and confidentiality standards required by the *Statistics Act* before they are released for public access. To ensure that these standards have been achieved, each PUMF goes through a formal review and approval process by an executive committee of Statistics Canada. Variables most likely to lead to identification of an individual are deleted from the data file or are collapsed to broader categories.

Note that there may be differences between the estimates published by Statistics Canada in its statistical products and the estimates reproduced by users with the public use microdata files, as Statistics Canada uses master files to produce its estimates.

To obtain a copy of the PUMF contact Statistics Canada at the address stated above.

The Data Liberation Initiative
The Data Liberation Initiative (DLI) Program enables students and researchers to use the public use microdata files in several universities across Canada. For more information, please consult the following web page: http://www.statcan.gc.ca/eng/dli/dli

Impact of the COVID-19 Pandemic on 2020 NTS PUMF

Due to the occurrence of the COVID-19 pandemic, collection for the National Travel Survey was temporarily suspended in 2020, resulting in no available data for March, April, May and June**. As this PUMF therefore only covers eight months, it cannot be used to calculate annual totals**. In addition, due to the low number of respondents reporting outbound travel in Q3 and Q4, all trip destinations for these quarters have been coded to "unknown". Only in Q1 which is covering only two months (January and February) are the names of individual countries published**.**

Table of Contents

1.0 Survey description

The National Travel Survey (NTS) was developed to fully replace the Travel Survey of Residents of Canada (TSRC record number 3810) and replace the Canadian resident component of the International Travel Survey (ITS record number 3152). The National Travel Survey collects information about the domestic and international travel of Canadian residents.

The National Travel Survey provides statistics on the activities of Canadian residents related to domestic and international tourism. It was developed to measure the volume, the characteristics and the economic impact of tourism. For the Canadian System of National Accounts, NTS measures the size of domestic travel in Canada from the demand side.

Users of the NTS data are Statistics Canada, Destination Canada, the provinces and tourism boards. Other users include the media, businesses, consultants and researchers.

This guide has been produced to facilitate the use of the PUMF of the survey results.

Any questions about the data set or its use should be directed to:

Statistics Canada
150 Tunney's Pasture Driveway, Ottawa, Ontario K1A 0T6
Telephone: 1-800-263-1136
Email : STATCAN.infostats-infostats.STATCAN@canada.ca

2.0 Concepts and definitions

2.1 National Travel Survey concepts and definitions

This section outlines concepts and definitions of interest. Users can also refer to the NTS questionnaire and the codebooks included in the survey release package. The codebooks present the complete information about each variable on the PUMF. The following can be included for each variable: the variable name, the description or definition, code lists with descriptions or alternatively the range of values that the variable can take on and their frequencies (weighted and/or unweighted).

2.1.1 Trip types covered by the survey

Trip types

The following trip are covered by the NTS:

- domestic trips (origin and destination in Canada) with no international leg
- domestic trips (origin and destination in Canada) with an international leg
- international trips (origin in Canada and destination outside Canada) with a domestic leg
- international trips (origin in Canada and destination outside Canada) with no domestic leg

Trips in Canada (domestic) can be regarded as entities with provincial components. A trip can cover one or more provinces. Domestic trips can therefore be classified as follows:

- domestic trips (origin and destination in Canada) with no international leg
    - intra-provincial (visits are all within the province of residence)
    - interprovincial (visits are all outside the province of residence)
- domestic trips (origin and destination in Canada) with an international leg
    - intra-provincial
    - interprovincial

Length of trip: same day or overnight

The length of a trip, defined as “same day” or “overnight”, is based on the total number of nights spent during the trip, whether the nights were spent in Canada or outside Canada.

The trips are classified as this in the PUMF:

Trip types
1 International Same day
2 International Overnight
3 Domestic Same day Interprovincial
4 Domestic Overnight Interprovincial
5 Domestic Same day Intraprovincial
6 Domestic Overnight Intraprovincial

Trip components: visits

For the visit analysis, domestic trips, international trips and the domestic and international legs are included.

Trips are divided into three types of components: origin of the trip, locations visited during the trip, and main destination of the trip. The trip's origin and main destination are always included as components and are unique. The locations visited during the trip are included if at least one night is spent in those locations. The number of "locations visited" components will be equal to the number of locations visited with an overnight stay (the main destination is excluded from the locations visited). Expenditures can be assigned to all component types. However, nights spent and accommodation types can only be assigned to the locations visited and the trip's main destination if applicable.

Airport components are added if the respondent reports taking a commercial aircraft to a Canadian destination or to return from their international destination and the commercial transportation expenditures are allocated to a Canadian carrier. There is one airport component for each origin, location visited and destination. Transportation expenditures are assigned to the airport components, but nights spent and accommodation types are not.

The location of each trip component is coded geographically by country (Canada or other countries), province, tourism region (TR), census metropolitan area (CMA), census division (CD) and census subdivision (CSD). During a trip, a visit to a particular geography is counted if at least one location visited or the main destination are reported in that geography. Trip origin and airports are not counted as visits. For example, a domestic trip will constitute one, and only one, province-visit to Quebec if at least one location visited in Quebec is reported or the main destination is in Quebec. By default, all domestic trips included in the visit analysis count as one Canada-visit, since they all have at least one location visited in Canada or the main destination reported is in Canada. This *geography*-visit concept applies to all of the survey's geographies: countries, province, group of provinces (Territories), tourism region, census metropolitan area and census division.

Under this concept of a visit, one visit to a particular geography may count as more than one visit to a smaller geography. For example, one province-visit may constitute two tourism region-visits if locations were visited in two of the province's tourism regions. The same concepts also apply to countries, CMAs and CDs.

Length of visit: same day or overnight

The number of nights spent during the visit is used to determine the length of the visit, regardless of the geography of the visit being analyzed (country, province, tourism region, etc.). If there are nights spent during a visit, the length of the visit is overnight; if not, the length of the visit is same day.

Trip expenditures

For each trip reported, the expenditures made during the trip are collected for various expenditure items (for example, commercial accommodations, vehicle rental, and local transportation). At the trip level, the expenditures reported are assigned to the trip's main destination. The expenditures reported for a trip are also reallocated to the various trip components (origin, visit, destination or airport) by expenditure item. Certain types of expenditures cannot be allocated to certain components.

2.1.2 Main definitions

Accommodation: Type of accommodation where nights were spent while on the *trip*. The data on the type of accommodation is collected for each of the nights spent on the *trip*. It includes paid (hotel, motel, resort, lodge, campgrounds, etc.) and unpaid (home of a friend or relative, private cottage, other unpaid) accommodations.

Activities: Activities in which *travellers* took part during the *trip* rather than during their *visit*. For this reason, an activity cannot always be associated to a precise location. For example, a person may have visited Vancouver and Whistler, and reported downhill skiing as an activity. It is impossible to know if the traveller skied in Vancouver, in Whistler or in both places. For a *same day trip*, it is the most important activity done while on the trip. For an *overnight trip*, it is all activities done while on the trip.

Country-visit: Each person visiting a country, Canada or another, is registered as having made one *person-visit* in the country.

Destination (main): Place reported by the traveller as the destination of a *trip*. If a *traveller* visited more than one place during a *trip*, the main destination is the place given by the *traveller* when the question was asked.

Domestic trip or visit: To be considered a domestic *trip* or *visit*, the destination of the *trip* or *visit* has to be in Canada.

Duration: A trip starts when the *traveller* leaves his/her usual residence and ends when he/she comes back to it. The trip duration is measured by the number of nights a *traveller* spends away from his/her usual residence.

Expenditures (reported): Reported expenditures are the expenditures made by the traveller and other household members who went on the *trip* plus expenditures covered by others who did not go on the trip. Expenditures made by members of other households who went on the trip are excluded. Expenditures are reported in Canadian dollars and include all taxes and tips. They are associated with a *trip* as a whole.

The following items are excluded from the reported expenditures: 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 categories:

- ***Accommodation expenditures:*** Total expenditures for nights spent in hotels, motels, resorts, cabins, rented or commercial cottages, campgrounds, etc. while on the *trip*. 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*.

- ***Vehicle rental expenditures:*** Amount of money spent on all operation, rental or users' fees incurred for the use of any vehicle such as an automobile, a truck, a motorcycle, a bicycle, a boat, a motor home, a snowmobile, etc. (including insurance) while on the *trip*. The amount reported includes all applicable taxes and tips.

- ***Vehicle operation expenditures:*** Amount of money spent related to the cost of operating a private or rental vehicle, namely gasoline, repairs and parking costs while on the *trip*. The amount reported includes all applicable taxes and tips.

- ***Local transportation expenditures:*** Amount of money spent on local transportation means within a city or metropolitan area, that is, intra-city transportation while on the *trip*. 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*.The amount reported includes all applicable taxes and tips.

- ***Commercial transportation expenditures****:* Amount of money spent on commercial transportation between urban and/or rural areas to get to or from the destination while on the *trip*. These transportation expenditures include ticket fares for intercity aircrafts, boats, hovercrafts, trains, buses as well as intercity ferries. The amount reported includes all applicable taxes and tips.

- ***Food or beverages purchased at restaurants or bars expenditures:*** Amount of money 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, while on the *trip*. The amount reported includes all applicable taxes and tips.

