Proceedings of Statistics Canada Symposium 2014
Beyond traditional survey taking: adapting to a changing world

Canadian Vehicle Use Study: Electronic Data Collection

Émile Allie[1]

Abstract

The Canadian Vehicle Use Study is a survey conducted by Transport Canada in partnership with Environment Canada, Natural Resources Canada and the provincial registrars. The study is divided in two components: the light vehicles like cars, minivans, SUVs and trucks with gross vehicle weight (GVW) less than 4.5 metric tons; the medium and heavy component with trucks of GVW of 4.5 metric tons and more. The study is the first that collects vehicle activity directly from the vehicle using electronic collection methods exclusively. This result in more information, which is very timely and reliable.

Key Words: electronic collection, vehicle, big data, Canada.

1. Introduction

1.1 Objectives

The Canadian Vehicle Use Study (CVUS) is divided in two components: the light vehicles (CVUS-L) and the medium and heavy vehicles (CVUS-H). The aim of the light component is to measure various light vehicle-related quantities such as vehicle-km traveled, passenger-km traveled, fuel consumption, speed, fuel consumption ratio, etc. The medium and heavy component looks more at the cargo-tonnage, tonnage-km, vehicle configuration, type of cargo, etc. The estimates are provided at national and provincial levels for the last four quarters to support analysis and policy development.

1.2 Canadian Vehicle Survey

The Canadian Vehicle Survey (CVS) was conducted by Statistics Canada under contract to Transport Canada and Natural Resources Canada between 1999 and 2009. The CVS was hampered by low participant response rates over its existence caused in large part by the burdensome paper collection method. The quality of the estimates was also weakened by significant errors in the way in which the on-road vehicle fleet was classified.

1.3 Canadian Vehicle Use Study

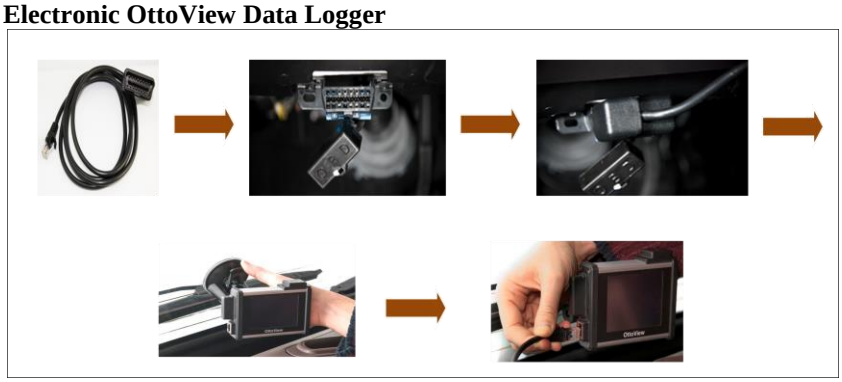
Transport Canada decided to conduct a revised CVS, with improved data collection method in collaboration with Natural Resources Canada, Environment Canada and the provincial registrars. Currently the study covers Prince Edward Island, Quebec, Ontario, Manitoba and Saskatchewan. This covers more than two-thirds of the fleet of vehicles in Canada. Following the recent changes in the *Motor Vehicle Safety Act*, we expect to cover all the provinces and territories starting the second quarter of 2015.

The survey is conducted each quarter and comprises two components: the light vehicles (cars, minivans, SUVs and trucks with a gross vehicle weight (GVW) of less than 4.5 metric tons), and the medium and heavy vehicles (GVW of 4.5metric tons and more). Section 2 will talk about the survey frame and Section 3 about the sampling design.

[1]Emile Allie, Transport Canada, 339 Sparks, Ottawa, Canada, K1A 0N5, emile.allie@tc.gc.ca.

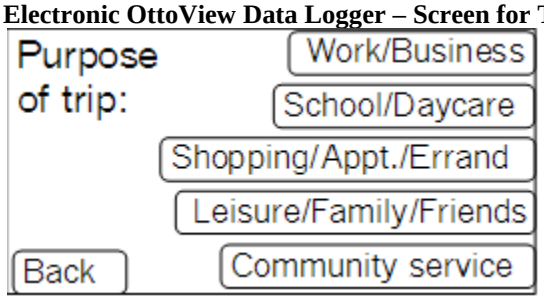
The study uses an electronic data logger to reduce reporting burden; to increase the accuracy and the amount of information collected; and to speed up the data processing. Figure 1.3.1 is an illustration of the electronic OttoView data loggers used in the study and how easy it is to install using the ODB-II/J1979 data port. To our knowledge, CVUS is the first study that collects vehicle activity directly from the vehicle using electronic collection methods exclusively.

