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

Challenges and Lessons Learned with the Implementation of Car Chips in the Fuel Consumption Survey

Agnes Waye, Serge Godbout, and Pierre Daoust[1]

Abstract

The National Fuel Consumption Survey (FCS) was created in 2013 and is a quarterly survey that is designed to analyze distance driven and fuel consumption for passenger cars and other vehicles weighing less than 4,500 kilograms. The sampling frame consists of vehicles extracted from the vehicle registration files, which are maintained by provincial ministries. For collection, FCS uses car chips for a part of the sampled units to collect information about the trips and the fuel consumed. There are numerous advantages to using this new technology, for example, reduction in response burden, collection costs and effects on data quality. For the quarters in 2013, the sampled units were surveyed 95% via paper questionnaires and 5% with car chips, and in Q1 2014, 40% of sampled units were surveyed with car chips. This study outlines the methodology of the survey process, examines the advantages and challenges in processing and imputation for the two collection modes, presents some initial results and concludes with a summary of the lessons learned.

Key Words: alternative collection, fuel consumption, response burden, collection costs, data quality

1. Introduction

In recent years, there has been an increasing interest in data related to environmental issues, such as fuel consumption. In 2013, Statistics Canada initiated the Fuel Consumption Survey (FCS), with the primary objective of providing accurate and reliable measures of kilometers travelled and fuel consumed by light vehicles (<4,500 kg) registered in Canada . This quarterly survey was conducted in 2013 and the first quarter of 2014. One distinctive feature of FCS is that it used two modes of collection: paper questionnaire and car chip. The electronic car chip records fuel consumed and distance driven while it is plugged into the respondent's car engine, so detailed data can be collected without increased respondent burden.

Until 2009, Statistics Canada used to conduct the Canadian Vehicle Survey (CVS), which required respondents to fill out a detailed trip log with information about each trip during the survey period. This survey provided very valuable and detailed information, but the respondent burden was very high. As a result of the low response rates, the CVS data quality was negatively impacted. With decreasing survey response rates in the recent years, it is important to minimize respondent burden, since this is one of the important factors that lead to nonresponse. FCS is able to collect detailed information for a subset of the respondents without the respondent burden associated with CVS due to its use of car chips.

This paper describes the methodology of FCS as well as the challenges and lessons learned from using the car chip as one of the modes of collection. There are numerous issues to consider when using car chips during all stages of the survey. In the remainder of the paper, we discuss the methodology of FCS and compare the two modes of collection. In section 2, a brief overview of FCS is given and the two modes of collection are described in detail. Section 3 describes the steps in the development of the survey, section 4 describes the methodology of FCS, section 5 gives some results and a comparison of the two modes, and finally, section 6 concludes this paper.

[1]Agnes Waye (agnes.waye@statcan.gc.ca), Serge Godbout (serge.godbout@statcan.gc.ca), Pierre Daoust (pierre.daoust@statcan.gc.ca); Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Canada, K1A 0T6.

2. Overview of the Fuel Consumption Survey

The objective of the survey is to provide reliable measures of total kilometers travelled and total fuel consumed by light vehicles registered in Canada. The target population for the survey is made up of all light vehicles that are registered in Canada. Examples of vehicles that are out of scope include trailers, tractors, large vans, etc. This is a voluntary survey.

2.1 Modes of Collection

As mentioned earlier, there are two modes of collection used in FCS: paper questionnaire and car chip. Each collection mode will be described in more detail in the following sections. As part of the pre-contact communication with the respondents, variables such as age, sex and whether or not the vehicle is used for commercial purposes are collected through computer assisted telephone interviews (CATI).

2.1.1 Paper Questionnaire

Respondents were asked to provide the following information in the questionnaire: odometer reading as well as fuel gauge reading at the beginning and end of the survey period, along with information about each fuel purchase (number of litres). Knowing the fuel tank size was also necessary for the survey's calculations, and sometimes this required manual research. Based on all this information, total distance travelled and total fuel consumed were derived. Respondents were also asked to provide information about their age, sex, as well as the percentage of distance driven for commercial purposes. The resulting data from questionnaires provided total fuel and total distance at the respondent level. Values of age, sex and commercial use from CATI were updated based on the values received from the paper questionnaire. The reason why these questions were asked in CATI was to collect information for these variables regardless of the mode of collection.

2.1.2 Car Chip

Car chips are able to collect information at the trip level for the respondents. Respondents were instructed to plug the car chip into the OBDII (On-Board Diagnostic) port of their vehicles for the duration of the survey period. The car chip takes readings from the engine every second while it is running, and stores information about the fuel consumed and distance travelled in its memory. The chip does not include any global positioning functionality, so there is no geographical information nor route collected. The car chip has some operational limitations: it does not work for vehicles with diesel or hybrid engines, as well as vehicles manufactured before 1996. The resulting raw data from the car chip were aggregated to the trip level. The resulting data from car chips provided total fuel and total distance at the trip level for each respondent, which is the distinguishing feature from data collected from paper questionnaires.