- ***Food or beverages purchased at stores during the trip expenditures:*** Amount of money spent on food and beverages 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. The amount reported includes all applicable taxes and tips.

- ***Recreation expenditures:*** Amount of money spent on sports or recreational activities as well as the rental of sporting/recreational equipment while on the *trip*. This category includes the costs of green fees, ski lift tickets, rental of bowling shoes and rental of sporting equipment such as skis, golf clubs and canoes. But it excludes purchases of recreational services for the season (e.g., season ski passes). The amount reported includes all applicable taxes and tips.

- ***Entertainment expenditures:*** Amount of money spent on cultural or entertainment activities and attractions while on the *trip*. This category includes cost of admission to theatres, art galleries, nightclubs or sporting events (e.g., attending a hockey game). It also includes gambling expenses, entrance fees to cultural or leisure activities (e.g., movies, museums, theme parks, go-carting, etc.) and spending on boat tours and balloon rides. But it excludes the costs of season tickets for leisure activities or spectacles (for example, season tickets to the theatre). The amount reported includes all applicable taxes and tips.

- ***Clothing expenditures:*** Includes any clothing, footwear or accessories purchase whether they are gifts or for personal use. The amount reported includes all applicable taxes and tips.

- ***Other expenditures:*** Amount of money spent on other items while on the *trip*. This category includes purchases of souvenirs, fabric and household items, registration fees for courses, conferences or conventions, customs duties, purchases of postcards and stamps, insurance fees, purchases of medication, books, craft supplies and films for cameras, costs for child care, house sitting and animal care, telephone charges, costs for renting facilities (e.g., seminar rooms, training rooms, etc.), purchases of season tickets for leisure activities or shows (e.g., season ski passes, season tickets to the theatre). The amount reported includes all applicable taxes and tips.

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.

International trip**:** To be considered an international *trip*, the *destination* of the trip has to be outside of Canada.

Inter-provincial trip or visit: *Trip* or *visit* to or in a province that is different from the province of origin of the trip.

Intra-provincial trip or visit: *Trip* or *visit* within the province of origin of the trip.

Main 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 first mode used is recorded.

Main reason or purpose: Main reason why the traveller went on a *trip*, regardless of the reason anyone else from the household had for taking the same trip. This is the reason without which the trip would not have happened.

Origin: The origin is the starting point of a *trip.* The trip must have originated in one of the ten provinces.

Overnight trip or visit: *Trip* or *visit* that includes at least one night away from home.

Package deal: A travel package covers at least part of the trip, and includes any combination of at least two services for which individual costs are not identified separately. **For example**, travel packages can include airfare and accommodation; airfare and cruise; accommodation, meals and entertainment; or all-inclusive vacations. A package deal is usually purchased from a travel agency or social organization.

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.

Person-trip: *Trip* taken by one person. If this person took more than one *trip* and/or traveled with other adult 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-visit: *Visit* taken by a *traveller* either single or traveling as a group. If four persons go on a *visit* together, it counts as four person-visits.

Person-visit-night: One night away from home in Canada or another country by a person taking a *trip* e.g. a person who takes a *trip* involving three nights away from home has a count of three person-visit-nights. The total number of person-visit-nights for a population is the count of the number of **nights** spent away from home **in Canada or another country**, by each person on each *trip* taken in the population. The levels of geography for which the person-visit-nights can be estimated in the NTS are the following: country, province, *tourism region*, *Census Metropolitan Area* and *Census Division*.

Province-visit: Each person visiting a province is registered as having made one *person-visit* in this province. Please note that if a *traveller* visits 2 provinces during the same *trip*, it will count for 2 province-visits but only one *country-visit.*

Quarter: Quarter during which the *trip* ended. The first quarter is from January to March, the second quarter is from April to June, the third quarter is from July to September and the fourth quarter is from October to December.

Reference month: Month in which the *trip* ended between the first and the last day of a calendar month; the *trip* may have started before the reference month

Reallocated expenditures: Refers to the process by which all of the travellers *reported expenditures* are redistributed to specific geographic regions where money was spent. The NTS uses an expenditure reallocation model by which money included in each expenditure category is redistributed to a geographic region according to specific rules. The levels of geography for which the expenditures are redistributed are country, province, tourism region, Census Metropolitan Area and Census Division.

Same-day trip or visit*:* To be considered a same-day trip, it has to be 40 km or more (one way) in which the *traveller* left and returned home on the same day.

Tourism: The definition of tourism follows that adopted by the World Tourism Organization and the United Nations Statistical Commission: “the activities of persons travelling to and staying in places outside their usual environment for not more than one consecutive year for leisure, business and other purposes”.

Tourism Region: The Tourism Regions (TR) boundaries are defined by each provincial Tourism Department and not by Statistics Canada. Every year these Departments have the opportunity to modify their TR boundaries, which are consequently applied to the NTS by Statistics Canada. Tourism Regions are built from CSD.

Traveller: Any person aged 18 or more who completes a *trip*. Any person who does not take a trip may be described as a non-traveller.

Trip: A trip must have *originated* in one of the ten provinces; have ended during the *reference month* and be less than 365 days/nights duration NTS collects all *overnight trips* that have a domestic or international destination and all *same day trips* of a distance that is 40km and over with a domestic or international destination. These trips are considered as “out-of-town” by the traveller.

In-scope trips include:

- all *trips* for purposes of pleasure, vacation or holiday
- all *trips* for visiting friends or relatives
- all business and work related *trips*, **except** routine travel which is a regular part of the job
- all *trips* for other reasons **except** regular household or grocery shopping, moving (or helping someone move) to a new residence (or school), commuting to school, regular medical or dental appointments or check-ups, regular attendance at religious observances/services, attendance at funerals and *trips* for various regular chores such as picking up someone at the arena

Trip distance: Refers to the one way distance between the *origin* of the *trip* and its *destination*, expressed in kilometers.

United Nation World Tourism Organization (UNWTO): The World Tourism Organization (WTO) is a specialized agency of the United Nations and the leading international organization in the field of tourism. It serves as a global forum for tourism policy issues and a practical source of tourism know-how.

Visit: A visit is defined as a location visited, either in Canada or another country. It is the location where the traveller has spent at least one night, in the case of an overnight visit, or the destination of the trip, in the case of a same-day visit. If the traveller travelled twice to the same location during the same trip, only one visit is recorded at that location.

2.1.3 NTS files description

Each month, three master files are created: a person file, a trip file and a visit file. Then, when the PUMF is created, the monthly files are merged together to created one annual files for each of the person, trip and visit.

Person data file

The person file provides information on travellers and non-travellers, a traveller being a person aged 18 and over who took at least one trip ending in the reference month and a non-traveller being someone who did not take a trip ending in the reference month. Each respondent to the NTS has one record on the person file.

The person microdata file includes basic socio-demographic information on both travellers and non-travellers. It can be used to produce simple socio-demographic profiles and to calculate travel incidences. For example, what is the age or gender 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 as a traveller only once.

Trip file

The trip file contains trip characteristics, for example, the origin, the main destination, the reason of the trip, the spending, etc. For each respondent on the person file, there is a trip record for every trip reported. If a respondent did not report any trips, there would be no trip records for that respondent on the trip file.

Visit file

The visit file 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 file also includes reallocated household expenditures information.

The visit concept is used to measure the number of person-visits to a specific location. The visit file has at least two records per visit location for every person-trip found on the-trip file (origin and destination).

2.2 Content development

The content of the National Travel Survey electronic questionnaire was drawn from the Travel Survey of Residents of Canada and the International Travel Survey, which were based on consultation with several tourism provincial organizations/departments. Statistics Canada System of National Accounts participated in the questionnaire design.

The questionnaire underwent cognitive testing in the form of in-depth interviews in both of Canada's

official languages, conducted by Statistics Canada's Questionnaire Design Resource Centre. The goal of the qualitative study was to test a new introduction to the survey and different trip definitions. There were two pilot tests done. The first pilot test done in February-March 2016 was used to evaluate multiple letter-based respondent selection methods. The conventional method of random selection was to select a household and use the application to select a respondent. The first pilot provided information on the ability of household members to interpret and comply with the random selection method described in the letter.

The second pilot test done in August 2017 was used to evaluate the on-line response application and to estimate the take-up rate. The second survey pilot was also used to evaluate multiple nonresponse follow-up strategies including mail out of letters, follow-up courtesy calls and phone calls to offer to complete the questionnaire over the phone.

3.0 Survey methodology

3.1 Target and survey population

The target population is the civilian, non-institutionalized population 18 years of age or older in Canada's ten provinces. Specifically excluded from the survey's coverage are: persons living in Indian reserves, persons living in the territories. Together, these groups represent an exclusion of less than 3% of the Canadian population aged 18 and older.

3.2 Sample design

The National Travel Survey is a sample survey with a cross-sectional design. A three-stage sample design is used: 1) Dwelling; 2) Person; 3) Trip.

For the first stage,

- The survey is based on the dwelling universe file (DUF) and the socioeconomic file (SEF). These data exclude residents of Yukon, the Northwest Territories and Nunavut, as well as persons living on Indian reserves and in parts of a collective dwelling that are not part of a dwelling associated with collective (DAWC).
- The sampling frame is then stratified by provinces, income and passport group. We use the passport data to stratify the frame in order to have more international traveller. For given months (January, February, October, November and December), an additional variable has been used in stratification group– TC regions. This variable refers to some specific CMA or group of CMAs. A simple random sample is drawn from each stratum. The total sample includes about 39,000 households monthly. Additional households could be added from other partners.