Figure 1.3.1
Electronic OttoView Data Logger



The reporting burden is reduced with the electronic OttoView data logger. This is done with a small sequence of questions on the driver-ID number, the number of passenger by age group, the purpose of the trip and if fuel was bought. All theses question are answered through the electronic OttoView data logger screen (Figure 1.3.2). All the remaining information about the trip is collected from the ODB-II/J1979 data port without the driver intervention.

Figure 1.3.2
Electronic OttoView Data Logger – Screen for Trip Purpose CVUS-L



The electronic data logger collects more accurate information because the driver does not have to rely on the diary information or the information he/she collects or recalls at the beginning and the end of a trip like the distance driven, the driving time and the type of road. All those information are collected automatically.

The electronic data logger provides more information because the information is collected **every second** as soon as the car is started. Also, it is linked to the car computer, as a result, it collects information that was not possible to collect otherwise, such as the instantaneous speed and fuel consumption, RPM, intake temperature, cooling temperature, idling, time of day (up to the second), etc. This could be useful to analyze driving behaviors, fuel efficiency, and air emissions. The electronic data logger also uses the GPS technology and records the spatial coordinates that could be used in analysis of traffic congestion, road safety and infrastructure planning. Because we collect information on trips every second, we end up each quarter with more than 90 million of records. Section 6 will talk about the total information available.

The data processing time is significantly reduced because all the information is collected electronically, skipping the step of data transfer from a paper diary to an electronic database. Usually, statistical reports are available 60 days after the end of a quarter. Section 4 will talk about data processing and Section 5 about the weighting procedure.

2. Canadian Vehicle Use Study Survey Frame

2.1 Survey Frame

The survey frame will consist of motor vehicle registration files provided by each jurisdiction before the beginning of a quarter in order to minimize changes of address and maximize the accuracy of the fleet information The information is limited to what is required to select the vehicles in the frame. The vehicles in the frame are then divided into two groups: the light-vehicles, and both medium and heavy vehicles. The division is based on the carrying capacity of the vehicle. All the cars and trucks with GVW less than 4.5 metric tons define the frame for CVUS-L. The medium and heavy trucks (GVW 4.5 metric tons and more) define the CVUS-H frame.

2.2 Out of Scope Population

The frame for the CVUS excludes some vehicles (out of scope) like motorcycles, trailers, cranes, buses, ambulances, fire trucks, farm equipment, motorhomes, police cars, etc. In order to eliminate the out of scope vehicles and assign vehicles either to CVUS-L or CVUS-H, we use the Vehicle Identification Number (VIN), a Polk VIN Decoder and other information in the registration file (vehicle type, model year and gross vehicle weight rating). The VIN consists of a seventeen character alpha-numeric code for all vehicles whose model year is 1981 and newer (pre-1981 vehicles use a different standard). The first eleven characters define the make, model and other characteristics of the vehicle while the last six digits uniquely identify each vehicle.

3. Sampling

3.1 Sampling Design

The sampling design is illustrated in Figure 3.1.1. In each jurisdiction, depending on the model year of the vehicles, the vehicles are classified as 3 years or less, 4 years to 8 years, 9 years or more.

In CVUS-L we make the distinction between passenger car, minivan, light truck (lower than 4.5 metric tons) and SUV. In CVUS-H we classify the vehicles into medium light trucks with GVW between 4.5 to 10 metric tons, medium heavy trucks with GVW between 10 to 15 metric tons and the heavy trucks with GVW of 15 metric tons and more. The discussion that follows will focus only on CVUS-L.