3. Steps in the Development of the Survey

There were two initial studies conducted before FCS was implemented in 2013. In 2012, a feasibility study was conducted to test the two collection modes (car chip and paper questionnaire). A small sample of units was instructed to install a car chip in their engines and also to complete a paper questionnaire in the same two week period. Based on the 125 responses received, the results from both collection modes were compared and analyzed. The findings showed that the distance reported from both modes were almost equal, but the fuel use reported from the questionnaire was 13% higher on average than those estimated through car chip. There were several reasons for this difference. First of all, the process for car chip data integration and derivation requires many input parameters. Car chip data has to be transformed as well as aggregated from the second by second level to the trip level and finally to the respondent level. Secondly, paper questionnaires are subject to non-sampling errors, such as inaccuracies in fuel gauge readings, errors in reading or transcribing values, data capture mistakes, misreporting fuel volume as cost of fuel, fill-ups from another car reported, etc. Next, a pilot survey was conducted in September 2012 with the objective of learning more about the collection systems, the length of the collection cycle, response rates and reasons for non-response. There were several difficulties encountered during the collection process. For

example, one issue was that interviewers did not have enough time to make contact with respondents, which resulted in a low contact rate. As a result, in 2013, FCS was implemented with improvements made to the collection process based on the experiences from the pilot survey.

4.0 Methodology of FCS

Given that FCS is a quarterly survey with a cross sectional design, the survey steps (sampling, edit and imputation and estimation) described in the following sections are repeated every quarter. The population parameters to be estimated are total distance ($t_{y_d} = \sum_{k \in u_d} y_k$) and total fuel ($t_{z_d} = \sum_{k \in u_d} z_k$), where u_d is the domain of interest.

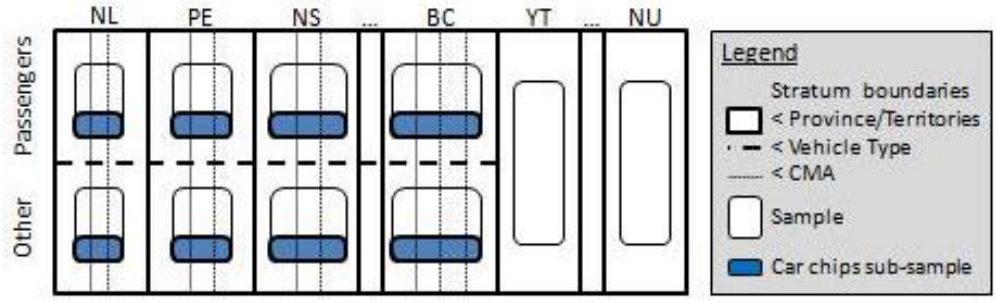
4.1 Sampling

The frame was created every quarter with the most recent provincial and territorial registration files, which contain contact information, make, model and year of the vehicle and other information on that vehicle. Each vehicle is identified by its unique Vehicle Identification Number (VIN), which contains information that describes the characteristics of that vehicle.

In 2013, the provinces covered by the survey were: Newfoundland and Labrador, Prince Edward Island, Nova Scotia, New Brunswick, Alberta, British Columbia, Yukon, Northwest Territories and Nunavut. In the first quarter of 2014, only Ontario was covered since Statistics Canada was interested in surveying a province with more urban areas. Vehicles were randomly sampled from each stratum, which were constructed based on province, census metropolitan area (CMA) and vehicle type. A certain percentage of the sample was allocated to the territories in 2013: 10% in the first quarter of 2013, and 4% for the rest of the year. The territories were not sub-stratified since there are no CMA's and the number of vehicles is quite low.

Every quarter, the split between the two collection modes were determined by budget constraints. The collection period was split into collection waves every quarter, with each wave being three weeks long, to manage collection resources. The provincial sample was equally allocated into waves, where each respondent was instructed to provide information for the duration of the wave. A percentage of the sample was allocated to car chip: 10% in the first quarter of 2013, and 5% for the rest of the year. 40% of the sample was allocated to car chip in the first quarter of 2014. The method of assigning the mode of collection was done by province and wave to facilitate managing the inventory of the car chips, taking into consideration vehicle limitations and targeted ratios of car chips described above. The purpose of assigning the modes of collection in this randomized method was to avoid bias. In some cases, the sampled unit requested to switch from car chip to paper questionnaire. This was allowed in FCS since it was acknowledged that some respondents were reluctant to install and use car chips. A visual representation of the sample design can be seen in table 4.1-1.