For the second stage, one adult per selected household will be chosen randomly using simple random sampling. A letter-based selection method “the age-order” is used. With these letter-based methods, the random selection of a person will be made before the person goes online, so only the selected person will go online to complete the electronic questionnaire.

For the third stage, the e-questionnaire asks the respondent for a brief description of all of his or her trips that ended in the reference month. The application then selects up to three trips using sequential Poisson sampling.

4.0 Data collection

For the year 2020, the data collection reference period was from January 1st 2020 to January 10th 2021. Responding to this survey is voluntary and the data are collected directly from survey respondents.

Selected households receive an invitation letter in the mail. The letter explains who, from the household, is selected to participate in the survey using the age selection method. A household may receive up to two mail reminders. The access code in the letters gives the respondent access to the electronic questionnaire. The electronic questionnaire is offered in the two official languages: French and English. The respondent must provide basic information on all of his or her trips (domestic and international) that ended in the reference month. The respondent then provides details on the trips selected. The average time required to complete the survey is 15 minutes.

Population projections produced by Statistics Canada's Demography Division are used at the weighting stage of the National Travel Survey (NTS). Person-level weights are calibrated based on these projections.

Estimates from Statistics Canada's Frontier Counts program are also used at the weighting stage of the NTS. Person-trip weights are calibrated to these estimates so that both data sources are consistent. Furthermore, calibration stabilizes the estimator of the number of international trips by Canadians and therefore improves the quality of the NTS estimates.

5.0 Data processing

Processing transforms survey responses obtained during collection into a form that is suitable for tabulation and data analysis. It includes all data handling activities – automated and manual – after collection and prior to estimation.

5.1 Data capture

For the electronic questionnaire, responses to survey questions are entered directly by the respondents. The electronic questionnaire reduces processing time and costs associated with data entry, transcription errors and data transmission. The responses were secure through industry standard encryption protocols, firewalls and encryption layers.

Some editing was done directly at the time the electronic questionnaire was completed. Where the information was outside the range (too large or small) of expected values, or inconsistent with the previous entries, the respondent was prompted, through message screens, to verify the information. However, the respondents had the option of bypassing the edits, and of skipping questions if they did not know the answer or refused to answer. Therefore, the data were subjected to further edit processes after they were submitted to head office. When the electronic data was received it was converted to readable text files.

5.2 Editing

Editing can occur at several points throughout the survey process and ranges from simple preliminary checks performed within the electronic questionnaire during the interview to more complex automated verifications performed by a computer program after the data have been captured. In general, edit rules are based upon what is logically or validly possible, based upon:

- expert knowledge of the subject matter;
- other related surveys or data;
- the structure of the questionnaire and its questions;
- statistical theory.

There are three main categories of processing edits applied: validity, consistency and distribution edits. Validity edits verify the syntax of responses and include such things as checking for non-numeric characters reported in numeric fields and checking for missing values. Validity edits can also check that the coded data lie within an allowed range of values. For example, a range edit might be put on the reported age of a respondent to ensure that it lies between 0 and 125 years.

Consistency edits verify that relationships between questions are respected. Consistency edits can be based on logical, legal, accounting or structural relationships between questions or parts of a question.

Distribution edits are performed by looking at data across questionnaires. These edits attempt to identify records that are outliers with respect to the distribution of the data. Distribution edits are sometimes referred to as statistical edits or outlier detection.

One example of a consistency edits would be that records with extreme weighted total trip spending are flagged and treated as outliers.

5.2.1 Computer-generated edits

As stated in a section above, some edits are done as the electronic questionnaire is completed by the respondent. When the information is outside the range of expected values (too large or too small), or inconsistent with previous entries, the respondent is prompted, through messages on the screen, to check the information. However, for some questions, the respondent may ignore the edits and skip questions if they do not know the answer or refuse to answer. For this reason, the response data undergo further edit and imputation processes after being received at the Head Office.

Range edits
These were built into EQ to deal with questions asking for numeric values. If values entered are outside the range, the system generates a pop-up window that states the error and instructs EQ respondent to make corrections to the appropriate question. For example, if the value entered into the computer for the number of nights away from home on the trip is high, a pop-up message will appear asking the respondent to confirm the answer.

Data are sent to Statistics Canada, where the information is processed in stages in preparation for dissemination. Data are checked to identify any inconsistencies. Trip records are validated to ensure that values in mandatory fields are acceptable. For some variables, a range of acceptable values is used. For example, we make sure that the number of nights falls within the logical range, that the type of trip is valid, etc.

For the majority of trip records, the geographic area is coded automatically. For a small number of records, coding is done manually at Statistics Canada's Head Office.

Flow pattern edits
All flow patterns were automatically built into the EQ. For example, if the trip was identified as an overnight trip, the applications would continue with a series of questions about the accommodations. If not, this series of questions are automatically skipped.

General consistency edits
Some consistency edits were included as part of the collection system to allow the EQ respondent to return to previous questions to correct inconsistencies. The system generates a pop-up window that states the error and instructs the respondent to return to the appropriate question to confirm the data and make corrections as required.

Several consistency edits are carried out on the data to verify the relationship between two or more variables. For example, the number of adults in a household who went on a trip cannot exceed the total number of adults in the household. If a city or other specific geographic location does not correspond to the province or other larger geographic area, only one location will be retained, depending on the question. For expenditure variables, several edit rules are applied to limit these values. If the value does not fall within the predetermined acceptable range, it will be imputed later.

5.2.2 Pre-edits

For all records where values were missing (blank) from the collection, the value of “9,” “99,” “999,” etc. was inserted to indicate that no information was collected. The “Don’t know” values returned by the both collection application as code “9” are changed to “7” in the pre-edits. As well, the “Mark all that apply” questions were unstrung and values converted to “Yes” (1) or “No” (2) responses. Finally, all text answers were removed from the processing file and set aside to be handled separately.

5.2.3 Flow edits

The flow edits replicate the flow patterns from the questionnaire. Variables that are skipped based on flows are converted from “Not stated” to “Valid skip” codes (“6,” “96,” “996,” etc.). For skips based on age or the answer to certain questions, skipped questions are set to “Valid skip.” For skips based on “Don't know” and “Refusal,” skipped questions are set to “Not stated”.

5.2.4 Consistency editing

After the flow edits were completed, consistency editing was carried out to verify the relationship between two or more variables. Decision tables are used to specify the consistency edits. LogiPlus software is used to input the decision tables and generate the SAS code. For more complex edits SAS programming is used. A report with the “before” and “after” counts of the variables is generated. Additionally, a report is generated providing the rule counts for each decision table.

5.3 Coding open-ended questions

A few data items on the NTS questionnaire were recorded by EQ respondents in an open-ended format. For example, if the carrier used for the cruise could not be found on the lookup list, the respondent can manually enter it. This text was coded using an in house system.

5.4 Creation of derived variables

A number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. For example MREASON identifies the main reason for taking the trip. Several questions are combined in order to create this new variable.

5.5 Imputation

Imputation is a process used to determine and assign replacement values to resolve problems of missing, invalid or inconsistent data.

5.5.1 Expenditures imputation for the number of accompanying people

The number of people who live in the same household as the respondent and accompanied the respondent on a trip is required for both analysis and calibration. It is also used in the expenditure imputation. If the values are missing (or if they contradict the reported household size) then they will need to be imputed.

The imputation is done using a donor imputation strategy within the BANFF software that relies on the nearest neighbour approach.

5.5.2 Expenditures imputation

The expenditure imputation system ensures that all selected trips have valid values for each of the relevant spending components. Spending components will be imputed for selected trips when there is a high probability that the respondent entered an incorrect value or if they did not enter a value at all.

The first step of the expenditure imputation system uses flags to identify values that need to be imputed. These flags are initially created during the edit step of the processing system.

After flagging the spending variables to impute for the trips, the random hot-deck imputation within a class is used as the imputation method. Instead of imputing a donor spending directly, the donor spending is divided by a divisor (ex. The number of nights) and multiplied back by the divisor of the recipient.

Once a value has been flagged for an imputation, the system will randomly select a donor record within the same imputation class as the recipient. The ratio (ex. Spending per day or spending per person per day) of the donor will be imputed to the recipient. Then, this ratio will be converted to a spending value by multiplying by the divisor. For example, if we impute spending per day to a given record, this value will have to be multiplied by the number of days. The table below shows the divisor variables, in order to create the ratio, for each spending category.

To ensure that extreme values would not be imputed, the top 95% and bottom 5% of values in the donor pool are excluded and cannot be used to impute records. Outliers were also excluded from the donor pool.

5.6 Package Cost Imputation

If a respondent indicated that a travel package was purchased for their trip and the total cost of the package was either $0 or not reported then the total cost of the package will be imputed using a donor. . In this case, the spending value of a donor (not a ratio) is imputed to the recipient. The imputation method used is random donor within classes. Hierarchical classes are built based on 8 characteristics. For example, if a donor for a domestic trip that had the same value for all of the 8 matching characteristic was not able to be found then the less correlated variable to package spending would be dropped. Then the system would look for a donor that matched on all remaining variables. If a donor still could not be found, then the second less correlated variable would be dropped and so on until a potential donor can be found.