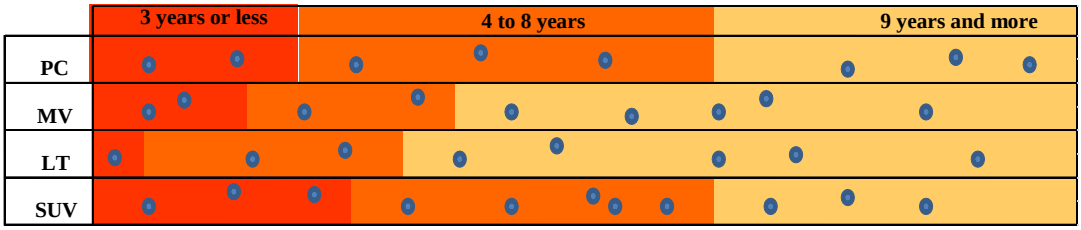
3.2 CVUS-L Sample Selection

In each quarter, there are 6,000 vehicles selected. The allocation to the four provinces participating is: 2,000 each for Quebec and Ontario and 1,000 each for Manitoba and Saskatchewan. Recently Prince Edward Island joined the study and a preliminary allocation to this province was 350 until we are able to assess their response rate.With the recent changes to the *Motor Vehicle Safety Act* we expect to include all the provinces and territories starting with the second quarter of 2015.

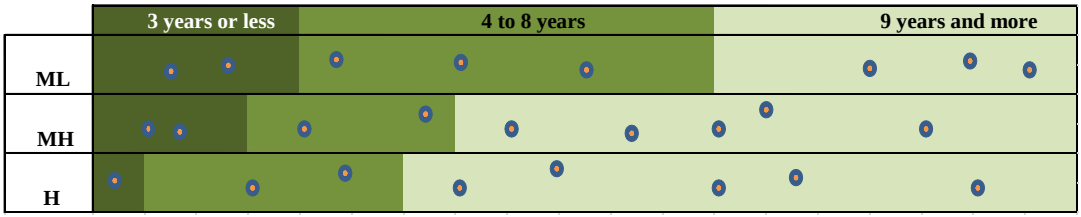
The sample of each province is stratified by type of vehicle and age (group) and the allocation to a stratum is proportional to the square root of the population in the stratum. At the Canadian level, the estimated final sample after non-response (around 1,000 per quarter) warranty a global confidence level of 95% and a 3% confidence interval. This will not be the case for each group in a province but the quality of data in groups with very small population reach a satisfactory level on an annual basis for most of the groups. The allocation to the stratum is proportional to the square root of the stratum population size.

Figure 3.1.1
Canadian Vehicle Use Study Sampling Design

Light Component



Medium and Heavy Component



PC: Personal Car
MV: Minivan
LT: Light Truck
SUV: Sport and Utility Vehicle

ML: GVW 4.5t to 10t
MH: GVW 10t to 15t
H: GVW 15t +

4. Canadian Vehicle Use Study Data Processing

4.1 Step1: Sample Selection

The data collection process starts with an agreement with a province and the vehicle registration service. One month before the end of a quarter, each participating province sends to Transport Canada a file with all the active vehicles. The file only includes the limited information required to stratify the sample. When the sample is selected, the list of the selected vehicles is sent to the province to get the contact information of the vehicle's owner.

4.2 Step2: Data Collection

A third party is hired to manage all the communications with the selected vehicle owners. Each sample is spread randomly in 13 batches to cover vehicle activities over the quarter. The owner of a selected vehicle receives an official letter from Transport Canada inviting him/her to participate in the survey. They are invited to respond via Transport Canada web site, by mail or through a tool free number and provide some basic information about the vehicle and the drivers (See Appendix C). If it is impossible to contact the owner, a telephone matching procedure is applied.

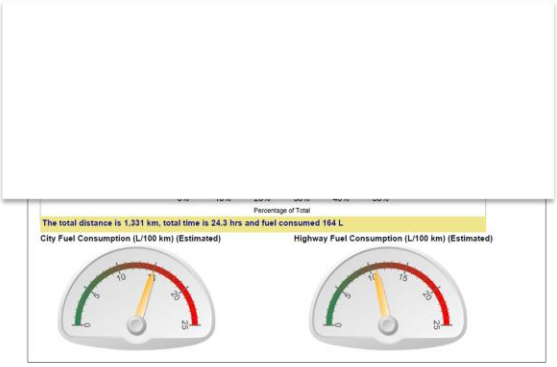
When the owners agree to participate, the consultant sends the logger and cables, an information kit and the date when to start and end the data collection. From the initiation of a first contact with a vehicle owner to the logger, it takes an average of 60 days (including the 21 survey days) for a full cycle.

