Use of Administrative Data to Increase the Efficiency of the Sample Design for the New National Travel Survey

Chunsan (Charles) Choi, Amanda Halladay, and Keven Bosa[1]

Abstract

As part of the Tourism Statistics Program redesign, Statistics Canada is developing the National Travel Survey (NTS) to collect travel information from Canadian travellers. This new survey will replace the Travel Survey of Residents of Canada and the Canadian resident component of the International Travel Survey. The NTS will take advantage of Statistics Canada's common sampling frames and common processing tools while maximizing the use of administrative data. This paper discusses the potential uses of administrative data such as Passport Canada files, Canada Border Service Agency files and Canada Revenue Agency files, to increase the efficiency of the NTS sample design.

Key Words: Administrative data; sample frame; sample design; stratification.

1. Introduction

1.1 Background

The National Travel Survey (NTS), which begins collecting information in February of 2018, will provide statistics on domestic and international travel of Canadian residents. Data from the NTS will be used to measure the size and status of Canada's tourism industry in terms of volume, characteristics and economic impact of domestic and international travel by Canadians. It will be an ongoing survey that provides tourism income and expenditures for Canada's System of National Accounts. The NTS will replace two components of the current Tourism Statistics Program at Statistics Canada: The Travel Survey of Residents of Canada (TSRC), which collects information from Canadians regarding their domestic travel; and the Canadian resident portion of the International Travel Survey (ITS), which collects information from Canadians who travel abroad. The non-Canadian resident portion of the ITS, which collects information from foreign visitors travelling to Canada, will be captured in a new survey, called the "Visitors Travel Survey". This new survey is still in development and will not be in the field until at least 2019. The foreign component of the ITS survey will fill in the gap between the implementation of the NTS in 2018 and the implementation of the VTS. Table 1-1 summarizes the restructuring of the Tourism Statistics Program.

Table 1-1
Current Surveys and New Surveys of the Tourism Statistics Program

Place of Residence	Destination of Trip	Current Survey	New Survey
Canada	Inside Canada	TSRC	NTS
Canada	Outside Canada	ITS	NTS
Outside Canada	Canada	ITS	VTS

---

[1] Chunsan (Charles) Choi, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6 (charles.choi@canada.ca); Amanda Halladay, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6 (amanda.halladay@canada.ca); Keven Bosa, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A 0T6 (keven.bosa@canada.ca)

1.2 Challenges with Current Travel Surveys

The TSRC is a live supplement to the Canadian Labour Force Survey (LFS), meaning that the TSRC questionnaire is asked to LFS respondents immediately after they submit their questionnaire (either with a telephone interviewer, or online via an electronic questionnaire). As a supplement of the LFS, the length of the TSRC interview is limited to 15 minutes. As such, some questions cannot be asked within the time constraints. For example, there is not enough time to ask respondents about their international trips. Also, since the sample frame for the TSRC comes directly from LFS respondents, the weighting is complex due to the complicated multi-stage sample design of the LFS. The LFS is a mandatory survey and has higher priority since respondents are obligated to participate. However, the TSRC is a voluntary survey. As a result, when there are time or resource constraints, the interviewers give priority to the LFS and they sometimes have to limit their efforts to collect TSRC data.

The main challenge with the ITS is that there is no sample frame from which to sample the Canadian resident portion of the survey. As a result, invitation cards are distributed at the border as Canadian residents return from abroad, which invites them to go online and voluntarily complete the ITS questionnaire. Another drawback of the ITS is its operational dependency on the Canada Border Services Agency (CBSA) to distribute the invitation cards. The CBSA is a federal agency that is responsible for ensuring national security at the border by monitoring the flow of people and goods entering Canada. Since the distribution of the ITS invitation cards is not a priority for the CBSA, often the invitation cards are not distributed as they should be, especially during rush hour in the busiest entry ports. Due to these limitations, the ITS has experienced a low rate of questionnaires returned to Statistics Canada.

1.3 Corporate Business Architecture

The Corporate Business Architecture (CBA) is a comprehensive review of the services delivered by Statistics Canada, whose aim is to foster an environment of integration and efficiency. For Statistics Canada surveys, this can be interpreted as increased efforts to maximize the use of common tools, electronic data collection and administrative data. Since the Tourism Statistics Program is being redesigned, efforts are being made to ensure that all travel surveys meet the CBA framework. For the NTS, this means making use of a common household frame, generalized sampling and estimation software, an electronic questionnaire for data collection and any administrative data sources that improve all survey steps when applicable. This paper will mainly focus on the potential benefits of the use of administrative data for the NTS sample stratification.