5.7 Disclosure control

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

For this reason, identifiers are not included on the file, as well as some socio-demographic and geographic variables which could have been used to identify respondents. For the socio-demographic variables that are on the PUMF, they have been grouped into broad categories.

6.0 Data quality

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

6.1 Non-sampling errors

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

6.1.1 Non-response

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

Non-response produces errors in the survey estimates in two ways. First, is that non-respondents often have different characteristics from respondents, which can result in biased survey estimates if non-response is not corrected properly. Secondly, it reduces the

effective size of the sample, since fewer units than expected answered the survey. As a result, the sampling variance increases and the precision of the estimate decrease.

The following table summarizes the response rates:

Province	Selected households	Out-of-scope households	Responding households	Response rates
Newfoundland and Labrador	24 685	1 241	4 150	17,7%
Prince Edward Island	16 428	1 537	2 881	19,3%
Nova Scotia	28 444	1 430	5 938	22,0%
New Brunswick	26 138	1 151	4 845	19,4%
Quebec	82 199	1 827	19 850	24,7%
Ontario	106 950	2 218	28 582	27,3%
Manitoba	38 138	686	9 645	25,8%
Saskatchewan	40 991	870	10 479	26,1%
Alberta	54 706	1 214	12 879	24,1%
British Columbia	57 517	1 338	14 803	26,3%
Canada	476 196	13 512	114 052	24,7%

6.1.2 Coverage errors

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

6.1.3 Measurement errors

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

The NTS electronic questionnaire incorporates features to serve to maximize the quality of data collection. There are edits built into the EQ that compare unusual responses and that check for logical inconsistencies. The EQ controls the sequence of questions in accordance with responses to previous questions. Other checks are also made during the questionnaire to reduce the number of errors attributable to typos and misunderstandings. For example, if the number of nights spent in various types of accommodation does not correspond to the total number of nights spent away from home, an edit message will appear on the screen. The respondent can then correct the mistake, and less editing has to be performed at the Statistics Canada head office. As well, for some questions, the respondent has the possibility to enter "Don't Know" or "Refused" as a valid response if he or she does not answer the question.

6.1.4 Processing errors

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

Data processing of the NTS was done in a number of steps including verification, coding, editing, imputation, etc. At each step, measures are taken to ensure quality data are produced and reports are created for verification.

6.2 Sampling errors

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

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

Factors affecting the magnitude of the sampling variance include:

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

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

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

It is more useful in many situations to assess the size of the standard error relative to the estimate of the characteristic being measured. The coefficient of variation (CV) provides such a measure. It is the ratio of the standard error of the survey estimate to the average value of the estimate itself, across all possible samples. The coefficient of variation is usually computed as the estimate of the standard error of the survey estimate to the estimate itself. This relative measure of sampling error is usually expressed as a percentage (10% instead of 0.1). It is very useful in comparing the precision of sample estimates, where their sizes or scale differ from one another.

6.2.1 Results for key estimates

For this reason, approximate CV tables have been provided for users to estimate the CV of the characteristics they are interested in. Appendix B explains how to use CV tables in various situations, and provides examples.

7.0 Weighting

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

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

This section provides the details of the method used to calculate sampling weights for the NTS.

7.1 Weighting procedures for the NTS

Three sets of weights are created and provided for the NTS: person weights, person-trip weights, and trip weights.

Person weights

The initial NTS weight is obtained by using the inverse of the probability of selection of the households selected. The households in which all members are less than 18 years old are out-of-scope for the NTS and dropped. In NTS-eligible households, one person aged 18 years or older is randomly selected.

The NTS weight is then adjusted for the selected household member by the following factors:

a) Non-response adjustment (household) : a factor is applied to the responding households within clusters where the response propensity is similar.
b) A calibration is done by province and household size with known totals of control.
c) Number of eligible persons (18+) in the household: this adjustment is capped at 6 to control the weight for respondents from rare households with many eligible people age 18 or over.
d) Non-response adjustment: To treat non-response, the NTS creates non-response classes within each province. These classes are created by modelling response probabilities using demographic variables, and using a clustering algorithm to create classes with similar response probabilities.

A calibration is done by province, age, sex and CMA (Census Metropolitan Area) with known totals of control.

Person-trip weight

The NTS sampling unit is the household (an individual aged 18 and over), but the major unit of analysis is a trip. A sampling unit may have taken any number of trips, including none at all. Furthermore, more than one individual may have taken the same trip together. In order to create a trip record with the appropriate weight attached to it, we first define a person trip as being any trip declared by a respondent (regardless how many other persons may have taken part in that trip).

a) The starting point for the person-trip weight is the person weight.
b) An adjustment for trips missing essential data: If the respondent reported both valid in-scope trips and in-scope trips missing essential data, the weight of trips missing essential data will be redistributed among the valid trips.

Trip weight

The person-trip weight is then adjusted for the number of persons aged 18 and over from the same household who went on that trip (including the respondent), to reflect the selection probability of the trip.

8.0 Guidelines for tabulation, analysis and release

This chapter of the documentation outlines the guidelines to be adhered to by users tabulating, analyzing, publishing or otherwise releasing any data derived from the survey microdata files. With the aid of these guidelines, users of microdata should be able to produce the similar 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. In some cases, the figures may differ because figures produced by Statistics Canada used the master files and not the PUMF.

8.1 Rounding guidelines

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

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

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

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

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

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

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

8.2 Sample weighting guidelines for tabulation

The sample design used for the NTS was not self-weighting. When producing simple estimates including the production of ordinary statistical tables, users **must** apply the proper survey weights.

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

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

8.3 Guidelines for statistical analysis

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

While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures may differ from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variances that are calculated are poor.

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

However, because the stratification and clustering of the sample's design are still not taken into account, the variance estimates calculated in this way are likely to be under-estimates.
The calculation of more precise variance estimates requires detailed knowledge of the design of the survey. Such detail cannot be given in this microdata file because of confidentiality. Variances that take the complete sample design into account can be calculated by Statistics Canada on a cost-recovery basis.

8.4 Coefficient of variation release guidelines

Before releasing and/or making inferences and/or publishing any estimates from the National Travel Survey, users should first determine the quality level of the estimate. The quality levels are *acceptable, marginal* and *unacceptable.* Data quality is affected by both sampling and non-sampling errors as discussed in Chapter 6.0. However for this purpose, the quality level of an estimate will be determined only on the basis of sampling error as reflected by the coefficient of variation as shown below. Nonetheless users should be sure to read Chapter 6.0 to be more fully aware of the quality characteristics of these data.

First, the number of respondents who contribute to the calculation of the estimate should be determined. By rule thumb, if this number is less than 30, the weighted estimate should be considered to be of unacceptable quality.

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

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

Quality level guidelines

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

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

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

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

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

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

Appendix A – Variance estimation for master files

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

Bootstrap method for variance estimation

Independently, in each stratum, a simple random sample of (*n*-1) of the *n* primary sampling units (PSUs) in the sample is selected with replacement, along with all of the households in the sample belonging to those PSUs. Note that since the selection is with replacement, a unit may be chosen more than once. An initial bootstrap weight is calculated for each sample unit in the stratum. This process of selecting simple random samples and recalculating weights for each stratum is repeated *B* times, where *B* is large, yielding *B* different initial bootstrap weights.

The description above is a description of the general bootstrap method. In the case of the TSRC, most strata contain a single PSU. Strata are therefore collapsed into super-strata, and *n*-1 strata are randomly selected from each super-stratum for each of the *B* bootstrap replicates. Coordination of bootstrap samples is also done, to allow for the variance to take into account the dependence between samples on two survey occasions. In the TSRC, the coordination is done because there is an overlap of samples between any 2 consecutive reference months (because any given interview contributes to the estimates of two consecutive reference months). To take this sample overlap into account, the same set of *B* selections of *n*-1 strata per super-stratum is carried forward from month to month.

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

Estimates of variance released by Statistics Canada are calculated using the bootstrap weights.

Statistical packages for variance estimation

Bootvar

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

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

SAS: https://www150.statcan.gc.ca/n1/pub/12-002-x/2014001/article/11901-eng.htm#a8

Other packages

Other than Bootvar, there are different commercial software packages that can carry out some design-based analysis for variance estimation; Stata 9 or above, SUDAAN and WesVar. Some descriptions about these software packages are provided here:

http://www.statcan.gc.ca/pub/12-002-x/2014001/article/11901-eng.htm#a4

Appendix B – Variance estimation for public use microdata files

Approximate sampling variability tables

In order to supply coefficients of variation (CV) which would be applicable to a wide variety of estimates produced from this microdata file and which could be readily accessed by the user, two sets of Approximate Sampling Variability Tables have been produced. These CV tables allow the user to obtain an approximate coefficient of variation based on the size of the estimate calculated from the survey data. Each of the two sets of tables should be used for different types of estimates. One set of tables is to be used for categorical estimates only, such as estimates of percentages of aggregates, while the other should be used for continuous estimates only, such as expenditures, person nights, person trips or household trips.