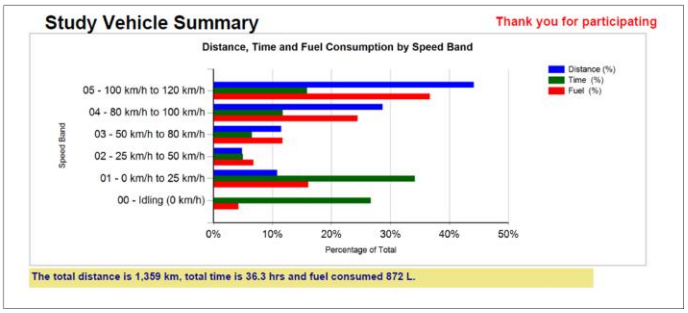
We provide incentives to participants to return the OttoView logger. One is a report on their use of the vehicle (Figure 4) and the other drivers. We also provide monetary incentives like an Early Bird draw ($500) for those returning the logger within 7 days after the end of the survey period; and a monthly draw ($1000) for those completing the survey and returning it within the month.
Figure 4: Extract from User Report

4.3 Step3: Production of Statistics

All the data cleansing and processing is done inside Transport Canada. The file available to the analyst does not contain private information about the vehicle owner. The private information is encrypted and accessible only to a limited number of persons. Figure 4.3.1 illustrates some of the information available in the participant report. Figure 4.3.2 is an example of the information shared with our partners. The information collected is used to produce summary tables shared with our partners. A reduced version is available on Transport Canada's website. The intra-trip information is based on a 10% random sample of the intra-trip file (around 90 million records).