Table 4.1-1
Diagram of Sample Design



4.2 Processing

After the collection stage, the data provided by the respondents was processed. The key variables (total fuel z_k consumed and total distance driven y_k for vehicle *k*) were calculated according to Table 4.2-1 for the two modes of collection. The paper questionnaire variables were based on odometer and fuel gauge information provided by the respondent, while for car chip, all the values of the variable for vehicle *k* and trip *i* (y_{ki}, z_{ki}) were summed up for all the trips belonging to vehicle *k*, denoted by t_k.

Table 4.2-1
Calculation of total distance and total fuel for the two modes of collection

Parameter	Paper Questionnaire	Car Chip
Total Distance	$y_k = Odometer_{end,k} - Odometer_{start,k}$	$y_k = \sum_{i \in t_k} y_{ki}$
Total Fuel	$z_k = \left(FuelGauge_{start,k} - FuelGauge_{end,k}\right) * TankCapacity_k + FuelFillups_k$	$z_k = \sum_{i \in t_k} z_{ki}$

4.3 Edit and Imputation

After the processing stage, the data was verified for consistency and errors by undergoing various detailed checks. In some cases, manual reviews were made to the data. Different sets of edit steps were necessary for each mode of collection, since the format of the data differed for each mode of collection. Any inconsistencies between CATI responses and values from the paper questionnaire were reconciled. Variables were checked to make sure their reported values were within acceptable ranges. In addition to the key derived variables such as distance driven and fuel consumed, other variables such as fuel efficiency were also calculated.

There were some units with item non-response. Missing items such as age, sex, percent commercial use of car, and fuel consumed were imputed using donor imputation within classes based on province and CMA groups. This process was implemented using BANFF, a Statistics Canada generalized edit and imputation system (Statistics Canada, 2014). Missing data for fuel consumed was imputed using a regression model with explanatory variables, such as total distance and fuel ratings based on a combination of highway and city fuel ratings from Natural Resources Canada. After this automated imputation, the dataset went through the consistency and validity edits again to ensure the data was completely free of error. All responding units provided distance so this field did not require any imputation.

The imputation rate for fuel consumed was 10.7% for 2013 and the first quarter of 2014. On average, imputation rates for age and sex were about 5% higher for car chip (27% for car chip compared to 22% for paper questionnaires). This was expected since paper questionnaires had two sources for these variables (the questionnaire itself as well as CATI), while car chip units only had CATI as a source.

4.4 Estimation

The design weights were calculated by dividing the stratum population count by the number of sampled units in the stratum. In order to account for the effects of unit total non-response, weighting adjustments were made. These adjustments were made by omitting the non-respondents and adjusting the design weights of respondents to account for the non-respondents. The design weights were also adjusted for total non-response by creating weighting classes based on province and CMA groups. These non-response adjusted weights were then calibrated to the provincial and CMA population counts of vehicles. These were the final weights w_k of the k^{th} vehicle, which were then used to compute the estimates. The estimates of interest, which were total distance driven and total fuel consumed, were respectively calculated for each quarter of 2013 as $\hat{t}_{y_d} = \sum_{k \in s_d} w_k y_k$ and $\hat{t}_{z_d} = \sum_{k \in s_d} w_k z_k$, combining both modes of collection. Quality indicators were produced, taking into account sampling and imputation variance estimates (Demnati, 2013). Annual estimates were produced as well for the year of 2013. For more details on the FCS, see Statistics Canada (2014).

5. Results and Comparison

5.1 Results

Table 5.1-1 shows the response rates of FCS for the 4 quarters of 2013 as well as the first quarter of 2014. The response rates were calculated in accordance with the Statistics Canada Standards on reporting response rates (Statistics Canada, 1998). As shown in this table, the response rates for paper questionnaires were higher for three of the five quarters. This is expected since the car chip is a relatively new technology and some people may be reluctant to try a mode of collection with which they are not familiar. Globally, the first quarter of 2014 had a higher response rate since only Ontario was covered, which is a province with significantly more urban areas compared to the provinces and territories covered in 2013, and therefore, the respondents are easier to locate.