How to Use the Coefficient of Variation Tables for Categorical Estimates
The following rules should enable the user to determine the approximate coefficients of variation from the Approximate Sampling Variability Tables for monthly estimates of the number, proportion or percentage of the surveyed population possessing a certain characteristic and for ratios and differences between such estimates. Note that these tables can only be used for monthly estimates.

Rule 1: Estimates of Numbers of Persons Possessing a Characteristic (Aggregates)
The coefficient of variation depends only on the size of the estimate itself. On the Approximate Sampling Variability Table for the appropriate geographic area, locate the estimated number in the left-most column of the table (headed "Numerator of Percentage") and follow the asterisks (if any) across to the first figure encountered. This figure is the approximate coefficient of variation.

Rule 2: Estimates of Proportions or Percentages of Persons Possessing a Characteristic
The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, for any given time period and geographic area, the proportion of women who travelled is more reliable than the estimated number of female travellers. (Note that in the tables the coefficients of variation decline in value when reading from left to right).

When the proportion or percentage is based upon the total population of the geographic area covered by the table, the CV of the proportion or percentage is the same as the CV of the numerator of the proportion or percentage. In this case, Rule 1 can be used.

When the proportion or percentage is based upon a subset of the total population (e.g. those in a particular sex or age group), reference should be made to the proportion or percentage (across the top of the table) and to the numerator of the proportion or percentage (down the left side of the table). The intersection of the appropriate row and column gives the coefficient of variation.

Rule 3: Estimates of Differences between Aggregates or Percentages
The standard error of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard error considered separately. That is, the standard error of a difference $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ is:

$$\sigma_{\hat{d}} = \sqrt{\left(\hat{X}_1\alpha_1\right)^2 + \left(\hat{X}_2\alpha_2\right)^2}$$

where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$. This formula is accurate for the difference between separate and uncorrelated characteristics, but is only approximate otherwise.

Rule 4: Estimates of Ratios
In the case where the numerator is a subset of the denominator, the ratio should be converted to a percentage and Rule 2 applied. This would apply, for example, to the case where the denominator is the number of travellers and the numerator is the number of overnight travellers.

In the case where the numerator is not a subset of the denominator, as for example, the ratio of the number of same-day travellers as compared to the number of overnight travellers, the standard error of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by $\hat{R}$. That is, the standard error of a ratio $\left(\hat{R} = \hat{X}_1 / \hat{X}_2\right)$ is:

$$\sigma_{\hat{R}} = \hat{R}\sqrt{\alpha_1^{\,2} + \alpha_2^{\,2}}$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. The coefficient of variation of $\hat{R}$ is given by $\sigma_{\hat{R}} / \hat{R}$. The formula will tend to overstate the error if $\hat{X}_1$ and $\hat{X}_2$ are positively correlated and understate the error if $\hat{X}_1$ and $\hat{X}_2$ are negatively correlated.

Rule 5: Estimates of Differences of Ratios
In this case, Rules 3 and 4 are combined. The CVs for the two ratios are first determined using Rule 4, and then the CV of their difference is found using Rule 3.

Examples of Using the Coefficient of Variation Tables for Categorical Estimates
The following examples are based on fictitious data from a domestic travel survey carried in 2009 in an imaginary state called “Pangea”. These examples are included to assist users in applying the foregoing rules. Please note that the data for these examples are different than the results obtained from the current survey and are only to be used as a guide.

Example 1: Estimates of Numbers of Persons Possessing a Characteristic (Aggregates)

Suppose that a user estimates that 5,414,335 persons travelled within Pangea in April 2009. How does the user determine the coefficient of variation of this estimate?

1) Refer to the Person coefficient of variation table for Pangea.

Approximate Sampling Variability Tables - Person weight
Pangea Domestic Travel Survey - 2009
Pangea

Numerator of Percentage ('000)	ESTIMATED PERCENTAGE													
	0.10%	1.00%	2.00%	5.00%	10.00%	15.00%	20.00%	25.00%	30.00%	35.00%	40.00%	50.00%	70.00%	90.00%
1	213.1	212.1	211.1	207.8	202.3	196.6	190.7	184.6	178.4	171.9	165.1	150.8	116.8	67.4
2	150.7	150.0	149.2	146.9	143.0	139.0	134.8	130.6	126.1	121.5	116.8	106.6	82.6	47.7
3	123.0	122.5	121.9	120.0	116.8	113.5	110.1	106.6	103.0	99.2	95.3	87.0	67.4	38.9
4	106.5	106.1	105.5	103.9	101.1	98.3	95.3	92.3	89.2	85.9	82.6	75.4	58.4	33.7
5	95.3	94.9	94.4	92.9	90.5	87.9	85.3	82.6	79.8	76.9	73.9	67.4	52.2	30.2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
75	********	24.5	24.4	24.0	23.4	22.7	22.0	21.3	20.6	19.8	19.1	17.4	13.5	7.8
80	********	23.7	23.6	23.2	22.6	22.0	21.3	20.6	19.9	19.2	18.5	16.9	13.1	7.5
85	********	23.0	22.9	22.5	21.9	21.3	20.7	20.0	19.3	18.6	17.9	16.4	12.7	7.3
90	********	22.4	22.2	21.9	21.3	20.7	20.1	19.5	18.8	18.1	17.4	15.9	12.3	7.1
95	********	21.8	21.7	21.3	20.8	20.2	19.6	18.9	18.3	17.6	16.9	15.5	12.0	6.9
100	********	21.2	21.1	20.8	20.2	19.7	19.1	18.5	17.8	17.2	16.5	15.1	11.7	6.7
125	********	19.0	18.9	18.6	18.1	17.6	17.1	16.5	16.0	15.4	14.8	13.5	10.4	6.0
150	********	17.3	17.2	17.0	16.5	16.0	15.6	15.1	14.6	14.0	13.5	12.3	9.5	5.5
200	********	15.0	14.9	14.7	14.3	13.9	13.5	13.1	12.6	12.2	11.7	10.7	8.3	4.8
250	********	********	13.3	13.1	12.8	12.4	12.1	11.7	11.3	10.9	10.4	9.5	7.4	4.3
300	********	********	12.2	12.0	11.7	11.3	11.0	10.7	10.3	9.9	9.5	8.7	6.7	3.9
350	********	********	11.3	11.1	10.8	10.5	10.2	9.9	9.5	9.2	8.8	8.1	6.2	3.6
400	********	********	10.6	10.4	10.1	9.8	9.5	9.2	8.9	8.6	8.3	7.5	5.8	3.4
450	********	********	9.9	9.8	9.5	9.3	9.0	8.7	8.4	8.1	7.8	7.1	5.5	3.2
500	********	********	********	9.3	9.0	8.8	8.5	8.3	8.0	7.7	7.4	6.7	5.2	3.0
750	********	********	********	7.6	7.4	7.2	7.0	6.7	6.5	6.3	6.0	5.5	4.3	2.5
1,000	********	********	********	********	6.4	6.2	6.0	5.8	5.6	5.4	5.2	4.8	3.7	2.1
1,500	********	********	********	********	5.2	5.1	4.9	4.8	4.6	4.4	4.3	3.9	3.0	1.7
2,000	********	********	********	********	********	4.4	4.3	4.1	4.0	3.8	3.7	3.4	2.6	1.5
3,000	********	********	********	********	********	********	3.4	3.3	3.2	3.1	3.0	2.8	2.1	1.2
4,000	********	********	********	********	********	********	3.0	2.9	2.8	2.7	2.6	2.4	1.8	1.1
5,000	********	********	********	********	********	********	********	2.6	2.5	2.4	2.3	2.1	1.7	1.0
6,000	********	********	********	********	********	********	********	********	2.3	2.2	2.1	1.9	1.5	0.9
7,000	********	********	********	********	********	********	********	********	********	2.1	2.0	1.8	1.4	0.8
8,000	********	********	********	********	********	********	********	********	********	********	1.8	1.7	1.3	0.8
9,000	********	********	********	********	********	********	********	********	********	********	1.7	1.6	1.2	0.7
10,000	********	********	********	********	********	********	********	********	********	********	********	1.5	1.2	0.7
12,500	********	********	********	********	********	********	********	********	********	********	********	1.5	1.0	0.6
15,000	********	********	********	********	********	********	********	********	********	********	********	********	1.0	0.6
20,000	********	********	********	********	********	********	********	********	********	********	********	********	********	0.5

NOTE: FOR CORRECT USAGE OF THESE TABLES PLEASE REFER TO THE MICRODATA DOCUMENTATION

2) The estimated aggregate (5,414,335) does not appear in the left-hand column (the “Numerator of Percentage” column), so it is necessary to use the figure closest to it, namely 5,000,000.

3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 2.6%.

4) So the approximate coefficient of variation of the estimate is 2.6%. The finding that there were 5,414,335 (to be rounded according to the rounding guidelines in Section 8.1) travellers in the reference period is publishable with no qualifications.

Example 2: Estimates of Proportions or Percentages of Persons Possessing a Characteristic

Suppose that the user estimates that 516,365 / 2,865,929 = 18% of men who travelled within Pangea in January 2009 went hiking during their trips. How does the user determine the coefficient of variation of this estimate?

1) Refer to the Person coefficient of variation table for Pangea (see above). The Pangea level table should be used because it is the smallest table that contains the domain of the estimate, all men who travelled in January 2009 in Pangea. If, for example, the percentage was referring to men who travelled within a certain province of Pangea, then the table for that province should have been used.