2. Use of Administrative Data for the National Travel Survey

2.1 Survey Design Strategy

The NTS data collection tool will be electronic questionnaire and the target population will be all non-institutionalized Canadian residents living in the 10 provinces who are 18 years of age or older. An invitation letter will be mailed to selected dwellings and then respondents must go online to complete the questionnaire. The survey frame that will be used to select the monthly samples is a common household frame managed by Statistics Canada known to have excellent coverage of the Canadian population (~97%), called the Dwelling Universe File (DUF). The DUF is a snapshot of the Address Register and is updated with data from the Census of Population, tax data and other administrative records. Another common frame managed by Statistics Canada is the Socioeconomic File (SEF), which contains socio-demographic information for approximately 80% of households on the DUF. This demographic information from the SEF will be used as auxiliary information in the stratification of the DUF. Additionally, a file of Canadian passport holders is anticipated to be provided by Passport Canada, which will also be useful for stratification in order to improve the hit rate (i.e. the proportions of travellers selected in the sample). The stratification strategy will consist of splitting the sample frame (DUF) into strata using the following stratification variables: geographic variable (10 provinces), pre-classified household income groups (9) and a passport flag, which identifies households that have at least one adult passport holder.

There are three stages to the sample allocation: an allocation by province, by the passport variable and by income group. For the first stage, the total sample will be allocated to each province using a power allocation method of the following form:

$$\gamma_p = \frac{N_p{}^q}{\sum_p N_p{}^q},$$

where N_p is the population in each province and q=1/3. For the second stage, an allocation based on the passport variable will be performed. The specific method for this second stage allocation is still being explored since a final dataset has not yet been received from Passport Canada. Finally, for the third stage, a Neyman allocation will be performed to allocate the sample to each of the income groups. From each sampled dwelling, one adult will selected to complete the NTS questionnaire.

2.2 Current Uses of Administrative Data

Several sources of administrative data are already used for the ITS and the TSRC, which will be inherited by the NTS and used for the same purposes. For example, the TSRC uses the monthly demographic population estimates from the Census of Population to calibrate to the necessary control totals for sex, age group and province. Similarly, the ITS uses administrative data provided by the CBSA, called frontier counts, which provide a full range of statistics for all international and domestic travellers who have been cleared for entry or re-entry into Canada (Statistics Canada, 2016). These frontier counts are used as control totals in the weighting stage of the ITS.

2.3 Tax Data Sources for Income Variable

A relationship between household income and frequency of travel has been observed from previous travel surveys. It is known that households with higher incomes have a higher likelihood of both international and domestic travel. Thus, in an effort to obtain more travellers in the sample for the NTS, household income is used as a stratification variable. In other words, households from strata with higher incomes will be, in general, selected with a higher probability.

The main source of household income comes from the T1 Personal Master File (T1 PMF), which is provided by the Canada Revenue Agency and has one record per person. Household income can be obtained from the T1 PMF by summing all records from the same dwelling and then merging with the DUF to obtain an income variable at the dwelling level. If a dwelling does not correspond to a T1 PMF record then the dwelling income can be obtained from another source, which is the SEF. The income that is available on the SEF comes from the T1 family file that is also provided by the Canada Revenue Agency and is updated annually. T1 PMF income is preferred to the SEF income because it is more up-to-date.

To evaluate the quality of the T1 PMF, a record linkage was performed to merge the file with the DUF based on the AR_UID and Social Insurance Number (SIN). The AR_UID is a unique dwelling identifier on the DUF and SEF. The results showed that 68% of the records on the sample frame (i.e., the DUF) were mapped to the T1 PMF. In addition, 16% of DUF records were mapped to the SEF to obtain household income for records that were not matched with the T1 PMF. Thus, household income is available for 84% of households on the sample frame (DUF).

2.4 Passport Data from Passport Canada

It is difficult to ensure that both domestic and international travellers are represented in the NTS monthly sample. In 2010, less than 3% of Canadians travelled internationally (excluding the United States) in a given month. Similarly, November 2013 TSRC data shows that approximately 30% of Canadians travelled within Canada during that month. In an effort to select more international travellers in the sample, a passport variable is derived to identify households where at least one adult member holds a valid Canadian passport. Only adult passport holders are considered since the target population for the NTS is Canadian residents who are 18 years of age or older.

Since a passport is required for international travel, the hope is that the likelihood of sampling international travellers (and potentially domestic travellers as well) will be higher for households that have at least one adult passport holder.

