

Modernizing Statistics Canada’s Transportation Surveys

Serge Godbout, Statistics Canada¹

Nathalie Hamel, Statistics Canada

1. Introduction

The movement of people and goods is critical to any economy, especially in a trade-reliant country like Canada. As such, there is a need to continually assess and monitor the transportation system which requires quality statistical information. The transportation statistics program at Statistics Canada measures various aspects of activities related to the movement of people and goods (Statistics Canada, 2017a, 2017b, 2017c, 2017d, 2017e) and covers:

- Transportation by air statistics based on surveys related to the movement of aircraft, passengers and cargo by air for both Canadian and foreign air carriers operating in Canada;
- Transportation by rail statistics based on surveys related to rail transportation of passengers and cargo in Canada and between the United States and Canada;
- Transportation by road statistics based on surveys relating to road transportation of passengers by bus and cargo by trucks in Canada.

This paper describes the methodology innovations that were implemented for six financial and operating transportation surveys. Section 2 presents the transportation statistics program and the remaining sections detail the innovations that the financial and operating survey redesign bring to the sampling frame creation, sampling, collection, data processing, estimation and dissemination methodologies.

2. Financial and Operating Surveys within the Transportation Statistics Program

The transportation statistics program currently comprises different components, including financial and operating surveys, transactional goods and passenger surveys, and administrative registers to provide the air, bus, trucking and rail statistics.

Table 1 – List of Financial and Operating Transportation Surveys by Mode

Mode	Survey Name	Year of Redesign Implementation
Air	Annual Civil Aviation Survey (ACAS)	2017
	Quarterly Civil Aviation Survey (QCAS)	2017
	Monthly Civil Aviation Survey (MCAS)	2017
Bus	Annual Passenger Bus and Urban Transit Survey (APBUT)	2015
	Monthly Passenger Bus and Urban Transit Survey (MPBUT)	2017
Rail	Annual Survey on Rail Transportation (ASRT)	2018

There are currently six financial and operating surveys as listed on Table 1. The three annual surveys share the same objective to collect detailed annual financial, operating and employment data on carriers operating in Canada. Excluding rail, they have corresponding subannual surveys targeting the key data from the largest companies to get timely signals on the growth and the performance of the air and bus industries. The resulting estimates from the financial and operating surveys are used as input to the Canadian System of National Accounts, by Transport Canada, by other federal and provincial departments, and by

¹ Presented at the 53rd Annual Meetings of the *Canadian Transportation Research Forum*, June 3-6, 2018 at Gatineau, Quebec

transportation companies, consulting firms, universities and foreign governments. The information is used for the analysis of transportation activities and for marketing and economic studies, as well as the calculation of industry performance measures (Statistics Canada, 2017a, 2017b, 2017c, 2017d, 2017e). Note there are no financial and operating surveys covering the trucking industry since the termination in 2012 of the Annual Trucking Survey and its corresponding quarterly components.

The transportation statistics program is following a traditional survey-centric approach, based on a set of independent surveys. Each component has its own specific objectives, designs and processes. Outside of the aviation statistics subprogram, the program components have little, or very limited, coherence. New data requirements have to fit into an existing survey, or require the development of a new one, resulting in a reduction of relevance and timeliness and increased costs and complexity. A modernized transportation statistics program would include consolidated, simplified, flexible, standardized and connected components which are able to work together to address new data needs in a responsive and timely manner.

As a first step of the transportation statistics program modernization, the redesign of the six financial and operating surveys started in 2015 and will be completed later in 2018. The objectives of the redesign were the development of electronic questionnaires, the standardization of the survey populations and concepts, the use of modern technology and generic methods for data processing and dissemination and increased operational efficiency. The common processing model makes full use of Statistics Canada's corporate tools such as the Business Register, the Business Collection Portal, the Integrated Business Statistics Program (IBSP) and the Economic Disclosure Control and Dissemination System.

The financial and operating surveys are complemented with surveys on detailed movements of vehicles, passengers and goods, within Canada and between Canada and the United States or other international areas. These surveys concerned with movement make full use of transactional data and require well-adapted processing tools connected to various data libraries and registers. The redesign of the other program components has been undertaken but will not be covered in this paper.

3. Modernizing the Frame Creation Methodology

With respect to frame creation, the innovations brought into the financial and operating transportation surveys are the maximized use of the Business Register and the standardization of the survey frames.