2) Because the estimate is a percentage which is based on a subset of the total population (i.e. men that travelled within Pangea in January 2009), it is necessary to use both the percentage (18%) and the numerator portion of the percentage (516,365) in determining the coefficient of variation.

3) The numerator, 516,365, does not appear in the left-hand column (the "Numerator of Percentage" column) so it is necessary to use the figure closest to it, namely 500,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the percentage closest to it, 20.0%.

4) The figure at the intersection of the row and column used, namely 8.5% is the coefficient of variation to be used.

5) So the approximate coefficient of variation of the estimate is 8.5%. The finding that 18% of men that travelled within Pangea in January 2009 went hiking during their trips can be published with no qualifications.

Example 3: Estimates of Differences between Aggregates or Percentages

Suppose that a user estimates that 2,548,406 / 11,814,359 = 21.6% of women travelled within Pangea in January 2009, compared to the estimate for men of 2,865,929 / 11,436,728 = 25.1%. How does the user determine the coefficient of variation of the difference between these two estimates?

1) Using the Person Pangea coefficient of variation table (see above) in the same manner as described in Example 2 gives the CV of the estimate for women as 3.4%, and the CV of the estimate for men as 3.3%.

2) Using Rule 3, the standard error of a difference $\left(\hat{d} = \hat{X}_1 - \hat{X}_2\right)$ is:

$$\sigma_{\hat{d}} = \sqrt{\left(\hat{X}_1\alpha_1\right)^2 + \left(\hat{X}_2\alpha_2\right)^2}$$

where $\hat{X}_1$ is estimate 1 (men), $\hat{X}_2$ is estimate 2 (women), and α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

3) That is, the standard error of the difference $\hat{d}$ = 0.251 – 0.216 = 0.035 is:

$$\sigma_{\hat{d}} = \sqrt{[(0.251)(0.033)]^2 + [(0.216)(0.034)]^2}$$
$$= \sqrt{0.000069 + 0.000054}$$
$$= 0.011$$

The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$ = 0.011 / 0.035 = 0.314.

So the approximate coefficient of variation of the difference between the estimates is 31.4%. The difference between the estimates is considered marginal and Statistics Canada recommends this estimate to be released with caution and accompanied by a warning to users regarding its accuracy.

Example 4: Estimates of Ratios

Suppose that the user estimates that 807,261 women had at least one overnight trip within Pangea in November 2009, while 968,511 men had at least an overnight trip within Pangea during the same period. The user is interested in comparing the estimate of women versus that of men in the form of a ratio. How does she/he determine the coefficient of variation of this estimate?

1) First of all, this estimate is a ratio estimate, where the numerator of the estimate ($\hat{X}_1$) is women that made at least one overnight trip within Pangea in November 2009. The denominator of the estimate ($\hat{X}_2$) is the number of men that made at least one overnight trip within Pangea during the same period.

2) Refer to the Person coefficient of variation table for Pangea.

3) The numerator of this ratio estimate is 807,261. The figure closest to it is 750,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 7.6% (see above).

4) The denominator of this ratio estimate is 968,511. The figure closest to it is 1,000,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 6.4%.

5) So the approximate coefficient of variation of the ratio estimate is given by Rule 4, which is:

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2}$$

where α_1 and α_2 are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively. That is:

$$\alpha_{\hat{R}} = \sqrt{(0.076)^2 + (0.064)^2}$$
$$= \sqrt{0.0058 + 0.0041}$$
$$= 0.099$$

6) The obtained ratio of women that made at least one overnight trip within Pangea in November 2009 versus men that made at least one overnight trip within Pangea in the same period is 807,261/ 968,511 which is 0.83 (to be rounded according to the rounding guidelines in Section 8.1). The coefficient of variation of this estimate is 9.9%, which makes the estimate releasable with no qualifications.

How to Use the Coefficient of Variation Tables to Obtain Confidence Limits

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

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

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

Confidence intervals for an estimate, $\hat{X}$, are generally expressed as two numbers, one below the estimate and one above the estimate, as $\left(\hat{X} - k, \hat{X} + k\right)$ where k is determined depending upon the level of confidence desired and the sampling error of the estimate.

Confidence intervals for an estimate can be calculated directly from the Approximate Sampling Variability Tables by first determining from the appropriate table the coefficient of variation of the estimate $\hat{X}$, and then using the following formula to convert to a confidence interval $(CI_{\hat{x}})$:

$$CI_{\hat{x}} = \left(\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}}\right)$$

where $\alpha_{\hat{x}}$ is the determined coefficient of variation of $\hat{X}$, and

t = 1 if a 68% confidence interval is desired;
t = 1.6 if a 90% confidence interval is desired;
t = 2 if a 95% confidence interval is desired;
t = 2.6 if a 99% confidence interval is desired.

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

Example of Using the Coefficient of Variation Tables to Obtain Confidence Limits
A 95% confidence interval for the estimated proportion of men who travelled within Pangea in January 2009 that went hiking during their trips (from Example 2) would be calculated as follows:

$\hat{X}$ = 18% (or expressed as a proportion 0.18)

t = 2

$\alpha_{\hat{x}}$ = 8.5% (0.085 expressed as a proportion) is the coefficient of variation of this estimate as determined from the tables.

$CI_{\hat{x}}$ = {0.18 - (2) (0.18) (0.085), 0.18 + (2) (0.18) (0.085)}

$CI_{\hat{x}}$ = (0.18 - 0.031, 0.18 + 0.031)

$CI_{\hat{x}}$ = (0.149, 0.211)

With 95% confidence it can be said that between 14.9% and 21.1% of the men who travelled within Pangea in January 2009 went hiking during their trips.

How to Use the Coefficient of Variation Tables to Do a T-test
Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The sample estimates can be numbers, averages, percentages, ratios, etc. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

Let $\hat{X}_1$ and $\hat{X}_2$ be sample estimates for two characteristics of interest. Let the standard error on the difference $\hat{X}_1 - \hat{X}_2$ be $\sigma_{\hat{d}}$.

If $t = \dfrac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$ is between -2 and 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. That is to say that the difference between the estimates is significant.

Example of Using the Coefficient of Variation Tables to Do a T-test
Let us suppose that the user wishes to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of women who travelled within Pangea in January 2009 and the proportion of men who travelled within the country during the same period. From Example 3, the standard error of the difference between these two estimates was found to be 0.011. Hence,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}} = \frac{0.251 - 0.216}{0.011} = \frac{0.035}{0.011} = 3.18$$

Since t = 3.18 is not between -2 and 2, it must be concluded that there is a significant difference between the two estimates at the 0.05 level of significance.

How to Use the Coefficient of Variation Tables for Continuous Estimates

The approximate coefficients of variation from the Approximate Sampling Variability Tables for estimates of the number of household trips, person trips, person nights and expenditures (i.e., numeric variables) can be obtained by using a different set of tables that works slightly differently from the ones presented for aggregates and percentages. Regarding the calculation of coefficients of variation of ratios or differences between numeric variables estimates, rules 3 to 5 presented above apply as well.

Estimates of Numeric Variables

Similarly to the cases discussed above, the coefficient of variation also depends only on the size of the estimate itself. On the Approximate Sampling Variability Table for the appropriate geographic area and time period, first locate the column that corresponds to the variable of interest (i.e., household trips, person trips, person nights or expenditures). The intersection of this column with the corresponding magnitude of the estimate itself should give the CV of interest. Note that there are two separate columns with different scales for the estimates. The last column on the right should be used for expenditures estimates only, while the first column on the left should be used for estimates of the remaining numeric variables.

Estimates of Differences for Numeric Variables

Refer to Rule 3 above for Estimates of Differences Between Aggregates or Percentages.

Estimates of Ratios

Refer to Rule 4 above for Estimates of Ratios.

Estimates of Differences of Ratios

Refer to Rule 5 above, for Estimates of Differences of Ratios.

Examples of Using the Coefficient of Variation Tables for Numeric Estimates

A series of examples from the same fictitious data for the 2009 “Pangea Domestic Travel Survey” are presented below to familiarize the user with the Numeric Estimates tables. Since the rules 3 to 5 that apply to estimates of aggregates are also valid for estimates of numeric variables, no examples of differences or ratios of estimates of numeric variables are presented in this manual. Further, no examples of Confidence Interval calculations or t-tests are presented either for the same reasons. Please note that the data for these examples are different than the results obtained from the current survey.

Example 5: Estimates of Expenditures

Suppose that a user estimates that Pangea residents spent $786 million on domestic travel in May 2009. How does the user determine the coefficient of variation of this estimate?

1) Refer to the Person coefficient of variation table for Pangea – May 2009 below. Note that, in contrast with the categorical estimates case, there are separate tables for each month. In addition, there are tables available for quarters and for the year as a whole.

2) Since this is an Expenditure estimate, the estimate value can be found in the last column on the right side of the table.

3) The value of the estimate itself ($786 million) is not on the list, so it should be rounded to $800 million.

4) By looking on the “Expenditures” column, the approximate coefficient of variation of the estimate is 10.3%. The finding that the Pangea population spent $786 million (to be rounded according to the rounding guidelines in Section 8.1) in domestic traveling in the reference period is publishable with no qualifications.