Table 5.1-1
Response rate of FCS by mode of collection

Time Period	Overall	Paper Questionnaire	Car Chip
Q1 2013	45.0%	44.1%	58.2%
Q2 2013	47.7%	48.0%	41.9%
Q3 2013	46.2%	46.1%	47.8%
Q4 2013	47.0%	47.5%	39.5%
Q1 2014	55.4%	59.3%	49.0%

Table 5.1-2 shows the estimates for average distance $\hat{t}_{y_d}/\widehat{N}_d$ and average fuel consumed $\hat{t}_{z_d}/\widehat{N}_d$ in the year of 2013 for paper questionnaires and car chips for the Atlantic provinces, Alberta and British Columbia. Fuel efficiency is also shown here, which is calculated as litres of fuel consumed per 100 kilometers driven ($100\hat{t}_{z_d}/\hat{t}_{y_d}$). The territories are not included in this table since only paper questionnaires were used. There are no major differences between paper and car chip in terms of fuel efficiency, and there is no systematic difference for this variable. This suggests that, although it has a slight impact on response rates, the mode of collection doesn't influence the response values. The differences could not be statistically tested from 2013 given the small sample size of car chips but we expect to test them for Q1 2014 data in 2015.

Table 5.1-2
2013 Estimates of average fuel and average distance comparison (per vehicle)

Province/Region	Paper Questionnaire			Car Chip		
	Distance (km)	Fuel (L)	Fuel efficiency (L/100 km)	Distance (km)	Fuel (L)	Fuel efficiency (L/100 km)
Atlantic	16,500	1,800	10.7	13,100	1,600	12.3
Alberta	13,800	1,700	12.2	17,000	2,500	14.8
British Columbia	12,400	1,500	11.8	11,400	1,200	10.6

5.2 Comparison of the Modes

After conducting FCS for five quarters, we experienced different types of challenges with both modes of collection. Each mode has its own advantages and disadvantages. There are several strong advantages of the car chip. One advantage is that it provides precise information, since it collects data directly from the car engine. Unlike questionnaire data, car chips can provide trip level data, which means respondents can see a detailed report of their trips. This is a good incentive for people to complete the survey if they are interested in seeing details of their fuel consumption. The rich details provided by car chip (for example, idle time) can also be used for deeper analysis. However, there are some drawbacks when using the car chip. The cost of purchasing and mailing the car chips is more expensive when compared to paper questionnaires. Since the cost is higher, this results in a bigger financial loss for units that are not returned, whereas for paper questionnaires there is a minimal monetary loss. Car chips require more data processing, as mentioned earlier. The volume of data for car chips is extremely high when compared to paper questionnaires, since the chip records information every second that the engine is running. Also,

the perception of risk may contribute to a lower response rate for units assigned to car chip, since some people may be reluctant to install an external device in their vehicles. One possible effect of using car chips is that driving behaviour and habits may be different since the respondent knows detailed information such as speed is being monitored. For example, the respondent may drive more below the speed limit when the car chip is installed, even if he normally drives above the speed limit.

As for paper questionnaires, one obvious advantage is its low cost of printing and of mailing. Unlike car chips, paper questionnaires are not restricted to certain types of vehicles. Using paper questionnaires can prevent bias by giving coverage to drivers who do not want to install car chips. On the other hand, paper questionnaires do not provide trip level information. They are also prone to non sampling errors, such as reporting errors by respondents (for example, reading the fuel gauge as ¼ full instead of ½ full.) As a result, more reviews are required for paper questionnaires to ensure values are consistent and free of errors.

To summarize, there are different types of effort associated with each mode of collection. Installing an external device in the engine may be extremely troublesome for some respondents, whereas other respondents may find it very difficult to fill in detailed information from their odometer and fuel gauge. The perception of burden would depend on the respondent's level of comfort with the mode used.

6. Conclusion

In conclusion, the car chip is a very effective tool for data collection in vehicle surveys. As discussed in this paper, there are many aspects to consider when using car chips from both a methodology and collection point of view. The car chip is an efficient and effective technology, especially for reducing respondent burden and for improving accuracy, but it has its limitations. There are several improvements that could be considered for similar surveys in the future. Sample allocation could take into account collection costs of the different modes of collection. (The collection cost consists of the cost of acquiring the device, the mail-out/mail-in cost, and the cost of processing the data.) The percentage of car chips could be increased in the survey in order to better compare the two collection modes and to have access to more detailed trip level data. The richness of trip level data can be used to conduct further analysis. One of the future studies to be conducted for FCS is to better assess the mode effect based on data from the first quarter of 2014, since the percentage of car chip makes up a significant part of the sample.

Acknowledgements

We would like to give Vincent Dubé, Lyne Guertin and Danielle Léger special thanks for all their help with this paper.

References

Demnati, Abdellatif. (2013), Fuel Consumption Survey: Variance Estimation, Statistics Canada Internal Document.

Statistics Canada. (1998), "Policy on Informing Survey Respondents", *Statistics Canada Policy Manual*.

Statistics Canada. (2014), "Fuel Consumption Survey", Integrated Metadata Base. http://www.statcan.gc.ca/imdb-bmdi/2749-eng.htm

Statistics Canada. (2014), "Functional Description of the BANFF System for Edit and Imputation – Version 2.06" (Technical Report).