Statistics Canada's Business Register is a frame for sampling and collection that was developed to be used by the economic survey program. It covers all units engaged in economic production within Canada. It provides a population of businesses from which a sample can be selected using criteria such as size, geography and industrial activities (Statistics Canada, 2017f). The redesign of the surveys uses the Business Register as the primary source of information for the survey populations and standardizes the frame creation process. All transportation financial and operating surveys extract their frame from the Business Register as the unique source of information. The industries from the aviation, rail and bus subsectors, targeted by the six financial and operating surveys, are listed in Table 2. The use of a common source for frame creation helps evaluate how the transportation industry is covered by the different surveys and identify potential gaps.

Besides the generic content, complementary information specific to surveys can be attached to operating entities corresponding to the survey frame using a standard feature developed by the Business Register called Survey-Specific Fields (SSF). For transportation, the main survey-specific information required to be added on the Business Register are survey in-scope flags, primary transportation activity code, revenue from transportation activities, legislation status and business identifiers shared with stakeholders. The SSF maintenance is synchronized by transportation mode (air, bus and rail), and not by survey, to improve coherence. Survey analysts monitor and update the business activity status, industrial and geographical classifications, size measures from revenue, operating structures and SSF content for all businesses

involved in the targeted industries. This way, the three civil aviation surveys create their frame from the same information which maximizes their consistency. Similarly, both annual and monthly bus surveys share the same frame and SSF content for their common portion.

Table 2 – List of Industries Covered by the Financial and Operating Surveys

Industry Subsector	Industry Code	Industry Description
481 – Air transportation	481110	Scheduled air transportation
	481214	Non-scheduled chartered air transportation
	481215	Non-scheduled specialty flying services
482 – Rail transportation	482112	Short-haul freight rail transportation
	482113	Mainline freight rail transportation
	482114	Passenger rail transportation
485 – Transit and ground passenger transportation	485110	Urban transit systems*
	485210	Interurban and rural bus transportation*
	485410	School and employee bus transportation
487 – Scenic and sightseeing transportation	485510	Charter bus industry
	485990	Other transit and ground passenger transportation
	487110	Scenic and sightseeing transportation, land

*: MPBUT only targets two industries: 485110 and 485210.

4. Modernizing the Sampling Methodology

With respect to sampling, the innovations brought into the financial and operating transportation surveys are the migration to a standardized sampling system with simplified methods and the realignment of some concepts toward survey objectives.

Table 3 – Sampling Design and Approximate Population and Sampling Sizes by Survey

Survey Name	Sampling Design	Population Size	Sample Size
Annual Civil Aviation Survey (ACAS)	Census	540	540
Quarterly Civil Aviation Survey (QCAS)	Census with threshold	540	28
Monthly Civil Aviation Survey (MCAS)	Census with threshold	540	8
Annual Passenger Bus and Urban Transit Survey (APBUT)	Sample	7000	1000
Monthly Passenger Bus and Urban Transit Survey (MPBUT)	Census with threshold	350	30
Annual Survey on Rail Transportation (ASRT)	Census	80	80

Among the six financial and operating surveys, only APBUT requires the selection of a probabilistic sample; ACAS and ASRT are population censuses, while the three subannual surveys are censuses of the largest population units (see Table 3). With the redesign, all six surveys migrated to a metadata-driven sampling system adaptable to various designs, powered by the generalized sampling system (G-Sam) developed by Statistics Canada. The use of a common sampling tool brings flexibility and reduced maintenance costs. Before its redesign, APBUT had a survey-specific tool that was already connected to G-Sam so the migration to a generic tool required only minimal changes to the global sampling strategy. However, this system allowed the adoption of available features to bring in targeted improvements, such as simpler sample stratification and allocation methods, without causing significant changes in the estimates.

While modernizing the sampling tools, some survey concepts have been realigned toward the survey objectives, primarily to improve their coverage.

- For both APBUT and ASRT, the statistical unit definition for businesses with complex operating structure was changed to target establishments (defined as profit centers) to improve the population coverage as some businesses in non-transportation sectors do have transportation operating entities;
- For APBUT, the cumulative revenue proportion of the smallest businesses excluded from sampling and collection increased from 5% to 10% to reduce their response burden. This nonsurveyed portion of the very small businesses is now taken into account in the survey estimates as described in section 7;
- Before its redesign, MPBUT was a fixed panel of 10 large urban transit companies and 2 large interurban and rural bus companies. To ensure a better population coverage and to provide more granular statistics at the sub-national level, the panel definition was redefined as the set of largest businesses that contribute to at least 75% to the cumulative total revenue by industry and province or territory. In 2017, the panel went from 12 to 30 businesses. The panel composition will be reviewed on a yearly basis;
- The sampling methodology of the three civil aviation surveys did not change in their redesign; the strategy remained the same with only the sampling system being changed.