Approximate Sampling Variability Tables
Pangea Domestic Travel Survey - May 2009
Pangea

Estimate ('000)	*Coefficient of variation for: Household Trips*	*Person Trips*	*Person Nights*	*Expenditures*	*Estimate ('000,000)*
5	70.2	80.2	142.3	43.7	5
10	57.4	64.4	109.8	35.8	10
20	47.0	51.8	84.7	29.4	20
30	41.7	45.6	72.7	26.2	30
40	38.4	41.6	65.3	24.1	40
50	36.0	38.8	60.1	22.6	50
60	34.1	36.6	56.1	21.5	60
70	32.6	34.9	52.9	20.6	70
80	31.4	33.4	50.4	19.8	80
90	30.3	32.2	48.2	19.1	90
100	29.4	31.2	46.3	18.6	100
200	24.1	25.0	35.7	15.2	200
300	21.4	22.0	30.7	13.6	300
400	19.7	20.1	27.5	12.5	400
500	18.4	18.7	25.3	11.7	500
600	17.5	17.7	23.7	11.1	600
700	16.7	16.9	22.3	10.7	700
800	16.1	16.2	21.2	10.3	800
900	15.5	15.6	20.3	9.9	900
1,000	15.1	15.1	19.5	9.6	1,000
2,000	12.3	12.1	15.1	7.9	2,000
3,000	11.0	10.6	12.9	7.0	3,000
4,000	10.1	9.7	11.6		4,000
5,000	9.4	9.1	10.7		5,000
6,000	9.0	8.6	10.0		6,000
7,000	8.6	8.1	9.4		7,000
8,000	8.2	7.8	9.0		8,000
9,000	8.0	7.5	8.6		9,000
10,000	7.7	7.3	8.2		10,000
15,000	6.9	6.4	7.1		15,000
20,000		5.8			20,000
25,000					25,000

NOTE: FOR CORRECT USAGE OF THESE TABLES PLEASE REFER TO THE MICRODATA DOCUMENTATION

Example 6: Estimates of Person Trips

Suppose that a user estimates that residents of Pangea between the ages of 18 and 24 took 63,226 same-day person trips in May 2009. How does he/she determine the coefficient of variation of this estimate?

1) Using the same table as Example 5, now the value of the estimate must be found in the first column to the left since it refers to person trips. This would be the case for estimates of person nights and household trips as well.

2) Similarly to the previous example, the value of the estimate itself (63,226) is not on the list, so it should be rounded to 60,000.

3) By looking on the “Person Trips” column, the approximate coefficient of variation of the estimate is 36.6%. This estimate is considered unacceptable and Statistics Canada recommends this estimate not to be released.

Appendix C – Differences between the Travel Survey of Resident of Canada and the National Travel Survey

Introduction
In 2015 a study was conducted with the objective to evaluate the feasibility of a new tourism travel survey of Canadian households to collect information on domestic and international travel from Canadian residents. This new survey, the National Travel Survey (NTS), aimed to meet the requirements of the System of National Accounts, to be financially autonomous, to be flexible to add new content and additional sample via cost recovery and to improve the quality of the Canadian outbound of the International Travel Survey. The result of the feasibility study produced clear recommendation to proceed with the new survey.

NTS is a replacement of the Travel Survey of Residents of Canada (domestic travel) and the Canadian component (Outbound) of the International Travel Survey. The new survey was implemented in February 2018 for the 2018 reference year.

The main issues of a survey redesign for stakeholders and users, both the System of National Accounts and external partners, are breaks in the time series and potential bias resulting from a single collection mode. Estimates from the 2018 results of NTS help to better understand the break in the time series and the extent of the bias when a single electronic collection mode is used.

This document will focus on domestic travel. The objective is to provide an overview of the methodology of the two surveys and the differences between the 2018 estimates produced with the NTS and the 2017 estimates obtained from the Travel Survey of Residents of Canada (TSRC). Observed differences in the estimates could be due to a true change in the population from 2017 to 2018. They could also be explained by the differences in the methodology of the two surveys. Or, it could be due to a mix of the two.

Methodology of NTS and TSRC
In the following sections, several aspects of the methodology of the two surveys will be described.

Sampling Design
Both surveys are household surveys using the same sample frame of Canadian dwellings. Although the TSRC is a supplement of the Labour Force Survey, and the NTS is an independent survey, when the NTS was developed, the design that was adopted for the survey was greatly inspired by the LFS design, since it is a proven and effective sampling design for a household survey. Because of the similar sampling design between the TSRC and the NTS, it is expected that sampling design will not be a contributing factor to the differences between the two surveys.

Target population and Coverage
In terms of coverage, there are no differences between the target populations of both surveys. The target population covers the Canadian Resident population 18 years old of age and over, living in the 10 provinces. Excluded from the survey's coverage are persons living on reserves and other Aboriginal settlements in the provinces, full-time members of the Canadian Forces and the institutionalized population.
Although the TSRC is a supplement of the LFS, where the LFS covers the population of Canadian Residents 15 year old and over, there is a selection for the TSRC that restrict the age group to 18 year old and over. So both target population are essentially the same.

Collection Mode and Response Rates
In regards to the non-response in the two surveys (more details on the rates in the next paragraph), an important difference between the two surveys that could explain the change in the response rates is that in the case of the TSRC, the survey was a supplement to the Labour Force Survey. Although not a mandatory survey, TSRC might have taken advantage of the fact that the LFS is a mandatory survey, and therefore the response rate might have been high because of this fact. Since respondents would have already participated in the LFS, there was a higher chance that they would continue to participate in the TSRC. Furthermore, the

TSRC response rate has probably benefited from the data collection process of the LFS which included Computer-Assisted Telephone Interviewing (CATI) and electronic questionnaire. The NTS on the other hand is not a mandatory survey and does not benefit from the LFS personal first contact. The first contact is made by letter, and there are no CATI follow-up made (due to budget constraints), which causes of a lower response rate.

As a supplement of the LFS, the TSRC response rate was approximately 75% on an annual basis. On an annual basis, approximately 105,000 households were selected for the TSRC, with about 80,000 respondent households. On the other hand, the annual response rate for the NTS is 28%. Starting from a sample of about 30,000 households monthly, the annual sample size is about 360,000 households, with about 108,000 responding households.

For both surveys, there is an adjustment factor for non-respondents. These adjustments are part of the weighting process. To treat non-response, the TSRC creates non-response classes within each province. These classes are created by modelling response probabilities using demographic variables, and using clustering algorithm to create classes with similar response probabilities. To treat non-response for NTS, a similar process is applied. The three demographic variables for NTS (age group, gender, household income) are part of the modelling of the response probabilities.

As the NTS does not include any CATI follow-up, there could be an opportunity to explore in further details the differences between respondents and non-respondents. Adding a CATI follow-up for non-respondents could confirm or negate our assumptions about non-respondent. As well, if there are some factors that could affect non-response, the follow-up could be tailored to specific units that could have potentially more impact on the estimates.

Data Processing

As for the processing errors, since the NTS processing system was greatly inspired by the TSRC processing system, there are no indication that this would be the cause of any major discrepancies between TSRC and NTS.

Finally, as both surveys are using an electronic questionnaire application, initial measurements errors should be similar between respondents to the TSRC and the NTS. The use of CATI follow-up for some of the TSRC respondents could reduce measurements errors, but this should not be a significant factor. Once again, the potential advantage of conducting specific targeted CATI follow-ups could confirm if respondent have issues with our concepts.