An NTS pilot study was conducted in February and March of 2016 in order to test the response to the electronic questionnaire as well as all operational procedures. For this pilot study, Passport Canada provided a simple random sample of one million records of Canadians that hold a valid passport (from their file with approximately 23 million records). For the main NTS survey in 2018, Statistics Canada anticipates to have the full file for all Canadian passport holders. The Passport Canada sample file was matched to dwellings on the sample frame (i.e. the DUF) and then a passport flag was used to identify households with at least one adult passport holder. This flag was set to 1 if there was at least one adult passport holder living in a given dwelling and 0 otherwise. This stratification strategy allows to increase the probability of having international travellers in the NTS sample. Similar to the T1 PMF, the quality of the Passport Canada file can be evaluated by performing a record linkage with the sample file. For the sample of one million records used in the NTS pilot, approximately 85% of adult records can be mapped to the NTS sample frame. The matching variables used for this linkage were: AR_UID, first name, last name, gender and date of birth (Yang, 2015).

2.5 Preliminary Results from the Pilot Study

As described in section 2.4, the NTS pilot study was stratified using the passport flag. Approximately half of the pilot sample was allocated to strata where this flag indicated the presence of an adult passport holder in the household. For the portion of the sample where the passport flag was 0, 9.3% of the respondents reported at least one instance of international travel and 31.8% reported at least one instance of domestic travel. For the portion of the sample where the passport flag was 1, 11.9% of the respondents reported at least one international trip and 35.6% reported at least one domestic trip. As expected, the proportion of travellers, both international and domestic, seems to be slightly larger in households where an adult owns a passport. The impact is most likely lowered by the fact that strata where the passport flag equals 0 does not imply that there is no adult member in the household who holds a valid passport, but rather that no adult from the household was matched with the sample of one million records provided by Passport Canada. A sample of less than 5% from the file of all Canadian passport holders was used to perform the stratification and it means that potentially 95% of the passport holders are living in dwellings where the derived passport flag is 0. It also means that the majority of the dwellings where the passport flag is 0 should be replaced by a value of 1. For all the stated reasons, the differences reported above should be larger if the full passport file is used, and even more for the international travellers.

2.6 Potential Sources of Administrative Data for Future Improvements

The CBSA has an administrative file for the vast majority of Canadians who entered Canada at all border crossings. If Statistics Canada can obtain this administrative file then Canadians returning from international travel can be more easily identified and there would be no need for the Passport Canada file. Hence, the sampling design can be improved in terms of stratification for international travellers. Ideally, Statistics Canada would like to obtain access to this file as it would greatly improve the sampling efficiency for the Tourism Statistics Program.

Another potential administrative data source which can help improve the Tourism Statistics Program is transaction-level credit card data. At the moment, Statistics Canada does not have access to this data but would like to explore the possibility of obtaining it and the potential improvements it could provide for the Tourism Statistics Program. For example, the spending patterns of people with characteristics similar to the respondent may be useful to impute unreported expenditures. It may also be possible to identify travellers based on the location of the transactions, if the information is available. More exploration into the potential uses and availability of credit card data needs to be explored but no matter what the intended use, Statistics Canada will respect the Privacy Act and is committed to protecting the rights and privacy of individuals and their personal information.

3. Conclusion

The development of the NTS satisfies the CBA goal of consolidating processes and standardizing systems to achieve cost savings while maintaining high standards of quality. The NTS improves on the existing surveys of the Tourism Statistics Program by using more administrative data (such as tax files and an administrative file provided by Passport Canada), making use of common tools, implementing an electronic questionnaire for data collection and incorporating generalized software to perform sampling and estimation. There are numerous potential uses of administrative data

sources in tourism surveys, such as the NTS. The sample design can be improved in terms of the stratification process by using data from Passport Canada, the CBSA and the Canada Revenue Agency. Frontier counts provided by the CBSA, along with projected demography totals, can be used in the calibration process in order to improve the quality of the estimates. Other administrative data sources yet to be explored, such as the CBSA border crossing file and transaction-level credit card information, present great potential to improve travel surveys at Statistics Canada. The feasibility of using these administrative data files in the future will be investigated in order to further improve the Tourism Statistics Program.

References

Statistics Canada (2016), International Travel Survey: Frontier Counts (ITS), available at: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5005&lang=en&db=imdb&adm=8&dis=2 (accessed May 19, 2015)

Yang, Ming Jie (2015), "NTS Record Linkage Project", unpublished report, Ottawa, Canada: Statistics Canada.