5. Modernizing the Collection Methodology

With respect to data collection, the innovations brought into the financial and operating transportation surveys are the review of questionnaire content, the development of electronic questionnaires and the implementation of active collection management.

Before their redesign, survey data were collected directly from respondents using mailed-out and mailed-back paper questionnaires or electronic templates sent by electronic file transfer. For their implementation year, the collection instrument of each survey was redeveloped into an electronic questionnaire from a standard application developed by Statistics Canada. The modular approach to questionnaire design is described in Table 4. For complete survey questionnaires, please refer to the survey documentation listed in the references (2017a, 2017b, 2017c, 2710d, 2017e).

All sampled businesses receive an email invitation, or a secure-access code letter, to log onto Statistics Canada’s collection portal and report their data. ASRT has exceptions: federally regulated companies are required by law to report their data directly to Transport Canada, which are shared with Statistics Canada as an administrative file. Most surveys adopted standard functionalities available in the electronic questionnaire applications to facilitate respondents’ work.

- Values outside a preset format of a questionnaire cell are rejected: text field, non-negative value, etc.;
- Skip patterns, derived from preset parameters or previously reported values, hide non relevant content;
- The application can calculate autosums, e.g. the sum of the total consumed fuel reported by province;
- Some surveys allow the option of attaching files such as financial statements to complement, replace or revise the information collected in the industry characteristics module;
- To ensure the submitted data are as consistent as possible at the time of capture, collection edits can inform respondents when expected rules are not observed or when the reported values are significantly different from the value reported in a previous cycle. The respondent usually has the option of changing the reported value, confirming it or ignoring the message.

Table 4 – List of Electronic Questionnaire Modules

Modules	Module Name	Module Description
1	Introduction	Provides key information to the respondent
2	Business or organization and contact information	Validate or collect information for the business or organization

3	Main activity	Verify or collect the current main business activity
4	Reporting period	Collect start and end date of reporting period
5	Industry Characteristics	Collect survey variables
6	Changes and events that affected the business or organization	Get any changes or events that affected the reported values, compared with the last reporting period
7	Contact person	Collect information on the contact person
8	Feedback	Get feedback from the respondent

The content of the questionnaires have been reviewed in depth, with the exception of MPBUT, which only collects two variables.

- Common concepts have been standardized across the three civil aviation surveys and some variables have been dropped or pushed from the monthly questionnaire to the quarterly, or from the quarterly to the annual, to reduce response burden;
- Similarly, the questionnaire content for ASRT has been drastically streamlined to remove many variables requested from federally regulated carriers but not included in the survey dissemination;
- Filter questions were added to APBUT to restrict provinces to those where the reporting unit operates.

In order to better manage collection costs and response burden, the active collection strategy for nonresponse and failed-edit follow-up has been revisited. Sampled units are prioritized based on their impact on the quality of key survey estimates, like total revenues or total number of employees by industry, if there is non-response or incorrect reporting. Nonresponses or reported values with failed-edits from highly weighted sampled units are followed up until quality targets are met or the collection period ends. These active collection strategies have been implemented into APBUT and ACAS. The prioritization step is omitted for the other four surveys as all of their sampled units are considered high-importance due to their small sample sizes.

6. Modernizing the Data Processing Methodology

With respect to data processing, the innovations brought into the financial and operating transportation surveys are the implementation of flexible data import tools, a multi-source data integration strategy, the development of automated edit and imputation strategies and an enhanced tracking of data value sources.

The six surveys have all migrated into a common processing platform developed as part of the IBSP. The IBSP processing model, commonly call Rolling Estimates, is based on iterative cycles throughout the collection and analysis period which produce a set of cleaned microdata and estimates with quality indicators attached. These outcomes are connected to the active collection strategy described in section 5 to prioritize the list of units for follow-up (Statistics Canada, 2015).