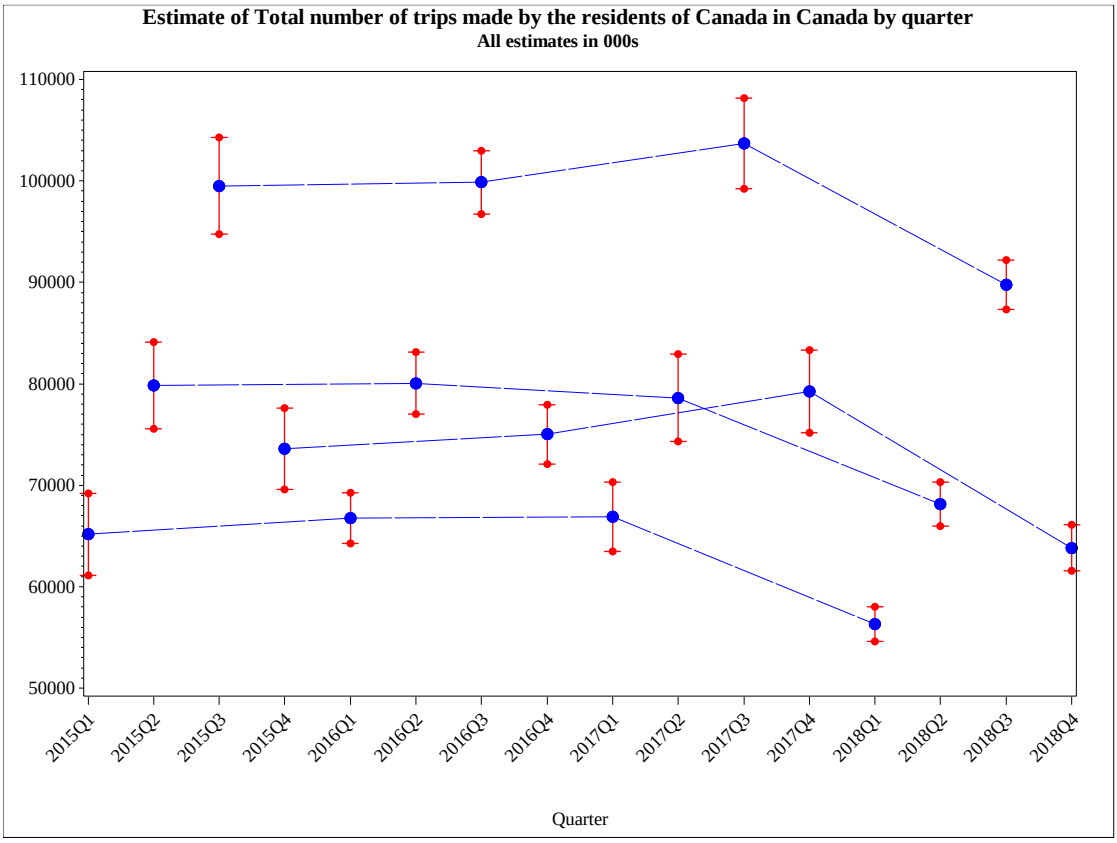
Detailed Comparison of the estimates

A detailed comparison between the NTS and the TSRC was undertaken. The analysis was done comparing three main variables: total number of nights, total spending and total number of trips. The analysis was done while taking into account the Coefficients of Variation (CV's) and the Confidence Intervals constructed around the estimations.

For the estimates of the total number of trips, the 2018 NTS estimates are lower and significantly different when compared to 2017 TSRC, and the estimates are also lower than the TSRC estimates for 2015 and 2016.

Domestic travel results: Total number of trips

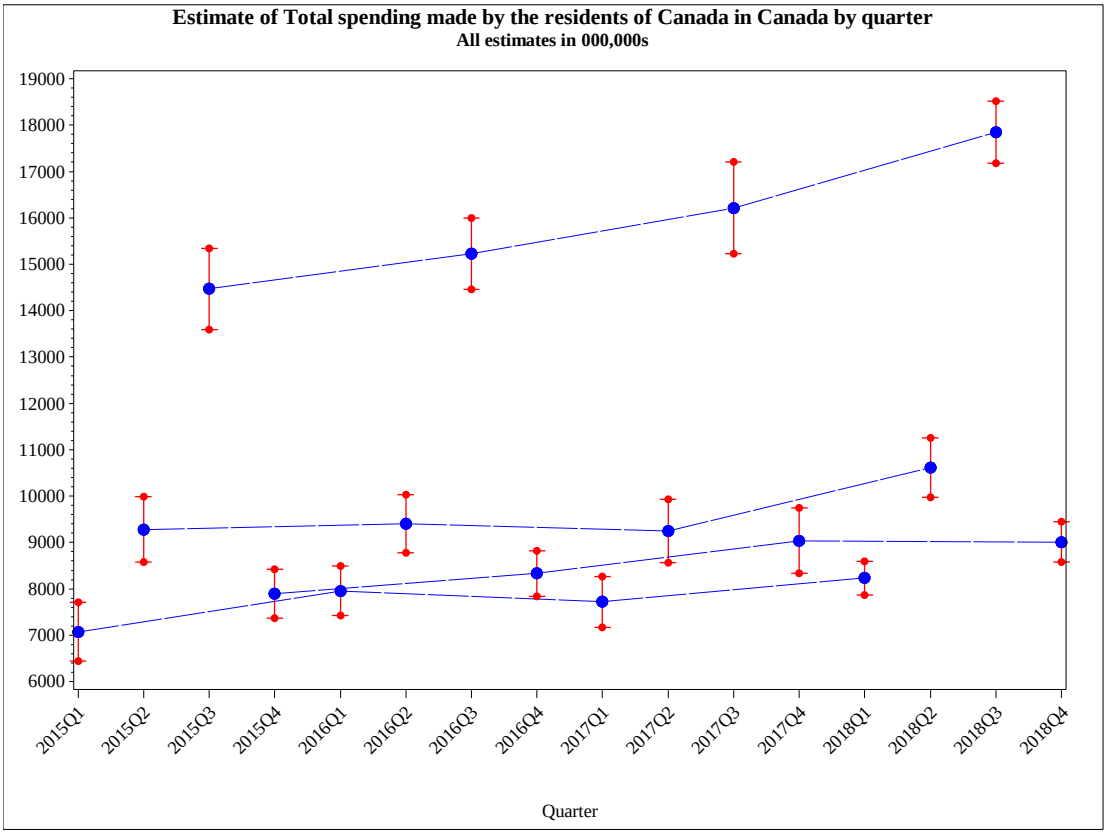
Note: results for Q1 2015 to Q4 2017 are obtained from TSRC
Results for Q1 2018 to Q4 2018 are obtained from NTS



For the estimates of the total spending, there are no statistical differences between the 2018 NTS estimates and the 2017 TSRC estimates. The estimates show an increase that is not statistically different, given that the confidence intervals overlap.

Domestic travel results: Total spending

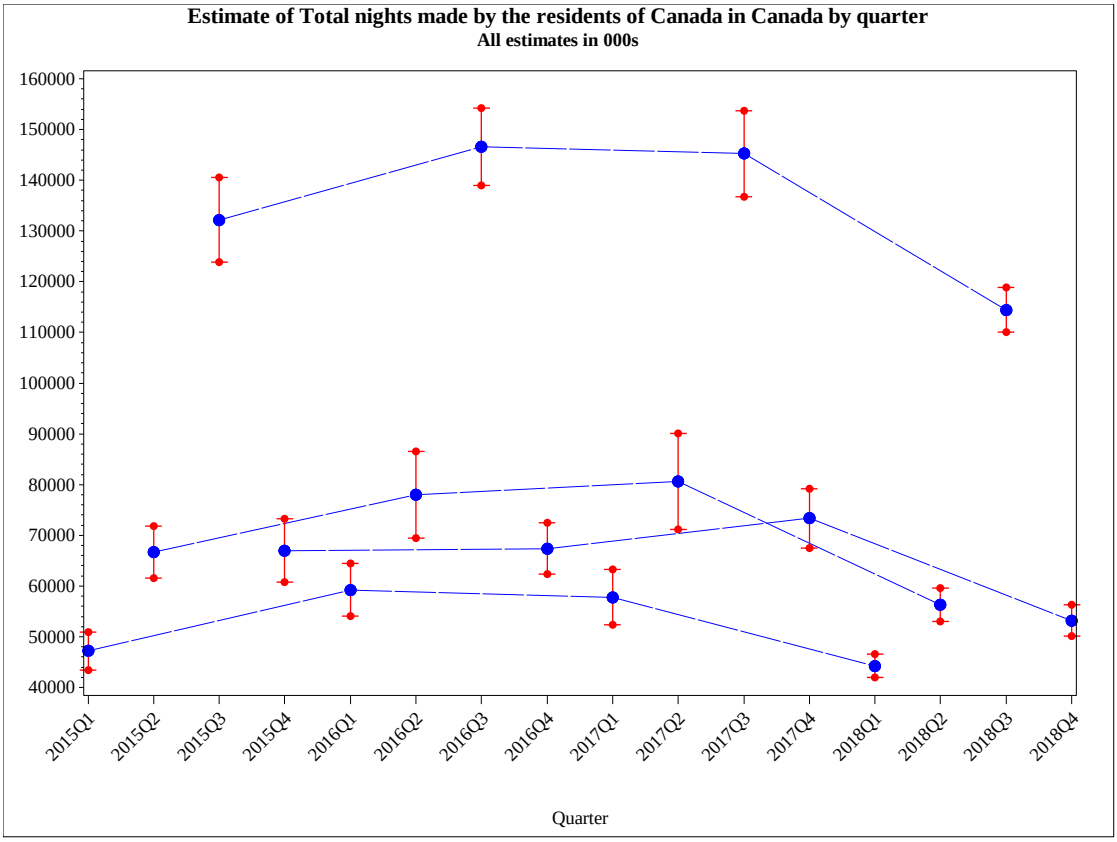
Note: results for Q1 2015 to Q4 2017 are obtained from TSRC
Results for Q1 2018 to Q4 2018 are obtained from NTS



For the estimates of the total number of nights, the 2018 NTS estimates for each quarter are lower and significantly different when compared to 2017 TSRC estimates, however they seem to be in line with the 2015 TSRC estimates.

Domestic travel results: Total number of nights

Note: results for Q1 2015 to Q4 2017 are obtained from TSRC
Results for Q1 2018 to Q4 2018 are obtained from NTS



A further analysis of the demographic variables for the respondent population for the TSRC and for the NTS will be done in the near future.

Conclusion

As it was described in the document, the Travel Survey of Resident of Canada and the National Travel Survey are two different surveys. Even though they target the same in-scope population, they yield different results, some of which are significantly different (total number of trips for instance) and some are not (total expenditures). Currently it is not possible to evaluate if these differences are a result of differences between the two surveys, or if the differences are reflecting true changes in the population between 2017 and 2018. Users should be careful when comparing results from TSRC and NTS because of the changes in the methodology between the two surveys. Statistics Canada does not recommend comparisons between the two surveys.