Figure 4.3.1
Sample of Participant Report



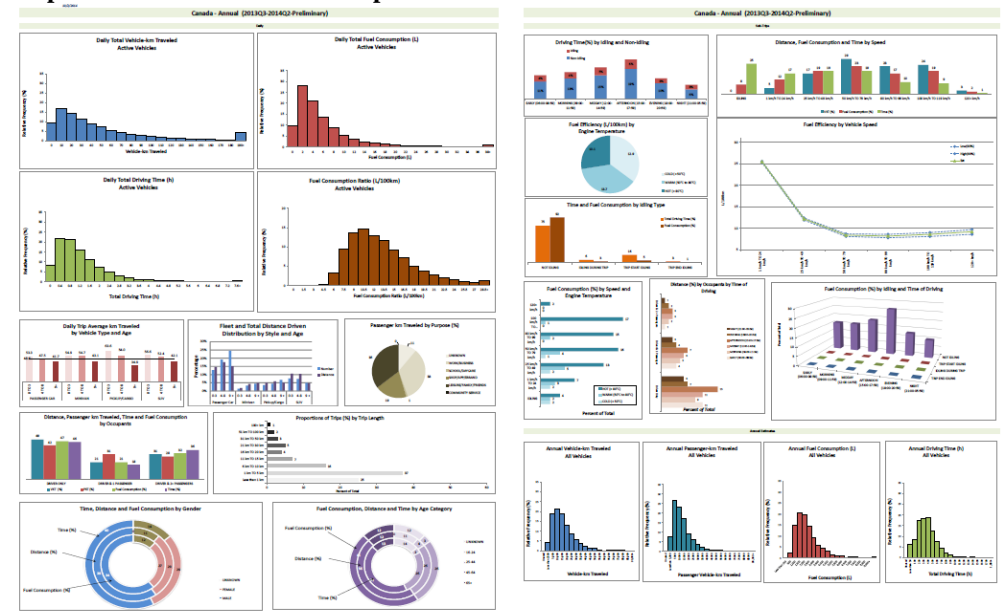


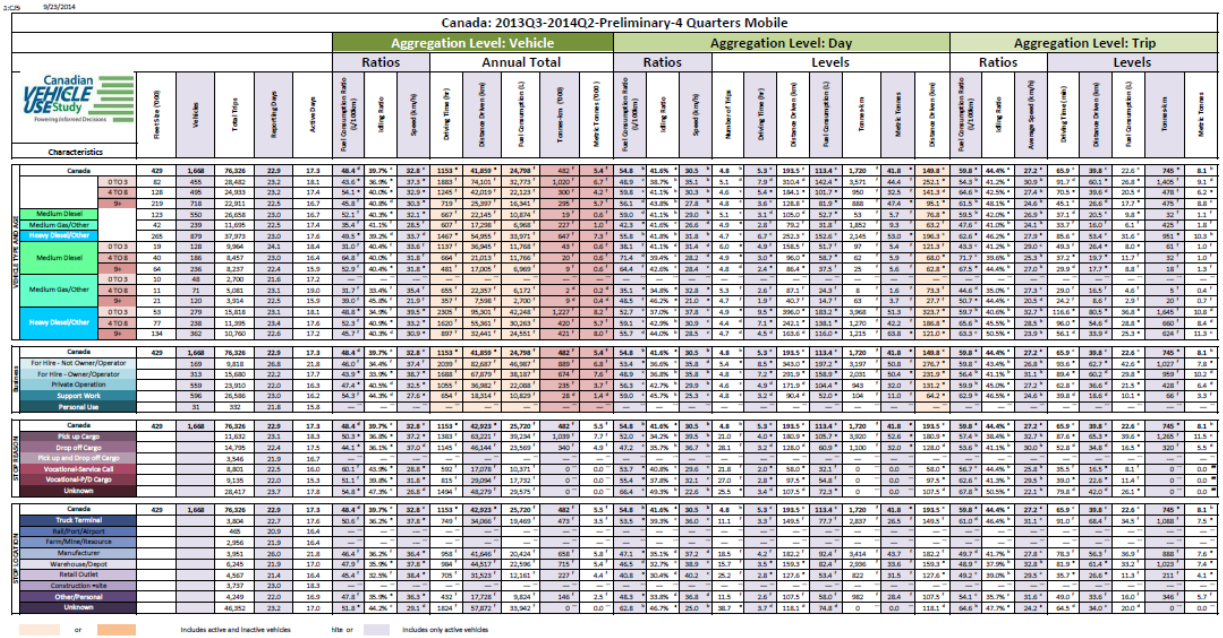
5. Weighting Procedure

Because the sample is stratified, all the selected observations have a weight which is the inverse of the selection probability within the stratum with adjustment for non-response. Supplementary adjustments to weights are done before using them for statistics production:

- A limit value of 21 was assigned to the weights meaning that a vehicle could not represent more than 21,000 vehicles within a quarter.
- For observations with a period of less than 21 days, the weights are reduced in proportion of the response period for a standardized period of 21 days e.g. if a respondent collect the information only for 5 days before dropping from the study, the associated weight, w, is reduced to w*5/21 because we do not know what is the driving behavior after the 5 days.
- Similarly, when the observation period is longer than 21 days, we do not want the observations associated to the vehicle to weight too much on the results. If the observation period is 30 days then the associated weight is reduced to w*21/30, given the standardized observation period of 21 days.

Figure 4.3.2
Sample of CVUS-L Statistical Report





6. A Large Amount of Information (Big Data)

CVUS produces each quarter a large amount of information. Five data sets are available to produce analysis. As an example, the intra-trip file described below includes around 90 million records. By linking those files, more than 200 variables may be available for analysis.

- **Registrar** files from the provinces that provide VIN.
- **Decoder** file that associate the VIN with a set of vehicle characteristics such as make, model, production date, engine displacement, number of cylinders, type of transmission, etc.

- **Contact** information which identifies the characteristics of the drivers like age class, gender. This information is collected when the participant sign in to the study. All other information is collected at the vehicle level.
- **Trip** information that comes from the questions answered at the beginning of a trip like the purpose and the age class of the passengers, the driver ID, the trip ID and the vehicle ID.
- **Intra-trip** information that provide a set of more than 65 variables registered each second like distance, fuel, geographic position, RPM, intake temperature, coolant temperature, etc.

7. Conclusion

The CVUS collects more information, more timely and more accurately with the use of an electronic data logger. The amount of information is more than 1,000 times the information collected using a paper log. Transport Canada expects to publish some research papers in a near future.