As the first processing step, the different data sources are imported into the processing database as new data are made available. Variables may be transformed if needed; for example, variables can be renamed, added or have a unit of measure change (e.g. kilograms to metric tonnes). The imported data sources, including collected, administrative, tax and historical data, are integrated with manual corrections following a preset hierarchy and combined with sampling and collection frames and paradata to fill in all the fields for each processing record. The flexible integration module allows some specificities:

- APBUT uses an administrative file provided by the Canadian Urban Transit Association as an alternative source for collected data for eligible records or used in the imputation strategy;
- The collected data from the non-federally regulated companies of ASRT are combined with the administrative data from the federally regulated companies provided by Transport Canada;

- In ACAS, some fields that are not collected for small carriers are imputed to ensure consistency and relationships among all variables as they are processed all together.

Following the data integration, some businesses are considered inactive with respect to the survey scope because they are misclassified or out-of-business; they are flagged and have their variables zero-filled. The active records may be totally or partially incomplete and both types of nonresponse are addressed through imputation. Before their redesign, only APBUT had an automated edit and imputation strategy but this was revisited to improve its performance; the imputation for the other five financial and operating surveys used to be done manually. For all six surveys, edit and imputation strategies have been consistently developed, making full use of the IBSP data processing tools which include Statistics Canada's generalized system for edit and imputation (Banff). From this perspective, the migration to the IBSP platform is considered a major innovation for the transportation statistics program.

- As a preliminary imputation step, deterministic rules, such as if-then-else statements and sum-of-parts algorithms, are applied to fill in missing values;
- Values are tested against editing rules; in case of edit failure, values are deleted and sent to imputation;
- The imputation strategy is designed in a waterfall approach to make sure all missing values or values failing edit rules are imputed at the end of a block. The preferred methods are historical imputation, with or without trends from an auxiliary variable, and average ratio imputation using respondents from the current period. Direct imputation from tax data is only used for some financial APBUT fields, as the other surveys share final microdata with external stakeholders without tax data access permission. As it often brings data instability, donor imputation is usually considered as a fallback method;
- After all missing values or values failing edit rules are imputed, prorating steps are applied to make sure the parts are equal to their totals.

Another key innovation for all six surveys is that the IBSP processing system systematically generates accurate source detail codes by variable for all processed records. This code specifies the source of the data (collection, administrative, tax, manual correction, historical, etc.) or the specific imputation method for imputed values, as well as the presence of prorating. Since all surveys share the same codeset, it facilitates data validation and analysis and makes sure quality is consistently evaluated.

7. Modernizing the Estimation Methodology

With respect to estimation, the innovations brought into the financial and operating transportation surveys are the migration to a standard estimation system, the estimation for the take-none portion, the integration of macro adjustments and the adoption of enhanced quality measures.

All six surveys have migrated to the estimation system imbedded into the IBSP Rolling Estimates model. It is a metadata-driven tool using the Statistics Canada's generalized estimation system (G-Est) to produce domain estimates and quality indicators from sample or census data. Estimates are computed at several levels of interest such as industry, geography or other industry-specific groupings. The IBSP estimation model can combine up to three different sources of estimates:

- Survey estimates are produced by summing up the values from the Collection Entities, weighted by their estimation weights. Given that there is no reweighting since total nonresponse is corrected through imputation, and that five out of the six surveys are censuses, the estimation strategy stays very simple;
- Take-none estimates are calculated by summing up data modeled from tax revenue for the nonsurveyed portion on the sampling frame called take-none (APBUT only);
- Macro adjustments can be provided by survey analysts in order to address conceptual, coverage or other type of errors that are not handled by other methods.

The final estimation totals are generated by adding take-none estimates and macro adjustments to the survey estimates. The final estimated ratios are then derived from the final estimated totals. By default, all domain estimates are provided with the following attributes and quality measures:

- Domain sample size, as the number of active collection units having the domain characteristics;
- Sampling variances are attached to all survey estimates in order to measure their variability, calculated from a simple random sampling design. For the five census surveys, the sampling variances are respectively set to null, as is the sampling variance for all take-none estimates and macro-adjustments;
- Coefficient of variation (CV) defined as the square root of the sampling variance divided by the estimated value. When the estimated value is 0, the CV is set to null;
- Item response rate of an estimated total, defined as the total of the reported values, weighted by the variable values and the estimation weights, divided by the actual estimated total. The item response rate for a ratio is defined as the minimum of the numerator's and the denominator's item response rates;
- The variance due to imputation is not yet available but should be implemented shortly in the IBSP tools.

8. Modernizing the Analysis and Dissemination Methodology

With respect to analysis and dissemination, the innovations brought into the financial and operating transportation surveys are the development of complete review and adjustment tools, the implementation of a consistent survey quality evaluation with standardized quality ratings, the standardization of confidentiality rules under a common platform and a strategy for breaks in a series.

Included in the IBSP platform, the Review and Adjustment Facility (RAF) is the main tool for survey analysts to perform a top-down approach for data analysis and validation. The RAF receives and presents survey results from data processing and estimation. The survey analysts use the RAF to review the survey data and, when needed, make manual corrections to microdata or add macro-adjustments, which will be fed back into the next Rolling Estimates iteration as explained in section 6 and 7. The RAF is also used to capture, review and edit survey metadata.

Table 5 – Definition of Quality Rating Codes

CV (%)	Item Response Rate (%)			
	90 – 100	67 – 89.99	40 – 66.99	0 – 39.99
0 – 4.99	A	B	C	E
5 – 9.99	B	C	D	E
10 – 14.99	C	D	E	E
15 – 24.99	D	E	E	F(*)
25 – 34.99	E	E	E	F(*)
35 – ∞	F(*)	F(*)	F(*)	F(*)

(*): Survey estimates with a quality rating of F are automatically suppressed from publication.

Throughout the collection and analysis phase, subject matter analysts examine the data, verify the results, perform coherence analyses, study changes across cycles, and compare results with other sources/surveys. Survey quality is evaluated by Survey Methodologists on an ongoing basis. At the dissemination phase, a quality report is produced at the end of the yearly cycle. Coverage and sampling rates are evaluated and compared with previous cycles while quality ratings are derived from CVs and imputation rates (see Table 5) and attached to every estimate.

Disclosure control is a priority for Statistics Canada. Various confidentiality rules implemented in Statistics Canada's generalized system for disclosure control (G-Confid) are applied to data 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. All six surveys have been migrated to the Economic Disclosure Control and Dissemination System (EDCDS) to finalize the publication or the data release.

Following the first publication of APBUT under the new design, a backcasting process was done as the data historically published had to be adjusted to take into account the changes in the survey coverage caused by the take-none estimation. The other five surveys had no need to revise their historical data from a backcasting perspective as the changes to the methodology did not cause breaks in the series for comparable concepts.

9. Conclusion

The six financial and operating transportation surveys play an important role within the transportation statistics program. The annual surveys build a global overview of various characteristics of their respective industries while the subannual surveys provide timely signals on the growth and the performance of key indicators from large companies. Their redesign is part of the modernization initiative of the transportation statistics program. The review and standardization of the survey content and methods, the implementation of new processing features and the migration to common systems significantly increased the quality of the survey outputs, in particular in terms of relevance, coherence, accuracy, timeliness and interpretability, while reducing maintenance costs and response burden. They also fit into the major modernization exercise initiated by Statistics Canada, especially in terms of delivering user-centric services, developing leading-edge methods and data integration and sharing and collaboration.

To some extent, the existing transportation statistics program at Statistics Canada retains a modal-based focus reminiscent of a more regulated industry. At that time, transportation statistics were largely modal-based to meet the administrative needs of a more regulated industry. The statistics program was designed to address the question of *how* the industry is doing. With regulatory reforms, data needs have changed and policy needs evolved to address the question of *what* the industry is doing. Therefore, the modernization of the transportation statistics program will continue. Transactional surveys on the movements of vehicles, goods and passengers, along with the administrative registers are being redesigned. At the same time, program components will be better consolidated among these surveys and connected to other programs from Statistics Canada. The use of new data sources for statistical and analytical purposes, like structured data collected from businesses, administrative files provided by external stakeholders and GPS data, are also under evaluation.

Reference

- Statistics Canada (2015). Integrated Business Statistics Program Overview. Statistics Canada Catalogue no. 68-515-X. Ottawa. <http://www.statcan.gc.ca/pub/68-515-x/68-515-x2015001-eng.htm>
- Statistics Canada (2017a). Annual Civil Aviation Survey, Integrated Metadata Base (IMDB). <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2713>
- Statistics Canada (2017b). Quarterly Civil Aviation Survey, Integrated Metadata Base (IMDB). <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2712>
- Statistics Canada (2017c). Monthly Civil Aviation Survey, Integrated Metadata Base (IMDB). <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5026>
- Statistics Canada (2017d). Annual Passenger Bus and Urban Transit Survey, Integrated Metadata Base (IMDB). <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2798>
- Statistics Canada (2017e). Monthly Passenger Bus and Urban Transit Survey, Integrated Metadata Base (IMDB). <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2745>
- Statistics Canada (2017f). Business Register (BR), Integrated Metadata Base (IMDB). <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=1105>